UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

What factors predict the evolution of consumer credit in Belgium ?

Identification of determinants and modelling

Auteur : SEYNAVE Charles
Promoteur(s) : Catherine D'Hondt
Année académique 2021-2022
Master en Sciences de Gestion à finalité spécialisée en Financial Management

Abstract

In this paper, we seek to identify variables that predict the amount of new consumer credit contracts. We identify 13 variables from the literature review. Seven macroeconomic variables: income, consumption, unemployment rate, interest rate, level of income redistribution, price level and returns on investment. Two microeconomic variables: repayment condition, and competition between credit providers. Three socio-demographic variables: level of dependency, level of education and household's size. Finally, we add a behavioral variable, namely consumer sentiment. We then built several econometric models to explain and predict consumer credit. The model chosen is an autoregressive model. The results obtained show a rough estimate of the expected data. The variables that are found to be useful in explaining the variable of interest are income, repayment terms, level of dependency and consumer sentiment.

Résumé

A travers ce mémoire, nous cherchons à identifier les variables capables de prédire le montant des nouveaux contrats de crédits à la consommation. Nous identifions 13 variables issues de la revue de littérature. Sept variables macro-économiques : les revenus, la consommation, le taux de chômage, le taux d'intérêts, le niveau de redistribution des revenus, le niveau des prix ainsi que les retours sur investissements. Deux variables microéconomiques : les conditions de re-paiement et la compétition entre fournisseurs de crédit. Trois variables socio-démographiques : le niveau de dépendance, le niveau d'éducation et la taille des ménages. Finalement, nous rajoutons une variable comportementale à savoir, le sentiment des consommateurs. Nous avons ensuite construit plusieurs modèles économétriques visant à expliquer et prédire le crédit à la consommation. Le modèle revenu est un modèle autorégressif. Les résultats obtenus montrent une estimation imprécise des données attendues. Les variables qui se révèlent être utiles pour expliquer la variable d'intérêt sont les revenus, les conditions de re-paiement, le niveau de dépendance et les sentiments des consommateurs.

Acknowledgements

First, I would like to express my gratitude to the promoter of this thesis, Mrs. Catherine D'Hondt, for her availability and her advice which guided me wisely. This has also contributed to an in-depth reflection throughout this work.

Secondly, I would like to thank the people at the Professional Union of Credit and Febelfin, for their wise advice and follow-up throughout this thesis.

My final thanks go directly to my family and friends who have supported and encouraged me throughout this thesis.

Table des matières

1. Introduction

Whether it is to finance the purchase of a car, to buy new furniture or to finance a holiday, these are all reasons why Belgians use consumer credit. This type of financing has become a common way to finance these expenses. As explained in Guillaume Derclaye's July 2022 article, the number of consumer credit contract is increasing.

Therefore, it is in the interest of many financial actors to predict the future amount of consumer credit. Indeed, in order to best adapt to the evolution of household financing, it is useful to have an estimate of future consumer credit figures.

The aim of this thesis is therefore to identify the parameters that could be used to predict consumer credit and to build a model capable of estimating future consumer credit data. This thesis is based on a request from the Professional Credit Union. PCU is a professional association representing the consumer credit sector, namely consumer credit and mortgage credit. PCU has more than 50 members, including banks, mortgage companies, credit insurance companies, distribution companies authorized to grant consumer credit and companies issuing credit and charge cards.

This work starts with a definition of the subject matter and an overview of the situation in Belgium regarding consumer credit.

Then, we made an overview of the literature to identify the factors behind consumer credit. This part of the work is important because we chose the determinant variables to construct the model that will forecast the consumer credit.

Section 5 explains the database and explains the methodology used throughout the modelling.

I constructed different type of model in the section 6. We established the forecast, analyzed, discussed and compared the different model.

Finally, we identified the limits of my work in section 7 and concluded the thesis in section 8.

2. Definition and context (framework)

Definition

In this study, the focus will be on the loans to households particularly on the consumer credit taken by Belgian private consumers for purpose other than housing loans. Consumer credit is personal and unsecured debt taken on to purchase goods and services. Consumer credit is extended by banks, retailers and others to enable consumers to purchase goods immediately and pay off the cost over time with interest (Kagan, 2021). Consumer credit can have big impacts on other spheres of the economy and it's why this topic is so valuable for economists. Three types of consumer credit will be studied: the credit opening, the installment loan and the installment sale.

Credit opening is defined as any agreement, whatever its qualification or form, at the end of which a purchasing power, a sum of money or any other means of payment is made available to the consumer, who may use it by making one or more credit withdrawals (drawdowns), specifically by means of payment or legitimation card or in any other way, and who undertakes to reimburse it following the conditions agreed upon (Economic Law Code). This kind of credit is the one which gives the more freedom to use, you can use it for different acts and with the liberty to reimburse the loan at a certain date (SPF Economie, 2021). The borrower often opts for a credit opening when the investment maturity is not certain or when the amount needed is uncertain (expenses likely to be more expensive than expected). In fact, payment terms are irregular, the interest rate can fluctuate over the borrowing period and the amount already reimbursed are reusable.

Installment loan is any credit agreement, regardless of its characterization or form, under which a sum of money or other means of payment is made available to a customer who agrees to repay the loan in periodic installments (Economic Law Code). People who opt for this form of loan, often have investment projects with a fixed maturity and a fixed target borrowing amount.

Installment sale is a credit agreement, regardless of its qualification or form, which normally involves the acquisition of tangible goods or the provision of services, sold by the lender or credit intermediary and for which the price is paid in periodic installments, in at least three payments, omit the deposit (Economic Law Code). This deposit is paid when the contract is signed, and it is the signal of the beginning of the credit.

This study doesn't focus on the leasing which is also a form of consumer credit (a lease with an option to buy provided in the contract) because this type of credit is no more commercialized.

Framework in Belgium

The impact of the consumer credit on the Belgian economy is clear with 10,5 million of consumer credit in 2021. In 2020, consumer credit represents 5,89% of Belgium's GDP and increase slightly over time (UPC, 2021). We can already suppose that consumer credit and GDP have a high degree of correlation. The number of consumer loans remains approximately stable over time. However, regarding the amounts of credit granted over time, we note a strong increase in general and some decline in crisis period (Huyghebaert, 2015). In 2021, we observed that a large part of a number of consumer credit is made through credit openings (69% of total consumer credit) and through installment loans (29% of total consumer credit) (UPC, 2022).

The outstanding debt per capita (calculated on the basis of these turnover data instead of production data) has increased primarily for installment loans (both in nominal and real terms). We also note that outstanding debt per capita increases mainly during boom periods. Indeed, in boom times there is a stronger increase in the fraction of consumption financed through consumer credit. Inversely, it was observed that households try to reduce their consumer credit intake when the business cycle is down (Huyghebaert, 2015).

The sectors most affected by consumer credit are furniture and home appliances, restaurants and hotels, culture and leisure, and automobiles.

Credit consumption is intuitively linked to the goods consumption. Unlike consumer credit, domestic consumption has an enormous importance for the Belgian economy. Prof. Dr. Huyghebaert (2015) observed in her study that consumption expenditure in Belgium is very stable. These expenditures are made in several sectors like food, beverages and tobacco (17,18% in 2012), transportation (11,98% in 2012), culture and entertainment (8,92% in 2012), restaurants and hotels (6% in 2012), furniture and household appliances (5,6% in 2012) or clothing and footwear (5,16% in 2012) (Huyghebaert, 2015).

Consumer credit is also linked to the disposable income. In 2020, disposable income represented about 79,24% of GDP at market prices. Sixty-one percent of the disposable income was consumed through households for the year 2020 and this proportion remain almost stable over time. We also observe that growth of disposable income is highly and positively correlated with the growth in consumption expenditure (Huyghebaert, 2015).

We also note that a growth in consumption is more likely explainable by an increase in credit than by an increase in household disposable income. Indeed, a 1% increase in disposable income in one semester increases consumption incrementally by 0,1899%, while a 1% increase in the amounts of credit in one semester increases consumption incrementally by 0,2916% (Huyghebaert, 2015).

The growth in the consumption of the category ‘furniture and household appliances’ is driven by credit (particularly installment loans and credit openings) rather than by an increase in disposable income. Installment loans also have a significant positive impact on the category ‘transport’ and installment sales and credit openings have a significant positive impact on the category ‘restaurants and hotels’ (Huyghebaert, 2015).

It is now important to reflect the state of mind of Belgian consumers regarding credit. In 2020, a study revealed that a major part of Belgian consumers (71%) doesn’t like to take on credit because they see credit as a burden. Fifty percent of these consumers are quite pessimistic about the economic situation in the future and consequently they don’t believe that it’s the right moment to enter into a credit. However, the confidence in credit providers (like banks or other credit providers) has increased in one year (+8%) (UPC, 2020).

Regarding the different categories of expenditures, the study emphasizes the most useful categories according to the consumers: Renovation and rebuilding, energy saving renovation, new vehicles, secondhand vehicles and other vehicles. Consumers see credit for vacation or travel as less useful than before (UPC, 2020).

When asking for a credit, personal contact is almost always preferred by consumers particularly for expensive categories. Questioning about the way the consumers prefer to sign their credit contracts, signing in a bank or in a financial institution is the most popular answer. We also note that digital signature (e-ID or via Itsme) is relatively popular (UPC, 2020).

The study highlights that 70% of households who enter into a credit contract has made a credit for a large purchase or a large expenditure. The type of consumer credit for each category is studied and we see for example that consumers prefer broadly installment loans for expenditures like vehicles, furniture and home appliances, renovation or technologies. For renovation and vehicles, consumers tend to take approximately 70% of the total amount on credit and they tend to take full amount on credit for cheaper purchases (UPC, 2020).

In 2020, the credit barometer estimates that 54% of the households expect to make a large purchase or expenditure in the next 12 months which is a lower percentage than one year before. People still plan to take on credit for categories like renovations and new vehicles (UPC, 2020).

3. Identifying the Factors Behind Consumer Credit

In this section, we will discuss over several factors that may be included in an explicative model of consumer credit. We will discuss each factor whether it is relevant to consider this variable and what could be the impact of this variable on the consumer credit demand.

3.1.Macroeconomic variables

3.1.1. Income

Intuitively, we can think that revenue will have an impact on the credit demand. In fact, we can see in the *Appendix 1* that the credit will be linked to the goods and services consumption and that this consumption will be impacted by the level of the household's revenue (Deaton, 1989 and Friedman, 1957). In this regard, the disposable revenue was the topic of a lot of research that is sometimes contradictory.

According to Burdeau (2015), households reduce their credit demand due to constraints on their balance sheets. It means that the credit demand depends notably on the assets and the cash flows of households. B. Bernanke and M. Gertler join this opinion by identifying the impact of the balance sheet on the consumer credit market particularly the impact of the net worth of borrowers. They observed that the greater the borrower's net worth the lower the external finance premium should be and thus the more attractive the loan seems. Besides, a stronger financial position enables the borrower to reduce her potential conflict of interest with the lender (Bernanke and Gertler, 1995).

Huyghebaert analyses in 2015 that consumer credit is procyclical. It means that households don't compensate for a decrease in disposable income by taking out or granting more credit. The correlation between growth of consumer loans and the growth of the household's disposable income is therefore significant and positive (56,57% in 2013).

However, J. Borowski, K. Jaworski and J. Olipra estimate that an increase in per capita income will lead to a decrease in the consumer credit. They noted that it is consistent with the permanent income hypothesis. In this study, the authors explained that in some EU country with a low per capita income, household expect that their income will increase in the future (due to the process of convergence to the EU average) and they are therefore likely to increase their consumption level and consequently increase their consumer credit intake. Finally, they find a result in contradiction to their beliefs: a positive correlation.

Other authors thought that households will take advantage of an increasing income by reducing their debt level. They thought in the same idea that in recessions, income is likely to decrease and that households will increase their demand for bank loans in order to smooth out the impact of lower income (Calza, Gartner & Sousa, 2001). Similarly, the households with a higher initial endowment of assets (high net wealth) have the capability to satisfy its needs without borrowing as proved in a study conducted in India (Swain, 2007).

Karlan and Zinman observed that loan size is more responsive to price (costs of borrowing, lending rates) for the higher-income individuals. Income has a part of permanent component and a part of transitory component. The permanent one is driven by the capital value or wealth, the economic activity and so on. And the transitory one is driven by cyclical fluctuations, chances or errors of measurement but specific to consumers (Friedman, 1957).

The credit consumption can be linked to the liquidity constraint which itself depends on the household's income. We thus must pay attention to the collinearity. Cash loan sizes can be relative to borrower income (Karlan & Zinman, 2005).

The potential explanatory variable to represent the disposable income could be the GDP per capita in purchasing power parity (PPS).

3.1.2. Consumption

Intuitively, we can think that consumption of goods will have impacts on the demand for consumer credit because more the consumption increases, more likely the consumer credit demand will increase too. We must be careful because of the link between consumption and economic growth and the revenue of households (*Appendix 2*) (Huyghebaert, 2015).

In fact, economic activity affects credit demand and for example, if economic activity grows, income and consumption will increase with respect of this growth. Consequently, robust economic growth enables households to support higher levels of indebtedness (Calza, Gartner, Sousa, 2001).

Calza, Gartner and Sousa presented this relation by a suite of links for both households and firms. In fact, if there is expectation of higher activity and productivity, it can lead to a larger number of projects becoming profitable in terms of expected net present value and consequently to a higher demand for credit to fund them.

The consumption factor reflects a cyclical factor via his transitory components like preferences, special opportunity to purchase or unusual sickness (Friedman, 1957).

The consumption of households can be measured by the GDP via the calculation through the consumption (C+I+G+X-M) (Carroll, Fuhrer and Wilcow, 1994). An index which helps to estimate the future evolution of the economy is the Purchasing Managers' Index (PMI) which reflect the sentiment of specialists about market conditions.

3.1.3. Unemployment Rate

Many authors studied the impact that the level of unemployment can have on the credit demand. It seems that the unsecured credit and the long-term average unemployment rate have a highly negative correlation. Over time we can see that unsecured loans increase while the long-term unemployment rate decrease (Bethune, Rocheteau and Rupert, 2015).

These previous authors also describe the link between the labor market and the credit market. The relation between these two markets can be explained by the fact that increasing debt limits cause an increase in borrowing amounts that will result in an increase in good consumption and production. Therefore, an increase in debt borrowing promotes job creation and reduce the unemployment rate. Consequently, they concluded that the differences in credit conditions have significant implications for unemployment in the long run (Bethune, Rocheteau and Rupert, 2015).

For instance, they calculated that a decrease in the probability to find a job will cause a decrease in the aggregate firm revenue (link with good market) and a decrease of the access to credit (link with credit markets). Their model predicts that credit differences can account for 69% of the total difference in the average unemployment rate between their period of study (Bethune, Rocheteau and Rupert, 2015).

There is a snowballing effect which creates amplifying consequences. If there is an exogenous decrease in production of goods, there is an increase in equilibrium unemployment (so the probability to fill job decrease) because firms post fewer vacancies, and the result is a rise in unemployment. This result has the effect of lowering households' access to unsecured credit. This causes a further drop in firm revenue and a larger increase in unemployment.

The employment factor can be a good determinant factor because it reflects the influence of the economic cycle on credit (Borowski, Jaworski and Olipra, 2017). To reflect this variable, we can take the unemployment rate in Belgium over time or also the probability to find a job over time.

3.1.4. Cost of Borrowing

The cost that the borrower must pay to obtain his credit will influence his decision and thus it will have an impact on the general credit demand. As seen in the Bank Lending Survey, the low general level of interest rates continued to have a positive impact on demand for all loans to households in 2021.

Interest rates represent the cost of borrowing for the consumers, and it is a cyclical factor that affects the consumer credit demand (*Appendix 5*). Borowski, Jaworski and Olipra demonstrate that interest rates are one of the main factors influencing cost of consumer credit and that they have an impact on creditworthiness of households even as credit availability.

Karlan and Zinman studied in 2005, the demand elasticity of credit and they emphasized the importance of the interest rates to increase or decrease the access to credit. They studied the consumer sensitivity to interest rates to nuance the previous results that said that elasticity are close to zero and they proved that this elasticity range from -0,73 to below unity. They also observed that a low initial interest rates will produce more future borrowing. They also studied the positive dependence between interest rates and default probabilities. Indeed, if interest rates increase due to bad shocks for instance, loan defaults will increase too. This is unprofitable for the lenders because of the decrease in the demand curve and the increase of loan losses.

Intuitively, we can estimate a negative dependence between the credit demand and the level of interest rates. Indeed, if the interest rates are increasing, the borrowers should be less likely to enter a credit contract. There are many studies that confirm this idea that a rise in interest rates provoke a loan contraction. Some research links the interest rate’s level to the decision of the monetary policy (Aiyar, Calomiris and Wieladek, 2016; Bernanke and Gertler, 1995). In their analysis, Bernanke and Gertler take the example of the federal fund rate which is a closely controlled and an overnight interest rate. They also discuss the effects of a change in interest rates over the financial position of borrowers if there is a tightening in the monetary policy.

One of the instruments of monetary policies is indeed the interest rates that central bank can control, and it will affect the bank lending. A decline in bank capital is one effect of an increase of rates and a loan contraction because of unexpected losses due to interest rate risk. If there is a low interest rate environment, banks are less risk averse and they will accept more credit demand and households are more likely to submit a credit demand (Aiyar, Calomiris and Wieladek, 2016). Monetary policy affects not only the general level of interest rates but also the size of the external finance premium (Bernanke and Gertler, 1995).

The variation in borrowing rates can also be endogenous to borrowing and repayment behavior through characteristics of the borrowers. This idea induces that exogenous variation of borrowing rates (via monetary policy for example) is rarely observed (Alan, Dumitrescu and Loranth, 2011). The interest rates can indeed depend on time or individual specific investment opportunities or financing alternatives.

Swain explains that in the cost of borrowing, we find different costs like the opportunity cost of funds, the administrative cost of servicing the loan or the risk premium to be charged to the borrowers (the main part of the cost).

For this variable, the negative relationship between demand for loans and their cost seems to be more consensual. We can now reflect about which type of cost we have to include into our model, interest rates from federal fund rate, bank lending rates, interest rate spreads or long-term bond yield.

3.1.5. Redistribution of Revenue

Redistribution of revenue is a principle to avoid the large inequality of income by redistribution via taxes, for example. Belgium seems to be one of the more equal countries in terms of redistribution. In fact, if we compare the distribution of gross revenue and disposable revenue (after taxes and social correction), we observe a decline in inequality (RTBF, 2014).

Bethune, Rocheteau and Rupert explained in 2015 that a general smoother labor income causes a decline in workers' desire to borrow. Redistribution allows lower income risk and lower default probability such that households will be less likely to enter into a loan contract.

Consensual, Borowski, Jaworski and Olipra (2017) explain that inequality leads to political pressure for redistribution. However, redistribution leads to a better access to credit for the low-income households and according to those authors, consumer credit is higher when there is redistribution.

Economies which find it more difficult to supply credit efficiently also implement policies that support stable labor income and compressed earnings. Besides, benefits of redistribution for individuals who are currently poor are smaller when these individuals can look forward to more strongly increasing expected income (Borowski, Jaworski and Olipra, 2017).

This factor could be measured via the employment protection indicator (index of labor market regulation) or via the Gini index that measures the level of inequality of the distribution of a variable in the population.

3.1.6. Inflation

Inflation is a strong determinant of the economic activity and reflects often the monetary policy strategy of a country. Inflation also has an impact on the interest rate, and it is why inflation can have a big influence on the consumer credit. European Central Bank observed a clear dependence between the level of inflation and the economic activity inside a country (Coralia-Emilia, 2013).

Inflation has also an evident influence on the way that consumers will buy goods. If the level of inflation is unanticipated, the dependence between inflation and purchasing attitudes of households is negative. If the unanticipated inflation increases, the purchasing attitudes and intentions will consequently decrease. Anticipated inflation has a positive influence on the expected purchases but less influence on the consumer sentiment so that the effect is weaker (Juster, Wachtel, Hymans and Duesenberry, 1972).

The credit boom observed in Greece since the mid-1990s is attributed to the disinflation process. It is a proof of the negative dependence of credit and inflation (Brissimis, Garganas and Hall, 2014). Inflation is often linked to interest rates so we will pay attention to collinearity between this two, but some authors explain that inflation nuance the effect of interest rates (Borowski, Jaworski and Olipra, 2017).

Inflation can be illustrated by the consumer price index like it was in the study of Coralia-Emilia in 2013.

3.1.7. Return on equity investment

Stocks returns can be a determinant factor for the consumer credit because they are related to changes in consumer confidence which impact the consumption expenditures and the credit (Huyghebaert, 2015). According to Wosko in 2015, the rate of return on stock market index is a good predictor of the overall economic sentiment. That can influence the bank loan policy as well as the demand for credit.

If the return on equity is high, the consumer sentiment will be good, and the households will be more favorable to take a credit to finance their needs.

3.2. Microeconomic variable

3.2.1. Repayment condition

The repayment conditions that bank offer can influence the credit demand. Some conditions will encourage the credit demand and others will dissuade consumers. Particularly when we face high-interest loans, borrowers must agree with several incentives to repay these risky loans. These carrots can take the form of decreasing prices over the repayment period or increasing future loans sizes following a good repayment behavior. However, some conditions are unattractive like reporting to credit bureaus, frequent phone calls from collection agents, court summons or wage garnishment. (Aiyar, Calomiris and Wieladek, 2016).

Credit demand will thus depend on collateral requirements. If general economic environment is risky, financial institutions will require more collateral for loans and then it will lower the desire to borrow.

Consumers who want to enter a loan contract are also sensitive to the repayment condition as they are sensitive to the price of the loan. A long maturity often implies lower stable repayment during a longer time which is attractive for households.

To illustrate this variable, reflect the tendency of the repayment condition can be a good estimator. The National Bank of Belgium provides indicators reflecting the opinion of consumers about global conditions of credit granting (collateral requested, clauses in loan contracts, maturity, etc.).

3.2.2. Credit Providers Competition

Consumer credit is driven by development and concentration of the credit providers and the financial sector over time. Indeed, higher concentration and development of the banking sector led to an increase in consumer credit (Borowski, Jaworski and Olipra, 2017). One explanation of this effect is that the concentration of banks leads to a decrease in the interest rates that leads to an increase in consumer credit demand (Weill, 2004). We can also think that more the credit sector is developed, more the households will find the right solution to take a credit and thus submit their demand for consumer credit.

A lot of credit providers in Belgium are not banks but the competition between the different other credit providers remains a possible determinant for the consumer credit.

Karlan and Zinman studied the price sensitivity of credit demand and they observed that when the cost of borrowing increases, the borrowers look for multiple solutions to find a credit and it amplifies the competition.

We can reflect about the variable which identifies this factor that can be the share of the five largest credit providers in total assets, the Herfindahl index for credit providers or the number of local units of credit providers per 100 000 inhabitants.

3.2.3. Substitute of consumer credit

The number of the substitute of consumer credit can influence the demand for this consumer credit. In fact, the households can have a lot of disposable substitutes to finance their project and it will impact the credit demand.

Substitute of consumer credit depends on the size of the consumer credit the households need and on the disposable substitute they have. If a household has enough disposable funds, one of them is the insurance advance which consists of a demand of advance on some specific life insurance. Advance on life insurance allows the borrower to dispose of an amount at the lowest interest than a credit. Another alternative to credit is to dip into your home savings plan. These first substitutes are linked to the saving behavior of households (Lambert, 2013).

André Babeau (2018) explains in his text that saving behavior of households are directly linked to their credit situation. He explains that the link between saving, and consumer credit was too often forgotten in the past and says that one of the main goals of saving is to reimburse a loan.

Households have also the possibility to use the pawnbroking. Indeed, if someone wants to borrow and have a valuable object like jewel or watch, he can deposit this object in a credit institution in exchange of the amount borrowed. Then the person must reimburse the amount with interest (Lambert, 2013).

It is also possible to borrow thanks to a family loan or a loan between individuals. Frequently, there is no interest in the family loan and low interest in the loan between individuals (Lambert, 2013).

There exists now some other alternative like instant loan on a digital platform named Lydia for example.

In high-rate environment, borrowers look for other solution than traditional credit or they wait for the lender's rates drop and this phenomenon leads to a decrease in the demand for consumer credit (Karlan and Zinman, 2005).

The impact of substitute to consumer credit can be measured by the proportion of people using substitute or by the part that alternative lending represent.

3.3.Sociodemographic variable

3.3.1. Age

Demographics can play an important role in the determination of the demand for credit. Higher share of young people in population leads to increase in consumer credit. This idea represents the life cycle hypothesis. Younger people, with expectations of rising income and high marginal utility of consumption are more interested in taking consumer loans than older people. Taking consumer credit allows to smooth the consumption which is valuable for young households. This positive dependence was proved in the study of Borowski, Jaworski and Olipra.

Age is one of the variables to describe the household's characteristics according to Swain. She finds that this variable is insignificant in the loan demand equation, but she considers another variable which represents a demographic indicator: the dependency ratio. This ratio is a measure of the number of dependents aged zero to 14 and over age of 65, compared with the total population aged 15 to 64. This indicator reflects the number of people of non-working age, compared with the number of those of working age (Hayes, 2021). This ratio also represents a characteristic of households, and he is significant and positive (Swain, 2007). A higher number of dependents imply a higher consumption need of the household but a lower per capita income which according to Swain leads to more credit demand.

Age is also linked to interest rates and financial sophistication (knowledge or education). Young people live longer with their parents in countries where credit is scarce (it's difficult to borrow) and where the labor market is inflexible (Bertola and Koeniger, 2004).

To reflect this variable, we can think about the share of young people in the total population or the dependency ratio.

3.3.2. Education

Levels of education positively affect consumer credit because a high level of education raises the probability to obtain a stable future income which leads to a rise in the credit demand (it also raises the creditworthiness and lowers the margin on consumer credit) (Swain, 2007).

Education is measured by taking the proportion of people aged between 15 and 64 that have completed upper secondary, post-secondary, tertiary education.

3.3.3. Size of household

An increase in the family size should decrease the demand for loans because a big family size reflect an increased number of earnings members in the family. And in this study, the author thinks that if households earn more money, they will be less likely to demand for a credit (Swain, 2007).

However, many authors do not agree with this statement about the earned income and the demand for credit. Moreover, we can think that the higher the household's size the higher the needs of consumption for the family and thus the higher the usefulness of credit.

4. Research Gap with the Literature

After having gone through the literature concerning the relation between consumer credit and macroeconomic, microeconomic and sociodemographic variables, we never see clear result about the relation between consumer credit and household's behavior. Many papers talk about the possible explanatory character of a household's behavior but none of them tested really if there is a concrete dependence.

BLS balances of opinions are frequently used as explanatory variables of the evolutions of credit flows but especially for firms' loans as in research of De Bondt, Maddaloni and Peydro.

The Bank Lending Survey has been set up by the Eurosystem in 2003. This is a qualitative survey implemented in each euro area country and developed to enhance monetary policymakers' knowledge on credit markets. There are qualitative questions about the future evolution of the credit markets and about the future household's behavior (Burdeau, 2015).

E. Burdeau observed that evolutions of balances of opinion in the BLS precede evolutions of quantitative data such as credit flows. This is the reason why this kind of indicator or variable must be delayed.

BLS collect a lot of expectation for the future evolution of credit markets and identify the developments in consumer confidence, the increased spending on durables and the low general level of interest to contribute positively to loan demand by households in the end of year 2021 (*Appendix 3 and 4*). Banks also expected that demand for consumer credit will have a moderate net increase.

However, the relation and the reliability of these surveys were never tested but just observed. What we can think is that consumer credit also depends on the change of consumer sentiment indicator. Indeed, if the optimism concerning the financial situation in the following months is high, the households' expenses will increase as well as their demand for credit.

We also know that households have a high sensitivity to high rates. It means that if interest rates increase slightly, credit demand decrease slightly too and vice versa for a decline (flat demand curves for consumer credit). Prospect theory from Kahneman and Tversky generates this pattern because consumers evaluate prices relative to their prior experiences and they weight losses more heavily than gains (Karlan and Zinman, 2005).

To illustrate this variable, the consumer survey or the bank lending survey reflect well the household's opinion and these indicators can be extracted out of the National Bank of Belgium (NBB).

One consumer survey proposes an indicator based on assessment of credit conditions. Households can evaluate the general credit conditions through their opinions about the interest rates, the volume of credit or the collateral required.

Another consumer survey reflects the confidence of consumers through their previsions of the economic situation, their prevision of unemployment, their prevision of the household's financial situation or also the prevision about the prices and the saving level situation.

5. Methodology

5.1. Data

In order to create a model to forecast future data, time-series data are relevant, and they are practical. Indeed time-series regression is often used to model or forecast economic, financial or engineering systems. Predict future responses based on the response history is a well-known statistical method also called autoregressive dynamics (MathWorks).

For example, in a time-series regression of consumer credit on GDP, the random variable CC is called the dependent variable, because the model shows how its mean depends on the vector of k regressors of independent or explanatory variable GDP. Time-series data reflect the same variables in different point of time no matter which group of individuals is selected.

I choose to select my variable with quarterly frequency because we want to forecast consumer credit more than one time in a year and four times a year was feasible in the point of view of the available data. The period we selected to run my model is 14 years from Q1 2007 to Q4 2020. It's a quite large period to draw trustable conclusions and it is recent period to be representative of the near future. Nevertheless, the number of observations is not big (48 observations + 8 observations) because we consider quarterly data. The reason is because we can't find all the variables monthly and the goal is not just to forecast the consumer credit with two or three variables, but it is also to identify which variables is useful for this purpose.

In the next section, the explanation of my dataset with the specification about the collection of the database.

5.1.1. Dependent variable

The dependent variable is the amount of new consumer credit contract. This number is available each month and expressed in billions of euros. This number of new consumer credit includes the three types of consumer credit explained in the definition's section. This variable is published by the statistics of UPC.

5.1.2. Macroeconomic variables

Variable full name (short name)	Unit	Definition	Source	Previous research
Gross wages and salaries (Income)	In billions €	Household's remuneration calculated in Gross Domestic Income (or Gross Internal Revenue). It doesn't consider social contributions.	NBB.Stat	Deaton, 1989 – Friedman, 1957 – Burdeau, 2015- Bernanke &Gertler, 1995 – Borowski, Jaworski&Olipra, 2013 – Calza, Gartner&Sousa, 2001 – Swain, 2007 – Karlan & Zinman, 2005
Final private consumption expenditure (GDP)	In billions €	Final private consumption expenditures calculated in Gross Domestic Product with the consumption approach.	NBB.Stat	Calza, Gartner & Sousa, 2001 – Friedman, 1957 – Carroll, Fuhrer & Wilcow, 1994
Harmonized unemployment rate (Unemployment)	In %	Unemployment rates of all age classes and all genders harmonized with official national figures from Eurostat.	NBB.Stat	Bethune, Rocheteau & Rupert, 2015 – Borowski, Jaworski & Olipra, 2017
MFI Interest Rate (IR)	In %	Bank interest rates for loans to households for consumption.	European Central Bank – Statistical Data Warehouse	Borowski, Jaworski &Olipra, 2017 – Karlan & Zinman, 2005 – Aiyar, Calomiris & Wieladek, 2016 – Bernanke & Gertler, 1995 – Alan, Dumitrescu & Loranth, 2001 – Swain, 2007

Gini Index (Gini)	In unit	Gini Index (representing the level of income redistribution) at survey year and considered as a year-fixed index.	EU-SILC, IWEPS	Bethune, Rocheteau & Rupert, 2015 – Borowski, Jaworski & Olipra, 2017
Global harmonized consumer price index (CPI)	In %	Consumer price index representing the price level. It includes taxes on products such as value-added tax and reflect seasonal balances.	NBB.Stat	Coralia-Emilia, 2013 – Juster, Wachtel, Hymans & Duesenberry, 1972 – Brissimis, Garganas & Hall, 2014 – Borowski, Jaworski & Olipra, 2017
BEL 20 stock return (Return)	In %	Stock return of BEL 20 which represent the 20 most important companies in Belgium	Boursorama	Wosko, 2015 – Huyghebaert, 2015

5.1.3. Microeconomic Variables

Changes in award criteria (Conditions)	Dummy	Scores of the Bank Lending Survey regarding the conditions to take on a credit. If cost of funds, balance sheet constraints, contract counterparts, motivation tighten (it's more difficult to obtain a credit), it equals to 1	European Central Bank – Statistical Data Warehouse.	Aiyar, Calomiris and Wieladek, 2016
Impact of bank competition (Competition)	Dummy	Bank Lending Survey's Scores. If bank competition ease credit standards, it equals to 1	European Central Bank – Statistical Data Warehouse.	Borowski, Jaworski & Olipra, 2017 – Weill, 2004 – Karlan & Zinman, 2005

5.1.4. Socio-demographic Variables

Dependency ratio (Dependency)	In %	Fixed-year ratio calculated by dividing number of people aged between 18 and 64 by people aged less than 18 and people aged more than 64	Statbel	Borowski, Jaworski & Olipra, 2017 – Swain, 2007 – Hayes, 2021 – Bertola & Koeniger, 2004
Level of education (Education)	In %	People with a level of education higher than secondary education (fixed-year effect)	Statbel	Swain, 2007
Size of households (Size)	In number of people	Mean number of people in a household (fixed-year number)	IBSA	Swain, 2007

5.1.5. Consumer sentiment

Consumer confidence index (Sentiment)	In unit	Index representing the future and past confidence of consumers	NBB.Stat	Burdeau, 2015 – Karlan & Zinman, 2005

The consumer confidence index is a good variable to represent the consumer sentiment and his beliefs about the general economy. This indicator is composed by some forecast about the economic situation, about the unemployment forecasting, about forecasting the financial situation of households or still the forecast of saving situation of the household. This indicator is interesting because it also considers the near past sentiment of consumers about the recent situation of economic life. Indeed, consumers are asked about the past and the future price level or about the past household's situation. They are also asked about their judgment if it is the right moment to make an important investment (Appendix 6). This variable is available monthly only, so we took the mean over 3 months to represent the trend over the quarter.

As observed, four variables are disposable only annually, so we assumed that these figures are stable in the same year and change only between years. Sometimes we only found monthly data and we transformed them in quarterly data by taking the mean on three months. This method gives the tendency over three months, but we considered them like the quarterly data.

5.2.Calculation Method

I build my model on a period of 12 years from 2007 to 2018, it will represent the in-sample part of the work. Then we will take a period of 2 years from 2019 to 2020 to test my model and it will represent the out-of-sample part of the work. Even more my model is called multivariate time series forecasting because we have multiple exogenous variables.

In a lot of types of models, external validity is paramount. The historical data must be able to predict future data. We must check the stationarity. A time series is stationary if the probability distribution does not change over time, the joint distribution doesn't depend on the time. Forecasting a stationarity series is easier and the forecast is more reliable. If the series are stationary, we have fewer difficulties to forecast because the autocorrelation will be fewer.

To create a stationary model, the best-known method is to differentiate the value. In other words, we subtracted the previous value from the current value. We must apply this differencing only for the variables that are non-stationary, not for the others. To know if a series is stationary, we performed an augmented Dickey-Fuller test and saw if the p-value is under 0,05 or not. If the p-value was fewer than 0.05, the series wasn't stationary, and we differentiated it.

We will also put importance to the correlation of series with their own lagged values (autocorrelation or serial correlation). To observe the correlation, an autocorrelation function (ACF) gives the correlation between 2 variables at different lags and indicates when it is significant or not.

A partial autocorrelation function gives the coefficient of a variable for different lags. For example, at lag 2, a PACF gives the pure contribution of β_2 in the equation $Y_t = \alpha + \beta_1 y_1 + \beta_2 y_2$. So the partial autocorrelation is the correlation between the series and its lag after excluding the contributions from the intermediate lags.

I compared different models by looking at an information criterion. With information criteria we trade off bias (too few lags) versus variance (too many lags). Bayes (BIC) and Akaike (AIC) are two information criteria that represent the information loss. The fewer they are the better.

My attention was also put on the R squared because it shows at which level my models can predict the future response. It determines the precision of my model and so the usability of my model to predict consumer credit. Obviously, we paid attention to the residuals of my models too.

5.2.1. Choice of the models

I tested some regression model like the Vector Auto-Regressive if the data are stationary or the Auto-Regressive Integrated Moving Average model or the Vector Error Correction Model.

Firstly, the vector auto-regressive model is often used in the literature to forecast macro and micro-economic variables such as output growth or inflation (Clements & Galvão, 2013). They can also be used for economic analysis (Lütkepohl, 2013). This model is identified as one of the most successful, flexible and easy-to-use models to forecast micro or macro-economic variables (Zivot & Wang, 2006). For example, the model to forecast the Swiss inflation has been shown to be useful to predict the future inflation level with a sample of 80 observations (Lack, 2006). Despite the short sample period we tried to use this model because of his success in the existing research (Levieuge, 2017).

Then, we performed a vector error correction model which is also a famous technique to forecast many types of variables. VECM gives better results of forecast with lower root mean squared errors (Anggraeni, Andri & Mahananto, 2017). VAR and VECM seems to be models useful to reflect marginal informational content of many indicators. VECM allows to investigate potential nonlinearity in credit dynamics. However, it appears difficult to accurately predict credit dynamics (Levieuge, 2017). The study leads by Andersson and Lauvsnes in 2007 shows that multivariate models predict better when cointegration is imposed.

Finally, we chose to run an ARIMA model because this model is very flexible and makes very few assumptions (Ho & Xie, 1998). Moreover, this model is a viable alternative that gives satisfactory results in term of its predictive performance. As shown by Noreen, Asif, Nisar and Qayyum in 2017, the bank credit to private sector can be model and forecast by an ARIMA model. This style of model is also used to forecast macro and microeconomic variables such as inflation or gold prices (Meyler, Kenny and Quinn in 1998 and Guha & Bandyopadhyay in

2016). However, care needs to be paid in the use of ARIMA models in a forecasting context because of some limitations of ARIMA approach (Stevenson, 2007).

6. Calculation and modeling

6.1.Causality

After the review of literature, we suppose that the selected variables will cause or at least will impact the dependent variable “CC.” However the impact of the variables is studied in theory, with a large dataset, over different situations and over a large period. In this study, we have a quite short time period with the specific situation of Belgium and specific kind of data between 2007 and 2018. Consequently, we must check the causality between each independent variable and the dependent variable “CC”.

It's why we perform a Granger Causality test to observe the relation between variables. The null hypothesis of this test is that the variable X does not Granger-cause the variable Y. We will also observe the p-value to see in which proportion each variable could cause the dependent variable “CC”. *Figure 1* is an example of a Granger Causality test result.

Figure 1:

```
> grangertest(Gross_Data$GDP~Gross_Data$CC,order = 4) # (4)=44%    (1)=56%  (2)=30%  (3)=37%
Granger causality test

Model 1: Gross_Data$GDP ~ Lags(Gross_Data$GDP, 1:4) + Lags(Gross_Data$CC, 1:4)
Model 2: Gross_Data$GDP ~ Lags(Gross_Data$GDP, 1:4)
  Res.Df Df      F Pr(>F)
1     35
2     39 -4 0.9594 0.4419
```

The Granger Causality test is a Wald test comparing the unrestricted model (model 1 : CC is explained by the lags of CC and GDP here) and the restricted model (model 2 : CC is only explained by the lags of CC). We notice that the p-value is equal to 44,19%, meaning that in 44,19% of the case, the null hypothesis can’t be rejected. The number of lags (here 4) specified in the formula correspond to the order of lags to include in the auxiliary regression.

The number of lags for each variable is chosen by performing a VAR selecting calculation on R. The number of lags for the autoregressive (AR) part is p and is often between 1 and 5 for our variables. The “VARselect” calculation gives four information criteria that show the ideal order. *Figure 2* is a summary of the different Granger Causality test for each variable and for some different lags.

Figure 2:

P-value of Granger test				
	Y = CC	Y = CC	Y = CC	Y = CC
X =	Order = 3	Order = 4	Order = 5	Best order
Income	25,72%	17,84%	82,35%	17,84%
GDP	36,74%	44,19%	93,64%	30% (2)
Unempl	23,09%	35,58%	71,96%	13% (2)
IR	3,79%	9,02%	7,87%	0,7% (1)
Gini	20,05%	37,05%	0,03%	0,03%
CPI	76,20%	70,38%	64,62%	40% (1)
Condition	16,31%	18,65%	26,26%	16,31%
Competition	23,49%	36,41%	67,98%	23,49%
Return	27%	45,49%	57,29%	25% (2)
Dependency	91,46%	94,63%	1,90%	1,90%
Education	94,16%	80,61%	81,55%	80,61%
Size	89,80%	83,98%	93,94%	71% (1)
Sentiment	28,73%	29,02%	37,35%	28,73%

We note that variables like “GDP”, “CPI”, “Competition”, “Return”, “Education” or “Size” have a quite high p-value which means that it is difficult to reject the null hypothesis of zero causality. However, it doesn’t necessarily mean that these variables don’t have any prediction’s power.

6.2.Correlation

The relation between the variables can have an impact in the results or in the performance of the models and it is why we perform beforehand a test of correlation between the variables. First, we calculate the correlation matrix (Appendix 7), and we notice particularly the correlation with a red flag (correlation>0,7). Among these correlations, we see a strong correlation between 3 main macroeconomic variables. In fact, the correlations “Income-GDP”, “Income-IR” and “GDP-IR” are respectively 88%, 72% and 87%. We notice the same problem with the variables “Dependency”, “Education” and “Size”.

I see that these correlations and particularly the relation of "GDP" or "Education" with the other variables can cause problems (like multicollinearity) when we perform models like VECM because a variable which is a linear multiple of another variable create disturbances in some model.

6.3. Autocorrelation

Like mentioned before, we perform autocorrelation test to see if some variables are correlated with their own lagged variables. We perform an Auto-Correlation Function (acf) test on each variable to show when the autocorrelation is significant and for which lag. The plots of acf test are in *Appendix 8*, they are called auto-correlograms.

I observe that the variables “CC”, “Condition”, “Competition” and “Return” have no autocorrelation tendency in their data. The degree of similarity between the times series and their lagged version are under 0,3.

We see a lot of variables with a declining tendency in their auto-correlogram. It means the autocorrelation decline when the number of lags grows.

We have to notice the fact that some variables have year-fixed effect in our assumptions and that can obviously influence their autocorrelation test. In fact, “Gini”, “Dependency”, “Education” and “Size” are variables with this type of data and we see a high autocorrelation in their first four lags.

The ACF test is also useful to figure out the order (q) of a Moving Average (MA) model. This model assumes that the current value is dependent on the error terms including the current error. And because error terms are random, there is no linear relationship between the current value and the error terms.

6.4. Stationarity

I continue the analysis by checking the stationary of the variables. Checking the stationarity is useful to test if the predictions will be consistent in the future. Stationary variables have a stable means and a stable variance over a period which is essential to forecast. Stationarity check is also useful to correct some problem of seasonality or trend.

To check the stationarity, we perform a Augmented Dicker Fuller test which will illustrate the phenomenon of stationarity. The null hypothesis of this test is that the tested variable is non-stationary. After performing this test, we look at the p-value and we want that this value is smaller than 5% to reject the null hypothesis and consider the variable as stationary. A result of two of these tests is shown in *Figure 3*.

Figure 3:

```
> apply(Set,2,adf.test)
$CC

	Augmented Dickey-Fuller Test

data:  newX[, i]
Dickey-Fuller = -3.1423, Lag order = 3, p-value = 0.118
alternative hypothesis: stationary


$Income

	Augmented Dickey-Fuller Test

data:  newX[, i]
Dickey-Fuller = -0.47263, Lag order = 3, p-value = 0.9793
alternative hypothesis: stationary
```

I notice that only the variable “Dependency” is stationary at a 5% level (stationary at a level or integrated of order 0). Consequently, we must take the first difference of the other variables, check again the stationarity and repeat this check until all the variables are stationary. After the calculation of the first difference, the variables “CC”, “Gini”, “CPI”, “Condition”, Competition”, “Return”, “Size”, “Sentiment” is stationary at a 5% level. We perform a second differentiation and obtain all the variables stationary. Indeed, all the p-values are less than 1% except for variable “GDP” (p-value is equal to 1,83%).

The consequence of these two steps of differentiation is length of the dataset, in fact, by differentiation, the data lost 2 observations. Now the period’s dataset run from 2007 Q3 to 2018 Q4.

6.5.Vector Auto Regression (VAR)

As explain in section 5.2.1., VAR model has shown some ability to forecast economic variables. In this section we explain how we choose the parameters and how to model a VAR model.

6.5.1. Short theory on VAR model

In a Vector Auto Regression each variable is a linear function of the past values of itself and the past values of all the other variables. For example, the equation of a multivariate VAR(1) is $y_1(t) = \alpha_1 + \beta_1 y_1(t-1) + \beta_2 y_2(t-1) + e_1(t-1)$ where $y_1(t)$ depends on the first lag of $y_1\ and\ y_2$. The estimation of a VAR is an equation-by-equation OLS (Ordinary Least Squares) where OLS are run on each equation. The VAR model is a kind of AR model where the multiple variable terms like y(t) = α + β*y(t-1) + e are accommodated by using vectors in this way:

$$\begin{matrix} y_1(t) \\ y_2(t) \end{matrix} = \begin{matrix} \alpha_1 \\ \alpha_2 \end{matrix} + \begin{matrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{matrix} * \begin{matrix} y_1(t-1) \\ y_2(t-1) \end{matrix} + \begin{matrix} e_1(t) \\ e_2(t) \end{matrix}$$

A multivariate VAR of order p can be represented like this:

$y_1(t) = \alpha + \beta_1 y(t-1) + .. + \beta_p y(t-p) + \varepsilon_t$

The term ε_t in the equation represents multivariate vector white noise. It means that ε_t should be a continuous random vector that satisfies the condition of the null expected value for the error vector and the condition that expected value of ε_t and ε_t' is equal to the standard deviation of the series (Lütkepohl, 2005).

We test a VAR model because it can understand and use the relationship between several variables. It is useful for describing the dynamic behavior of macroeconomic data. However, to use VAR model we must use the stationary time series (Aishwarya, 2018).

6.5.2. VAR (2) implementation

I started to make my series stationary, and we looked for the p order to use in my model. To make all the data stationary we performed an Augmented Dicker Fuller test, and we differentiated my data once or twice if necessary. After differentiating, my time series was stationary and ready to be used.

Data not differentiated	Dependency
Data differentiated once	CC, Gini, CPI, Condition, Return, Size
Data differentiated twice	Income, Unempl, IR, Competition, Education, Sentiment
Data differentiated three times	GDP

To find the order p that we used for my model we used the “VARselect” formula to see which lag seems to provide the smaller information criteria. Following this calculation, the number of lag p tends to be 3 but one information criterion gives a lag 2.

First, we tried the equation with all the variables and a p-value equals to 2. It means that in our equation there are each variable lagged once and each variable lagged twice.

The equation FirstVAR (2) looks like:

CC_t= -2,541 **- 0,649 CC_{t-1} - 0,095 $Income_{t-1}$** - 0,019 GDP_{t-1} - 0,172 $Unempl_{t-1}$ - 0,096 IR_{t-1} - 9,301 $Gini_{t-1}$ + 0,032 CPI_{t-1} **- 0,294 $Condition_{t-1}$** + 0,078 $Competition_{t-1}$ - 0,005 $Return_{t-1}$ **- 0,671 $Dependency_{t-1}$** - 0,014 $Edyucation_{t-1}$ + 5,902 $Size_{t-1}$ + 0,004 $Sentiment_{t-1}$ **- 0,438 CC_{t-2} - 0,079 $Income_{t-2}$** - 0,009 GDP_{t-2} - 0,049 $Unempl_{t-2}$ - 0,012 IR_{t-2} - 31,925 $Gini_{t-2}$ - 0,083 CPI_{t-2} - 0,08 $Condition_{t-2}$ + 0,086 $Competition_{t-2}$ **-0,012 $Return_{t-2}$ + 0,717 $Dependency_{t-2}$** - 0,126 $Edyucation_{t-2}$ + 49,895 $Size_{t-2}$ **- 0,029 $Sentiment_{t-2}$**

I observed (*Appendix 9*) the number of significant variables (5 significant variables at the 5% level and 4 at the 10% level, there are highlighted in green). It means that nine lagged variables really contribute to estimate my model. For the other variables, the null hypothesis can't be rejected, and we can think that these variables don't help to model the variable "CC".

The R-squared is equal to 75,55% but the adjusted R-squared is quite weak with a level of 35,27%. The most relevant information we must look at is the adjusted R-squared because it is a corrected goodness-of-fit measure for linear models, and it shows the model accuracy. Adjusted R-squared identifies the percentage of variance in the target field that is explained by the inputs. This measure can go upper if the non-significant variables are removed out of the model. The formula of this measure is: $\frac{Residual\ Mean\ Square\ Error}{Total\ Mean\ Square\ Error} - 1$. The adjusted R-squared take into account the number of additional variables that are included into the model.

The residual standard error is equal to 0,2289.

Then we tested to remove some non-significant variables out of the model FirstVAR (2). We removed variables progressively and observed the changes in adjusted R-squared. Because of the big correlation between GDP and IR, we tried to remove one and then the other variable. Consequently, we found better characteristic for the model with GDP. So, we decided to remove IR to lower the correlation in my model. We also removed fixed-year variables as Gini, Education or Size because their p-value is quite high and because of their nature in this kind of model. In fact, fixed-year variables are not easy to interpret in a VAR model because their figures are fixed for one year (4 quarters) and the lag used is 2.

Finally, we observed (*Appendix 10*) after removing the sus-mentioned variables that there are nine significant variables, a higher adjusted R-squared (44,69%) and a lower residual standard error (0,2116).

The equation SecondVAR (2) looks like:

CC_t= -1,26 - **0,624** $\boldsymbol{CC_{t-1}}$ - **0,05** $\boldsymbol{Income_{t-1}}$ - 0,009 GDP_{t-1} - **0,17** $\boldsymbol{Unempl_{t-1}}$ + 0,018 CPI_{t-1} - **0,19** $\boldsymbol{Condition_{t-1}}$ + 0,072 $Competition_{t-1}$ - 0,008 $Return_{t-1}$ - **0,44** $\boldsymbol{Dependency_{t-1}}$ + 0,01 $Sentiment_{t-1}$ - **0,397** $\boldsymbol{CC_{t-2}}$ - **0,045** $\boldsymbol{Income_{t-2}}$ - 0,01 GDP_{t-2} - 0,039 $Unempl_{t-2}$ - 0,09 CPI_{t-2} - 0,07 $Condition_{t-2}$ + 0,068 $Competition_{t-2}$ -0,009 $Return_{t-2}$ + **0,46** $\boldsymbol{Dependency_{t-2}}$ - **0,02** $\boldsymbol{Sentiment_{t-2}}$

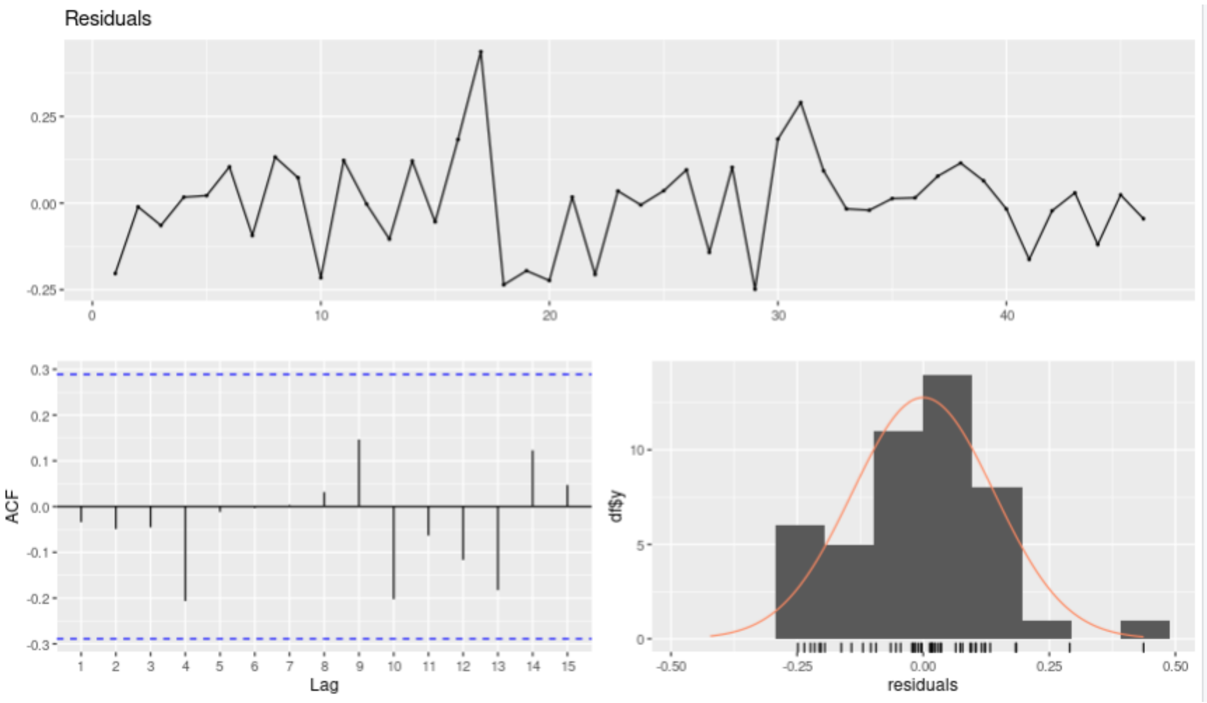
6.5.3. VAR (2) Analysis

To analyze the error of the FirstVAR and check if the residuals are white noise, we checked residuals on R with three plots of residuals (*Figure 4*). The first upper plot shows the chart of the residuals over the 46 observations.

The second left-hand plot shows the ACF plot of the residuals, and we observed that there is no significant autocorrelation between the residuals for the 15th first lags.

The right-hand plot represents the distribution function of the residuals in function of the values of "CC". The graph seems to be distributed following a kind of normal distribution (indicated with the orange line). We also verified the normal distribution of the residuals by doing a Shapiro normality test which gives a p-value of 0,087. The null hypothesis of this test is that the residuals are normally distributed. Here, we can't reject H0 at a confidence level of 5%. However, we can reject this null hypothesis at a 10% level of confidence. In other words, the residuals are normally distributed at the 5% level of confidence.

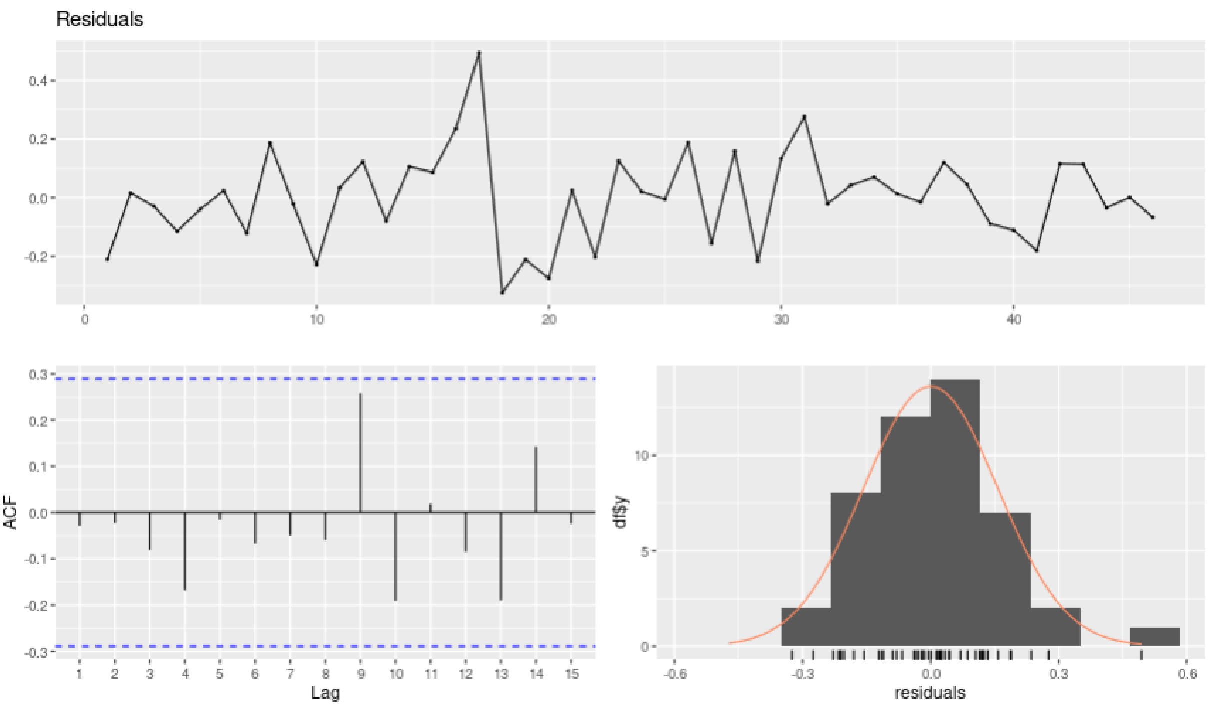
Figure 4:




```
> shapiro.test(residuals(FirstVAR[["varresult"]][["DCC"]]))

	Shapiro-Wilk normality test

data:  residuals(FirstVAR[["varresult"]][["DCC"]])
W = 0.95694, p-value = 0.08715
```

Then we also checked the residuals for the SecondVAR (2) model (*Figure 5*). We noticed that there is no significant autocorrelation in the residuals in the ACF plot. Moreover, the distribution of the residuals of this model is a normal distribution as shown with the Shapiro-Wilk normality test.

Figure 5:




```
> shapiro.test(residuals(SecondVAR[["varresult"]][["DCC"]]))

        Shapiro-Wilk normality test

data:  residuals(SecondVAR[["varresult"]][["DCC"]])
W = 0.97542, p-value = 0.4329
```

To summarize, these two models show that their residuals are normally distributed. It means that their mean is very close to zero and the variance is constant. This is a good point in the analysis of the model and in order to forecast the consumer credit. However, we performed a test to verify that the residuals are white noise and this test revealed that the two models are not. The test of Box-Pierce verifies if the residuals are independently and identically distributed. The null hypothesis suggest that they are independently distributed and thus that there is no autocorrelation in the residuals. For the two models, the results (*Appendix 22*) show that the residuals are not white noise. This point is problematic in order to build strong forecast.

6.5.4. VAR (3) implementation

I tried to run the same model with three lags but there were too many variables to estimate all the results. We observed that we must keep only nine variables added to the variable 'CC'. We decided to remove the variables Gini, Education and Size because we observed the high p-values in the previous model with two lags. The Granger causality test (Section 6.1.) also identified these three variables as weak predictors. Moreover, these variables are fixed-year variables which is not very relevant in this model.

I also observed something unexpected with the previous models because the p-values of "IR" are high and maybe it is not a good predictor. We continued to remove it out of the VAR (3) model and preferred to keep GDP as potential predictor. From the previous models we kept in mind that these two variables (GDP and IR) could create correlation inside a model. It is why we made the same selection as previously.

I tried to run the model with and without the variable CPI because this variable was not significant in the previous model. The following model is the one without CPI.

The equation for this ThirdVAR (3) model is:

$CC_t = - 0{,}92 - \mathbf{0{,}44\ CC_{t-1}} - \mathbf{0{,}09\ Income_{t-1}} - 0{,}01\ GDP_{t-1} - 0{,}086\ Unempl_{t-1} - \mathbf{0{,}28\ Condition_{t-1}} + 0{,}02\ Competition_{t-1} - 0{,}004\ Return_{t-1} - 0{,}14\ Dependency_{t-1} + 0{,}016\ Sentiment_{t-1} - 0{,}55\ CC_{t-2} - 0{,}13\ Income_{t-2} - 0{,}005\ GDP_{t-2} + 0{,}115\ Unempl_{t-2} - 0{,}17\ Condition_{t-2} - 0{,}04\ Competition_{t-2} - 0{,}004\ Return_{t-2} + 0{,}18\ Dependency_{t-2} - 0{,}018\ Sentiment_{t-2} - \mathbf{0{,}22\ CC_{t-3}} - \mathbf{0{,}13\ Income_{t-3}} - 0{,}07\ GDP_{t-3} + 0{,}04\ Unempl_{t-3} - 0{,}04\ Condition_{t-3} - 0{,}12\ Competition_{t-3} - 0{,}003\ Return_{t-3} - 0{,}03\ Dependency_{t-3} + 0{,}011\ Sentiment_{t-3}$

I also noticed that this model (*Appendix 11*) presents even less significant variables (only 5 in total). Only three different variables are significant: CC, Income and Condition. The adjusted R-squared is still weak (43,25%).

I tried to replace the variable GDP by the variable CPI because they have the same kind of p-value in the previous VAR (2) model. We kept the other variables unchanged and ran the model again.

The resulting equation is:

CC_t= - 0,3 - **0,52** $\boldsymbol{CC_{t-1}}$ - 0,03 $Income_{t-1}$ + 0,07 CPI_{t-1} - 0,01 $Unempl_{t-1}$ **- 0,24** $\boldsymbol{Condition_{t-1}}$ - 0,01 $Competition_{t-1}$ - 0,007 $Return_{t-1}$ - 0,28 $Dependency_{t-1}$ + 0,015 $Sentiment_{t-1}$ **- 0,59** $\boldsymbol{CC_{t-2}}$ - 0,046 $Income_{t-2}$ - 0,13 CPI_{t-2} + 0,05 $Unempl_{t-2}$ - 0,22 $Condition_{t-2}$ - 0,08 $Competition_{t-2}$ -0,007 $Return_{t-2}$ + 0,25 $Dependency_{t-2}$ **- 0,02** $\boldsymbol{Sentiment_{t-2}}$ - 0,13 CC_{t-3} – 0,01 $Income_{t-3}$ + 0,01 CPI_{t-3} + 0,04 $Unempl_{t-3}$ + 0,02 $Condition_{t-3}$ - 0,12 $Competition_{t-3}$ -0,01 $Return_{t-3}$ + 0,03 $Dependency_{t-3}$ + 0,006 $Sentiment_{t-3}$

I always observe (*Appendix 12*) a weak significant variable (4 in total). This weak number of significant variables is worrying because if the explanatory power of my model is weak, the prediction power could be weak too. The adjusted R-squared is higher with 44,94% which is good point for the model accuracy.

As we did previously, we tried to remove out of the model the variables that are not significant. First, we kept only 6 variables added to the dependent variable CC, so we kept Income, CPI, GDP, Condition, Return, Sentiment. The variables that are not included are Unempl, Competition and Dependency added to the other variables already removed. We chose to remove them because they had the higher p-value.

The result (*Appendix 13*) of this test is better with a better adjusted R-squared and always lower residuals (residual standard error equals to 0,1985).

Here is the equation of FifthVAR (3):

CC_t= 0,01 **- 0,63** $\boldsymbol{CC_{t-1}}$ + 0,006 $Income_{t-1}$ **+ 0,1** $\boldsymbol{CPI_{t-1}}$ - 0,005 GDP_{t-1} - 0,15 $Condition_{t-1}$ **- 0,01** $\boldsymbol{Return_{t-1}}$ + 0,01 $Sentiment_{t-1}$ **- 0,62** $\boldsymbol{CC_{t-2}}$ + 0,06 $Income_{t-2}$ **- 0,14** $\boldsymbol{CPI_{t-2}}$ + 0,03 GDP_{t-2} - 0,15 $Condition_{t-2}$ - 0,01 $Return_{t-2}$ - 0,01 $Sentiment_{t-2}$ - 0,18 CC_{t-3} + 0,03 $Income_{t-3}$ - 0,01 CPI_{t-3} + 0,02 GDP_{t-3} **+ 0,21** $\boldsymbol{Condition_{t-3}}$ **- 0,01** $\boldsymbol{Return_{t-3}}$ - 0,001 $Sentiment_{t-3}$

In fact, we observed an adjusted R-squared that equals to 53,45% and an R-squared quite similar to before removing variables (76,19%). We also noticed that the number of significant variables grew a little with 7 significant variables. We continued to try to remove the variables GDP because it had a high p-value could always create problems of correlation.

Here is the equation of the new SixthVAR (3) without GDP:

$CC_t =$ 0,01 **- 0,64** $\boldsymbol{CC_{t-1}}$ - 0,02 $Income_{t-1}$ **+ 0,1** $\boldsymbol{CPI_{t-1}}$ **- 0,17** $\boldsymbol{Condition_{t-1}}$ **- 0,01** $\boldsymbol{Return_{t-1}}$ + 0,01 $Sentiment_{t-1}$ **- 0,63** $\boldsymbol{CC_{t-2}}$ - 0,006 $Income_{t-2}$ **- 0,15** $\boldsymbol{CPI_{t-2}}$ - 0,18 $Condition_{t-2}$ **- 0,01** $\boldsymbol{Return_{t-2}}$ - 0,01 $Sentiment_{t-2}$ - 0,19 CC_{t-3} + 0,01 $Income_{t-3}$ - 0,002 CPI_{t-3} **+ 0,2** $\boldsymbol{Condition_{t-3}}$ **- 0,01** $\boldsymbol{Return_{t-3}}$ + 0,001 $Sentiment_{t-3}$

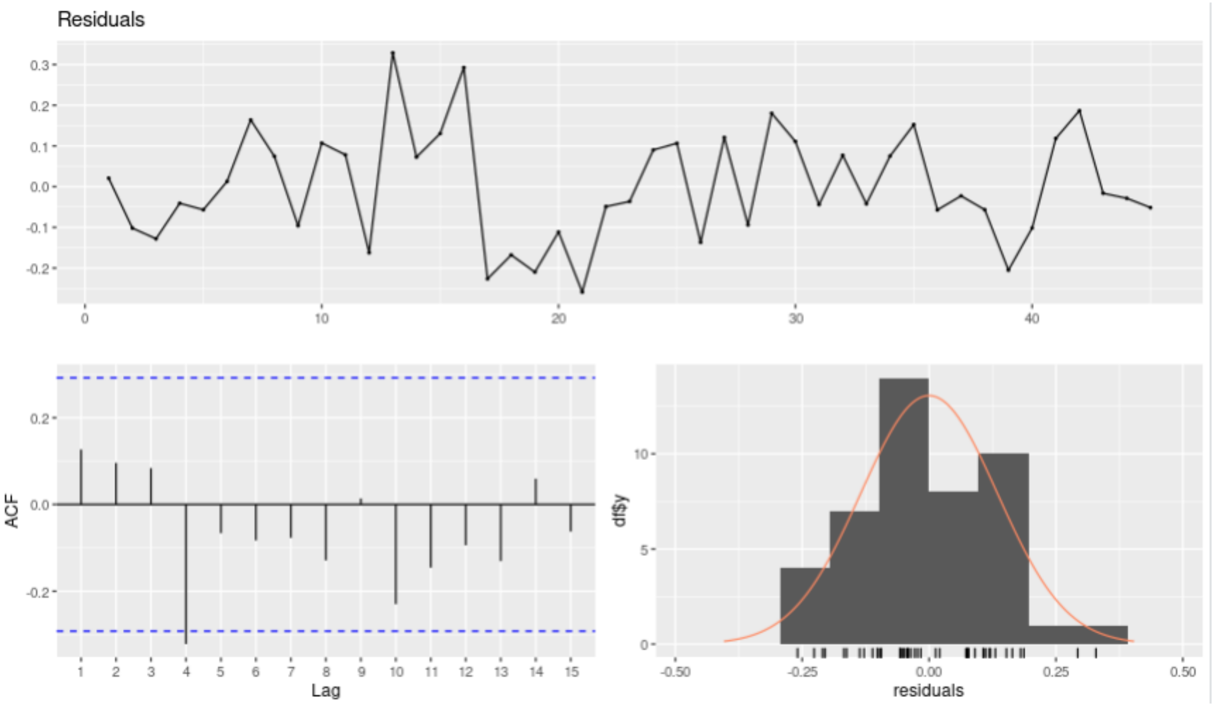
By removing the variable GDP (*Appendix 14*), the adjusted R-squared continued to grow (57,61%) and the R-squared remain approximately stable (75,35%). We also noted that there are nine variables that are significant at the 10% level and still six at the 5% level of confidence.

When we removed the variable Sentiment or Income, the adjusted R-squared and the R-squared both went down so we decided to keep these variables in my model. Despite their significance is not proved in this model, we saw previously that these variables could help to explain the dependent variable.

6.5.5. VAR (3) analysis

As previously, we must check the residuals to evaluate the ability of the models to forecast consumer credit. In *Figure 6*, the analysis of residuals of ThirdVAR model are shown. We saw that for lag 4 there is some significant autocorrelation. Regarding the normality of residual's distribution, it seems that the distribution follows a normal distribution. The Shapiro-Wilk normality test shows that the residuals are normally distributed.

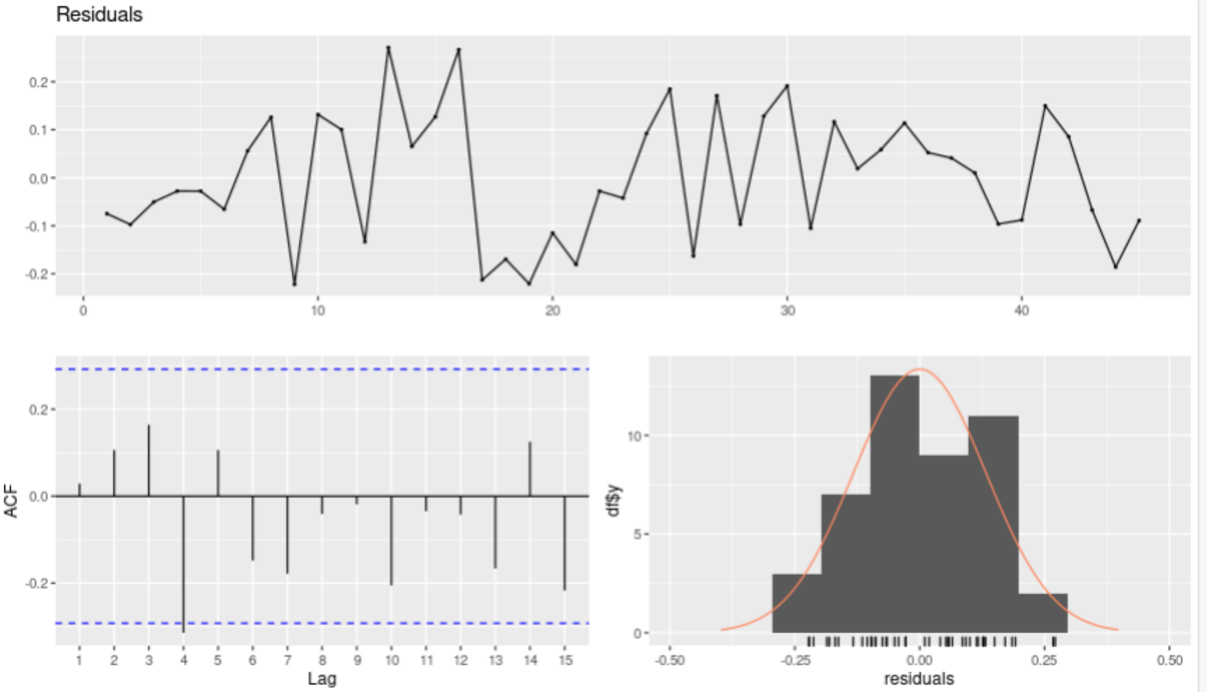
Figure 6:




```
> shapiro.test(residuals(ThirdVAR[["varresult"]][["DCC"]]))

	Shapiro-Wilk normality test

data:  residuals(ThirdVAR[["varresult"]][["DCC"]])
W = 0.97821, p-value = 0.5488
```

Then, we checked the residuals of the FourVAR (*Figure 7*). We noted the significant autocorrelation of residuals at lag 4. This significant autocorrelation announced that the residuals are not independently distributed. We observed and confirmed the normality of residuals with the high p-value of normality test.

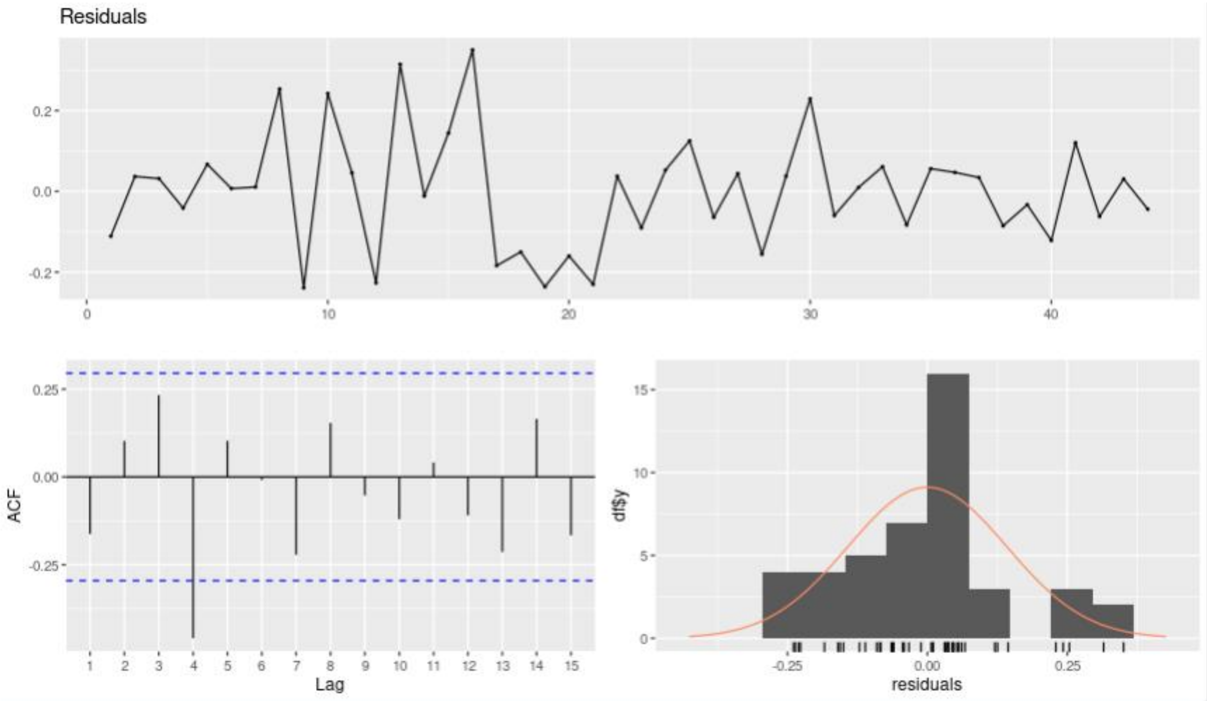
Figure 7:




```
> shapiro.test(residuals(FourVAR[["varresult"]][["DCC"]]))

	Shapiro-Wilk normality test

data:  residuals(FourVAR[["varresult"]][["DCC"]])
W = 0.96751, p-value = 0.2347
```

The residuals of model FifthVAR are serially correlated at lag 4 (*Figure 8*). Despite the kurtosis that we noted on the graph, the distribution of error terms is normally distributed.

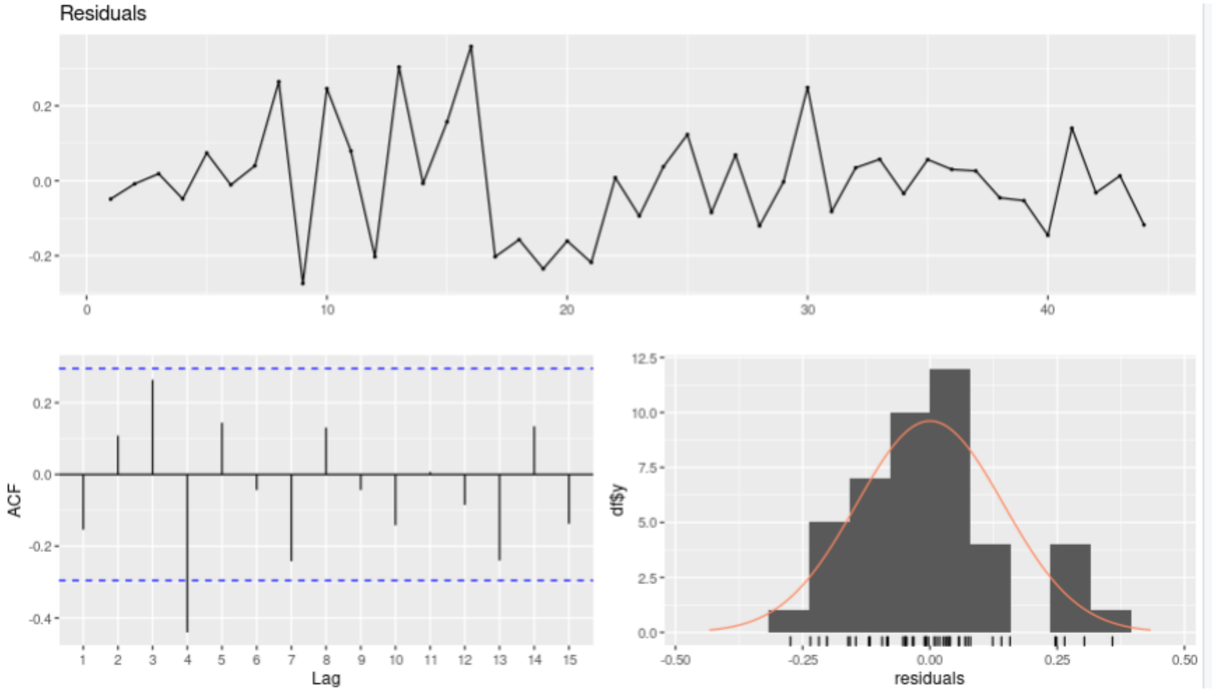
Figure 8:




```
> shapiro.test(residuals(FifthVAR[["varresult"]][["DCC"]]))

	Shapiro-Wilk normality test

data:  residuals(FifthVAR[["varresult"]][["DCC"]])
W = 0.96003, p-value = 0.1305
```

Finally, we checked the residuals of the SixthVAR (*Figure 9*). The results show that residuals are serially correlated at lag 4. The distribution of error terms for this model is normally distributed as shown with the normality test.

Figure 9:




```
> shapiro.test(residuals(SixthVAR[["varresult"]][["DCC"]]))

	Shapiro-Wilk normality test

data:  residuals(SixthVAR[["varresult"]][["DCC"]])
W = 0.97016, p-value = 0.3067
```

To summarize, the models tested in this section are all normally distributed. This information is good to know because we know that the mean of the error terms is close to zero which means that model capture well the information to explain the consumer credit. Nevertheless, we noticed that the residuals are serially correlated. Indeed, the Box-Pierce tests (*Appendix 22*) all revealed that the residuals are not independently distributed. This point is problematic to construct a robust model.

6.5.6. VAR forecasting

Finally, we wanted to forecast data over my testing period. From 2019 Q1 to 2020 Q4 we have the real data, and we compared this data with the forecasted data given by the model (*Appendix 15*).

we noted that ThirdVAR model and FifthVAR model have quite irrelevant figures. Let's remember that we observed here the forecasted figures for the variable "DCC" meaning that the forecasted values are values that are differentiated. To obtain the amount of new consumer credit in euro, we must add the differentiated figures to the previous amount of new consumer credit. *Figure 10* represents the forecasted figures obtained for FirstVAR model and FifthVAR model.

Figure 10:

			FifthVAR model	= 1,030 + (-0,28992855)	FirstVAR model	
	Real figures	Differentiated real figures	Forecast of differentiated figures	Forecast figures based on year t-1	Forecast of differentiated figures	Forecast figures based on year t-1
2018Q4	0,887					
2019Q1	1,224	0,33632754	0,43554321	1,323	0,263935247	1,151
2019Q2	1,030	- 0,19344726	-0,06233865	1,161	-0,213841042	1,010
2019Q3	0,893	- 0,13699922	-0,28992855	0,740	0,004762871	1,035
2019Q4	0,836	- 0,05774148	-0,09828863	0,795	-0,009824547	0,883
2020Q1	0,900	0,06463746	0,34950654	1,185	-0,027020341	0,809
2020Q2	0,994	0,09390079	0,2161076	1,116	0,154822246	1,055
2020Q3	1,032	0,03832891	-0,38644782	0,608	-0,154466786	0,840
2020Q4	0,849	- 0,18318049	-0,29347995	0,739	-0,075833115	0,957

In this table, we recalculated the real figures based on the previous quarter. We can already say that there are some differences in each forecasted data between real figures and forecasted figures. We also observed that the forecasted data go in the same direction as the real data which is a good point. We must analyze how the forecasted data fit the real data to have a good opinion of my model.

For each series of forecast, we calculated the root mean squared error. The formula to calculate it is: $\sqrt{(mean(forecasted\ data - observed\ data)^2)}$

Name of model	**RMSE**
FirstVAR (2)	0.1053893
SecondVAR (2)	0.1312606
ThirdVAR (3)	2.158985
FourVAR (3)	0.2379964
FifthVAR(3)	1.529254
SixthVAR (3)	0.2063828

I observed that the lower RMSE are those for the FirstVAR and SecondVAR models. As expected, the RMSE of ThirdVAR and FifthVAR models are very high in comparison with the other models.

6.5.7. Summary of VAR forecasting

In this section, we summarized all the information given by the implementation and analysis. The table below shows the different model, the R-squared and the adjusted R-squared, the RMSE are represented in the fourth column and in the last, there are the significant variables with their impact on the consumer credit (positive or negative).

Name	**R-squared**	**Adjusted R-squared**	**RMSE**	**Significant variables**
FirstVAR (2)	75,55%	35,27%	0,1053893	- : CC_{t-1}, $Income_{t-1}$, $Condition_{t-1}$, $Dependency_{t-1}$, CC_{t-2}, $Income_{t-2}$, $Return_{t-2}$, $Sentiment_{t-2}$ +: $Dependency_{t-2}$
SecondVAR (2)	69,27%	44,69%	0,1312606	- : CC_{t-1}, $Income_{t-1}$, $Unempl_{t-1}$, $Condition_{t-1}$, $Dependency_{t-1}$, CC_{t-2}, $Income_{t-2}$, $Sentiment_{t-2}$ +: $Dependency_{t-2}$
ThirdVAR (3)	78,07%	43,25%	2,158985	- : CC_{t-1}, $Income_{t-1}$, $Condition_{t-1}$, CC_{t-2}, $Income_{t-2}$
FourVAR (3)	78,73%	44,94%	0,2379964	- : CC_{t-1}, $Condition_{t-1}$, CC_{t-2}, $Sentiment_{t-2}$
FifthVAR (3)	76,19%	53,45%	1,529554	-: CC_{t-1}, $Return_{t-1}$, CC_{t-2}, CPI_{t-2}, $Return_{t-3}$ +: CPI_{t-1}, $Condition_{t-3}$
SixthVAR (3)	75,35%	57,61	0,2063828	-: CC_{t-1}, $Condition_{t-1}$, $Return_{t-1}$, CC_{t-2}, CPI_{t-2}, $Return_{t-2}$, $Return_{t-3}$ +: CPI_{t-1}, $Condition_{t-3}$

I can already reject the Third and the Fifth -VAR models because of their too high root mean squared error. We can also reject the FourVAR model because it approximately has the same adjusted R-squared as the SecondVAR model and moreover this model has less significant variables and a higher RMSE.

I can now envisage the three last model that have approximately all the same number of significant variables. FirstVAR model has the lowest RMSE but a weak adjusted R-squared. SecondVAR has a better adjusted R-squared but has a quite higher RMSE. Finally, SixthVAR has a RMSE twice as big as FirstVAR but has approximately an adjusted R-squared twice as big as FirstVAR.

6.6. Vector Error Correction Model (VECM)

In this section, we test the ability of a VECM to explain and to forecast the consumer credit. As explained in section 5.2.1., this sort of model could improve the result for my goal in this study. We make a quick review of the theory about VECM, we explain how to implement this model and how to analyze the different result obtained.

6.6.1. Short theory on VECM

In our database, we show that some variables are stationary and other one not stationary. It means that some variables have a trend, or a seasonality and others have none of them. To deal with this problem, an econometric concept called cointegration can resolve this problem. Cointegration shows that sometimes, between variables that are different in terms of stationarity, there is some common element. Cointegration is the representation of a (long-term) comoving between very different variables. A model of non-stationary variables can result in a stationary model thanks to the concept of cointegration. There is sometimes a linear combination between variables that tend to move together and tends to be bounded together (Lütkepohl, 2005).

For example, when x_t and y_t are non-stationary but there exists a θ such that $y_t - \theta x_t$ is stationary, x_t and y_t are cointegrated. In other words, although each time series is non-stationary, their linear combination may be stationary. In this kind of model, the short-run dynamic is augmented with the long-run relationship.

The variables used in a VECM must be integrated of order 1 or weakly stationary after one differentiation. It is a condition to find a stationary relationship between two non-stationary variables.

In an error correction model, the long-term path or long-term trend is identified between each series. Even if the variables have a common trend or path, there are some deviations to this trend and the main goal of the VECM is to correct these variations (Lütkepohl, 2005). A VECM consists of a VAR model with stationary growth variables and error-correcting equations with non-stationary level variables. This model describes a short-run dynamic by vector autoregression as well as a long-run equilibrium by error correcting equations.

A conventional Vector Error Correction Model looks like:

$$\Delta y_t = \beta_0 + \sum_{i=1}^{n} \beta_i \Delta y_{t-i} + \sum_{i=0}^{n} \delta_i \Delta x_{t-i} + \varphi z_{t-1} + \mu_t$$

Where z is the Error Correction Term and is the residuals from this long-run cointegrating regression $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$.

6.6.2. VECM implementation

To realize this model, first, we must measure the level of cointegration that there is between all my variables. To this end, we realize a Johansen Cointegration test to see at which level my variables have a common path. Before we must identify the level of lag for an autoregressive model like VAR to see the information criterion that will identify the ideal lag to model my equation.

The result of the "VARselect" formula (*Figure 11*) shows that lag 4 is the good order to use in the cointegration formula. It means that the number of lags to use in a Vector Auto-Regression is p equals to 4.

Figure 11:

```
> Data<-cbind(CC,Income,CPI,Gini,Unempl,IR,Cond,Return,Dep,Sent)
> lag<-VARselect(Data)
> lag$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     4      4      4      4
```

After identifying the order p for my whole non-stationary database, we can perform the calculation of the cointegration test. This Johansen test is either a trace test or a maximum eigenvalue test. Both approaches show the level of cointegration, in other words, how many cointegrating equations there are.

For this test a maximum of 10 variables is accepted by the econometric tool. It is why we made a first choice of ten variables based on the Granger Causality tests and based on the stationarity. We rejected the variables "GDP", "Competition", "Education" and "Size". The variable "GDP" causes always a problem because of the problem of the very low stationarity of this variable if integration is at level 1 or 2. In fact, like explained before, my variables must be stationary or weakly stationary after the first difference to fit with this kind of VEC model.

Figure 12 shows that at the level of 5%, there is a cointegration relationship in my database of 5 because the statistic value of the test at a level 6 is 23,41 and the corresponding critical value is 27,14. At level 6, there is thus no more cointegration and there is cointegration until level 5.

Figure 12:

```
> cointegration<-ca.jo(Gross_Data10, K = 3)
> summary(cointegration)

######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
 [1] 0.914944545 0.872518279 0.747273426 0.720647120 0.675605842 0.626523279 0.405555221 0.258215923 0.108137144
[10] 0.002044575

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 9 |   0.09  6.50  8.18 11.65
r <= 8 |   5.15 12.91 14.90 19.19
r <= 7 |  13.44 18.90 21.07 25.75
r <= 6 |  23.41 24.78 27.14 32.14
r <= 5 |  44.32 30.84 33.32 38.78
r <= 4 |  50.66 36.25 39.43 44.59
r <= 3 |  57.39 42.06 44.91 51.30
r <= 2 |  61.90 48.43 51.07 57.07
r <= 1 |  92.69 54.01 57.00 63.37
r = 0  | 110.90 59.00 62.42 68.61
```

After finding the number of cointegrating equation there are in my database, we calculated the VECM using a lag order p (equals to 3 and not 4 to take into account the order of integration to be stationary) and the cointegration level is 5. In the *Appendix 16*, the cointegrating equations are observable. In this equation, there are the long run cointegrating relationships but no other parameters which are obtained later. Bellows here is an example of a cointegrating equation. In this example, CC has a negative long-run relationship with Income, CPI, Condition, Return, Dependency and Sentiment. CC has a small but positive relationship with Unempl and IR.

$$
\begin{aligned}
ECT_{t-1} = 1 * CC_{t-1} - 1{,}11e^{-16} * Income_{t-1} + 1{,}66e^{-16} * Unempl_{t-1} + 2{,}49e^{-16} * IR_{t-1} \\
- 0.0066 * CPI_{t-1} - 0{,}876 * Condition_{t-1} - 0{,}0069 * Return_{t-1} - 0{,}043 \\
* Dependency_{t-1} - 0{,}0069 * Sentiment_{t-1}
\end{aligned}
$$

This Error Correction Term relates to the fact that last period deviation from long-run equilibrium (the error) influences the short-run dynamics of the dependent variable.

I also observe (*Appendix 17*) the equation of VECM in which there are five error correction terms (ECT) to correct the deviations of the variables to have a long-run common trend and path between all the variables of the model. We also see the intercept and each variable in their three lags.

I also see that a quite poor number of variables is statistically significant (highlighted in green) for the equation of "CC" which is one of my preoccupations here.

$$\begin{aligned} CC_t = & \mathbf{-2,63\ ECT_{t-1}} - 0{,}18\ ECT_{t-2} - 0{,}12\ ECT_{t-3} \mathbf{-0,74\ ECT_{t-4}} \\ & + 3{,}4\ ECT_{t-5} \mathbf{+32,52} + 0{,}92\ CC_{t-1} + 0{,}15\ Income_{t-1} + 0{,}27\ Unempl_{t-1} \\ & + 0{,}75 \mathbf{+0,76\ IR_{t-1}} + 0{,}9\ Gini_{t-1} + 0{,}03\ CPI_{t-1} - 0{,}32 \\ & + 0{,}03\ Return_{t-1} + 0{,}24\ Dep_{t-1} - 0{,}01\ Sent_{t-1} + 0{,}39\ CC_{t-2} \\ & \mathbf{+0,17\ Income_{t-2}} + 0{,}13\ Unempl_{t-2} + 0{,}32\ IR_{t-2} + 18{,}48\ Gini_{t-2} \\ & + 0{,}05\ CPI_{t-2} - 0{,}46\ Cond_{t-2} + 0{,}02\ Return_{t-2} - 0{,}62\ Dep_{t-2} \\ & + 0{,}01\ Sentiment_{t-2} + 0{,}1\ CC_{t-3} \mathbf{+0,13\ Income_{t-3}} + 0{,}09\ Unempl_{t-3} \\ & + 0{,}33\ IR_{t-3} + 0{,}07\ Gini_{t-3} - 0{,}05\ CPI_{t-3} - 0{,}32\ Cond_{t-3} \\ & + 0{,}003\ Return_{t-3} - 0{,}05\ Dep_{t-3} + 0{,}002\ Sent_{t-3} \end{aligned}$$

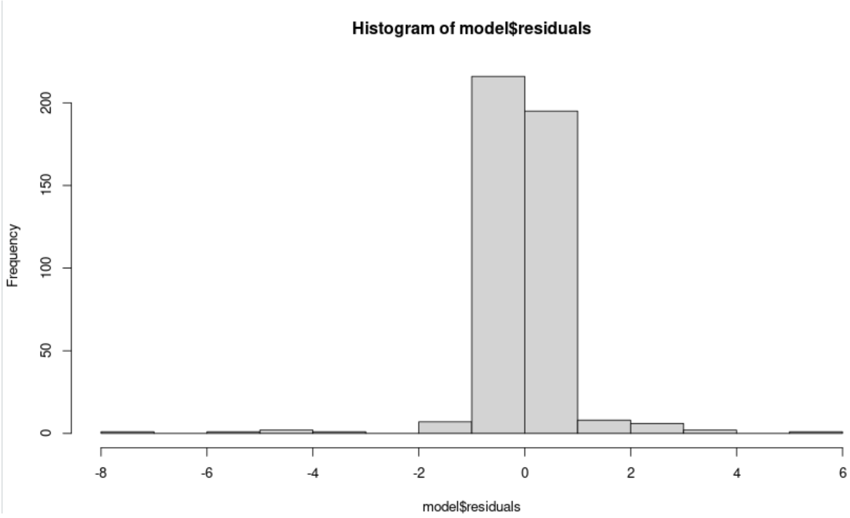
The coefficients of the ECT's are the speed of adjustment because it measures the speed at which dependent variable returns to equilibrium after a change in other variables. We noticed that the first and the fourth ECT's are significant meaning that there are long-run relationships. We always focused my attention on the first row "equation CC" and observed each coefficient (the first figure) and saw the different standard error (in brackets).

I tried to perform other models including variables Competition and removing variables like Gini or Dependency. The resulting models always have higher Akaike Information Criterion. Furthermore, these models didn't have higher number of significant variables.

6.6.3. VECM analysis

After observing the model, we plotted the histogram (*Figure 13*) of the residuals of this model, and we see that it didn't look like a normal distribution. Except the mean, which is near zero, the distribution is not equally distributed around the mean.

Figure 13:



I verified this sentiment by performing a Shapiro-Wilk normality test (*Figure 14*) and it confirmed that the residuals are not normally distributed because of the small p-value.

Figure 14:

```
> shapiro.test(model$residuals)

        Shapiro-Wilk normality test

data:  model$residuals
W = 0.50589, p-value < 2.2e-16
```

I also wanted to know if the residuals of the model are white noise. It is why we performed a Ljung-Box test (*Figure 15*). Ljung-Box test is similar as the Box-Pierce test performed before for the VAR model. The p-value of this test is very small so we could conclude that the residuals are not white noise for this model. Indeed, the null hypothesis is that the residuals are independently and identically distributed but here this hypothesis is rejected.

Figure 15:

```
> ljung.wge(resid_model)
Obs 0.02988623 0.1050648 0.1627946 -0.5297668 -0.1207597 -0.2025978 -0.3741727 0.178356 0.09897231 0.05829628 0.3250918 -0.02318145 -0.09533457 0.01914373 -0.
2039251 -0.07879723 0.1191419 -0.06533884 0.1109021 0.1238703 -0.1388307 0.1027045 -0.04899801 -0.2443948
$test
[1] "Ljung-Box test"

$K
[1] 24

$chi.square
[1] 52.35626

$df
[1] 24

$pval
[1] 0.0007028733
```

6.6.4. VECM forecast

Despite the poor chances to obtain a good model and despite the difficulty to perform diagnostic tests on a VEC model in R, we tried to forecast it to see if he fits well with the available data from 2019 to 2020 (8 observations). *Figure 16* provides the results of the forecasting process.

Figure 16:

```
> Model_fit<-predict(model,n.ahead = 8)
> summary(Model_fit)
       CC                Income          Unempl              IR             Gini             CPI             Cond              Return
 Min.   :-0.05935   Min.   :41.09   Min.   :-3.0585   Min.   :2.199   Min.   :0.2392   Min.   :5.301   Min.   :-1.69271   Min.   :-54.4257
 1st Qu.: 0.51540   1st Qu.:44.19   1st Qu.:-0.5788   1st Qu.:3.126   1st Qu.:0.2411   1st Qu.:6.086   1st Qu.:-0.71843   1st Qu.:-16.8341
 Median : 1.39910   Median :46.79   Median : 0.8712   Median :4.057   Median :0.2430   Median :7.263   Median :-0.11781   Median :  0.5938
 Mean   : 1.07087   Mean   :47.59   Mean   : 0.6861   Mean   :3.740   Mean   :0.2452   Mean   :7.541   Mean   :-0.01829   Mean   : -2.4578
 3rd Qu.: 1.61635   3rd Qu.:51.57   3rd Qu.: 2.1544   3rd Qu.:4.348   3rd Qu.:0.2469   3rd Qu.:9.280   3rd Qu.: 1.09151   3rd Qu.: 19.0323
 Max.   : 1.74214   Max.   :55.36   Max.   : 3.6220   Max.   :4.613   Max.   :0.2561   Max.   :9.902   Max.   : 1.26082   Max.   : 27.7907
      Dep             Sent
 Min.   :56.86   Min.   :-6.246
 1st Qu.:57.49   1st Qu.: 9.445
 Median :57.87   Median :18.073
 Mean   :57.73   Mean   :14.893
 3rd Qu.:58.08   3rd Qu.:21.091
 Max.   :58.27   Max.   :31.222
Connected to your session in progress, last started 2022-Jun-29 13:59:38 UTC (8 minutes ago)
> print(Model_fit)
           CC   Income     Unempl       IR      Gini      CPI        Cond     Return      Dep      Sent
49  1.69732903 41.08915  3.6219522 2.198949 0.2561320 5.300741 -1.13049102   1.377948 58.26988 20.849499
50  0.63115741 48.69245  2.6845771 3.170644 0.2537302 7.655033 -0.08523551  27.790694 58.01283 19.954778
51  1.29317669 42.37742  1.0073497 2.992790 0.2441309 9.901567 -0.15038375 -16.912328 57.89067 31.221889
52  0.16814060 51.36282  0.7350548 4.140995 0.2446073 9.279645  1.26081799 -16.807993 58.26198  4.142510
53  1.74213723 44.79212  1.9776405 3.972135 0.2406311 9.279856 -0.58107131 -54.425703 57.83971 11.211954
54  1.50501969 52.19846 -1.0611062 4.612794 0.2412517 6.152130  1.06666550  -0.190255 57.63576 -6.246185
55  1.58935649 44.88267 -0.4179705 4.553407 0.2418963 6.871492 -1.69271352  21.193122 56.85720 21.814637
56 -0.05935487 55.35974 -3.0584760 4.279597 0.2391648 5.888948  1.16605292  18.311986 57.06901 16.191895
```

I directly noticed that some figures are non-relevant (like the negative figure). This problem is probably due to the imperfect stationarity after one differentiation of some variables of my model.

I observed a large difference between the theoretical values of “CC” and those given by the forecast of the model. However, in *Appendix 18* the general tendency is good but the range between the two red lines is too big and maybe quite imprecise. We calculated the root mean squared error and the result is equal to 0,6219 which is quite high when comparing to previous models.

6.6.5. VECM Transformation to VAR

For many diagnostic tests in R, the formulas accept only objects from a “var” or a “vec2var” object. Consequently, we transformed the VECM in differences (with error correction terms) into a VAR in levels thanks to the "vec2var" formula. Thanks to this transformation, the long-run relationships (the cointegration relationships) are included into the VAR equation. The VAR equation includes thus the cointegration. From this "vec2var" formula, there results a Vector Auto-Regression which is stationary at level meaning that there is no differentiation needed.

Here is the equation of this VAR (3):

$$
\begin{aligned}
CC_t = {} & +0{,}07 - 0{,}2\, CC_{t-1} + 0{,}06\, Income_{t-1} + 0{,}01\, Unempl_{t-1} \boldsymbol{- 0{,}09\, IR_{t-1}} \\
& + 45{,}75\, Gini_{t-1} + 0{,}12\, CPI_{t-1} + 0{,}12\, CPI_{t-1} - 0{,}01\, Return_{t-1} \\
& - 3{,}9\, Dep_{t-1} - 0{,}005\, Sent_{t-1} - 0{,}23\, CC_{t-2} \boldsymbol{+ 0{,}04\, Income_{t-2}} \\
& + 0{,}02\, Unempl_{t-2} + 0{,}05\, IR_{t-2} - 25{,}16\, Gini_{t-2} - 0{,}17\, CPI_{t-2} \\
& - 0{,}10\, Cond_{t-2} - 0{,}004\, Return_{t-2} - 0{,}07\, Dep_{t-2} + 0{,}01\, Sentiment_{t-2} \\
& - 0{,}10\, CC_{t-3} \boldsymbol{+ 0{,}01\, Income_{t-3}} + 0{,}08\, Unempl_{t-3} - 0{,}01\, IR_{t-3} \\
& + 6{,}41\, Gini_{t-3} + 0{,}07\, CPI_{t-3} + 0{,}45\, Cond_{t-3} - 0{,}004\, Return_{t-3} \\
& + 0{,}30\, Dep_{t-3} + 0{,}007\, Sent_{t-3}
\end{aligned}
$$

In Appendix 20, we observed the results of the cointegrated VAR.

6.6.6. VAR analysis

For this model, we must calculate the R-squared manually. So, we calculated the total sun of squares (tss) which is equal to $\sum(true\ y - mean\ true\ y)^2$. We calculated the residual sum of squared errors (rss) which is equal to $\sum(true\ y - y\ observed)^2$. Then the R-squared is equal to $1 - \frac{rss}{tss}$. The R-squared of this model is equal to 68,31%.

The adjusted R-squared is calculated with this formula: adj R^2=$1 - \left[\left(\frac{n-1}{n-k-1}\right)\right] * (1 - R^2)$ where n is the number of observations (45) and k the number of independent variables (9). The adjusted R-squared is thus 60,16%.

For this model we could test the serial correlation (*Figure 17*). We performed a serial test to see if the residuals of my model are correlated over time. We noticed that the p-value is greater than 5%, so we concluded that this model is not serially correlated. Indeed, The null hypothesis of no autocorrelation in residuals can't be rejected at a level of 5%.

Figure 17:

```
> serial1<-serial.test(model1VAR, lags.pt = 16, type = "PT.asymptotic")
> serial1

        Portmanteau Test (asymptotic)

data:  Residuals of VAR object model1VAR
Chi-squared = 1394.7, df = 1310, p-value = 0.05114
```

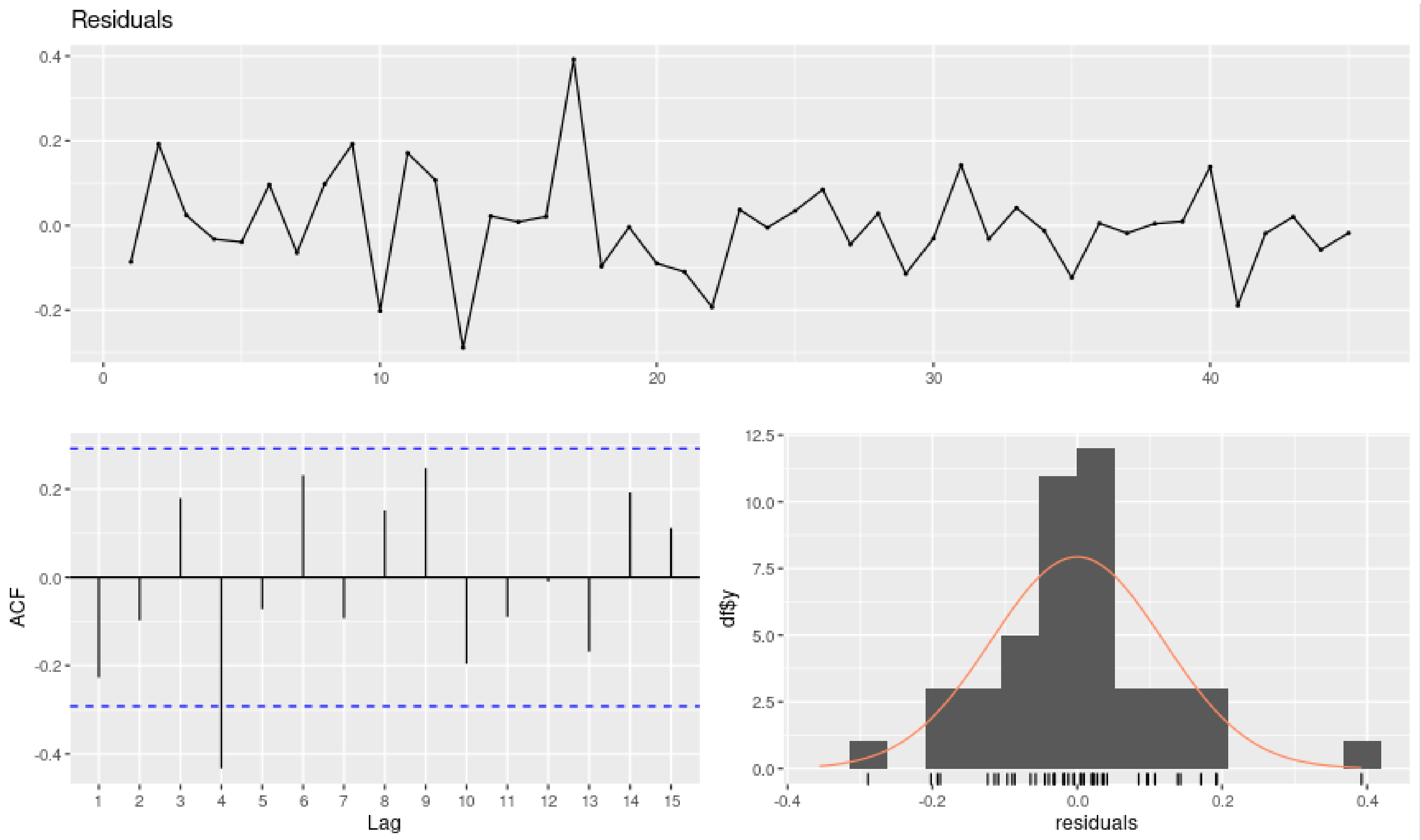
Then we checked the normality of the residuals of model1VAR. *Figure 18* shows that the model is normally distributed via the p-value of the Jarque-Bera test. In fact, p-value is greater than 5% so we can't reject the null hypothesis of normality. We noticed in the same sense that there is no skewness and no kurtosis.

Figure 18:

```
> Norm1<-normality.test(model1VAR, multivariate.only = TRUE)
> Norm1
$JB

        JB-Test (multivariate)

data:  Residuals of VAR object model1VAR
Chi-squared = 28.884, df = 20, p-value = 0.09007


$Skewness

        Skewness only (multivariate)

data:  Residuals of VAR object model1VAR
Chi-squared = 17.53, df = 10, p-value = 0.06343


$Kurtosis

        Kurtosis only (multivariate)

data:  Residuals of VAR object model1VAR
Chi-squared = 11.354, df = 10, p-value = 0.3306
```

Nevertheless, we observed in *Figure* 19, the autocorrelation plot and saw that he ACF plot shows a significant autocorrelation at lag 4. This is contradictory with the previous information from the serial test.

Figure 19:



6.6.7. VAR forecast

With this model, we could expect more relevant results in the forecast of the model. We tried to forecast this model over my training period. As expected, the results in the *Figure 20* are much more relevant than before.

Figure 20:

```
> VEC2VAR_forecast<-predict(model1VAR,n.ahead = 8)
> summary(VEC2VAR_forecast)
         Length Class   Mode
fcst      10    -none-  list
endog     480   -none-  numeric
model     12    vec2var list
exo.fcst   0    -none-  NULL
> print(VEC2VAR_forecast)
$CC
         fcst     lower    upper        CI
[1,] 1.826227 1.5963752 2.056078 0.2298513
[2,] 1.283215 0.9583597 1.608070 0.3248554
[3,] 1.053245 0.7088199 1.397670 0.3444249
[4,] 1.039090 0.6156373 1.462543 0.4234529
[5,] 1.785719 1.3427566 2.228681 0.4429624
[6,] 1.278154 0.8118160 1.744493 0.4663385
[7,] 1.031712 0.5450014 1.518422 0.4867102
[8,] 1.226372 0.7215961 1.731147 0.5047756
```

The RMSE of this model is equal to 0,4331.

6.6.8. Summary of VECM forecast

In this section we compared the two models that we obtained thanks to the VEC theory. The first one is the VECM which is quite difficult to diagnose, and which gives non-relevant data over the training period. The second one is the VAR model derived from the VECM which is easier to diagnose, and which gives more relevant data.

To compare these models, we calculated the Root Mean Squared Error. The RMSE is lower for the cointegrated VAR model in comparison with the VECM.

6.7. Auto Regressive Integrated Moving Average Model (ARIMA Model)

This model can be useful in econometrics because any non-seasonal time series that show the trend and is a random white noise can be modeled with ARIMA models. The autoregressive (AR) term means that the model results in a linear regression model that uses its own lags as predictors. Linear regression model includes thus that the result can only get better if the predictors are not correlated and are independent of each other (Katchova, 2013).

First, a pure Auto-Regressive model is a model where Yt depends only on its own lags.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

In this formula, the p is the value of the number of lags needed to model the better the Yt. In other words, p is the number of lags that impact the current value of Y. Consequently, it is necessary to find which number of lags to use for this AR part of the equation (Katchova, 2013).

Secondly, a pure Moving Average model is a model where Yt depends only on the lagged forecast errors.

$$Y_t = \alpha + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + .. + \varphi_q \varepsilon_{t-q}$$

Where the error terms are the errors of the autoregressive models of the respective lags. The error ε_{t-1} is the error term of this equation for example $Y_{t-1} = \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + .. + \beta_0 Y_0 + \varepsilon_{t-1}$. In the formula of moving average model of Yt, the q is the number of lags that still impact the current value of Y. As the value p, we need to find this value q (Katchova, 2013).

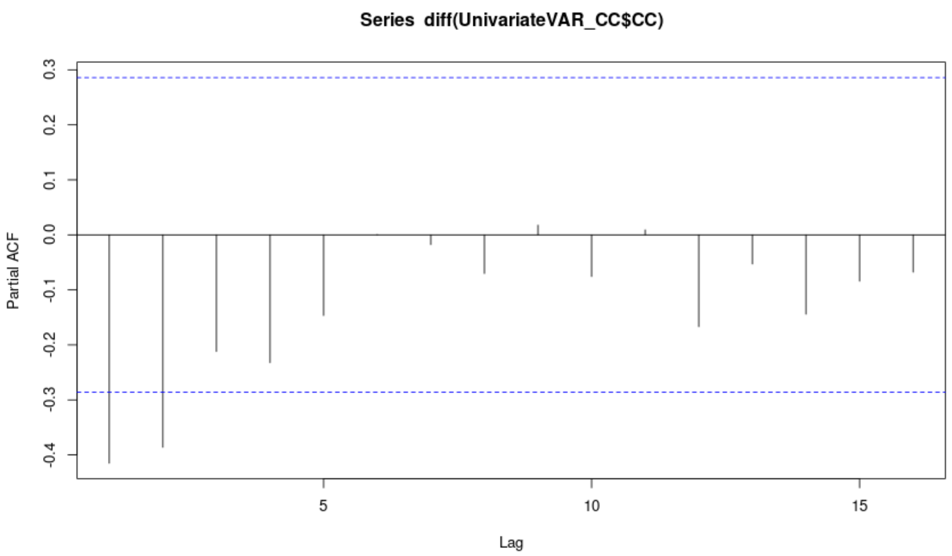
To model the CC, we must add to my model the different exogenous variables. This kind of ARIMA model is called multivariate ARIMA model or ARIMAX model (the letter X representing exogenous variables). However, in R Studio the exogenous variables are not lagged. Consequently, we must prepare a lagged data with the independent variables which should be stationary too (Katchova, 2013).

Finally, we can now write and imagine the equation of a multivariate ARIMA model (or ARIMAX model) where the time series was differenced at least once to make it stationary, and you combine the AR and the MA terms. The equation looks like : $Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + X_{1\,t-1} + \ldots + X_{k\,t-1} + \epsilon_t + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} + .. + \varphi_q \epsilon_{t-q}$, where α is the constant or intercept and where there are a linear combination lags of Y, k lagged exogenous variables and a linear combination of lagged forecast errors (Katchova, 2013).

As said before, to run an ARIMA(p,d,q) model we must find the order p for the autoregressive part of the model, the order d of differentiation for stationarity and finally the order q for the Moving Average part of the model.

To find the order of the autoregressive part, we proceeded like we do for the VAR model before. In fact, we look for the ideal number of lags to use as predictors. To find this number we use the existing formula in R called VARselect. We perform this formula on the stationary dependent variables. We saw different information criteria that indicate me the number of lags for the autoregressive model. The AIC indicates 1 as well as the HQ, the SC and the FPE. Consequently, we tried by selecting an order p equals to 1. We could also look at the Partial Auto-Correlation function and see which lags are the most significant (or impact the most the current value). By doing this method, we found 2 significant lags (*Figure 21*) meaning that order p could equal to 1 or 2.

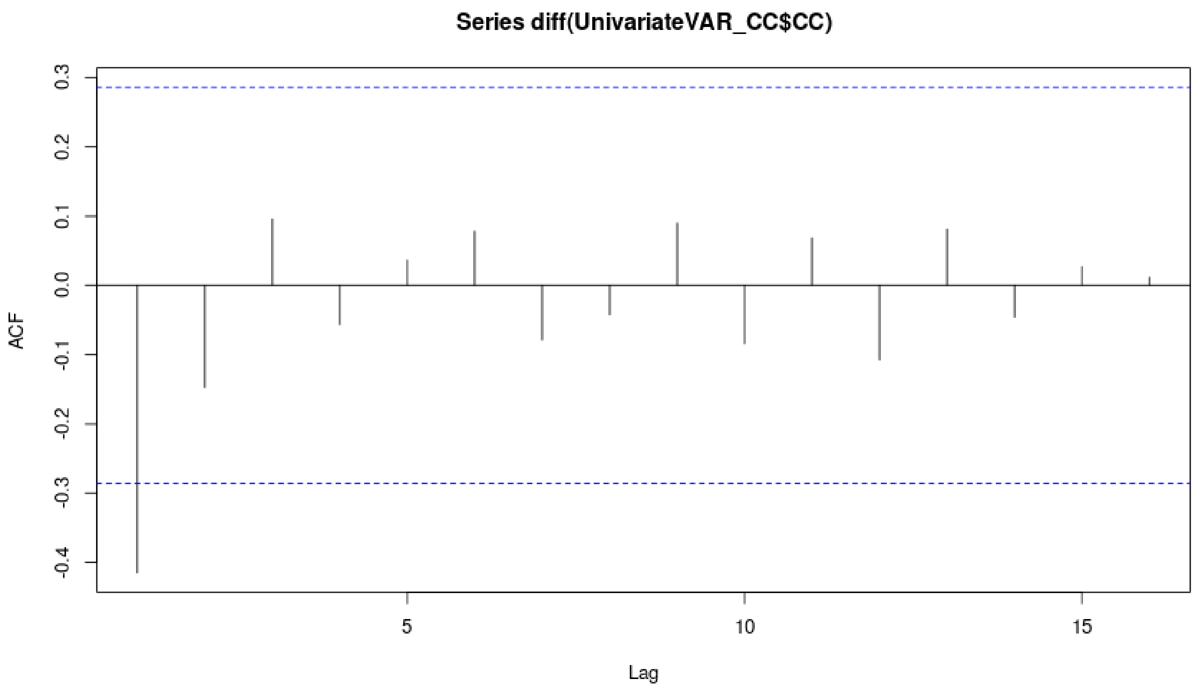
Figure 21:



After, we looked for the parameter d that represents the minimum number of times the series have to be differentiated to be stationary. The letter "I" in ARIMA (meaning integrated) refers to the fact that in this model the series should not necessarily be stationary and must be differentiated to become stationary. Like we explained in the point 6.4., we performed an Augmented Dicker Fuller test to see the stationarity. After performing this test and observed that the series needed one differentiation to become stationary, we fixed the order d as 1.

Finally, to find the last needed parameter, the one for the moving average part, we looked for the ideal number of lagged forecast errors. We tested to find this number by looking at the autocorrelation function to see at which lag the correlation is the most significant and at which lag there is no autocorrelation. *Figure 22* helped me to see that for the lag 1, the autocorrelation is significant. Thus, the ideal order q for my model should be 1.

Figure 22:



After these few steps, we had all the required parameters and could calculate the ARIMA model by using the function arima in R and by choosing the parameters. We tested several models with different parameters like ARIMA (2,1,1), ARIMA (1,1,1), ARIMA (2,1,0) or ARIMA (1,1,0).

6.7.1. ARIMA Implementation

To compare the model, we observed the different AIC of each model to see which model had the less information loss. In *Appendix 19*, the coefficients for the ARIMA (2,1,1) model is observable. The exogenous variables are lagged once. There are two AR coefficients, one MA coefficient and 13 coefficients for each lagged variable. The AIC for this model is equal to -2,87.

$$CC_t = -0{,}5359AR_1 - 0{,}2903AR_2 - 0{,}9999MA_1 - 0{,}0019\ Income_{t-1} + 0{,}0091\ GDP_{t-1} - 0{,}0392\ Unempl_{t-1} + 14{,}9326\ Gini_{t-1} - 0{,}2028\ IR_{t-1} - 0{,}0563\ CPI_{t-1} - 0{,}0148\ Return_{t-1} - 0{,}0319\ Condition_{t-1} - 0{,}0639\ Competition_{t-1} - 0{,}0705\ Depedency_{t-1} - 0{,}1231\ Education_{t-1} - 14{,}8164\ Size_{t-1} + 0{,}0061\ Sentiment_{t-1}$$

However, we observed some problems to observe the fitted values and thus we couldn't calculate the R-squared. We tried to remove variables like GDP or year-fixed variables but the model couldn't give the fitted values.

The ARIMA (1,1,1) model gives a higher AIC (-0,91) which shows that this model will not fit better than the previous one. We observed the same conclusions for the two other models ARIMA (2,1,0) and ARIMA (1,1,0).

All the models tested above didn't give fitted value due to a possible convergence in the data. Without the possibility to check the R-squared of these models, it's difficult to diagnose them and to trust in these models to forecast.

I also tried to run the auto.arima formula which is a formula who does the work explained before and choose the best parameters for the time series. The software runs a KPSS test to see how much differentiation the dependent variable needs to be stationary. A KPSS test is similar to the ADF test. The goal is in fact to check the stationarity of time series. However, this test checks the stationarity in presence of a deterministic trend meaning that the null hypothesis of stationarity will not necessarily be rejected if there is a trend.

Then the software test different values for p and q and identified the best model as the one with the lower AIC. We observed that this model is an ARIMA (0,0,0). This result is unexpected because of the parameters that we found. The parameters d should be at least 1 because the variable CC need to be differentiated once to be stationary (performing the KPSS test gives the same conclusion as performing the ADF test in this case) (Prabhakaran, 2019).

In fact, by running this formula in R Studio (*Appendix 21*), the best model identified is the ARIMA (0,0,0). In fact, the AIC equals to -3,27. This model has thus no AR or MA coefficient and has just coefficients for the lagged exogenous variables. This model is more like a VAR model.

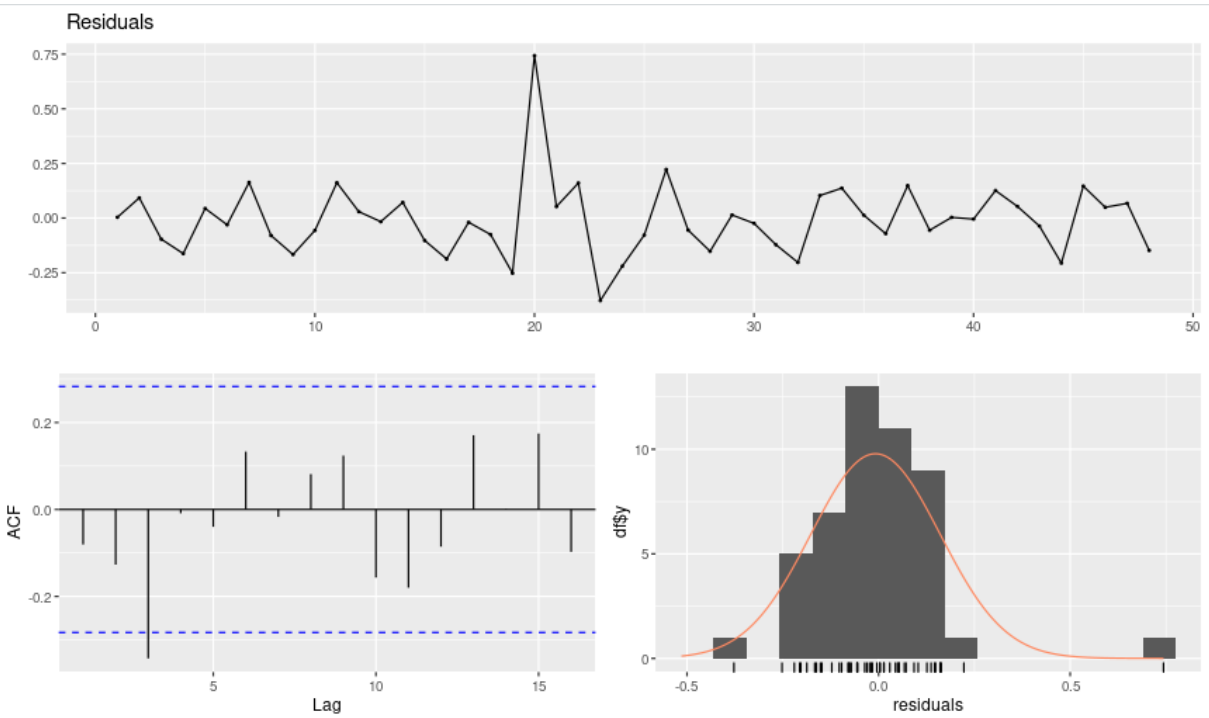
Because of this conclusion, we decided to not consider this ARIMA (0,0,0) model. Indeed, since there is no MA or AR factor to reflect the moving average process or the autoregressive process, the ARIMA model is not useful and didn't bring anything to the modelling.

6.7.2. ARIMA analysis

The next step was to test if the ARIMA (2,1,1) model is trustworthy. For this purpose, we checked the residuals to see if they were white noise. Check the residuals is useful to see if the model has adequately captured the information in the data.

First, we plotted the histogram of residuals of ARIMA (2,1,1) model (*Figure 23*). We noticed that the mean of residuals is nearly zero which is a good sign (the mean of model7's residuals is equal to -0,0071).

Figure 23:



I also calculated the autocorrelation function of residuals (*Figure 23*). We noted the significant autocorrelation at lag 4. This autocorrelation between residuals means that there is information left in the residuals. Because of the information left in the residuals, the model can't fit perfectly with the observations.

Finally, we run a Ljung-Box test to verify the serial autocorrelation of residuals. This test determines thus whether errors are independent and identically distributed (or white noise).

The large p-value in *Figure 24* suggests that there is no significant autocorrelation in the residuals.

Figure 24:

```
> Box.test(model7$residuals,type = "Ljung-Box")

	Box-Ljung test

data:  model7$residuals
X-squared = 0.33192, df = 1, p-value = 0.5645
```

I can consider that the residuals of this model are white noise which is a good information to trust in the model.

6.7.3. ARIMA forecast

Figure 25 shows the result we obtained when we tried to estimate the future responses of my model7 for the 8 coming periods. We compared this result with my test period for which we had already the results of the consumer credit until Q4 2020 (56^{th} observations).

I noted the 8 forecasted values knowing the figures for the exogenous variables. Indeed, in the database "Test_lagged_ARIMA" there are the figures of exogenous variables for the 8 periods ahead.

After forecasting these values, we calculated the Root Mean Squared Errors between the forecasted values ("CCTrain") and the observed values ("CCTest"). We saw a value of 0,197 for this result which means that the model fits quite well with the real observations in comparison with the other models seen previously.

Figure 25:

```
> model_fit7=predict(model7,n.ahead = 8,newxreg = Test_lagged_ARIMA)
> print(model_fit7)
$pred
Time Series:
Start = 49
End = 56
Frequency = 1
[1] 1.2561230 1.0896130 1.1367813 1.0802530 1.1278427 1.3051808 1.1938957 0.9649193

$se
Time Series:
Start = 49
End = 56
Frequency = 1
[1] 0.1604862 0.1804849 0.1804813 0.1824507 0.1827886 0.1827871 0.1828690 0.1828595

> CCTest<-cbind(1.2237152,1.0302679,0.8932687,0.8355272,0.9001647,0.9940655,1.0323944,0.8492139)
> CCTrain<-cbind(1.2561230,1.0896130,1.1367813,1.0802530,1.1278427,1.3051808,1.1938957,0.9649193)
> rmse(CCTest,CCTrain)
[1] 0.1974404
```

6.7.4. Summary of ARIMA forecast

The ARIMA model is not very easy to use and to diagnose. In fact, the fitted values for the ARIMA (2,1,1) are not given by the model. We could just diagnose with the calculation of the RMSE,with AIC and with the analysis of error terms. Consequently, for this model we only considered the ARIMA (2,1,1) but with precautionary because of the lack of diagnose.

6.8. Summary of all the models

In the table below, we gather all the models that we consider to be able to forecast consumer credit. These models have the better diagnostic parameters. In terms of Adjusted R-squared, the cointegrated VAR model gives the best result. This demonstrates that this model has the best explanation power. In terms of root mean squared error, the FirstVAR model gives the best result. This is the main goal of this study in order to predict the consumer credit. This model has the higher prediction power.

Name	**R-squared**	**Adjusted R-squared**	**RMSE**
FirstVAR (2)	75,55%	35,27%	0,1053893
SixthVAR (3)	75,35%	57,61	0,2063828
Model1VAR or cointegrated VAR	68,31%	60,16%	0,4331
Model7 or ARIMA (2,1,1) model	/	/	0,197

Based on this table, we can say that the better model to predict the consumer credit in Belgium is the VAR (2) model estimated in section 6.5. The goal of this thesis was to identify determinant factors behind consumer credit, to construct a predictive model but also forecast the consumer credit.

The potential predictors are consumer credit, income, repayment conditions, dependency ratio at lag 1 with a negative impact. The consumer credit, the income, the return on investment and consumer sentiment at lag 2 impact negatively the consumer credit. And the dependency ratio impact positively the dependent variable.

For the coming years we do not have access to the exogenous variables meaning that the model will forecast in the long run without knowing the exogenous variables. There is a possibility to forecast the independent variables or to have access to estimation for these future variables. However, in the short run, the model offers the possibility to predict the data of the next period.

In this table, we used the model selected to predict the new contract of consumer credit in euros. The forecasted data are given by the econometric tool and by a function that forecast the dependent variable based on the model.

	Forecast of differentiated figures	**Forecast of real figures**	**Forecast of real figures**
2018Q4			
2019Q1	0,43554321	1,323	1.322.930.840,000
2019Q2	-0,06233865	1,161	1.161.376.523,380
2019Q3	-0,28992855	0,740	740.339.364,380
2019Q4	-0,09828863	0,795	794.980.067,410
2020Q1	0,34950654	1,185	1.185.033.754,300
2020Q2	0,2161076	1,116	1.116.272.276,720
2020Q3	-0,38644782	0,608	607.617.650,580
2020Q4	-0,29347995	0,739	738.914.432,720
2021Q1	0,030630181	0,880	879.844.068,870
2021Q2	0,210986922	1,091	1.090.830.990,870
2021Q3	-0,303910013	0,787	786.920.977,870
2021Q4	0,056984254	0,844	843.905.231,870
2022Q1	0,030992539	0,875	874.897.770,870
2022Q2	0,032595714	0,907	907.493.484,870
2022Q3	-0,071061227	0,836	836.432.257,870
2022Q4	-0,089540267	0,747	746.891.990,870
2023Q1	0,066232274	0,813	813.124.264,870
2023Q2	0,004749427	0,818	817.873.691,870
2023Q3	-0,033601363	0,784	784.272.328,870
2023Q4	-0,100923827	0,683	683.348.501,870
2024Q1	0,03813975	0,721	721.488.251,870
2024Q2	0,031790479	0,753	753.278.730,870
2024Q3	-0,084307619	0,669	668.971.111,870
2024Q4	-0,02125642	0,648	647.714.691,870

The first 8 observations are forecasted knowing the exogenous variables and the 16 last observations are forecasted based on the model and the formula proposed by the econometric tool.

We can observe a declining tendency over the years. Indeed, for 2024, the amount of new consumer credit is estimated to 647.714.691,87 €.

7. Limitations of the work

The findings of this study must be seen in light of some limitations. Indeed, each work of this type contains some bias whether in terms of database or procedure.

One limitation of this master thesis is the choice of the variables. Indeed, we selected 13 variables after reading many papers on the topic. Previous studies demonstrated the explanation power of factors behind consumer credit but very few studies demonstrated the prediction power of my variables. We selected some variables that didn't help me to predict and to explain consumer credit. Moreover, we didn't include in my models some variables like saving behavior of households or impact of mortgage loans. The reasons why we didn't select these variables are the unavailability of data for some variables or the lack of research about the subject. Making choices is an obligation in this type of work. We based my choices on the literature and on the availability of data.

Another big limitation of this work is the limitations concerning the size of the database. In fact, the variables selected are often available quarterly but not monthly. This lack of available data has the effect of reducing my data sample. The data sample is weak, and this causes a problem for the robustness of the models. In a lot of studies, the data sample is superior to 100 observations. Fixing a model on a such short sample period is difficult and maybe biased. A point of improvement for a future study is to collect data over a period of minimum 100 observations to construct and train the model on 80 observations and test the model on the last 20 observations.

The choice of the model used to explain and forecast the dependent variable is also a limitation. In fact, we chose to use 3 kinds of model, but some other research probably uses other types of models. This choice is once again based on the literature review. Use a model without a lot of modification to make whether on the database or on the parameters is practice using again. It's also easier to adapt the model in the future with other available data.

Linked to the previous limitations, the development of the model is limited by the capacity of the econometric tool. Here, R Studio sometimes limited the work in the sense that the number of accepted variables is limited, or that the data are not suitable (scale of unit too large for example). We chose to work in the R environment because it is the tool for which we had the basics.

Another limitation is linked to the forecasting in the long term. The exogenous variables are unknow since we look to the out-of-sample period. The estimated models are build using lagged variables. It means that since we dispose of the data of a period, we can easily predict the new consumer credit for the next period. However, the difficulty is to predict consumer credit for large future period. Despite, the fact that some variables are predicted by others financial actors, base robust forecast with data that are themselves forecasted is extremely complicated. The forecast or the estimation of the exogenous variables could potentially constitute a parallel work.

8. Conclusion

Throughout this master thesis we identified different variables in order to predict the new amount of consumer credit. We tried to construct the econometric model using macroeconomic variables such as the income of households, their consumption, the unemployment rate, the interest rate, the level of revenue's redistribution, the consumer price index or the return on investment. we also identified microeconomic variables as potential predictors: the repayment condition and the competition between credit providers. We also included socio-demographic variables like the dependency ratio, the level of education or the size of households. We identified all these variables thanks the existing literature linked to this subject. Nevertheless, we tried to include in my models a variable which represent the consumer sentiment.

I realize few tests to diagnose the database. The Granger causality test give us a first idea of the variables that could cause the consumer credit. Indeed, the interest rate, the Gini index, the dependency ratio, the unemployment rate or the condition of repayment are the variables that seems to Granger cause the dependent variable.

We also noticed that a major part of the data is not stationary. It means that the figures show sign of trend or doesn't present a stable mean and a stable variance. We also observed the serial correlation of some variables.

After that, thanks to different models, we observed that the variables which are the most often useful to explain the consumer credit are the household's income, the evolution of repayment condition, the return on investment, the dependency ratio and the consumer sentiment. The lagged consumer credit is also very useful to explain and predict the new consumer credit. The different predictors sus-mentioned impact the consumer credit positively or negatively depending on the number of lags.

I constructed the models based on three theories: the vector autoregression model, the vector error correction model and the autoregressive integrated moving average model. Among these three models, the first one is the type that we chose to forecast the consumer credit. This selected model gives us the better robustness sign parameter. My personal opinion is that this is also the easiest model to use, it is easy to understand the functioning. The VECM is quite difficult to interpret in his pure form. It is also not the easiest to use before the transformation to VAR. The ARIMA model gives us irrelevant parameters. It is also not the most easy-to-use model because of the manual lagging of exogenous variables.

After constructing the models, we forecasted the consumer credit over the training period of 8 observations. We evaluated the capacity to forecast and finally we forecasted the consumer credit until 2024.

This master thesis served to highlight the factors that can influence the future response of the consumer credit. With this thesis, we realized that the size of the sample period is not easy to ensure because it's not always easy to get data. We also observed and judge the manipulation of econometric models.

9. Bibliography

Andersson, J. & Lauvsnes, S.O. (2007). Forecasting stock index prices and domestic credit : Does cointegration help?. *Tartu University Institute of Technology Conference*.

Anggraeni, W., Andri, K.B. & Mahananto, F. (2017). The performance of ARIMAX model and Vector Autoregressive (VAR) model in forecasting strategic commodity price in Indonesia. *Procedia Computer Science*, 124, pp. 189-196.

Aishwarya, S. (2018). A Multivariate Time Series Guide to Forecasting and Modeling. Analytics Vidhya. Consulted on June 21 via https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/

Aiyar, S., Calomiris, C. and Wieladek, T. (2016). How does credit supply respond to monetary policy and bank minimum capital requirements? *European Economic Review*, Vol. 82, pp. 142-165

Babeau, A. (2018). Les relations entre épargne et crédit : causes et conséquences d'un oubli bicentenaire. Revue d'économie financière, 130, pp. 337-350.

Bernanke, B. and Gertler, M. (1995). Inside the Black Box: The Credit Channel of Monetary Policy Transmission. *Journal of Economic Perspectives*, Vol. 9 N°4, pp. 27-48

Bertola, G. and Koeniger, W. (2004). Consumption Smoothing and the Structure of Labor and Credit Markets. *IZA DP*. N°1052

Bethune, Z., Rocheteau, G. and Rupert, P. (2015). Aggregate unemployment and household unsecured debt. *Review of Economic Dynamics,* Vol. *18, pp. 77-100*

Borowski, J., Jaworski, K. and Olipra, J. (2017). Economic, institutional and socio-cultural determinants of consumer credit in the context of monetary integration. *NBP Working Pape*r. N°254

Burdeau, E. (2015). Assessing dynamics of credit supply and demand for French SMEs, an estimation based on the Bank Lending Survey. *IFC Bulletin*. N°39

Brissimis, S. N., Garganas, E. N. & Hall, S. G. (2014). Consumer credit in an area of financial liberalization : an overreaction to repressed demand? , Applied Economics, 46:2, 139-152.

Calza, A., Grtner, C. and Sousa, J. (2001). Modelling the demand for loans to the private sector in the euro area. *European Central Bank*. Working Paper N°55

Carroll, C., Fuhrer, J. and Wilcox, W. (1994). Does Consumer Sentiment Forecast Household Spending? If So, Why? *The American Economic Review*, Vol. 84 N°5, pp. 1397-1408

Clements, M.P. & Galvão, A.B. (2013). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, 29(4), pp. 698-714.

Coralia-Emilia, P. (2013). The influences of inflation on the monetary policy interest. University of Craiova. Faculty of Economics and Business Administration.

Deaton, A. (1989). Saving and liquidity constraints. *NBER* Working Paper N°3196

Derclaye, G. (2022). Les crédits à la consommation en augmentation : hausse vertigineuse ou simple retour à la normale ? Le Soir

European Central Bank. (2021). Bank Lending Survey (BLS). Consulted on December 12 via https://www.ecb.europa.eu/stats/ecb_surveys/bank_lending_survey/html/ecb.blssurvey2021q3~57cc722cfb.en.html#toc5

Firedman, M. (1957). A Theory of the Consumption Function. *NBER*, pp. 20-37

Friedman, B., Kuttner, K., Bernanke, B. and Gertler, M. (1993). Economic Activity and the Short-Term Credit Markets: An Analysis of Prices and Quantities. *Brooking Papers on Economic Activity*, Vol. 1993 N°2, pp. 193-283

Guha, B. & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Sciences*, 4(2).

Hayes, A. (2020). Capital Adequacy Ration – CAR. *Investopedia.* Consulted on January 23 via https://www.investopedia.com/terms/c/capitaladequacyratio.asp

Hayes, A. (2021). Dependency Ration. *Investopedia*. Consulted on January 22 via https://www.investopedia.com/terms/d/dependencyratio.asp

Ho, S.L. & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. Computers & industrial engineering, 35(2), pp. 213-216.

Huyghebaert, N. (2015). Studie naar het belang van het consumentenkrediet en het hypothecair kredit voor de Belgische economie. KU Leuven.

Hyndman, R.J. & Athanasopoulos, G. (2018). Forecasting: Principles and Practice, 2nd edition, Monash University, Australia. Consulted on June 25 via https://otexts.com/fpp2/

Juster, F. T., Wachtel, P., Hymans, S. & Duesenberry, J. (1972). Inflation and the Consumer. *Brookings Papers on Economic Activity*, 1972(1), 71-121.

Kagan, J. (2021). Consumer credit. *Investopedia.* Consulted on January 21 via https://www.investopedia.com/terms/c/consumercredit.asp

Karlan, D. and Zinman, J. (2005). Elasticities of Demand for Consumer Credit. *Center Discussion Paper*. N°926

Katchova, A. (2003). Time Series ARIMA models – Time Series ARIMA Models ANAlysis. Korea University. pp. 1-22

Lack, C. (2006). Forecasting Swiss inflation using VAR models (No. 2006-02). Swiss National Bank.

Lambert, A. (2013). Les alternatives aux crédits classiques. Le Monde. Consulted on February 12 via https://www.lemonde.fr/economie/article/2013/12/07/les-alternatives-aux-credits-classiques_3527286_3234.html

Levieuge, G. (2017). Explaining and forecasting bank loans. Good times and crisis. *Applied Economics*, 49(8), pp. 823-843.

Lown, C., Morgan, D. and Rohatgi, S. (2000). Listening to Loan Officers: The Impact of Commercial Credit Standards on ending and Output. *FRBNY Economic Policy Review*

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

Lütkepohl, H. (2013). Vector autoregressive models. In *Handbook of Research Methods and Applications in Empirical Macroeconomics* (pp. 139-164). Edward Elgar Publishing.

Meyer, A., Kenny, G. & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models. Central Bank and Financial Services Authority of Ireland Technical Paper Series, vol. 1998 No 3, pp. 1-48.

Noreen, A., Asif, R., Nisar, S. & Qayyum, N. (2017). Model building and forecasting of bank credit to public and private sector. *Universal Journal of Accounting and Finance*, 5(4), 2017.

Prabhakaran, S. (2019). KPSS Test for Stationarity. Machine Learning + . Consulted on July 14 via https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/

Prabhakaran, S. (2021). ARIMA Model – Complete Guide to Time Series Forecasting in Python. Machine Learning + . Consulted on July 20 via https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

SPF Economie. (2021). Tarifs maximaux. Consulted on January 26 via https://economie.fgov.be/fr/themes/services-financiers/credit-la-consommation/cout-du-credit/tarifs-maximaux

Stevenson, S. (2007). A comparison of the forecasting ability of ARIMA models. *Journal of Property Investment & Finance*.

Swain, R. (2007). The demand and supply of credit for households. *Applied Economics*.

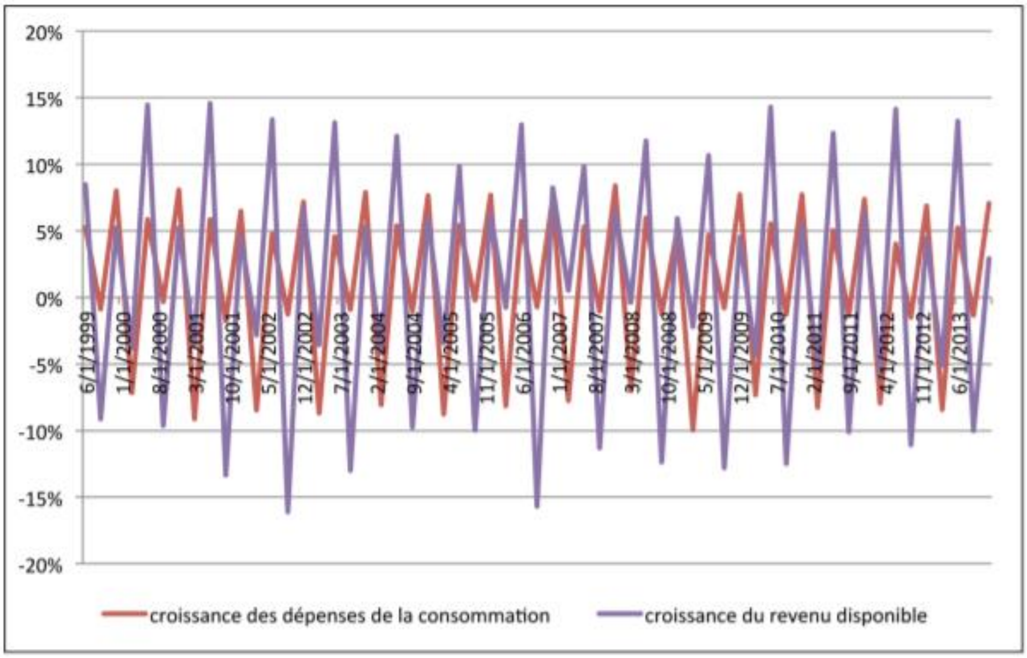
Wosko, Z. (2015). Modelling credit growth in commercial banks with the use of data from Senior Loan Officers Opinion Survey. *Economic Institute Warsaw*. NBP Working Paper N°210

Zivot, E. & Wang, J. (2006). Vector autoregressive models for multivariate time series. Modeling financial time series with S-PLUS, pp. 385-429.

10. Appendix

Appendix 1

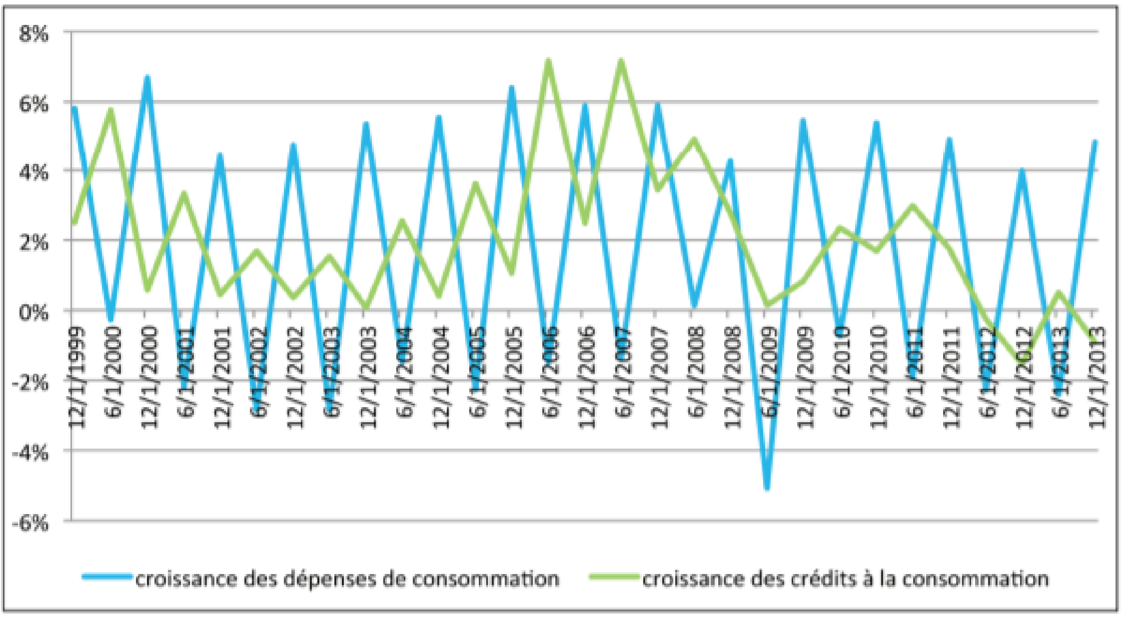
Figure 19 - Corrélation entre la croissance de la consommation des ménages et la croissance du revenu disponible, données semestrielles



Huyghebaert, 2015

Appendix 2

Figure 20 - Corrélation entre la croissance de la consommation des ménages et la croissance du crédit à la consommation, données semestrielles



Huyghebaert, 2015

Appendix 3

Table 12

Factors contributing to changes in demand for consumer credit and other lending to households

(net percentages of banks)

Country	Spending on durable goods		Consumer confidence		Consumption exp. (real estate)		General level of interest rates		Use of alternative finance	
	Q2 2021	Q3 2021	Q2 2021	Q3 2021	Q2 2021	Q3 2021	Q2 2021	Q3 2021	Q2 2021	Q3 2021
Euro area	8	4	12	8	1	-1	5	3	-1	0
Germany	7	10	3	7	3	0	7	0	-1	-1
Spain	30	0	40	10	0	0	10	0	-3	0
France	-8	-5	-2	1	0	0	0	7	-3	4
Italy	22	15	22	23	0	-8	11	8	4	3

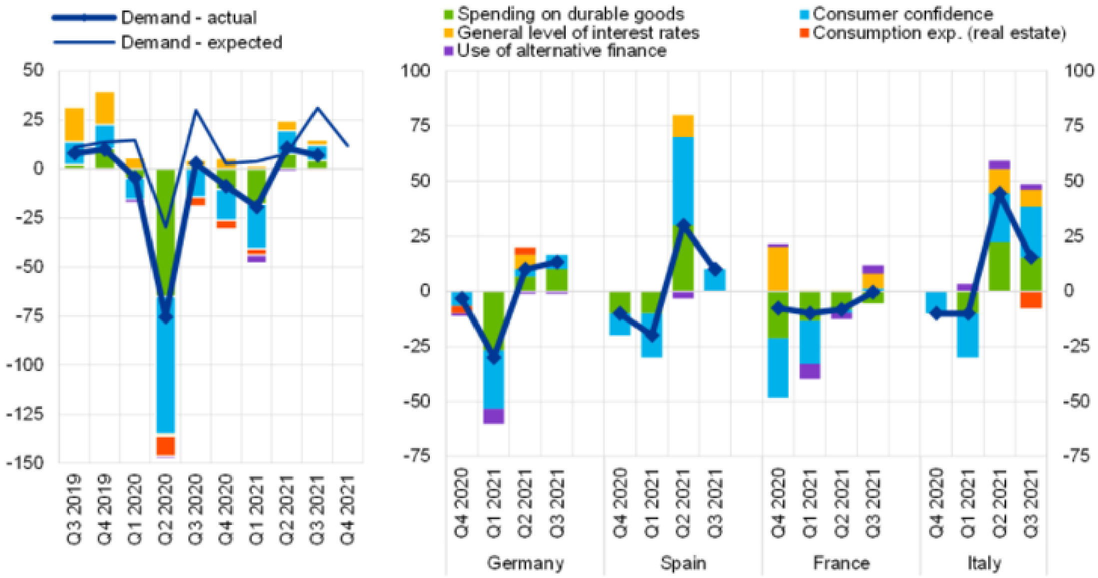
Note: See the notes on Chart 12.

European Central Bank, 2021

Appendix 4

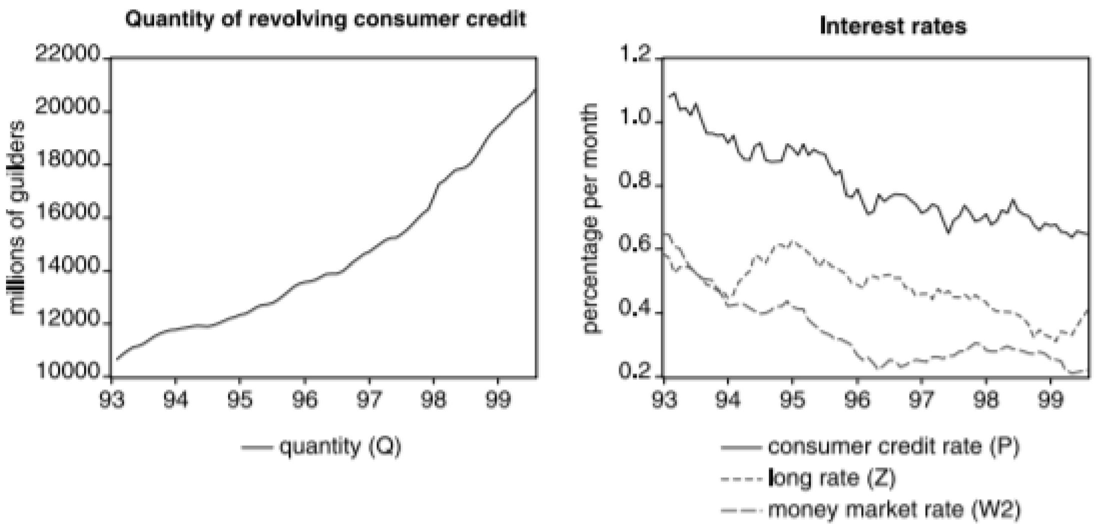
Changes in demand for consumer credit and other lending to households, and contributing factors

(net percentages of banks reporting an increase in demand, and contributing factors)



European Central Bank, 2021

Appendix 5



Toolsema, L. A. (2002). Competition in the Dutch consumer credit market. *Journal of banking & finance*, *26*(11), 2215-2229.

Appendix 6

CONSUMER SURVEY INERVIEW

1. Sex of respondent
- male
- female

2. How old are you?
- 18 to 29 years
- 30 to 49 years
- 50 to 64 years
- 65 or over

3. Do you work?
- yes
- no

4. In the case of "work" is it a full-time or a part-time job?
- full-time
- part-time

5a. In the case of "no work", how would you best describe your situation?
- retired
- on early retirement
- on sick leave
- student
- housewife/house-husband
- unemployed
- other

5b. How would you best describe your occupational category?
- white-collar worker or executive
- skilled worker - - -
blue-collar worker
- self-employed or business manager (including company director)
- farmer
- other

6. In your opinion, over the past twelve months, has the general economic situation in Belgium become:
- much better
- slightly better
- remained the same
- slightly worse
- much worse

7. How do you expect the general economic situation in Belgium to develop over the next twelve months?

It will....
- get a lot better
- get a little better
- stay the same
- get a little worse
- get a lot worse

8. How do you think that consumer prices have developed over the last twelve months? They have...
- risen a lot
- risen moderately
- risen slightly
- stayed about the same
- fallen

9. By how many per cent do you think that consumer prices have gone up/down over the last twelve months? (Please give a single figure estimate).

Consumer prices have

increased/decreased by: ....%.

10. By comparison with the past twelve months, how do you expect that comsumer prices will develop in the next twelve months? They will...:
- increase more rapidly
- increase at the same rate
- increase at a slower rate
- stay about the same
- fall

11. By how many per cent do you expect consumer prices to go up/down change in the next twelve months? (Please give a single figure estimate).

Consumer prices will increase/decrease

by .... %.

12. What do you think will happen to unemployment in Belgium over the next twelve months? Do you believe that it will rise, fall or remain steady?:
- increase sharply
- increase slightly
- remain the same
- fall slightly
- fall sharply

13. As regards furniture, a washing machine, a television, a computer and other durables, do you think that now is a good or bad time for people to make such a major purchase, or neither good nor bad?

- good
- not good but not bad either
- bad

14. As regards major purchases such as furniture, a washing machine, a television or other durables, do you think that over the next twelve months your household will be spending more, less or the same on such durable goods as over the past twelve months?

- much more
- a bit more
- the same
- a bit less
- much less

15. Do you expect to buy a car within the next twelve months?

- yes, definitely
- yes, possibly
- probably not
- definitely not

16. Are you planning to buy or build a home over the next twelve months (to live in yourself, for a member of your family, as a holiday home, to let...)

- yes, definitely
- possibly
- probably not
- definitely not

17. How likely are you to spend any large sums of money on home improvements or renovations over the next twelve months?

- very likely
- fairly likely
- not likely
- not at all likely

18. Would you say that your financial situation has got better or worse or remained unchanged over the past twelve months?

- much better
- slightly better
- remained unchanged
- slightly worse
- much worse

19. And what about your financial situation at the moment? Is your household in a financial situation where you are short of money, or have money to spare, or can just get by?

- plenty over, saving
- saving a bit
- just getting by
- using savings
- getting into debt

20. How do you expect the financial position of your household to change over the next twelve months? It will...

- get a lot better
- get a little better
- stay the same
- get a little worse
- get a lot worse

21. Do you think that you will be able to put any money by, i.e. save, over the next twelve months?

- definitely
- yes, possibly
- probably not
- definitely not

22. In view of the general situation, do you think that now is...?

- a very good moment to save
- a fairly good moment to save
- not a good moment to save
- a very bad moment to save

23. For a survey such as this we need to be able to place people in broad income groups. May I therefore ask you to class yourself in one of the following income groups (= total net household income per month in EUR)?

- under EUR 1,000 per month
- between EUR 1,000 and 2,500 per month
- between EUR 2,500 and 4,000per month
- over EUR 4,000 per month

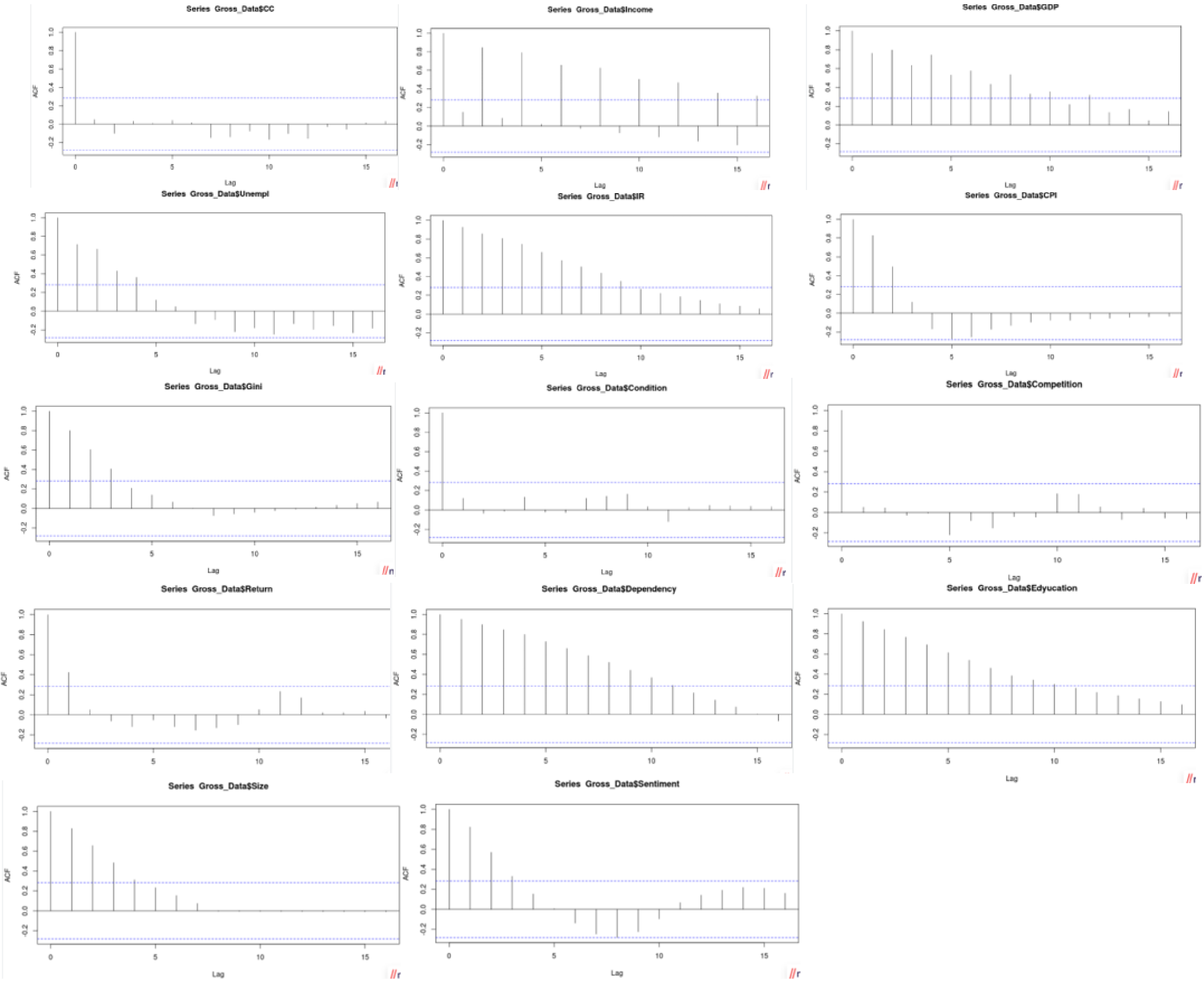
24. And finally, what is the highest standard of education that you attained?

- elementary school
- lower intermediate
- higher intermediate
- higher non-university education
- university

Appendix 7

	CC	Income	GDP	Unempl	IR	Gini	CPI	Condition	Competition	Return	Dependency	Edyucation	Size	Sentiment
CC	1.00000000	0.04283493	0.01889749	-0.10988539	-0.21637819	0.03100308	0.16397335	-0.04213300	-0.02403758	-0.04232765	0.05663260	0.08379375	-0.05823153	0.14294532
Income	0.04283493	1.00000000	0.88625068	-0.31228889	-0.72799661	-0.45271411	-0.07296225	-0.22742265	0.13060817	0.03308298	0.62867192	0.72999240	-0.59584485	0.21879673
GDP	0.01889749	0.88625068	1.00000000	-0.26213077	-0.87509205	-0.55220751	-0.09732617	-0.32158432	0.13120861	0.07734620	0.79473849	0.91098670	-0.72697144	0.33034561
Unempl	-0.10988539	-0.31228889	-0.26213077	1.00000000	0.14882947	0.32500408	-0.51958984	-0.09707882	-0.31804559	0.36268377	-0.31542273	-0.23135514	0.43092173	-0.36182990
IR	-0.21637819	-0.72799661	-0.87509205	0.14882947	1.00000000	0.48148789	0.18497912	0.28833170	-0.09396857	-0.22580757	-0.83065840	-0.94139354	0.67102739	-0.48036726
Gini	0.03100308	-0.45271411	-0.55220751	0.32500408	0.48148789	1.00000000	0.05795723	0.16329061	-0.22876842	0.14183520	-0.67181791	-0.55978500	0.52336905	-0.32051939
CPI	0.16397335	-0.07296225	-0.09732617	-0.51958984	0.18497912	0.05795723	1.00000000	0.01311791	0.18158915	-0.59829649	-0.14345513	-0.15284922	-0.10110987	0.12846747
Condition	-0.04213300	-0.22742265	-0.32158432	-0.09707882	0.28833170	0.16329061	0.01311791	1.00000000	0.15517288	-0.09458491	-0.23752848	-0.22174987	0.07559289	-0.26145982
Competition	-0.02403758	0.13060817	0.13120861	-0.31804559	-0.09396857	-0.22876842	0.18158915	0.15517288	1.00000000	-0.07948172	0.19643381	0.14272478	-0.20946371	0.25892773
Return	-0.04232765	0.03308298	0.07734620	0.36268377	-0.22580757	0.14183520	-0.59829649	-0.09458491	-0.07948172	1.00000000	0.03462615	0.16861545	0.02045352	0.06231283
Dependency	0.05663260	0.62867192	0.79473849	-0.31542273	-0.83065840	-0.67181791	-0.14345513	-0.23752848	0.19643381	0.03462615	1.00000000	0.84337543	-0.59239110	0.55315649
Edyucation	0.08379375	0.72999240	0.91098670	-0.23135514	-0.94139354	-0.55978500	-0.15284922	-0.22174987	0.14272478	0.16861545	0.84337543	1.00000000	-0.83323166	0.39551896
Size	-0.05823153	-0.59584485	-0.72697144	0.43092173	0.67102739	0.52336905	-0.10110987	0.07559289	-0.20946371	0.02045352	-0.59239110	-0.83323166	1.00000000	-0.23985227
Sentiment	0.14294532	0.21879673	0.33034561	-0.36182990	-0.48036726	-0.32051939	0.12846747	-0.26145982	0.25892773	0.06231283	0.55315649	0.39551896	-0.23985227	1.00000000

Appendix 8


Series Gross_Data$CC
Series Gross_Data$Income
Series Gross_Data$GDP
Series Gross_Data$Unempl
Series Gross_Data$IR
Series Gross_Data$CPI
Series Gross_Data$Gini
Series Gross_Data$Condition
Series Gross_Data$Competition
Series Gross_Data$Return
Series Gross_Data$Dependency
Series Gross_Data$Edyucation
Series Gross_Data$Size
Series Gross_Data$Sentiment
ACF
Lag

Appendix 9

```
> FirstVAR<-VAR(Base1,p = 2,type = "const",ic = "AIC")
> summary(FirstVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DGDP, DUnempl, DIR, DGini, DCPI, DCondition, DCompetition, DReturn, DDependency, DEducation, DSize, DSentiment
Deterministic variables: const
Sample size: 46
Log Likelihood: 357.561
Roots of the characteristic polynomial:
1.031 0.9822 0.9423 0.9423 0.9289 0.9289 0.8601 0.8601 0.8551 0.8551 0.8155 0.8155 0.801 0.801 0.798 0.798 0.716 0.6748 0.6732 0.6732 0.6399 0.6399 0.6025 0.60
25 0.5906 0.5906 0.2485 0.2485
Call:
VAR(y = Base1, p = 2, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DGDP.l1 + DUnempl.l1 + DIR.l1 + DGini.l1 + DCPI.l1 + DCondition.l1 + DCompetition.l1 + DReturn.l1 + DDependency.l1 + DEducation.l1
+ DSize.l1 + DSentiment.l1 + DCC.l2 + DIncome.l2 + DGDP.l2 + DUnempl.l2 + DIR.l2 + DGini.l2 + DCPI.l2 + DCondition.l2 + DCompetition.l2 + DReturn.l2 + DDepende
ncy.l2 + DEducation.l2 + DSize.l2 + DSentiment.l2 + const

                 Estimate Std. Error t value Pr(>|t|)  
DCC.l1          -0.648925   0.224614  -2.889   0.0102 *
DIncome.l1      -0.095063   0.036865  -2.579   0.0195 *
DGDP.l1         -0.018601   0.018274  -1.018   0.3230  
DUnempl.l1      -0.171620   0.099256  -1.729   0.1019  
DIR.l1          -0.096319   0.136413  -0.706   0.4897  
DGini.l1        -9.301142  28.300668  -0.329   0.7464  
DCPI.l1          0.031937   0.065859   0.485   0.6339  
DCondition.l1   -0.294114   0.138209  -2.128   0.0483 *
DCompetition.l1  0.078114   0.071104   1.099   0.2873  
DReturn.l1      -0.005439   0.006721  -0.809   0.4295  
DDependency.l1  -0.671305   0.332060  -2.022   0.0592 .
DEducation.l1   -0.014550   0.100534  -0.145   0.8866  
DSize.l1         5.901788  25.421627   0.232   0.8192  
DSentiment.l1    0.004134   0.010563   0.391   0.7004  
DCC.l2          -0.437813   0.242256  -1.807   0.0885 .
DIncome.l2      -0.079082   0.033050  -2.393   0.0285 *
DGDP.l2         -0.009534   0.018749  -0.508   0.6176  
DUnempl.l2      -0.049105   0.121560  -0.404   0.6913  
DIR.l2          -0.011802   0.119915  -0.098   0.9227  
DGini.l2       -31.924915  31.090181  -1.027   0.3189  
DCPI.l2         -0.082646   0.076105  -1.086   0.2927  
DCondition.l2   -0.080036   0.124533  -0.643   0.5290  
DCompetition.l2  0.086432   0.060656   1.425   0.1723  
DReturn.l2      -0.012169   0.006795  -1.791   0.0912 .
DDependency.l2   0.717410   0.350979   2.044   0.0568 .
DEducation.l2   -0.126393   0.122220  -1.034   0.3156  
DSize.l2        49.894927  34.841102   1.432   0.1703  
DSentiment.l2   -0.029118   0.011413  -2.551   0.0207 *
const           -2.541593   2.445553  -1.039   0.3132  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.2289 on 17 degrees of freedom
Multiple R-Squared: 0.7555,	Adjusted R-squared: 0.3527 
F-statistic: 1.876 on 28 and 17 DF,  p-value: 0.08861
```

Appendix 10

```
> SecondVAR<-VAR(Base1x,p = 2,type = "const",ic = "AIC")
> summary(SecondVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DUnempl, DGDP, DCPI, DCondition, DCompetition, DReturn, DDependency, DSentiment
Deterministic variables: const
Sample size: 46
Log Likelihood: -427.266
Roots of the characteristic polynomial:
1.028 1.009 1.009 0.8152 0.8152 0.7617 0.7617 0.7413 0.7413 0.6942 0.6877 0.6877 0.6846 0.6846 0.6494 0.6494 0.4314 0.3106 0.3106 0.2495
Call:
VAR(y = Base1x, p = 2, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DUnempl.l1 + DGDP.l1 + DCPI.l1 + DCondition.l1 + DCompetition.l1 + DReturn.l1 + DDependency.l1 + DSentiment.l1 + DCC.l2 + DIncome.l
2 + DUnempl.l2 + DGDP.l2 + DCPI.l2 + DCondition.l2 + DCompetition.l2 + DReturn.l2 + DDependency.l2 + DSentiment.l2 + const

                 Estimate Std. Error t value Pr(>|t|)
DCC.l1          -0.624354  0.174397  -3.580  0.00144 **
DIncome.l1      -0.051227  0.021919  -2.337  0.02774 *
DUnempl.l1      -0.167729  0.078247  -2.144  0.04198 *
DGDP.l1         -0.009360  0.012239  -0.765  0.45158
DCPI.l1          0.017737  0.054324   0.326  0.74677
DCondition.l1   -0.197269  0.104096  -1.895  0.06970 .
DCompetition.l1  0.071904  0.056876   1.264  0.21781
DReturn.l1      -0.008236  0.005402  -1.524  0.13995
DDependency.l1  -0.440202  0.213775  -2.059  0.05004 .
DSentiment.l1    0.010698  0.008664   1.235  0.22836
DCC.l2          -0.396926  0.169699  -2.339  0.02763 *
DIncome.l2      -0.044841  0.022917  -1.957  0.06164 .
DUnempl.l2      -0.039481  0.090537  -0.436  0.66652
DGDP.l2         -0.011678  0.012976  -0.900  0.37671
DCPI.l2         -0.091286  0.056008  -1.630  0.11566
DCondition.l2   -0.071347  0.093476  -0.763  0.45245
DCompetition.l2  0.068728  0.050023   1.374  0.18166
DReturn.l2      -0.009145  0.005861  -1.560  0.13128
DDependency.l2   0.462942  0.226728   2.042  0.05185 .
DSentiment.l2   -0.020879  0.009470  -2.205  0.03689 *
const           -1.257478  2.010092  -0.626  0.53726
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.2116 on 25 degrees of freedom
Multiple R-Squared: 0.6927,     Adjusted R-squared: 0.4469
F-statistic: 2.818 on 20 and 25 DF,  p-value: 0.007676
```

Appendix 11

```
> ThirdVAR<-VAR(Base2,p = 3,type = "const", ic = "AIC")
> summary(ThirdVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DGDP, DUnempl, DCondition, DCompetition, DReturn, DDependency, DSentiment
Deterministic variables: const
Sample size: 45
Log Likelihood: -214.915
Roots of the characteristic polynomial:
1.087 1.087  1.02  1.02 0.9769 0.9742 0.9742 0.9324 0.9324 0.9157 0.9157 0.9145 0.9145 0.855 0.855 0.8535 0.8535 0.8472 0.8472 0.8452 0.8452 0.7726 0.7726 0.62
61 0.6261 0.434 0.3283
Call:
VAR(y = Base2, p = 3, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DGDP.l1 + DUnempl.l1 + DCondition.l1 + DCompetition.l1 + DReturn.l1 + DDependency.l1 + DSentiment.l1 + DCC.l2 + DIncome.l2 + DGDP.l
2 + DUnempl.l2 + DCondition.l2 + DCompetition.l2 + DReturn.l2 + DDependency.l2 + DSentiment.l2 + DCC.l3 + DIncome.l3 + DGDP.l3 + DUnempl.l3 + DCondition.l3 + D
Competition.l3 + DReturn.l3 + DDependency.l3 + DSentiment.l3 + const

                 Estimate Std. Error t value Pr(>|t|)
DCC.l1          -0.441354   0.223178  -1.978   0.0644 .
DIncome.l1      -0.088274   0.035457  -2.490   0.0234 *
DGDP.l1         -0.011278   0.016973  -0.665   0.5153
DUnempl.l1      -0.086482   0.129777  -0.666   0.5141
DCondition.l1   -0.280367   0.129371  -2.167   0.0447 *
DCompetition.l1  0.020345   0.088248   0.231   0.8204
DReturn.l1      -0.003862   0.006474  -0.596   0.5587
DDependency.l1  -0.143159   0.319765  -0.448   0.6600
DSentiment.l1    0.015878   0.009405   1.688   0.1096
DCC.l2          -0.549057   0.231826  -2.368   0.0300 *
DIncome.l2      -0.128954   0.059407  -2.171   0.0444 *
DGDP.l2         -0.004842   0.020667  -0.234   0.8176
DUnempl.l2       0.115529   0.127931   0.903   0.3791
DCondition.l2   -0.175847   0.165520  -1.062   0.3029
DCompetition.l2 -0.038935   0.116880  -0.333   0.7431
DReturn.l2      -0.003844   0.006184  -0.622   0.5424
DDependency.l2   0.189532   0.508154   0.373   0.7138
DSentiment.l2   -0.018580   0.011083  -1.676   0.1119
DCC.l3          -0.217944   0.210484  -1.035   0.3150
DIncome.l3      -0.134169   0.086405  -1.553   0.1389
DGDP.l3         -0.075490   0.050963  -1.481   0.1568
DUnempl.l3       0.039728   0.104134   0.382   0.7076
DCondition.l3   -0.045919   0.136804  -0.336   0.7412
DCompetition.l3 -0.124190   0.096177  -1.291   0.2139
DReturn.l3      -0.003493   0.006745  -0.518   0.6113
DDependency.l3  -0.029874   0.307786  -0.097   0.9238
DSentiment.l3    0.011596   0.012106   0.958   0.3515
const           -0.917938   2.317452  -0.396   0.6970
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.2167 on 17 degrees of freedom
Multiple R-Squared: 0.7807,     Adjusted R-squared: 0.4325
F-statistic: 2.242 on 27 and 17 DF,  p-value: 0.04314
```

Appendix 12

```
> FourVAR<-VAR(Base3,p = 3,type = "const", ic = "AIC")
> summary(FourVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DCPI, DUnempl, DCondition, DCompetition, DReturn, DDependency, DSentiment
Deterministic variables: const
Sample size: 45
Log Likelihood: -192.257
Roots of the characteristic polynomial:
1.029 1.014 1.014 0.9966 0.9759 0.9759 0.9039 0.9039 0.8633 0.8633 0.8545 0.8545  0.84  0.84 0.8374 0.8374 0.8343 0.8343 0.7811 0.7811 0.7811 0.7811
512 0.7083 0.4896 0.4896
Call:
VAR(y = Base3, p = 3, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DCPI.l1 + DUnempl.l1 + DCondition.l1 + DCompetition.l1 + DReturn.l1 + DDependency.l1 + DSentiment.l1 + DCC.l2 + DIncome.l
2 + DUnempl.l2 + DCondition.l2 + DCompetition.l2 + DReturn.l2 + DDependency.l2 + DSentiment.l2 + DCC.l3 + DIncome.l3 + DCPI.l3 + DUnempl.l3 + DCondit
Competition.l3 + DReturn.l3 + DDependency.l3 + DSentiment.l3 + const

                 Estimate Std. Error t value Pr(>|t|)
DCC.l1          -0.525246   0.229172  -2.292   0.0349 *
DIncome.l1      -0.035551   0.025308  -1.405   0.1781
DCPI.l1          0.071051   0.081371   0.873   0.3947
DUnempl.l1      -0.010744   0.136200  -0.079   0.9380
DCondition.l1   -0.239500   0.124084  -1.930   0.0704 .
DCompetition.l1 -0.013531   0.076065  -0.178   0.8609
DReturn.l1      -0.006616   0.007028  -0.941   0.3597
DDependency.l1  -0.279725   0.356052  -0.786   0.4429
DSentiment.l1    0.014903   0.010276   1.450   0.1652
DCC.l2          -0.586626   0.229687  -2.554   0.0205 *
DIncome.l2      -0.045818   0.039960  -1.147   0.2674
DCPI.l2         -0.133409   0.079962  -1.668   0.1135
DUnempl.l2       0.049038   0.131386   0.373   0.7136
DCondition.l2   -0.222428   0.167477  -1.328   0.2017
DCompetition.l2 -0.078318   0.103504  -0.757   0.4596
DReturn.l2      -0.007623   0.008402  -0.907   0.3770
DDependency.l2   0.252237   0.532858   0.473   0.6420
DSentiment.l2   -0.019017   0.010864  -1.750   0.0981 .
DCC.l3          -0.132751   0.199080  -0.667   0.5138
DIncome.l3      -0.011341   0.032295  -0.351   0.7298
DCPI.l3          0.009901   0.068728   0.144   0.8871
DUnempl.l3       0.037800   0.105909   0.357   0.7256
DCondition.l3    0.022687   0.149725   0.152   0.8813
DCompetition.l3 -0.118191   0.085813  -1.377   0.1863
DReturn.l3      -0.008805   0.007889  -1.116   0.2799
DDependency.l3   0.033160   0.295363   0.112   0.9119
DSentiment.l3    0.005758   0.011360   0.507   0.6188
const           -0.303512   2.186927  -0.139   0.8913
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.2134 on 17 degrees of freedom
Multiple R-Squared: 0.7873,     Adjusted R-squared: 0.4494
F-statistic:  2.33 on 27 and 17 DF,  p-value: 0.03636
```

Appendix 13

```
> FifthVAR<-VAR(Base4,p = 3,type = "const", ic = "AIC")
> summary(FifthVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DCPI, DGDP, DCondition, DReturn, DSentiment
Deterministic variables: const
Sample size: 44
Log Likelihood: -337.474
Roots of the characteristic polynomial:
1.192 1.192 1.016 1.016 1.007 1.007 0.9324 0.9324 0.8701 0.8701 0.8364 0.8364 0.8182 0.8182 0.7994 0.7994 0.7938 0.7938 0.5346 0.1603 0.003006
Call:
VAR(y = Base4, p = 3, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DCPI.l1 + DGDP.l1 + DCondition.l1 + DReturn.l1 + DSentiment.l1 + DCC.l2 + DIncome.l2 + DCPI.l2 + DGDP.l2 + DCondition.l2 + DReturn.
l2 + DSentiment.l2 + DCC.l3 + DIncome.l3 + DCPI.l3 + DGDP.l3 + DCondition.l3 + DReturn.l3 + DSentiment.l3 + const

               Estimate Std. Error t value Pr(>|t|)
DCC.l1        -0.633742   0.170664  -3.713  0.00121 **
DIncome.l1     0.006727   0.058252   0.115  0.90911
DCPI.l1        0.104858   0.054325   1.930  0.06657 .
DGDP.l1       -0.004871   0.012039  -0.405  0.68967
DCondition.l1 -0.154334   0.106007  -1.456  0.15955
DReturn.l1    -0.009168   0.005133  -1.786  0.08787 .
DSentiment.l1  0.012423   0.009006   1.379  0.18163
DCC.l2        -0.616607   0.188626  -3.269  0.00351 **
DIncome.l2     0.056588   0.103322   0.548  0.58942
DCPI.l2       -0.138600   0.060310  -2.298  0.03143 *
DGDP.l2        0.026432   0.043757   0.604  0.55198
DCondition.l2 -0.150852   0.132803  -1.136  0.26822
DReturn.l2    -0.009023   0.006405  -1.409  0.17286
DSentiment.l2 -0.013989   0.009908  -1.412  0.17196
DCC.l3        -0.180366   0.186872  -0.965  0.34494
DIncome.l3     0.034549   0.064587   0.535  0.59807
DCPI.l3       -0.009587   0.055318  -0.173  0.86399
DGDP.l3        0.017926   0.044635   0.402  0.69185
DCondition.l3  0.213280   0.104090   2.049  0.05258 .
DReturn.l3    -0.011657   0.005183  -2.249  0.03484 *
DSentiment.l3 -0.001031   0.009647  -0.107  0.91585
const          0.010306   0.030288   0.340  0.73688
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 0.1985 on 22 degrees of freedom
Multiple R-Squared: 0.7619,     Adjusted R-squared: 0.5345
F-statistic: 3.351 on 21 and 22 DF,  p-value: 0.003378
```

Appendix 14

```
> SixthVAR<-VAR(Base5,p = 3,type = "const", ic = "AIC")
> summary(SixthVAR)

VAR Estimation Results:
=========================
Endogenous variables: DCC, DIncome, DCPI, DCondition, DReturn, DSentiment
Deterministic variables: const
Sample size: 44
Log Likelihood: -295.163
Roots of the characteristic polynomial:
1.029 1.029 0.9955 0.8662 0.8662 0.8477 0.8157 0.8157 0.8044 0.8044 0.7784 0.7784 0.7678 0.7678 0.7352 0.7352 0.3348 0.3348
Call:
VAR(y = Base5, p = 3, type = "const", ic = "AIC")


Estimation results for equation DCC:
====================================
DCC = DCC.l1 + DIncome.l1 + DCPI.l1 + DCondition.l1 + DReturn.l1 + DSentiment.l1 + DCC.l2 + DIncome.l2 + DCPI.l2 + DCondition.l2 + DReturn.l2 + DSentiment.l2 +
DCC.l3 + DIncome.l3 + DCPI.l3 + DCondition.l3 + DReturn.l3 + DSentiment.l3 + const

                Estimate Std. Error t value Pr(>|t|)    
DCC.l1        -0.6441509  0.1621883  -3.972 0.000533 ***
DIncome.l1    -0.0225205  0.0172390  -1.306 0.203318    
DCPI.l1        0.1040451  0.0509100   2.044 0.051655 .  
DCondition.l1 -0.1711468  0.0956127  -1.790 0.085574 .  
DReturn.l1    -0.0100216  0.0047393  -2.115 0.044610 *  
DSentiment.l1  0.0123206  0.0083161   1.482 0.150959    
DCC.l2        -0.6317952  0.1791359  -3.527 0.001650 ** 
DIncome.l2    -0.0066011  0.0266187  -0.248 0.806171    
DCPI.l2       -0.1498896  0.0559677  -2.678 0.012893 *  
DCondition.l2 -0.1803401  0.1165474  -1.547 0.134345    
DReturn.l2    -0.0098435  0.0056400  -1.745 0.093217 .  
DSentiment.l2 -0.0123568  0.0091090  -1.357 0.187047    
DCC.l3        -0.1940641  0.1745245  -1.112 0.276735    
DIncome.l3     0.0146794  0.0190890   0.769 0.449099    
DCPI.l3       -0.0018949  0.0515457  -0.037 0.970967    
DCondition.l3  0.2006530  0.0959110   2.092 0.046748 *  
DReturn.l3    -0.0115112  0.0048357  -2.380 0.025228 *  
DSentiment.l3  0.0008958  0.0083212   0.108 0.915130    
const          0.0089345  0.0286904   0.311 0.758071    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 0.1894 on 25 degrees of freedom
Multiple R-Squared: 0.7535,     Adjusted R-squared: 0.5761 
F-statistic: 4.246 on 18 and 25 DF,  p-value: 0.0005121
```

Appendix 15

```
> print(Fcst_First)
$DCC
             fcst      lower     upper        CI
[1,]  0.263935247 -0.1847714 0.7126419 0.4487066
[2,] -0.213841042 -0.8118355 0.3841534 0.5979944
[3,]  0.004762871 -0.7012557 0.7107815 0.7060186
[4,] -0.009824547 -0.7560197 0.7363706 0.7461952
[5,] -0.027020341 -0.8124998 0.7584592 0.7854795
[6,]  0.154822246 -0.6535840 0.9632285 0.8084063
[7,] -0.154466786 -0.9870665 0.6781329 0.8325997
[8,] -0.075833115 -0.9179987 0.7663325 0.8421656
```

```
> print(Fcst_Second)
$DCC
            fcst      lower     upper        CI
[1,]  0.25000386 -0.1647971 0.6648048 0.4148010
[2,] -0.04797723 -0.5913701 0.4954157 0.5433929
[3,] -0.07396279 -0.7203481 0.5724225 0.6463853
[4,] -0.16688267 -0.8396695 0.5059042 0.6727868
[5,]  0.27044225 -0.4367965 0.9776810 0.7072387
[6,] -0.12937856 -0.8535446 0.5947875 0.7241660
[7,]  0.06818395 -0.6672909 0.8036588 0.7354749
[8,] -0.16783326 -0.9142156 0.5785491 0.7463824
```

```
> print(Fcst_Third)
$DCC
           fcst      lower      upper        CI
[1,] -2.0249540 -2.4496289 -1.6002792 0.4246749
[2,]  1.3632419  0.8461587  1.8803251 0.5170832
[3,] -0.9781842 -1.5674601 -0.3889082 0.5892760
[4,]  1.9117099  1.1098528  2.7135669 0.8018571
[5,] -1.5373580 -2.6641136 -0.4106024 1.1267556
[6,]  1.0974397 -0.1557708  2.3506501 1.2532105
[7,] -3.1281001 -4.5040798 -1.7521205 1.3759796
[8,]  3.1491944  1.6996820  4.5987068 1.4495124
```

```
> print(Fcst_Four)
$DCC
            fcst      lower     upper        CI
[1,]  0.12306284 -0.2952204 0.5413460 0.4182832
[2,] -0.05120215 -0.5661368 0.4637325 0.5149347
[3,] -0.16788640 -0.7722527 0.4364799 0.6043663
[4,] -0.01927212 -0.6991722 0.6606280 0.6799001
[5,]  0.18113561 -0.5696054 0.9318767 0.7507410
[6,]  0.11103938 -0.6669847 0.8890635 0.7780241
[7,] -0.45347800 -1.2555402 0.3485842 0.8020622
[8,]  0.17633190 -0.6514511 1.0041149 0.8277830
```

```
> print(Fcst_Fifth)
$DCC
           fcst      lower      upper        CI
[1,]  1.2573571  0.8683167  1.6463974 0.3890404
[2,] -0.6002803 -1.0857728 -0.1147877 0.4854926
[3,] -0.8496337 -1.4255991 -0.2736684 0.5759653
[4,] -0.4346952 -1.1756449  0.3062544 0.7409496
[5,]  2.7865536  1.9218647  3.6512424 0.8646889
[6,] -0.9756296 -1.8758650 -0.0753943 0.9002353
[7,] -2.7144326 -3.8984990 -1.5303662 1.1840664
[8,]  0.7732567 -0.5211960  2.0677093 1.2944527
```

```
> print(Fcst_Sixth)
$DCC
            fcst       lower     upper        CI
[1,]  0.43554321  0.06426341 0.8068230 0.3712798
[2,] -0.06233865 -0.52543213 0.4007548 0.4630935
[3,] -0.28992855 -0.81955283 0.2396957 0.5296243
[4,] -0.09828863 -0.72449930 0.5279220 0.6262107
[5,]  0.34950654 -0.32089310 1.0199062 0.6703996
[6,]  0.21610760 -0.46610482 0.8983200 0.6822124
[7,] -0.38644782 -1.08909139 0.3161958 0.7026436
[8,] -0.29347995 -1.00644263 0.4194827 0.7129627
```

Appendix 16

```
> model<-VECM(Gross_Data10, 3, r = 5, estim = "ML")
> summary(model)
#############
###Model VECM
#############
Full sample size: 48    End sample size: 44
Number of variables: 10        Number of estimated slope parameters 360
AIC -9026.79    BIC -8339.877   SSR 270.3882
Cointegrating vector (estimated by ML):
               CC        Income        Unempl           IR         Gini          CPI   Condition       Return
r1  1.000000e+00 -1.110223e-16  1.665335e-16 2.498002e-16 0.000000e+00 -0.006633940 -0.87601592 -0.0069371571
r2 -7.105427e-15  1.000000e+00  2.220446e-15 1.998401e-15 1.136868e-13 -5.220345620  5.09738985 -0.9243483475
r3 -1.776357e-15 -1.665335e-16  1.000000e+00 8.326673e-17 2.842171e-14  2.020819072 -0.93566102  0.1739347283
r4  0.000000e+00  1.110223e-16 -6.661338e-16 1.000000e+00 0.000000e+00  1.171077452  1.62416888  0.2938894366
r5  0.000000e+00 -3.469447e-18  5.204170e-18 6.071532e-18 1.000000e+00  0.003685674 -0.03794464  0.0006078592
     Dependency     Sentiment
r1 -0.043613933 -0.0069606574
r2 -4.728320137  0.5539359919
r3  0.986201649  0.0397327543
r4  1.560947698 -0.1219658120
r5  0.001186513 -0.0004926636
```

Appendix 17

	ECT1	ECT2	ECT3	ECT4	ECT5	Intercept
Equation CC	-2.6307(0.7850)*	-0.1784(0.1099)	-0.1217(0.1162)	-0.7441(0.2885)*	3.3977(31.5812)	32.5155(15.9092).
Equation Income	0.6370(1.5056)	-0.1273(0.2107)	0.1533(0.2230)	-0.3277(0.5534)	-51.7133(60.5712)	10.0709(30.5131)
Equation Unempl	0.8936(0.6927)	-0.0614(0.0970)	-0.4386(0.1026)**	0.1612(0.2546)	35.0723(27.8674)	-11.9349(14.0384)
Equation IR	-0.8667(1.3147)	-0.0170(0.1840)	-0.0271(0.1947)	-0.0706(0.4832)	28.3481(52.8896)	-6.4111(26.6435)
Equation Gini	0.0167(0.0061)*	0.0004(0.0009)	-0.0007(0.0009)	0.0007(0.0022)	-0.2138(0.2454)	0.1631(0.1236)
Equation CPI	-2.3388(2.8026)	-0.1997(0.3923)	-0.2806(0.4150)	-0.7176(1.0301)	-41.5420(112.7484)	50.5115(56.7978)
Equation Cond	1.7203(1.7806)	0.1935(0.2492)	0.0574(0.2637)	0.5587(0.6544)	40.8498(71.6331)	-22.4132(36.0857)
Equation Return	12.5396(23.6846)	-9.2272(3.3150)*	-5.3391(3.5072)	-20.6225(8.7053)*	-2193.5590(952.8434).	878.4772(480.0015)
Equation Dep	-0.3527(0.6035)	-0.0433(0.0845)	-0.0226(0.0894)	-0.3365(0.2218)	23.2228(24.2810)	15.7555(12.2317)
Equation Sent	-23.6358(6.2049)**	-7.3194(0.8685)***	-5.7424(0.9188)***	-18.0567(2.2806)***	-1363.8181(249.6281)***	787.3653(125.7519)***

	CC -1	Income -1	Unempl -1	IR -1	Gini -1	CPI -1
Equation CC	0.9167(0.7424)	0.1448(0.0817)	0.2703(0.1737)	0.7559(0.3581).	0.8894(44.7699)	0.0347(0.1201)
Equation Income	-1.1283(1.4238)	-0.9683(0.1567)***	-0.5288(0.3332)	0.0326(0.6868)	33.0844(85.8666)	-0.1166(0.2304)
Equation Unempl	-0.2186(0.6551)	-0.0940(0.0721)	-0.3532(0.1533).	-0.1518(0.3160)	-59.4032(39.5052)	0.0399(0.1060)
Equation IR	1.4090(1.2433)	0.0325(0.1368)	0.1866(0.2909)	0.3607(0.5997)	-18.8816(74.9771)	-0.1518(0.2012)
Equation Gini	-0.0142(0.0058)*	-0.0001(0.0006)	-0.0001(0.0013)	0.0016(0.0028)	-0.3473(0.3479)	0.0014(0.0009)
Equation CPI	1.6621(2.6504)	0.2652(0.2917)	0.2507(0.6202)	0.2726(1.2784)	177.5423(159.8337)	0.6930(0.4289)
Equation Cond	-1.6163(1.6839)	-0.0712(0.1853)	-0.3234(0.3940)	-0.0595(0.8122)	-121.4252(101.5482)	-0.3458(0.2725)
Equation Return	-7.4412(22.3984)	3.3193(2.4653)	2.3445(5.2413)	-1.6239(10.8042)	1147.9438(1350.7647)	-3.5519(3.6244)
Equation Dep	0.1035(0.5708)	0.0308(0.0628)	0.1676(0.1336)	0.1031(0.2753)	-57.6200(34.4212)	0.1896(0.0924).
Equation Sent	11.8386(5.8680).	3.8488(0.6459)***	1.7011(1.3731)	9.0769(2.8305)*	1587.1478(353.8764)**	-0.3662(0.9495)

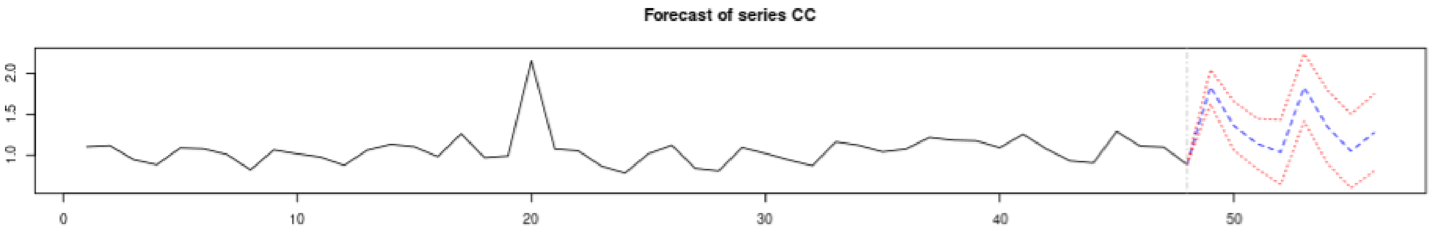
	Cond -1	Return -1	Dep -1	Sent -1	CC -2	Income -2
Equation CC	-0.3172(0.4701)	0.0289(0.0193)	0.2403(0.3602)	-0.0088(0.0176)	0.3939(0.6229)	0.1678(0.0882).
Equation Income	-0.3229(0.9015)	-0.0029(0.0371)	-0.5738(0.6908)	-6.4e-05(0.0338)	-1.0474(1.1946)	-0.9929(0.1692)***
Equation Unempl	1.2483(0.4148)*	-0.0263(0.0171)	0.1591(0.3178)	0.0001(0.0155)	0.1062(0.5496)	-0.0730(0.0779)
Equation IR	0.0528(0.7872)	-0.0114(0.0324)	-0.2195(0.6032)	0.0045(0.0295)	1.3556(1.0431)	0.2877(0.1478).
Equation Gini	0.0019(0.0037)	0.0003(0.0002)	0.0030(0.0028)	0.0001(0.0001)	-0.0101(0.0048).	0.0001(0.0007)
Equation CPI	-1.8340(1.6782)	0.1021(0.0691)	-0.1938(1.2859)	-0.0088(0.0628)	0.2804(2.2237)	0.4106(0.3150)
Equation Cond	-0.0192(1.0662)	-0.0399(0.0439)	0.1731(0.8170)	0.0084(0.0399)	-1.2577(1.4128)	-0.1123(0.2001)
Equation Return	9.0381(14.1822)	-0.7513(0.5837)	-24.8489(10.8676).	1.1020(0.5310).	-3.2530(18.7929)	-0.4015(2.6623)
Equation Dep	1.1413(0.3614)*	0.0429(0.0149)*	-0.3214(0.2769)	-0.0036(0.0135)	-0.0413(0.4789)	-0.0106(0.0678)
Equation Sent	-5.2817(3.7155)	0.2187(0.1529)	5.6429(2.8471).	0.8382(0.1391)***	6.7436(4.9234)	1.0403(0.6975)

	Unempl -2	IR -2	Gini -2	CPI -2	Cond -2	Return -2
Equation CC	0.1348(0.1512)	0.3246(0.2417)	18.4815(42.1615)	0.0528(0.1553)	-0.4623(0.3892)	0.0179(0.0158)
Equation Income	-0.2605(0.2900)	-0.1068(0.4636)	-5.0600(80.8637)	-0.0436(0.2978)	-0.5410(0.7464)	-0.0057(0.0304)
Equation Unempl	-0.2007(0.1334)	-0.0083(0.2133)	-50.0312(37.2035)	-0.2170(0.1370)	1.4618(0.3434)**	-0.0437(0.0140)*
Equation IR	0.1393(0.2532)	-0.2406(0.4048)	36.7332(70.6087)	-0.1536(0.2601)	0.3041(0.6517)	-0.0099(0.0265)
Equation Gini	0.0006(0.0012)	-0.0013(0.0019)	-0.6177(0.3276).	0.0011(0.0012)	0.0009(0.0030)	0.0001(0.0001)
Equation CPI	-0.3867(0.5398)	1.0389(0.8629)	-40.8650(150.5214)	0.9373(0.5544)	-1.9529(1.3893)	0.1043(0.0565)
Equation Cond	-0.2874(0.3429)	-1.1511(0.5482).	22.7815(95.6317)	-0.2224(0.3522)	0.0995(0.8827)	-0.0378(0.0359)
Equation Return	9.7355(4.5617).	6.2672(7.2926)	2010.6358(1272.0654)	-7.0190(4.6855)	3.2245(11.7412)	-0.5981(0.4777)
Equation Dep	0.3105(0.1162)*	0.2469(0.1858)	-103.3373(32.4157)*	-0.0352(0.1194)	0.8955(0.2992)*	0.0240(0.0122).
Equation Sent	1.1946(1.1951)	10.4506(1.9105)***	793.9861(333.2586)*	3.5915(1.2275)*	-9.0329(3.0760)*	0.2509(0.1251).

	Dep -2	Sent -2	CC -3	Income -3	Unempl -3	IR -3
Equation CC	-0.6218(0.4671)	0.0050(0.0185)	0.0986(0.4561)	0.1293(0.0634).	0.0935(0.2138)	0.3349(0.2450)
Equation Income	-0.5913(0.8959)	-0.0135(0.0354)	-0.7757(0.8749)	-1.0026(0.1215)***	-0.0828(0.4100)	-0.0712(0.4698)
Equation Unempl	-0.3200(0.4122)	0.0646(0.0163)**	0.5458(0.4025)	0.0334(0.0559)	-0.7600(0.1886)**	0.1549(0.2162)
Equation IR	-0.2330(0.7823)	0.0090(0.0309)	0.6180(0.7639)	0.2291(0.1061).	-0.4918(0.3580)	0.1577(0.4103)
Equation Gini	-0.0015(0.0036)	-6.5e-05(0.0001)	-0.0075(0.0035).	-3.9e-05(0.0005)	0.0001(0.0017)	-0.0012(0.0019)
Equation CPI	1.2145(1.6677)	-0.0253(0.0660)	-0.8101(1.6285)	0.3032(0.2262)	0.1491(0.7632)	0.3629(0.8746)
Equation Cond	-1.0922(1.0595)	0.0096(0.0419)	-0.1221(1.0346)	-0.1720(0.1437)	-0.4063(0.4849)	-0.1796(0.5556)
Equation Return	-0.6499(14.0937)	0.7679(0.5575)	13.0922(13.7625)	0.7545(1.9117)	3.9626(6.4497)	12.8013(7.3911)
Equation Dep	-0.2522(0.3591)	0.0049(0.0142)	-0.0238(0.3507)	-0.0086(0.0487)	0.2517(0.1644)	-0.1560(0.1883)
Equation Sent	0.5903(3.6923)	0.2905(0.1461).	5.1564(3.6055)	1.5466(0.5008)*	7.5759(1.6897)**	9.9151(1.9363)***

	Gini -3	CPI -3	Cond -3	Return -3	Dep -3	Sent -3
Equation CC	0.0786(30.4466)	-0.0517(0.1112)	-0.3207(0.3044)	0.0034(0.0105)	-0.0467(0.5492)	0.0017(0.0187)
Equation Income	-60.6065(58.3951)	-0.0359(0.2133)	-0.2889(0.5839)	-0.0115(0.0202)	-0.9279(1.0534)	0.0149(0.0358)
Equation Unempl	-0.7214(26.8662)	0.1710(0.0981)	0.6207(0.2686)*	0.0028(0.0093)	-0.7250(0.4846)	0.0118(0.0165)
Equation IR	32.0212(50.9895)	0.1723(0.1862)	-0.4588(0.5098)	0.0305(0.0176)	0.2139(0.9198)	-0.0166(0.0313)
Equation Gini	-0.3489(0.2366)	0.0007(0.0009)	0.0008(0.0024)	5.7e-05(8.2e-05)	-0.0026(0.0043)	0.0001(0.0001)
Equation CPI	-13.4680(108.6978)	0.1915(0.3970)	-0.7083(1.0868)	0.0689(0.0376)	2.5022(1.9608)	0.0062(0.0667)
Equation Cond	-16.5763(69.0597)	0.1109(0.2522)	-0.5591(0.6905)	-0.0242(0.0239)	-1.0065(1.2458)	-0.0129(0.0424)
Equation Return	2131.1980(918.6121)*	-2.7330(3.3549)	6.6796(9.1850)	-0.5782(0.3177)	11.3354(16.5708)	0.8729(0.5638)
Equation Dep	-24.6545(23.4087)	-0.1219(0.0855)	0.6096(0.2341)*	0.0060(0.0081)	-0.6901(0.4223)	0.0188(0.0144)
Equation Sent	291.8589(240.6601)	-1.9146(0.8789).	-2.0502(2.4063)	-0.1976(0.0832)*	10.4864(4.3412)*	0.5051(0.1477)**

Appendix 18



Appendix 19

```
> model7<-arima(UnivariateVAR_CC,order = c(2,1,1),xreg = IndVar_lagged_)
Warning messages:
1: In log(s2) : NaNs produced
2: In stats::arima(x = x, order = order, seasonal = seasonal, xreg = xreg,  :
  possible convergence problem: optim gave code = 1
> summary(model7)

Call:
arima(x = UnivariateVAR_CC, order = c(2, 1, 1), xreg = IndVar_lagged_)

Coefficients:
          ar1      ar2      ma1   Income     GDP   Unempl       IR     Gini      CPI  Condition  Competition   Return  Dependency  Edyucation      Size
      -0.5359  -0.2903  -0.9999  -0.0019  0.0091  -0.0392  -0.2028  14.9326  -0.0563    -0.0319      -0.0639  -0.0148     -0.0705     -0.1231  -14.8164
s.e.   0.1673   0.1556   0.0554   0.0160  0.0190   0.0383   0.0456   7.2356   0.0243     0.0790       0.0661   0.0038      0.0352      0.0271       NaN
      Sentiment
         0.0061
s.e.     0.0038

sigma^2 estimated as 0.02522:  log likelihood = 17.09,  aic = -2.18

Training set error measures:
              ME RMSE MAE MPE MAPE
Training set NaN  NaN NaN NaN  NaN
Warning messages:
1: In sqrt(diag(x$var.coef)) : NaNs produced
2: In trainingaccuracy(object, test, d, D) :
  test elements must be within sample
```

Appendix 20

```
> model1VAR<-vec2var(cointegration, r = 5)
> summary(model1VAR)
              Length Class  Mode
deterministic   10   -none- numeric
A                3   -none- list
p                1   -none- numeric
K                1   -none- numeric
y              400   -none- numeric
obs              1   -none- numeric
totobs           1   -none- numeric
call             3   -none- call
vecm             1   ca.jo  S4
datamat       1845   -none- numeric
resid          450   -none- numeric
r                1   -none- numeric
> model1VAR

Coefficient matrix of lagged endogenous variables:

A1:
             CC.l1     Income.l1       CPI.l1       Gini.l1    Unempl.l1        IR.l1       Cond.l1     Return.l1        Dep.l1       Sent.l1
CC     -0.197885524  0.0609725735  0.117211157    45.7554935  0.009852926 -0.093925596 -0.1056770960 -0.0146329327 -3.903119e-01 -0.0051172586
Income -0.200233008 -0.2434307823  0.644387593   -97.3561043  1.197464097 -1.684951766 -0.1600691964  0.0615885709 -1.908979e+00 -0.1208782707
CPI     0.514444284 -0.1006211146  1.673598283    96.1289332 -0.775440448 -0.960114078 -0.0062344333  0.0427418752 -1.452732e+00 -0.0556375230
Gini    0.003395698 -0.0002958102 -0.001449041     0.5375402  0.001148549 -0.000194275  0.0006630015 -0.0001221349  4.566487e-04  0.0001748146
Unempl -0.056276129 -0.1154838651 -0.355879153   -37.5035442  0.439866830  0.038043127 -0.4221436778 -0.0160834005  7.224940e-01 -0.0418560355
IR      0.378963558 -0.1283195526 -0.138570606   -38.8170938 -0.337286607  1.022480058 -0.4530448805 -0.0145029986  8.215237e-01  0.0083945670
Cond   -0.351045395  0.0789524716 -0.375990622   -36.6703448 -0.068045038  0.230734492 -0.3410033981 -0.0468111268  6.203382e-01  0.0313720286
Return -2.214057237 -0.0661287026 -6.978382327 -1616.0625696  0.936630630 -3.485288328 -2.1893158613  0.0039336355 -1.937739e+01  0.2514291534
Dep     0.048154421  0.0797663172 -0.003612464    10.3857813  0.096221917  0.062024421  0.0465132329  0.0042997239  8.250938e-01 -0.0144699542
Sent   -3.528485198  0.6941598447  0.180744115   587.5868838 -3.307294402 -1.655890383  2.4795101228  0.1571001588 -5.767915e+00  0.8612502272


A2:
             CC.l2     Income.l2        CPI.l2       Gini.l2     Unempl.l2        IR.l2      Cond.l2     Return.l2       Dep.l2       Sent.l2
CC     -0.226404167  0.0394211986 -0.1693926471   -25.1577249  2.333402e-02  0.047058250 -0.103262907 -4.518785e-03 -0.071322776  0.0083761291
Income  0.876958915  0.4653269058 -0.2949430871    10.5885919 -5.839019e-01  2.478514132  0.214435286  6.675936e-03  2.687406593  0.1541817668
CPI    -0.031430735  0.0447893426 -0.5039683368  -188.8230338  4.426395e-01  1.464789290 -0.112557513  8.527522e-03  2.272657022  0.0341886668
Gini    0.005074609 -0.0002027036  0.0004157102     0.1657369 -1.804646e-05 -0.001190811 -0.000566394 -5.725513e-05 -0.003450046 -0.0001204047
Unempl  0.220774667 -0.0042426694  0.1721437778    33.7835874 -1.773463e-02 -0.053789804  0.053815610 -1.616622e-02 -1.903017639  0.0425861420
IR     -0.120296326 -0.0375964164 -0.0413479036    24.2943615  1.709656e-01 -0.881018816 -0.137140757 -8.950008e-03 -0.795457881 -0.0101609393
Cond   -0.321278920 -0.0445956345  0.0385938779    75.3241588 -3.920109e-01 -0.715991910 -0.190134420 -1.846462e-02 -1.104783220 -0.0059845942
Return  0.184771230 -0.5897708627 -0.9401512523  1551.5387641  5.050291e+00 -7.727174044 -3.494962926 -3.523242e-01  5.720553215 -0.1662256477
Dep    -0.344910443 -0.0243277206  0.1198267798   -21.0835366 -1.813880e-01 -0.022390869  0.122912867  5.837197e-03  0.017395138  0.0130895100
Sent    1.992251387 -0.4612244692  1.6004849508 -1060.2459014  1.521238e+00  0.137544364  0.293489439  4.213881e-02  2.099143963 -0.6366912915


A3:
             CC.l3     Income.l3        CPI.l3      Gini.l3   Unempl.l3        IR.l3      Cond.l3     Return.l3       Dep.l3       Sent.l3
CC     -0.104838471 -1.030007e-02  6.889943e-02    6.4075256  0.07752548 -0.015021185  0.446353353 -4.429363e-03  0.298211330  6.628001e-03
Income -0.015688158 -2.708497e-01  5.424845e-01 -224.5952579 -0.32738085 -2.859724047  1.021178639  3.871783e-02 -0.195658064 -1.205068e-01
CPI     0.982568887  8.694891e-02 -3.287203e-02   33.6570213  0.60009233 -0.265653253  1.538965200  1.847644e-02 -0.515258060  2.753105e-04
Gini    0.002161488  9.788572e-05 -4.764017e-06    0.1935738 -0.00131238  0.001252879 -0.001473594  7.046802e-05  0.003030414 -1.745638e-06
Unempl  0.301271002  2.289068e-01 -1.319594e-01   58.5355961  0.16904497  0.318760163 -0.719973539  1.592745e-02  1.088440713 -3.433012e-02
IR      0.033820592  5.278042e-02  7.525905e-02   36.0676935 -0.23127083  0.780594646 -0.889097863  4.738076e-03 -0.061527015 -3.718749e-03
Cond    0.325661397 -9.744332e-02 -3.680393e-02   27.5498769 -0.10270576  0.285934312 -0.687808084 -1.476463e-02  0.160379339 -3.589269e-02
Return  6.734201305 -1.254015e+00  2.601373e+00  141.3788845 -7.95615959  3.299248422 -7.948626466 -3.646187e-02  8.431216660 -8.236278e-02
Dep    -0.127310051 -5.808588e-02 -1.000371e-01   -1.1145670  0.17047142 -0.172491491  0.187609568 -1.164337e-02  0.121191115 -1.108547e-02
Sent   -2.106716718 -1.049903e+00 -9.798348e-01 -107.4214691  2.94684916 -1.908810267  5.784127784  8.338225e-02  4.241834545  1.728796e-01

Coefficient matrix of deterministic regressor(s).

           constant
CC       0.07241391
Income  95.38251590
CPI     -8.67741410
Gini     0.03303727
Unempl -11.92534842
IR       3.93108527
Cond    10.14732662
Return 418.52933723
Dep      5.75577006
Sent   158.54054373
```

Figure 21

```
> Indvar_lag<-cbind(IndVar_lagged_$Income,IndVar_lagged_$GDP,IndVar_lagged_$Unempl,IndVar_lagged_$IR,IndVar_lagged_$Gini,IndVar_lagged_$CPI,IndVar_lagged_$Con
dition,IndVar_lagged_$Competition,IndVar_lagged_$Return,IndVar_lagged_$Edyucation,IndVar_lagged_$Size,IndVar_lagged_$Dependency,IndVar_lagged_$Sentiment)
> mod<-auto.arima(UnivariateVAR_CC,xreg = Indvar_lag)
> mod
Series: UnivariateVAR_CC
Regression with ARIMA(0,0,0) errors

Coefficients:
       xreg1   xreg2   xreg3    xreg4    xreg5    xreg6   xreg7    xreg8    xreg9   xreg10  xreg11   xreg12  xreg13
      0.0046  0.0043  0.0263  -0.0965  10.1375  -0.0073  0.0236  -0.0616  -0.0109  -0.0300  1.2597  -0.0626  0.0061
s.e.  0.0153  0.0200  0.0496   0.0611   9.8957   0.0305  0.0823   0.0694   0.0044   0.0434  2.0366   0.0581  0.0056

sigma^2 = 0.04186:  log likelihood = 15.64
AIC=-3.27   AICc=9.46   BIC=22.93
```

Appendix 22

```
> Box.test(residuals(FirstVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(FirstVAR[["varresult"]][["DCC"]])
X-squared = 0.053244, df = 1, p-value = 0.8175

> Box.test(residuals(SecondVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(SecondVAR[["varresult"]][["DCC"]])
X-squared = 0.038849, df = 1, p-value = 0.8437

> Box.test(residuals(ThirdVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(ThirdVAR[["varresult"]][["DCC"]])
X-squared = 0.72288, df = 1, p-value = 0.3952

> Box.test(residuals(FourVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(FourVAR[["varresult"]][["DCC"]])
X-squared = 0.037401, df = 1, p-value = 0.8467

> Box.test(residuals(FifthVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(FifthVAR[["varresult"]][["DCC"]])
X-squared = 1.1601, df = 1, p-value = 0.2814

> Box.test(residuals(SixthVAR[["varresult"]][["DCC"]]))

        Box-Pierce test

data:  residuals(SixthVAR[["varresult"]][["DCC"]])
X-squared = 1.0429, df = 1, p-value = 0.3071
```