What can credit vintages tell us about non-performing loans?

By: Santiago Gamba-Santamaria
Luis Fernando Melo-Velandia
Camilo Orozco-Vanegas

Borradores de
ECONOMÍA

No. 1154
2021

BANCO DE LA REPÚBLICA
COLOMBIA

tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

¿Qué nos pueden decir las cosechas de crédito sobre la cartera en mora?*

Santiago Gamba-Santamaria† Luis Fernando Melo-Velandia‡
Camilo Orozco-Vanegas§

Resumen

Usando información de cosechas de crédito, en este documento descomponemos la cartera en mora en un componente que captura la evolución de la capacidad de pago de los deudores y otro componente que captura los cambios en la toma de riesgo de crédito del sistema financiero al momento del desembolso. Utilizamos estimadores intrínsecos y técnicas de regresión penalizadas para solucionar el problema de multicolinealidad perfecta asociado a la estimación de los parámetros de los modelos. Encontramos que estos dos tipos de componentes han evolucionado de manera diferente a lo largo del tiempo y que buenas condiciones económicas y condiciones financieras laxas mejoran la capacidad de pago de los deudores para cumplir con sus obligaciones y, a su vez, tienden a coincidir con el otorgamiento de préstamos de mayor riesgo por parte del sistema financiero. Finalmente, recomendamos el uso de esta metodología como herramienta de política de fácil aplicación por parte de las autoridades financieras y económicas que disponen de un flujo constante de información de cosechas de crédito. A través de ella las autoridades podrían identificar el origen de la materialización del riesgo crediticio y contener la toma de riesgo del sistema financiero.

Clasificación JEL: C13; C20; G21.

Palabras clave: cosechas de crédito; cartera en mora; regresiones penalizadas; estimadores intrínsecos.

*Agradecemos los comentarios y discusiones con Kenneth Land, Daniel Osorio, Camilo Sánchez, Wilmar Cabrera y Stev Abril. También agradecemos a la Superintendencia Financiera de Colombia por proveernos la información usada en este artículo. Como es usual, los errores y omisiones son de exlusiva responsabilidad de los autores.

†Economista en el Banco de la República, Colombia. Correo: sgambasa@banrep.gov.co.
‡Econometrista Principal en el Banco de la República, Colombia. Correo: lmelovel@banrep.gov.co.
§Asistente de investigación en el Banco de la República, Colombia. Correo: caorozcov@unal.edu.co.

What can credit vintages tell us about non-performing loans?*

Santiago Gamba-Santamaria† Luis Fernando Melo-Velandia‡
Camilo Orozco-Vanegas§

Abstract

Using Colombian credit vintage data, we decompose the non-performing loans into one component that captures the evolution of the payment capacity of borrowers, and other component that captures changes in the credit risk taken by the financial system at the time of loan disbursement. We use intrinsic estimators and penalized regression techniques to overcome the perfect multicollinearity problem that the model entails. We find that these two type of components have evolved differently over time, and that good economic conditions and loose financial conditions improve the payment capacity of borrowers to meet their obligations, and in turn, they tend to coincide with the financial system engaging in riskier loans. Finally, we advocate the use of this methodology as a policy tool that is easy to apply by financial and economic authorities that dispose of a constant flow of credit vintage information. Through it, they will be able to identify the origin of the credit risk materialization and curb the risk taken by the financial system.

JEL code: C13; C20; G21.

Key words: credit vintages; non-performing loans; elastic net regressions; intrinsic estimators.

*This article benefited greatly from the useful comments and discussions with Kenneth Land, Daniel Osorio, Camilo Sánchez, Wilmar Cabrera and Stev Abril. We are grateful to the Financial Superintendence of Colombia for making the information available for us. The errors and omissions are the sole responsibility of the authors.

†Economist at Banco de la República (Central Bank of Colombia, Colombia. E-mail: sgambasa@banrep.gov.co.

‡Senior Econometrician at the Banco de la República (Central Bank of Colombia). E-mail: lmelovel@banrep.gov.co.

§Intern at Banco de la República (Central Bank of Colombia). E-mail: caorozcov@unal.edu.co.

1 Introduction

Credit is a crucial activity for the proper function of any economy based on saving and investment, typically performed by financial intermediaries that accept deposits and grant loans. One of the main risks they face is credit risk, that is the risk of borrowers defaulting on their loans. This is the reason why there is wide research and regulation on credit risk management. Most of the literature focuses on studying the ex-ante assessment of loans. Likewise, credit risk regulation is generally based on designing credit risk management systems like provisioning schemes.

In contrast, ex-post materialization of credit risk has been studied with lesser intensity, despite the fact that non-performing loans regularly increase during economic recessions and financial crises. At first, defaults signal the reduction on household's and firm's incomes associated to economic downturns. However, defaults generate losses to financial intermediaries that, depending on their magnitude, might create obstacles for new credit supply. Therefore, defaults play a shock amplifier role in the economy and the financial system.

Moreover, credit risk materialization does not only emerge as a consequence of economic downturn, it can also be the cause of financial crises when it is generated from loans granted when credit conditions are too loose. An example of this can be found in the Global Financial Crisis of 2008-2009, that originated, among other reasons, from the systematic granting of mortgage loans to borrowers that did not have enough payment capacity or income stability to honor their financial obligations.

Hence, during crises, non-performing loans serve as a measure of the impact on both the bank's loses and the income reduction of households and firms. However, few research articles have studied their determinants. This literature, using aggregate data, bank-level data and loan-level data, has found two types of determinants. First, macroeconomic variables impact credit quality in the short run, mainly by changing the payment capacity of borrowers. On the other hand, bank's variables associated to lending standards affect the credit risk materialization in the longer run.

Alternatively, we employ credit vintage information to perform a decomposition of the non-performing loan's ratios (henceforward NPL). This decomposition allows us to track and study the components that capture each one of the two groups of mechanisms found in the literature separately. Specifically, in the first stage, we apply an *Age-Period-Cohort* methodology (henceforward APC) to credit vintages of the Colombian financial system. This enables us to decompose the dynamics of NPL into three components: Vintage Component, Period Component and Age Component. The first two components are indexed by time so they will be analyzed as time series in the next stage. In this decomposition the first component captures all characteristics associated to the loan disbursement (credit standards and bank's risk taking), while the second reflects the evolution of the payment capacity of borrowers throughout the lifetime of loans. In the second stage, using a vector autoregression model, we estimate the dynamic relationship between NPL components and key macroeconomic variables.

The contribution of this article is to offer a methodology to track and disentangle the NPL dynamics. Since the vintage information tends to be available more frequently, this methodology is more suitable than the panel data estimations found in literature for monitoring the credit risk. For instance, authorities that receive a constant flow of this type of vintage information can use this methodology as a policy tool, applying it continuously to monitor the credit risk in the financial system and its causes. Additionally, in the first stage we employ the Intrinsic Estimators and Elastic Net regression techniques, two methods that overcome the over-specification problems associated to APC models.

This article continues as follows. The next section presents a review on the NPL determinants literature and the use of estimation techniques to deal with high multicollinearity. The third section describes the data we employ and details the features of vintage information. Section 3 presents the methodology and results of both the proposed decomposition and the relation of the components to key macroeconomic variables. In Section 6 we discuss our interpretation of the decomposition and Section 7 concludes.

2 Literature Review

Most of the literature that studies the determinants of non-performing loans relies on panel data of the NPL ratios for a set of banks over a period of time. This approach identifies key macroeconomic variables correlated to NPLs and exploits cross-sectional variation among banks to identify bank-level characteristics that explain NPLs dynamics.

Among the most recurrent results, literature finds that variables associated to the business cycle, such as GDP growth and unemployment rates, are correlated with the credit defaults. The former presents a positive relationship, while the latest exhibit a positive one. Articles like Salas and Saurina (2002), Dimitrios et al. (2016) and Ghosh (2015) argue that the slow growth of economic activity and unemployment reduces the cash inflow of firms and households, hence, rendering more borrowers unable to meet their obligations.

Other macroeconomic variables have been studied as determinants. Rosch and Scheule (2004) find that the industrial production index has a relation to the NPLs. Additionally, Ghosh (2015) identifies that housing prices are inversely related to NPLs, arguing that an increment in the collateral value increases the borrowers worthiness. Similarly, Dimitrios et al. (2016) find that a greater tax burden reduces the disposable income.

Conversely, there is no consensus on the results of inflation as a determinant of NPLs. Klein (2013) and Greenidge and Grosvenor (2010) identify a positive relation of inflation and NPLs, while Rosch and Scheule (2004) find a negative correlation between these two variables. Ghosh (2015) encounters both of these results in different estimations, they state that the negative relation could reflect the reduction of the real value of the debt, but if wages are sticky, inflation would reduce the disposable income of borrowers.

On the other hand, Klein (2013) state that the level of interest rates can make loans more costly for borrowers to pay off. Nonetheless, he does not find a conclusive result, while Greenidge and Grosvenor (2010) do find the positive relation.

Similarly, literature has identified consensus as well as ambiguity concerning the bank-level results. Starting with the consensus, variables that measure the bank's intensity on a loan with respect to assets or deposits, and credit growth indicators exhibit a positive correlation with NPLs. These variables might reflect the risk attitude of banks so that lax credit standards induce high levels of NPLs.

Furthermore, profitability and efficiency turn out to be negatively related to NPLs. Some of the authors mentioned above and Makri et al. (2014) state a *moral hazard hypothesis* that consists of more profitable banks facing fewer incentives to take risks, while desperate or inefficient banks engage in unsafe behavior.

Likewise, many articles find evidence supporting the same hypothesis with bank's capitalization, however, Ghosh (2015) estimations show a negative relation between this variable and NPLs, suggesting a *too big to fail hypothesis*, in which better capitalized banks assume greater risks. In contrast, Salas and Saurina (2002) find that the size of banks is negatively correlated to NPLs.

Finally, Ghosh (2015) and Salas and Saurina (2002) examine variables related to borrowers worthiness such as loan loss provisions and the share of loans without collateral, both articles find a positive relation with NPLs.

All of the papers discussed above employ ordinary estimations that result in finding some determinants correlated with NPLs, but that none of them design an identification strategy to achieve causal evidence on their results.

Notably, the mechanisms behind the results in this literature can be classified into two groups. In the first group variables reflect the borrower capacity to pay off the loans, whereas the ones of second group are associated with the risks taken by banks. More or less, the first consists of the macroeconomic determinants and the second is comprised of the bank-level determinants. Additionally, most of this literature uses the default ratio of the whole loan portfolio at infrequent time periods, which may include loans granted last year as well as more than five years ago. Moreover, each group of mechanisms discussed would operate at different timing on the NPL. The bank-level mechanisms are supposed to affect the NPL at the moment of the loans origination, while the macroeconomic channels are supposed to impact the current performance of loans.

Our methodology exploits credit vintage data to break down the NPL dynamics into three components, two of which can be interpreted as time series. We argue that each of these components captures each of the groups of mechanisms described above. Even though vintage information is in some way similar to panel data, this methodology should be interpreted as kind of a filter that returns time series components. These series of components allow us to separately analyze the determinants associated with the loan's origination and the performance of current loans. This is to say that our methodology can identify changes to the risk-taking positions of banks, as well as changes to the payment capacity of borrowers.

The vintage analysis, as Siarka (2011) explains, has been taken from the world of wine, where the wine's taste depends on the year when the grapes were harvested. The wine's quality is determined by the natural conditions in a particular year and the properties intrinsic to the crops. So, it is possible to previously know if a given wine should be stored longer to obtain the perfect taste and texture. In the bank context, several authors have considered the analysis of the loan's quality given the period when they were disbursed, the conditions proper to them, and the influence of exogenous determinants.

In particular, Bosman (2012) explores how the cohort analysis, which comes from other sciences like demography, sociology and biostatistics, improves the vintage analysis techniques. He describes the identification problem in the cohort analysis and the distinct models developed for its solution. In fact, he considers the Intrinsic Estimator approach when the APC model is introduced. In the same way, Forster and Sudjianto (2013) consider the problem of decomposing the NPL data series into the exogenous variability (period), the maturity (age or time since origination) as a simile of the APC model. They explain the relation between these variables and the difficulties that modeling entails.

Other authors like Breeden (2016) and Breeden and Thomas (2016) consider the APC model and its implementation in stress tests on retail lending and as a mechanism to forecast the loan performance at a point-in-time for the loan's lifetime. Particularly, they consider an APC model where the Period Component is in function of other macroeconomic variables to avoid a specification error. The influence of the macroeconomic variables in the NPL occurrence has been a topic of large study in vintage literature.

Our methodology differs from the applications above by relying only on vintage information and aggregate macroeconomic time series to achieve a comprehensive decomposition of the NPL dynamics. In this way, our methodology offers a simple tool to track changes in the payment capacity of borrowers as well as the credit standards of bank's overtime, and to identify the macroeconomic environments related to each one of these components.

3 Methodology

The methodology is divided into two stages. The first stage involves the estimation of the Age, Period, and Vintage (cohort) components on the ratio of non-performing loans using two methods: Intrinsic Estimators and Penalized Regressions. In the second stage, we estimate a VAR model for the Period and Vintage components and a set of macroeconomic variables.

3.1 First stage: Age-Period-Cohort decomposition

The Age-Period-Cohort analysis has been used in many areas such as health science, biostatistics, and demography, among others. This methodology allows us to decompose a response variable in these three components. This type of analysis can be used to track subjects during a period of time since the date they were born. In that way, at each period of time we can observe various subjects that were born at different moments. Specifically, every observation is indexed by: the age of the subjects, the period that corresponds to the calendar time, and the cohort which is the moment of the subject's birth.

In our case, we use credits as subjects. The endogenous variable is the non-performing loans ratios (NPL). The objective of this methodology is to decompose the NPL of a credit into the Age, Period and Cohort (Vintage) components as follows:

$$NPL_{i,n,t} = \mu + \alpha_i + \beta_t + \gamma_n + \varepsilon_{i,n,t} \tag{1}$$

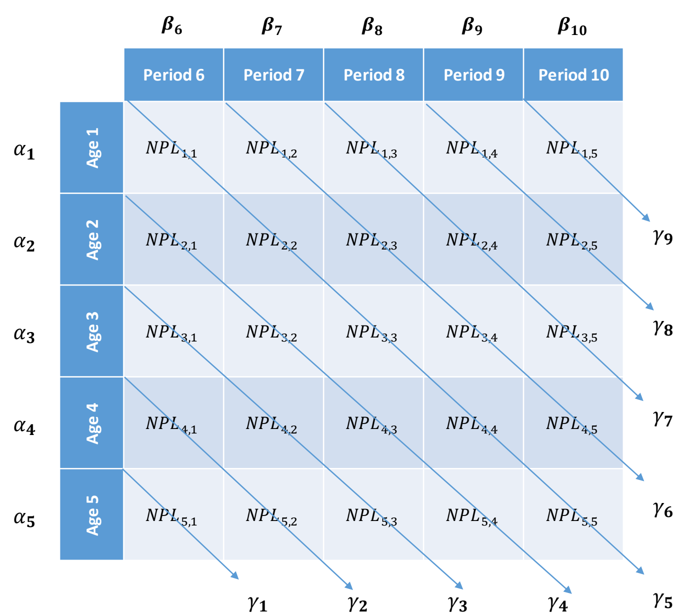
Where $NPL_{i,n,t}$ corresponds to a credit disbursed at vintage n, observed in the period t when it is i months old. μ is a constant effect, while α_i, β_t and γ_n correspond to age, period and cohort effects, respectively. Finally, $\varepsilon_{i,n,t}$ is an *iid* error term.

There is a linear relationship among the previous three components. An easy way to see this association is by using the 'Lexis Diagram' in Figure 1. In this table, the NPL information is presented as a matrix where the rows represent the ages (i), the columns are the periods (t), and the table diagonals are associated with the vintages (n) with $n = t - i$. Finally, the element of each cell represents the observation of the response variable for the corresponding age, period and cohort.

In this example there are five ages (1 to 5), $\bar{A}$, and five periods (6 to 10), $\bar{P}$. Then, the total number of vintages is $\bar{C} = \bar{A} + \bar{P} - 1 = 9$. Given the relationship among the three components, we only need to define two dimensions in order to obtain the vintage information. For example, if we select the cell $(1,1)$ the $NPL_{1,1}$ is 1 unit of time old and we are observing it at period 6, which implies that the NPL belongs to the vintage 5.

According with equation (1), the labels of the rows, columns and diagonals of Figure 1 represent the coefficients of age (α_i), period (β_t) and vintage (γ_n), respectively. These coefficients refer to the average effect of the corresponding age, period, or vintage on the response variable.

Figure 1: Lexis Diagram



The APC literature decomposes the response variable in several factors. Model (1) includes all three factors. However, there are other representations of the decomposition. The most parsimonious one is the single-factor model, which only include either age (A), period (P), or cohort (C) . The second kind of models uses two factors, either age-period (AP), age-cohort (AC), or period-cohort (PC).

Specifically, the single-factor models are:

$$NPL_{i,n,t} = \mu + \alpha_i + \varepsilon_{i,n,t}, \tag{2}$$
$$NPL_{i,n,t} = \mu + \beta_t + \varepsilon_{i,n,t}, \tag{3}$$
$$NPL_{i,n,t} = \mu + \gamma_n + \varepsilon_{i,n,t}, \tag{4}$$

where $\varepsilon_{i,n,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$. The total number of parameters of (2), (3) and (4) is $\bar{A}+1$, $\bar{P}+1$, and $\bar{C}+1$, respectively.

On the other hand, the models with two factors are:

$$NPL_{i,n,t} = \mu + \alpha_i + \beta_t + \varepsilon_{i,n,t} \tag{5}$$
$$NPL_{i,n,t} = \mu + \alpha_i + \gamma_n + \varepsilon_{i,n,t} \tag{6}$$
$$NPL_{i,n,t} = \mu + \beta_t + \gamma_n + \varepsilon_{i,n,t} \tag{7}$$

with $\varepsilon_{i,n,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$.

Finally, the three factors model is represented by equation (1) assuming $\varepsilon_{i,n,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$.

The model (1) can also be represented as:

$$NPL_{i,n,t} = \mu + \sum_{i=1}^{A} \alpha_i \, age_i + \sum_{n=1}^{C} \gamma_n \, vintage_n + \sum_{t=1}^{P} \beta_t \, period_t + \varepsilon_{i,n,t}, \tag{8}$$

where age_i, $vintage_n$ and $period_t$ are a set of dummies that indicate the corresponding age, cohort and period, respectively. Models (2) to (7) can be specified in a similar way.

The estimation of the parameters of equations (2) to (7) can be performed through traditional techniques like OLS with restrictions such as centralization of the parameters. Nevertheless, the estimation for model (8) with three factors is not plausible using these techniques. In this case, the linear relationship among the variables causes perfect multicollinearity.

There are several approaches to solve this estimation problem. In this paper, we use two techniques: the first one corresponds to the Intrinsic Estimators method (IE), proposed by Fu (2000); and the second approach is penalized regressions (Zou and Hastie, 2005). In the following sections we briefly explain these techniques.

3.1.1 Intrinsic estimators

This approach is a constraint-based method developed by Fu (2000) as a way to derivate the APC estimators. This method was introduced to solve the perfect multicollinearity problem of this model by means of estimable functions and the singular value decomposition of matrices.

According to Yang et al. (2008), equations (1) to (7) can be treated as fixed-effects generalized linear models after centering the parameters to zero[1]. This restriction is used to overcome the overparameterization presented in these equations, specifically:

$$\sum_{i=1}^{\bar{A}} \alpha_i = 0; \qquad \sum_{t=1}^{\bar{P}} \beta_t = 0; \qquad \sum_{n=1}^{\bar{C}} \gamma_n = 0$$

Following Yang et al. (2008), after the reparameterization with the centralization condition, we can express model (1) in a matrix form:

$$\mathbf{Y} = \mathbf{XB} + \varepsilon \tag{9}$$

Where $\mathbf{Y}$ is the vector that includes the values of the response variable, in our case NPL. $\mathbf{X}$ is the regression design matrix composed of $m-$dummy variable column vectors $\left(m = 1 + [\bar{A}-1] + [\bar{P}-1] + [\bar{A}+\bar{P}-2]\right)$, and $\mathbf{B}$ a vector with the parameters of the model:

$$\mathbf{B} = \left(\mu, \alpha_1, \ldots, \alpha_{\bar{A}-1}, \beta_1, \ldots, \beta_{\bar{P}-1}, \gamma_1, \ldots, \gamma_{\bar{A}+\bar{P}-2}\right)^T \tag{10}$$

Note that the linear relationship among age, period and cohort causes the design matrix $\mathbf{X}$ to be less than full rank; hence, the inverse associated with traditional estimator of $\mathbf{B}$ does not exist.

Yang et al. (2008) remove the influence of the design matrix on the coefficient estimates through a principal component regression estimation. Specifically, IE estimates a constrained parameter vector $\mathbf{B}_0$, which is the

[1] Fu (2018) shows that the coefficient centralization restriction is the one that yields the smallest variance of parameter estimates.

projection of the unconstrained parameter vector **B**, and applies an inverse orthonormal transformation to it. The inverse transformation makes the IE parameter estimates directly interpretable as age, period, and cohort effects. A detailed algorithm of the IE estimation procedure is presented in Fu (2018).

The IE approach has several advantages that make it suitable for the APC analysis. First, as Fu (2018) notes, it provides consistent estimates of the APC model parameters.[2] Second, this author also proposes an additional algorithm for calculating the standard error of the parameters.

Nevertheless, this methodology has an important disadvantage, it requires the data to be balanced in the age and period dimensions. To explain this restriction, consider Figure (1). The Lexis Diagram implies a balanced design if there are $\bar{P}$ observations in each age category, and there are $\bar{A}$ observations for each period, where $\bar{A}$ and $\bar{P}$ are the total age and period categories, respectively. So, the data is unbalanced if there are any missing values in one or more cells of the Lexis Diagram. In this case, it is necessary to reduce the sample ignoring some period or age categories to obtain a balanced matrix.

3.1.2 Penalized regressions

Another technique to overcome the multicollinearity problem associated with model (8) is to use penalized regressions. A general case of these methodologies is Elastic Net,. This technique can be viewed as an extension of OLS methodology adding parameter restrictions in order to penalize the model for having too many variables. In this case, the loss function is the following:

$$\mathscr{L}(\mathbf{B},\vartheta,\lambda) = \|\mathbf{Y}-\mathbf{XB}\|_2^2 + \lambda\left[(1-\vartheta)\tfrac{1}{2}\|\mathbf{B}\|_2^2 + \vartheta\|\mathbf{B}\|_1\right] \tag{11}$$

Where **y** is the vector with the response variable values, NPL. **X** is the design matrix, and **B** the vector with the parameters. In terms of equation (8), **X** is a matrix including columns of dummies related to the different periods, ages and cohorts, and **B** includes all the parameters of this equation. $\lambda \in [0,\infty)$ and $\vartheta \in [0,1]$ are the penalizing parameters, and $\|\cdot\|_p$ is the $p-$norm.

The first term of the right hand side of (11) corresponds to OLS loss function and the second term to the penalization. In this methodology the regression penalty can take two particular forms depending on the value of ϑ. $\vartheta = 0$ corresponds to the Ridge penalty, and $\vartheta = 1$ to the Lasso penalty.

According to Friedman et al. (2010), Ridge estimator is characterized by shrinking the coefficients of the correlated predictors so that their effect on the endogenous variable is shared. This implies that in the case of m identical predictors, each estimated parameter will have a size of $\frac{1}{m}$ of their value if they were estimated individually. However, this method has the disadvantage of not being parsimonious and not having the property of sparsity, meaning that irrelevant variables may have estimated values other than zero.

On the other hand, Lasso estimator satisfies the sparsity property. Nonetheless, this method depends on the number of variables to estimate. According to Zou and Hastie (2005) and van Wieringen (2015), when working with high-dimensional data, in other words, more variables (k) than observations (N), Lasso will select a maximum of N coefficients different from zero, leaving the remaining estimated parameters equal to zero.

[2]The properties of IE estimator are analyzed in Fu (2016).

3.2 Second Stage: Time-Series Analysis

Given the results of the APC decomposition for the response variable using the methodologies described in previous sections, we obtain the age, period, and cohort (vintage hereafter) components. Additionally, it is important to note that the Period and Vintage components can be seen as time-series.

As a second stage of the methodology, we can relate the Period and Vintage components with some macroeconomic variables. In this regard, we estimate a vector auto-regressive model for each of the two mentioned components along with a set of key macroeconomic variables:

$$\mathbf{Y}_t = \mathbf{a}_0 + \sum_{i=i}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} + \boldsymbol{\varepsilon}_t \tag{12}$$

Where $\mathbf{Y}_t$ is a vector containing k macroeconomic variables and the $Component_t$ at time t. $Component_t$ refers to either $\widehat{\beta}_t$ or $\widehat{\gamma}_t$, the Period and Vintage Components estimated in the first stage, $\boldsymbol{\varepsilon}_t$ is a $k-$dimensional white noise error, $\mathbf{a}_0$ is a $(k \times 1)$ vector of intercepts, and $\mathbf{A}_i$ are fixed $(k \times k)$ coefficient matrices.

Since we expect that the economic variables affect the components but not the other way around, the coefficients of the VAR model (12) are restricted as follows:

$$\mathbf{A}_i = \begin{bmatrix} a^i_{1,1} & \cdots & a^i_{1,k} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ a^i_{k,1} & \cdots & a^i_{k,k} & 0 \\ a^i_{k+1,1} & \cdots & a^i_{k+1,k} & a^i_{k+1,k+1} \end{bmatrix}, \quad i = 1, \dots, p \tag{13}$$

In this case, the $(k+1)$ row corresponds to the $Component_t$ variable.

Once the restricted VAR is estimated, we can analyze the corresponding impulse-response functions associated with the $Component_t$.

4 Data

Non-performing loan ratios (NPL) are the shares of loans that have defaulted on the total loan portfolio. Figure 2 presents the aggregate NPL of the Colombian financial system over the last two decades. Notably, NPL exhibit a cyclical behavior. For instance, around 2008 and 2009 this series reaches its peak while the Colombian economy went through an episode of output deceleration right after a period of rapid credit expansion. This episode raises the question of what the cause of the increase in NPL was, the decrease of the economic activity or the credit boom that occurred before? Analyzing credit vintage data allows us to answer that question below.

Vintage information consists of a monthly register of the non-performing loan ratios from the loan portfolio discriminated by the month the loans were granted. Therefore, the unit of analysis is the non-performing loan ratio of credit over time by the origination period (i.e. by cohort). This article employs the vintage information of consumer loans originating from January 2005 to December 2019 collected by the Financial Superintendence of Colombia.

Figure 2: Aggregate NPL

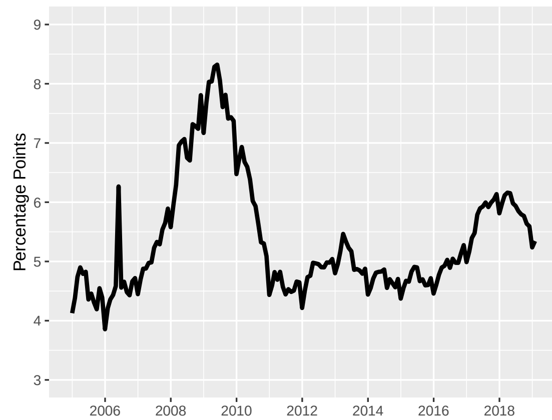
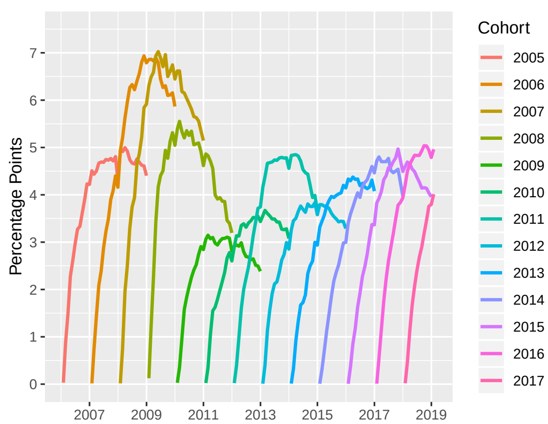


Figure 3: NPL Loan Vintages

Panel A Examples of two NPL Loan Vintages



Panel B Average Dynamics of NPL Vintages

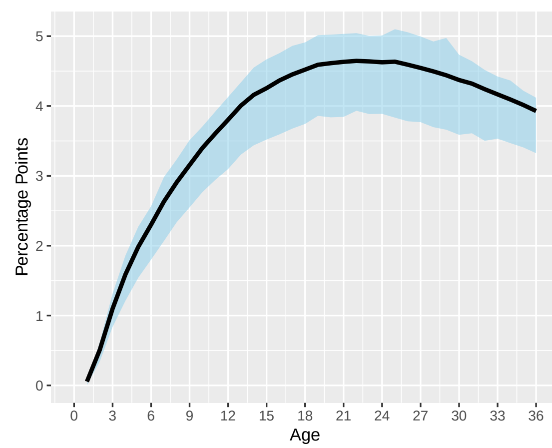


Figure 3 Panel A presents a subset of the vintage information employed in this article. There are the NPL series for some credit vintages from 2005 to 2018. Each line shows the NPL of a credit vintage from the month they are originated to their 36^{th} month of life, which is an age at which every vintage has passed its peak and it is close to the average consumer loan's maturity. Section 3 explains how to deal with data with this kind of structure. There is a regular pattern that is summarized in Panel B that shows the average dynamics of all loan vintages and the shaded area covers the 25^{th} and 75^{th} quantiles. This pattern consists of a rapid and sustained increase of the NPL during the first months of the credit. At this point every vintage reaches its peak and starts to slightly decrease. Nonetheless, Panel A shows that, over time, there is wide heterogeneity of behaviors. For example, consistent to what is registered in Figure 2, the vintages that were active between 2008 and 2009 present higher NPLs than other vintages.

However, it can be noticed that vintages deviate from this pattern in multiple ways. For example, they can present different average delinquency ratios and different paces of growth before reaching their maximum, which implies that some vintages deteriorate faster than others. All these features might respond to changes in the payment capacity of debtors through the life of the credit, or rather they are related with the debtor's intrinsic risk and the original conditions of loans. Arguably, the methodology presented above provides a decomposition of this information in these two characteristics.

In the second stage of this methodology, we study the relationship between some components of the NPL dynamics and a set of macroeconomic variables to estimate two vector autoregressions. Figure 4 presents that set of variables: the economic activity index (ISE, for its acronym in Spanish) which is a monthly proxy for the GDP, the unemployment rate, the real inter-bank interest rate, the spread between the inter-bank interest rate and the average interest rate of consumer loans and the real aggregate consumer credit. ISE and aggregate credit are presented here in real annual growth rates, but will be introduced in the model in real monthly growth rates. This information is made public by the Statistics National Administrative Department (DANE for its acronym in Spanish), the Central Bank of Colombia (Banco de la República) and the Financial Superintendence of Colombia.

5 Empirical exercise

In this section we present the results of the two stage methodology. In the first stage, we decompose the NPL series into three components, Age, Period and Vintage. In the second step, we estimate a VAR model for the estimated Period and Vintage components and the set of the macroeconomic variables described in the previous section.

5.1 First Stage

As a first step we perform the APC decomposition of the NPL series using intrinsic estimators and penalized regressions, described in sections 3.1.1 and 3.1.2.

We use several specifications for estimating the APC decomposition throughout the IE methodology. In particular, we estimate models (1) to (7) and choose the model with the smallest Bayesian Information Criteria (BIC). However, we use a restricted sample of the data since this technique requires a balanced information matrix. Hence, we use information with periods between January-2008 to January-2019, and ages from 1 to 36 months.

The results of the IE estimations are presented in Table 1 in Appendix A for models (1) to (7). This table shows that the smallest BIC is obtained for model (1), the three-factor model.

At the next step, we estimate the decomposition using Penalized Regressions. Given that this method does not require a balanced data, we can use the complete data set from January-2005 to January-2019. In this case, we have to specify the penalizing parameters λ and ϑ of the loss function (equation (11)). Following Härdle and Prastyo (2017), we select λ by cross-validation for a grid of ϑ, $\{0, 0.1, \cdots, 0.9, 1\}$.[3] Then, we estimate the parameters of models (1) to (7) using this set-up and the best model is chosen according to the smallest BIC.

Table 2 of Appendix A presents the estimated results for models (1) to (7) using the penalized regression method. Again, the best model according to BIC is the three-factor model (APC model) with $\widehat{\vartheta} = 0$, that corresponds to the Ridge regression case.

[3] Specifically, we use $k-$fold cross-validation for $k = \{5, 10, 20\}$.

Figure 4: Macroeconomic Variables

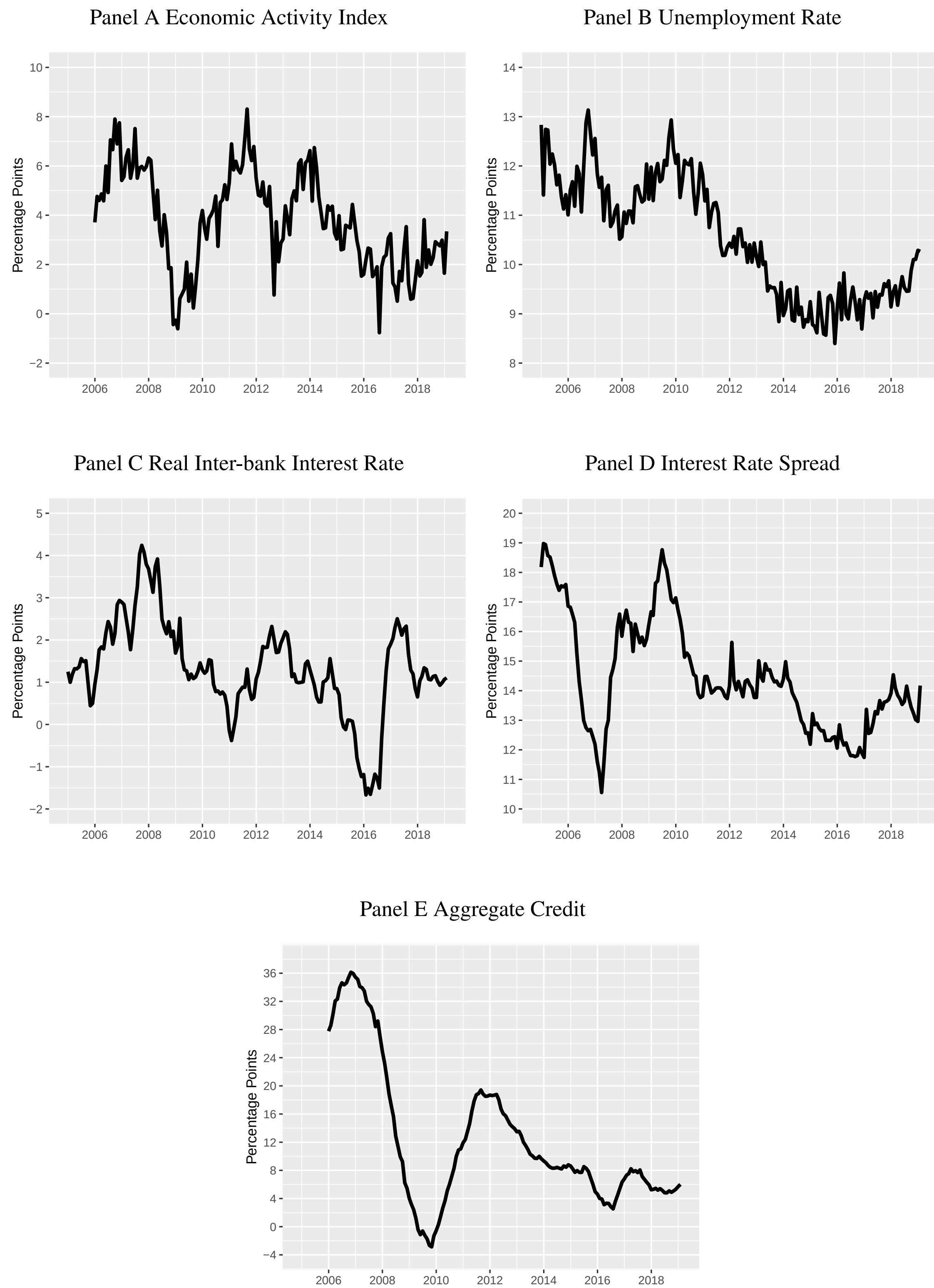
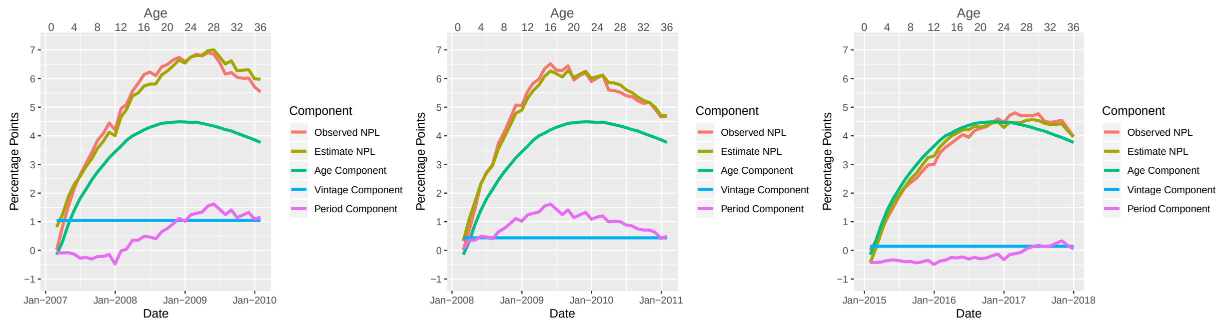


Figure 5: NPL Decomposition of Loan Vintages

Panel A Vintage of January 2007 Panel B Vintage of January 2008 Panel C Vintage of December 2014



The NPL decomposition of the APC model for the IE and Ridge regression methodologies are presented in Figures 10 and 11 of Appendix A, respectively. Given that the dynamic of the estimated coefficients of both methodologies shows similar behavior for the three components and that the Ridge regression method allows us to estimate the Period Component for a longer sample, we use the result of the latter method hereafter.

The result of this decomposition can be studied from multiple perspectives. However, to better understand it, Figure 5 presents the loan vintage of January 2007, January 2008 and December 2014, each one with their respective components. Each Panel shows the evolution of the respective credit vintage, hence, they can be tracked using their age (upper horizontal axis) as well as the periods they were alive through (lower horizontal axis).

Each panel of this figure contains five variables. First, the age component in red computed as the sum of $\widehat{\mu}$ and $\widehat{\alpha}$ coefficients of (1), they capture the general dynamics of all the vintages in the sample. Hence, they display the average aging of any loan vintage. Second, the Vintage Component in light blue represents the $\widehat{\gamma}$ coefficient that describes the average default ratio of the respective vintage, and therefore, it is constant for each cohort. Third, the Period Component in dark blue embodies the $\widehat{\beta}$ coefficients, which captures the average behavior of all the vintages that were alive at specific periods and faced the same economic conditions. Finally, the yellow and green lines represent the observed NPL and estimated response series $\widehat{NPL}$.

For instance, take June 2009 when the aggregate default's ratio of the financial system was increasing dramatically and correspondingly many loan vintages too. The January 2007 vintage reached that date when it was 27 months old, while the January 2008 vintage did it at 15 months old. However, both vintages exhibited an increase at that moment, which suggests that both default's increases are related with something that happened that month rather than the risk inherent to these cohorts. That is why the Period Component rose around mid-2009 and that affected all the loan vintages that were active in those periods, even though each one went through those times at different ages.

From this perspective, $\widehat{\gamma}$ coefficients can be though of as the fixed effects corresponding to each loan vintage resembling the individual's panel data characteristics. However, these effects are associated with specific periods just like the $\widehat{\beta}$ coefficients but spanning different dates. As a result, $\widehat{\gamma}$ and $\widehat{\beta}$ coefficients can display two time series corresponding to different components of the NPL that spans different time periods. The first one captures every characteristic that will remain throughout the life of the vintage, such as the interest rate, maturity, ex-ante debtors' payment capacity and loans approval standards of banks; it will be referred to here on as the Vintage Component. As discussed before, the second component contains every mechanism that affects the debtors' payment capacity after the time loans are granted; referred to now on as the Period Component.

Figure 6 presents the three components for all the estimation sample and makes the fact explicit that the Vintage and the Period components are indexed by time, but spanning different dates. Hence, both of them can be interpreted as time series. Panel C shows the age component, which is the average behavior of any loan vintage. Panel A shows that the Vintage Component has varied over time, presenting its highest peak around 2006 and 2007, and its lowest trough around 2009-2010. This suggests that loans granted between 2006 and 2007 were riskier than the ones granted at other moments, and that credit standards were loose at that time. Later on, around 2010 the financial system showed little risk taking by granting the safest loans in the sample, followed by a fast risk taking increase in 2012. Then, since 2013 the risk taken by the financial system through credit risk has remained roughly constant.

On the other hand, the Period Component in Panel B in Figure 6 exhibits two main incremental episodes, 2009 and 2017. These episodes coincide with the two peaks observed in the aggregate dynamics of NPL in Figure 2.

Nonetheless, using this decomposition we can accurately explain the causes of the behavior observed in the aggregate NPL. We can confidently state that while the high delinquency ratios observed during 2008-2009 were affected by certain economic conditions they were preceded by an episode of risky loans granting. In other words, the financial system took a risky position during 2006 and 2007, building up a credit risk vulnerability which materialized in 2008 and 2009 along with the negative economic conditions of those years. On the contrary, the high levels of NPL registered since 2016 respond only to the economic conditions of the time, but they are not explained by previous risk taking by the financial system.

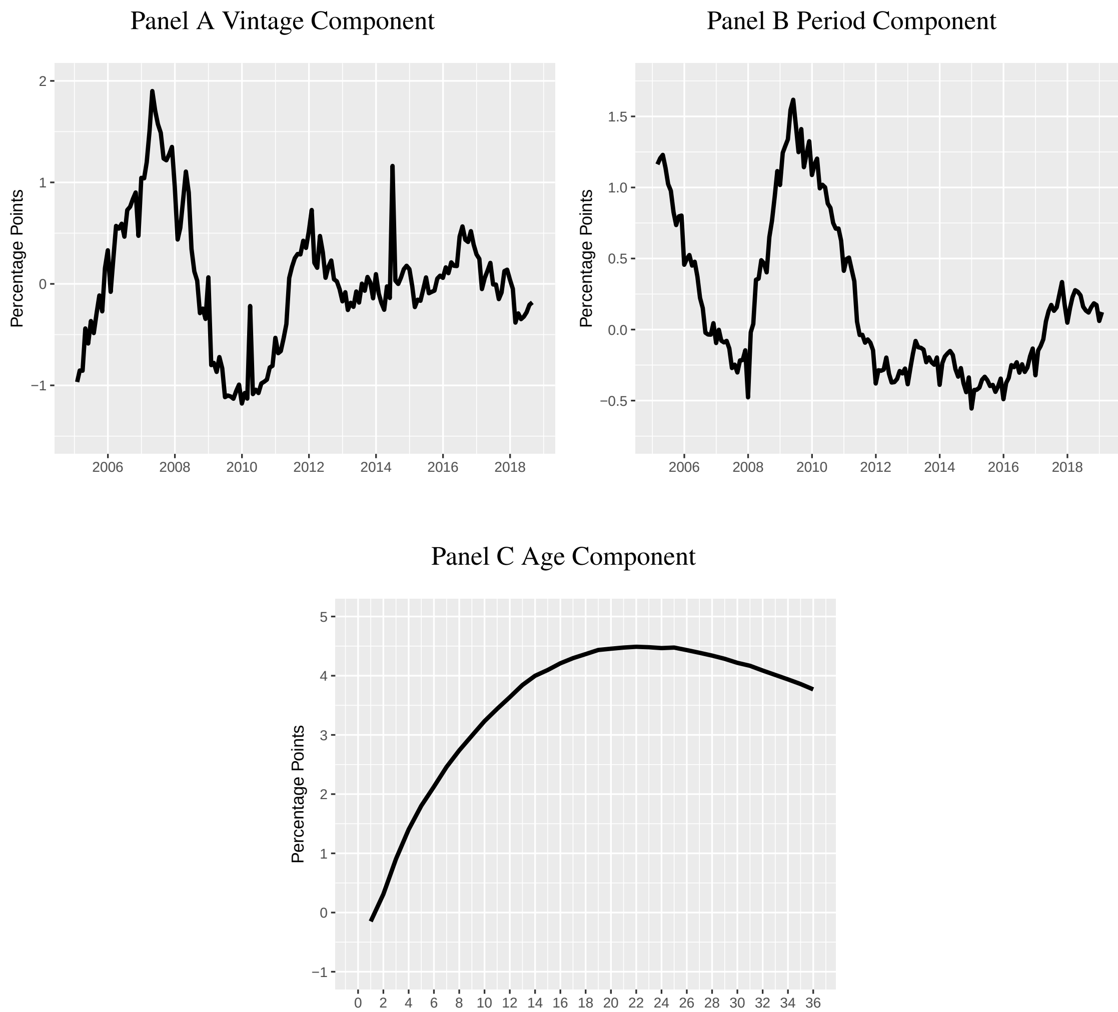
The cyclical behavior of the Vintage and Period components suggests that their behavior is associated with the real and the financial cycles. For instance, around the peak of the Period Component between 2008 and 2009, the Colombian economy went through an intense deceleration of economic activity. Conversely, during the peak of the Vintage Component in 2007 the Colombian financial system was in the middle of a phase of rapid credit expansion. The second stage of this methodology is to offer better understanding of the macroeconomic and financial conditions that explain the behavior of these two components.

5.2 Second Stage

In the second step of our methodology we estimate the two restricted VAR models specified in section 3.2. The first one includes the Period Component estimated in the first stage and the set of macroeconomic variables, while the second one includes the estimated Vintage Component and the same macroeconomic series. As explained in section 4, we include the following macroeconomic variables: ISE and unemployment rate describe economic and labor market activity, real inter-bank interest rate contains the monetary policy stance over time, the interest rate spread between the consumer loans and the monetary policy interest rates that can be interpreted as the credit risk perceived by the financial system and aggregate consumer credit tracks credit cycle dynamics[4]. According to the results of Tables 3 and 4 of Appendix B all the series are integrated of order one and they are not cointegrated. Then, we estimate the following VAR model for the first difference

[4]ISE and aggregate credit are log-transformed, so they are expressed in monthly growth rates when differentiated. Additionally, ISE and unemployment time series are seasonally adjusted using ARIMA-Seats.

Figure 6: Components of NPL Loan Vintages



of the series:[5]

Additionally, the number of lags, p, was chosen as the minimum lag for which the Q-statistics indicates no residual auto-correlation. Table 5 of Appendix B presents the diagnostic tests computed for the residuals of these two models. These results show that there is no evidence of heteroscedasticity and auto-correlation. However, the normality assumption fails. For that reason, the confidence intervals of the impulse response functions are estimated using bootstrap methods.

The impulse-response function analysis of the Period and Vintage VAR models are presented in Figures 7 and 8, respectively.[6] Although, these models are estimated with the components in differences, we present the cumulative impulse-response functions to study the effect on the level of each component. It might be more intuitive to begin interpreting the results for the Period Component model that captures the evolution of the loan quality during its lifetime. To this end, one must have in mind a debtor that is currently facing periodical payments of a loan that was acquired in the past.

Panels A and B of Figure 7 show that following a drop in the economic activity or an increase in unemployment, a rise in credit defaults is expected. This make sense since these two variables are related to the debtors' income and therefore affects their payment capacity. Results suggest that the impact of economic activity on the Period Component is perceived almost immediately, while the changes in the unemployment rate affect NPL three months after the shock. Both effects appear to be highly persistent.

On the other hand, the inter-bank interest rate does not show a significant impact on NPLs dynamics. In contrast, the interest rates spread displays a positive and persistent impact on the performance of existing credits. Since this variable describes the interest rate spread of new loans, existing loans should not be affected by an increase in this variable unless they are credits indexed to a variable interest rate. Nonetheless, this variable can also be interpreted as a measure of the credit risk perceived by the financial system. Following that interpretation, it is expected that we find a strong correlation between this variable and the NPL, as credit default rises banks might increase their risk perception. Banks might even be able to adjust their risk perception before credit defaults begin to materialize using the private information they have on the quality of debtors.

Last, an increase in the credit growth rates induces a drop in the NPLs of existing loans. This result does not exhibit an direct interpretation since an expansion of the loan portfolio of banks does not directly affect the payment capacity of existing loans, but one can argue that this negative effect might be associated with the indirect effect of a credit expansion on the aggregate income of the economy.

On the other hand, Figure 8 presents the impulse response functions for the Vintage component. To interpret this result, instead of a borrower paying back its credit, it is useful to have in mind a new debtor and a bank determining the conditions of a new loan. For this reason, the Vintage Component is easily associated to the credit standards of banks. In contrast to the Period Component, the Vintage Component is negatively related to interest rate spreads, which means that when interest rates come closer to the monetary policy rate, riskier loans are granted. This may be counter-intuitive at first, because one would expect that a loan that is agreed upon at a low interest rate would concede with greater payment capacity from debtors. However, this result

[5] This model is estimated under the restriction (13) specified in section 3.2

[6] For the identification of the impulse-response function we use a Cholesky decomposition that implies the following contemporaneous order of the shocks, where the first one is the most exogenous: ISE, unemployment, inter-bank interest rate, interest rate spread, aggregate credit, Component. Where component stands for either Period or Vintage Component depending on the VAR model.

Figure 7: Impulse-Response on Period Component

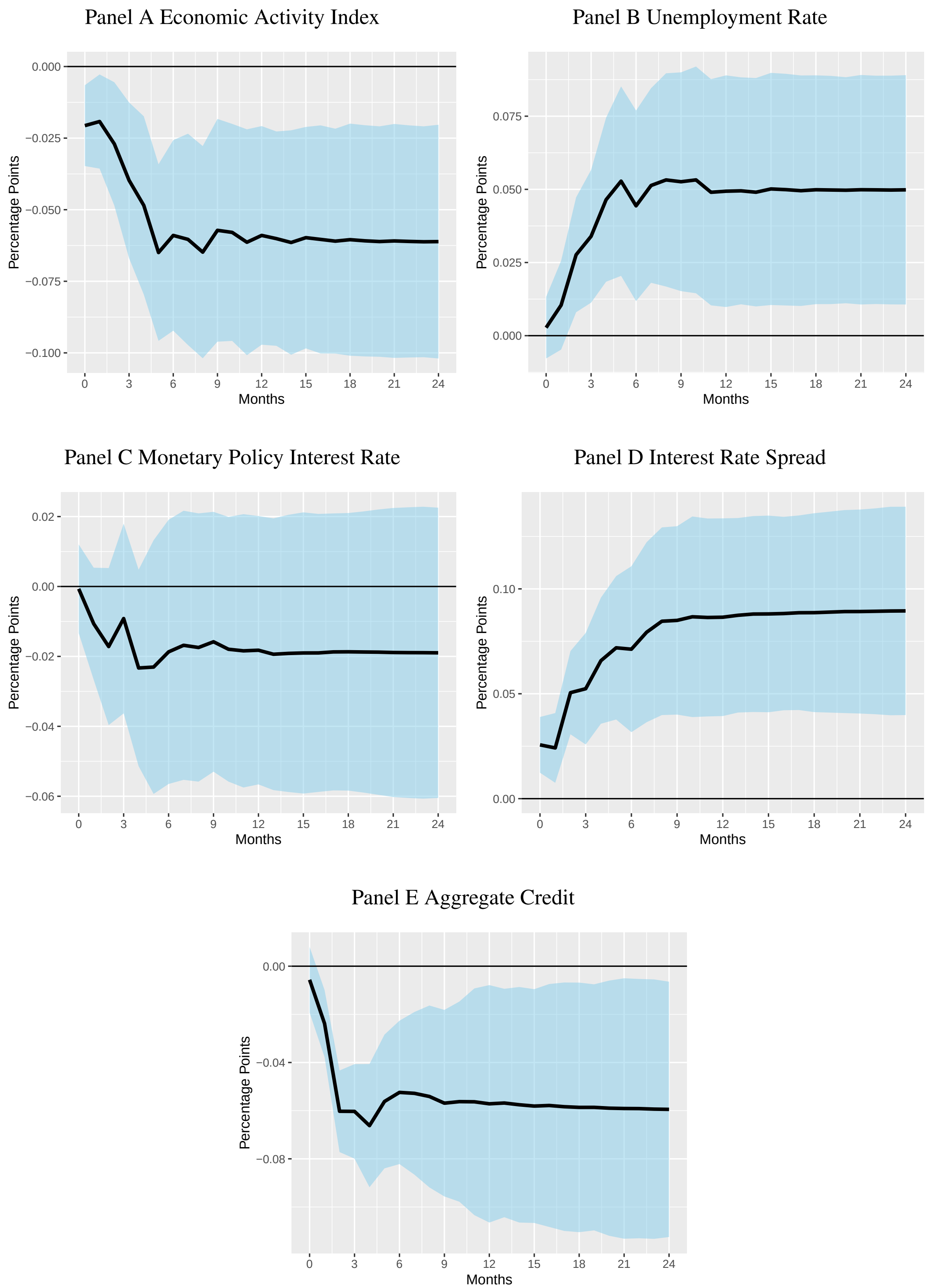
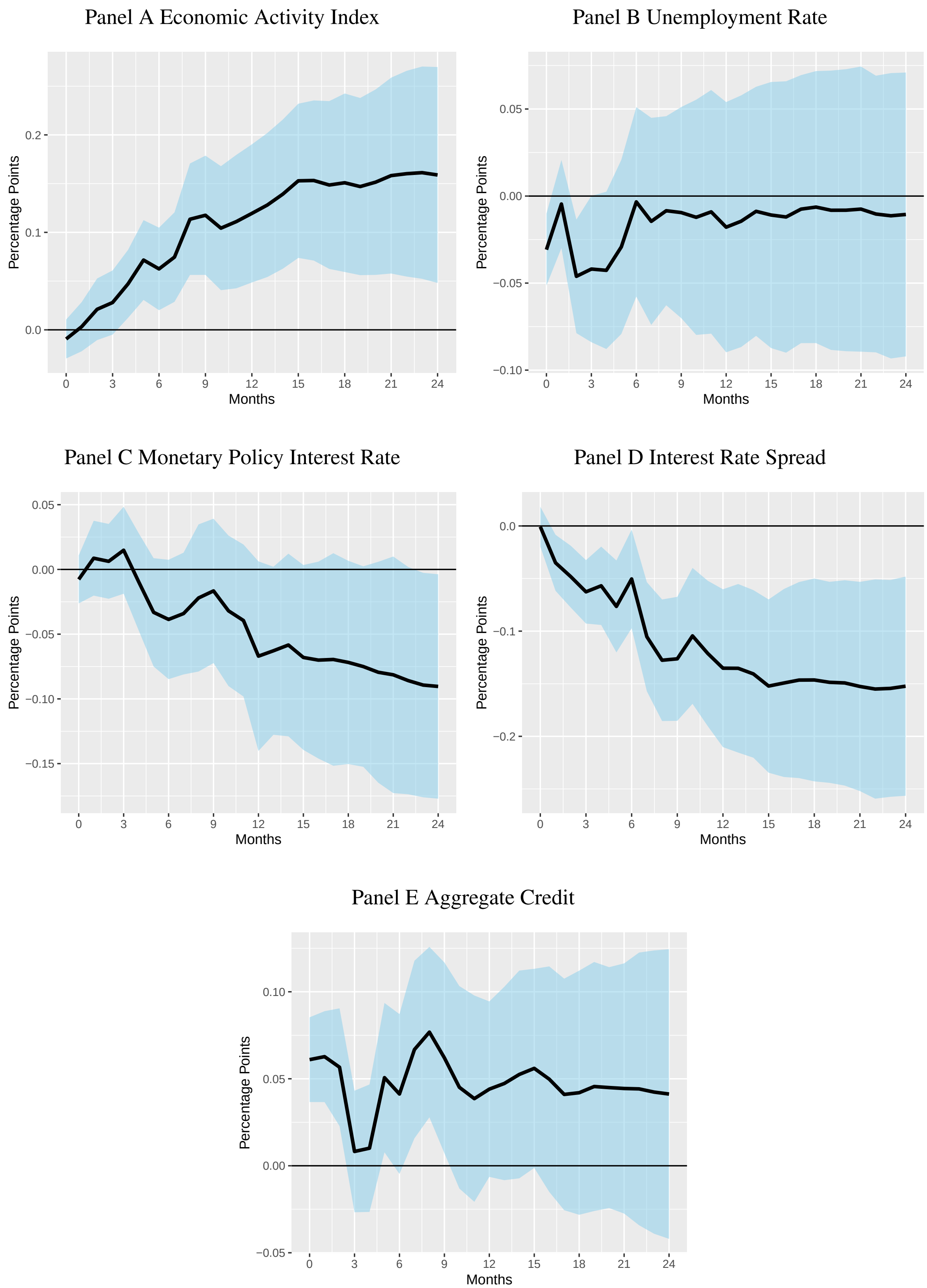


Figure 8: Impulse-Response on Vintage Component



indicates an underpricing of the risk of loans granted. Then, when interpreting this variable as a measure of credit risk perception, we suggest that periods of low interest rate spreads are associated to higher risk taking of banks which is reflected in the granting of riskier loans. Hence, this variable should be interpreted as a measure of risk appetite of the financial system and not necessarily indicating its real risk perception. For instance, episodes of falling interest rates might be motivated by high competition among banks. In an episode alike, lower interest rates do not necessarily mean that banks are perceiving lesser credit risk. Instead, this would mean that they are deciding to take greater risks by under-pricing their new loans.

Similarly, the impact of ISE and unemployment on the Vintage Component is opposite to their impact on the Period Component. Results counter-intuitively suggest that loans granted during periods of high economic activity and employment turn out to be riskier. Then again, the credit risk perception of banks appears to decrease when economic growth improves and unemployment decreases, leading banks to assume riskier positions.

The positive relation between Vintage Component and aggregate consumer credit suggests that riskier loans are granted during times of fast credit expansion. These last results let one infer that at moments of good macroeconomic and loose financial conditions, the financial system tends to take more credit risk than the real credit quality of borrowers would have supported. In contrast, monetary policy does not seem to directly affect the Vintage Component.

In synthesis, the Period Component appears to naturally increase with the deterioration of macroeconomic conditions (slower economic activity and higher unemployment) and the tightening of financial conditions (slower credit growth and larger interest rate spreads). On the other hand, the Vintage Component turns out to increase with better macroeconomic conditions and loose financial conditions. This suggests that during good times the financial system tends to take more credit risk, that is to say that banks exhibit an over-optimistic behavior when the economy is experiencing healthy conditions.

Finally, Figure 9 depicts the historical shock decomposition for both NPL components. Notice that this analysis is present with the components in levels even though the VAR models are estimated with the components in differences. Therefore, these figures present the cumulative contribution each shock on the components dynamics in the sample, leaving aside the starting point of each components. As a result, this exercise allows for studying the changes in each one of the components, but the levels registered in Figure 9 does not coincide with the ones presented in Figure 6[7].

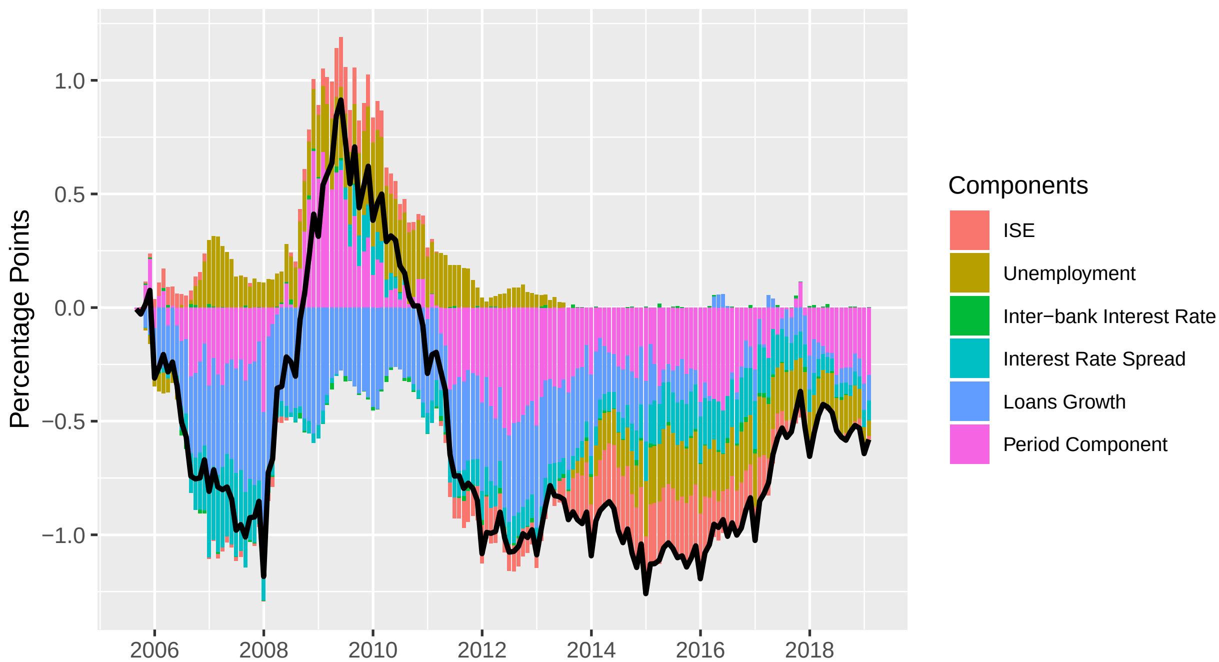
This exercise offers a clear narrative of the macroeconomic conditions associated with the cycles observed in Vintage and Period Components. For example, notice that ISE and unemployment played an important role in explaining the peak of the Period Component around 2009. Similarly, in the most recent rise of this component, the reduction of economic activity along with the increase in the interest rate spreads are identified as the key drivers.

Regarding the shock decomposition for the Vintage Component, we argue that the peak of 2007 was vastly explained by the low interest rate spreads exhibited at the time. In addition, economic activity, unemployment and aggregate credit growth also contributed to that peak. ISE and aggregate credit also explain other peaks and valleys. The economic recovery of 2011 and 2012 led to the disbursement of risky loans, while the economic deceleration observed from 2017 to 2018, led to a decrease in the risk taken by the financial

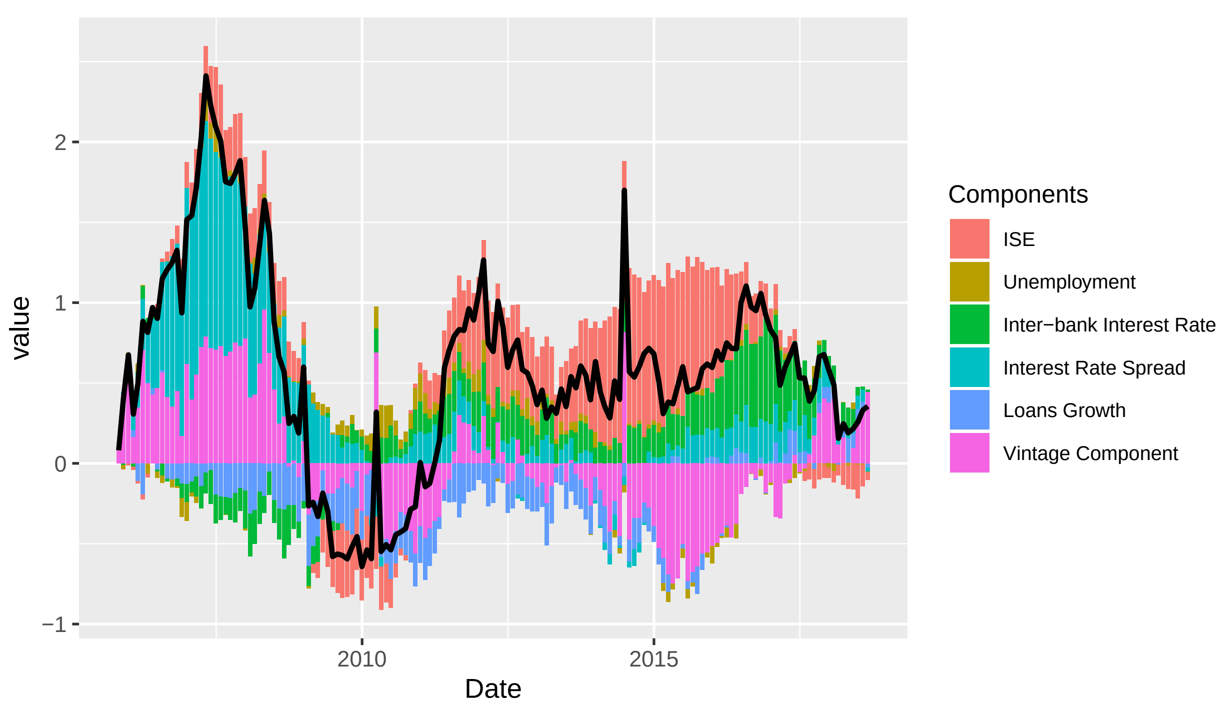
[7] The bars corresponding to the respective components contains both the contributions of the components shocks and the constant of each model

Figure 9: Historical Shock Decomposition of NPL Components

Panel A Period Component



Panel B Vintage Component



system.

6 Discussion

This section aims to discuss the rationale behind our interpretation of the components estimated in the first stage. The interpretation of the age component is more straightforward than the other two, each α_i coefficient that forms this component (summed to the constant μ) represents the average NPL of any vintage when it is i months old.

In contrast, the interpretation of the Period and the Vintage components require more discussion. Notice that one can interpret them as deviations from the general dynamics described by the age component. First, the Vintage Component is conformed by the series γ_n coefficients, which captures the deviation from the age component that is prevalent to every observation of the respective vintage n. Hence, this component can be interpreted as the average NPL level of each credit vintage. On the other hand, the Period Component is defined as the series of β_t coefficients. They embody the deviation from the general dynamics that is common for every credit vintage that lived during the period t. Therefore, the changes in this component should be understood as changes in the NPL levels of all credit vintages.

Given this precision on what the components technically capture, we argue that banks have more to do with the Vintage Component, whereas the Period Component is more related to the borrowers behavior. Typically, banks do not have much maneuverability on the payment of loans, unless they agree to restructure the loan's conditions. Instead, the payment of loans almost entirely depends on the borrowers behavior, and since credit reporting agencies generate incentives not to default loans, the payment capacity of borrowers must be the main force behind credit defaults[8]. Considering that the Period Component captures any shock or mechanism that affects all credit vintages at specific periods of time independently of when those loan vintages were granted, we argue that this component describes changes in the general payment capacity of the whole set of borrowers.

Alternatively, since the Vintage Component represents a deviation of the NPL level of each vintage that remains constant throughout its lifetime, this component can be interpreted as a measure of the intrinsic risk of the respective vintage. This intrinsic risk refers to the latent payment capacity of the borrowers or the conditions of the loans granted in that vintage, in general, any characteristic that remains constant through the life of the vintage. Although, the payment capacity of borrowers is still important for this component, we argue that banks play an active role in determining the type of borrowers that receives the loans and their conditions (interest rate and maturity). For these reasons, the Vintage Component reflects the stance of credit standards over time, hence, it describes the credit risk taken by the financial system. Even though this component might be capturing changes in the general risk level of the pool of debtors in the credit market instead of banks deliberately taking more risk, we argue that this component should be still be interpreted as a measure of banks risk taking. Even if that is the case, and the risk of the whole pool of debtors increased, in the end, banks took riskier positions.

Regarding the second stage of our methodology, it is important to note that this exercise does not aim to find the exact causal mechanisms that explain Period and Vintage components. Instead, this exercise explores the serial correlation between macroeconomic variables and components of the non-performing loan's dynamics and therefore provides some idea of the behavior of bank's lending standards and debtor's default

[8] Some literature has studied strategic default, which might appear when the collateral value falls below the loan value. We argue this should not be present in this study since consumer loans in Colombia are generally unsecured.

when facing different macroeconomic scenarios.

Finally, we compare the results of the second stage with the results found in previous literature of NPL determinants. The VAR model for the Period Component confirms most of the macroeconomic determinants found in literature, such as low economic growth and high unemployment causing increments in NPL.

On the other hand, the other group of determinants found in literature are identified using bank-level and credit-level data, and correspond to bank's characteristics that are related to different risk taking mechanisms. Instead of relying on cross-sectional variation to identify these mechanisms, our decomposition offers the Vintage component, which can over time track changes of the credit risk taken by the financial system as a whole. Additionally our approach makes it possible to study the interaction of the macroeconomic variables and the risk taken by banks.

7 Conclusion

The methodology proposed in this article achieves a comprehensible way to decompose the non-performing loan's dynamics into the Period and Vintage components. The Period Component describes the changes in the payment capacity of borrowers, while the Vintage Component captures the credit risk taken by the financial system at the loan's origination. In addition to confirming the results found in previous literature on NPL determinants, our methodology returns an indicator that tracks the evolution of the credit risk taken by the financial system.

Moreover, the model behind the decomposition is simple and straightforward, and can be applied to any credit vintage information structure. Nonetheless, using techniques like intrinsic estimators or penalized regression is necessary to overcome the perfect multicollinearity problem. From here, the two most relevant components can be analyzed with any time series approach, we propose two restricted VAR models to examine their relationship with key macroeconomic variables. However, at this stage a wide variety of models can be applied, for instance, we leave a joint model of the two components for future work.

We found that the borrower's payment capacity and the credit risk taken by the financial system have changed differently over time and they are contrarily affected by aggregate macroeconomic and financial variables. In general, we found that loose economic and financial conditions alleviate the burden of borrower not meeting their obligations, but in turn, they tend to coincide with the financial system engaging in riskier loans. This result stresses the importance of the stability objectives of macroeconomic and financial authorities. At the same time, when authorities deal with difficult situations such us the current recession, these results help to keep track of the risks entailed in a policy response that aims to get through the recession.

Finally, this decomposition is a useful tool for authorities that monitor credit risk because it allows them to identify the origin of the credit risk materialization. For instance, financial authorities can monitor the Vintage Component to design financial regulation to curb the risk taken by the financial system. Conversely, macroeconomic authorities can focus on the Period Component to continuously track how the current economic conditions are affecting the defaults ratios. This methodology can also be used in stress testing exercises, making it possible to distinguish between crisis scenarios (stressing the Period Component) from risks building up scenarios (stressing the Vintage Component).

References

Bosman, M. (2012). The potential of cohort analysis for vintage analysis. an exploration. Master's thesis, University of Twente.

Breeden, J. L. (2016). Incorporating lifecycle and environment in loan-level forecasts and stress tests. *European Journal of Operational Research*, 255(2):649–658.

Breeden, J. L. and Thomas, L. (2016). Solutions to specification errors in stress testing models. *Journal of the Operational Research Society*, 67(6):830–840.

Dimitrios, A., Helen, L., and Mike, T. (2016). Determinants of non-performing loans: Evidence from euro-area countries. *Finance research letters*, 18:116–119.

Forster, J. J. and Sudjianto, A. (2013). Modelling time and vintage variability in retail credit portfolios: the decomposition approach. *arXiv preprint arXiv:1305.2815*.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Fu, W. (2016). Constrained estimators and consistency of a regression model on a lexis diagram. *Journal of the American Statistical Association*, 111(513):180–199.

Fu, W. (2018). *A Practical Guide to Age-Period-Cohort Analysis: The Identification Problem and Beyond*. Chapman and Hall/CRC.

Fu, W. J. (2000). Ridge estimator in singulah oesiun with application to age-period-cohort analysis of disease rates. *Communications in statistics-Theory and Methods*, 29(2):263–278.

Ghosh, A. (2015). Banking-industry specific and regional economic determinants of non-performing loans: Evidence from us states. *Journal of financial stability*, 20:93–104.

Greenidge, K. and Grosvenor, T. (2010). Forecasting non-performing loans in barbados. *Journal of Business, Finance & Economics in Emerging Economies*, 5(1).

Härdle, W. K. and Prastyo, D. (2017). Default risk calculation based on predictor selection for the southeast asian industry.

Klein, N. (2013). *Non-performing loans in CESEE: Determinants and impact on macroeconomic performance*. Number 13-72. International Monetary Fund.

Makri, V., Tsagkanos, A., and Bellas, A. (2014). Determinants of non-performing loans: The case of eurozone. *Panoeconomicus*, 61(2):193–206.

Rosch, D. and Scheule, H. (2004). Forecasting retail portfolio credit risk. *Journal of Risk Finance*, 5(2):16–32.

Salas, V. and Saurina, J. (2002). Credit risk in two institutional regimes: Spanish commercial and savings banks. *Journal of Financial Services Research*, 22(3):203–224.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric theory*, 10(1):91–115.

Siarka, P. (2011). Vintage analysis as a basic tool for monitoring credit risk. *Mathematical Economics*, (7 (14)):213–228.

van Wieringen, W. N. (2015). Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.

Yang, Y., Schulhofer-Wohl, S., Fu, W. J., and Land, K. C. (2008). The intrinsic estimator for age-period-cohort analysis: what it is and how to use it. *American Journal of Sociology*, 113(6):1697–1736.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Appendices

A Age-Period-Cohort decomposition for NPL using Intrinsic Estimators and Ridge regression

Table 1: Intrinsic Estimator results for different specifications

Model	**Adjusted R^2**	**BIC**
APC	0.997	$-16,003.1$
A	0.937	$-3,233.3$
P	0.876	807.2
C	0.912	-619.8
AP	0.977	$-7,193.6$
AC	0.991	$-11,545.9$
PC	0.961	$8,529$

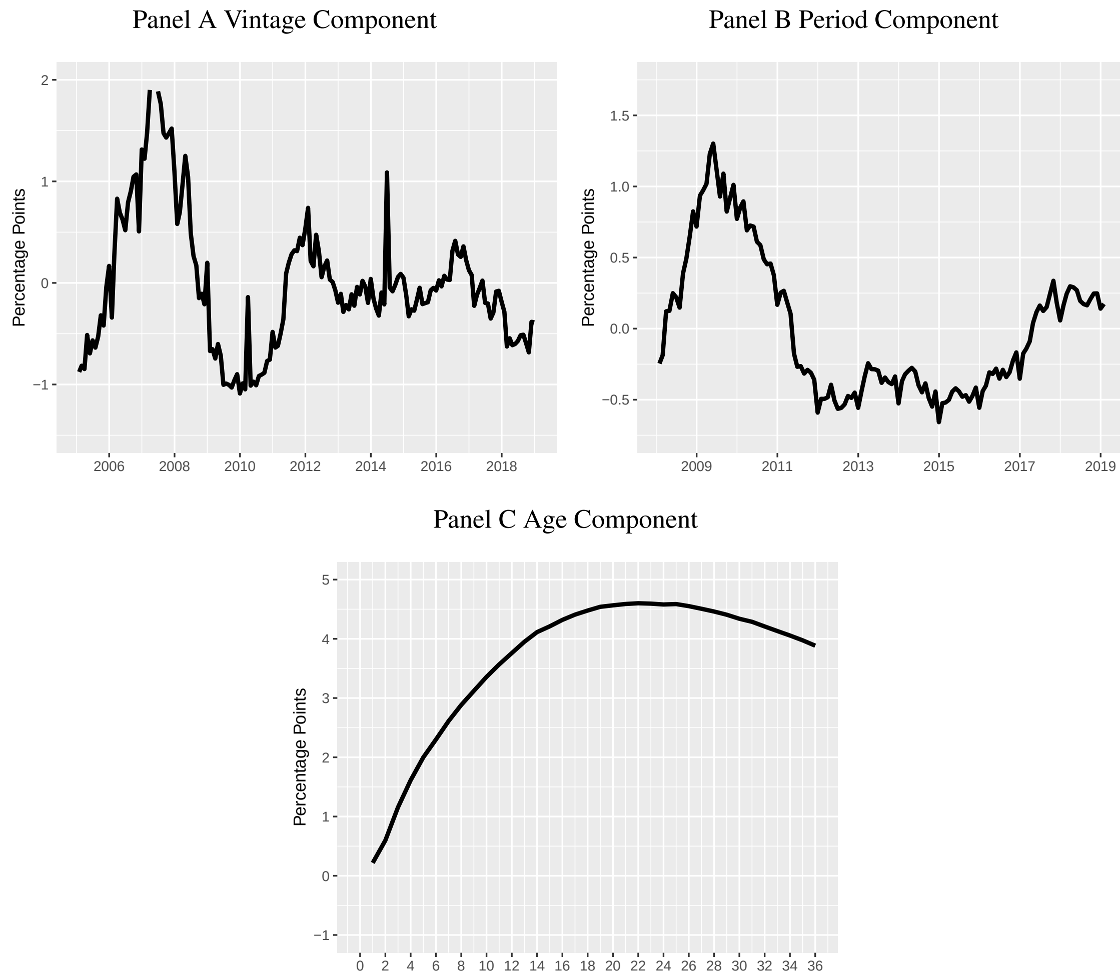
Note: APC (Age-Period-Cohort), A, P, C, AP, AC and PC refer to models (1), (2), (3), (4), (5), (6) and (7), respectively.

Table 2: Penalized regressions results for different specifications

Model	$\widehat{\lambda}$	$\widehat{\vartheta}$	**Adjusted R^2**	**BIC**
APC	$1.864e-05$	0	0.978	$-16,182.6$
A	$2.509e-05$	0	0.634	$-3,464.4$
P	$6.162e-05$	0	0.239	$1,509.9$
C	$1.788e-04$	0	0.306	$1,017.1$
AP	$7.849e-05$	0	0.859	$-7,379.2$
AC	$7.946e-06$	0	0.924	$-1,0741.5$
PC	$1.138e-06$	0	0.702	$-2,301.7$

Note: APC (Age-Period-Cohort), A, P, C, AP, AC and PC refer to models (1), (2), (3), (4), (5), (6) and (7), respectively. $\widehat{\lambda}$ and $\widehat{\vartheta}$ are the estimated penalizing coefficients of model (11).

Figure 10: Age-Period-Cohort decomposition of NPL using Intrinsic Estimators for model APC.



B Second Stage: Unit Root, cointegration and VAR residuals tests

Figure 11: Age-Period-Cohort decomposition of NPL using Ridge regression for model APC.

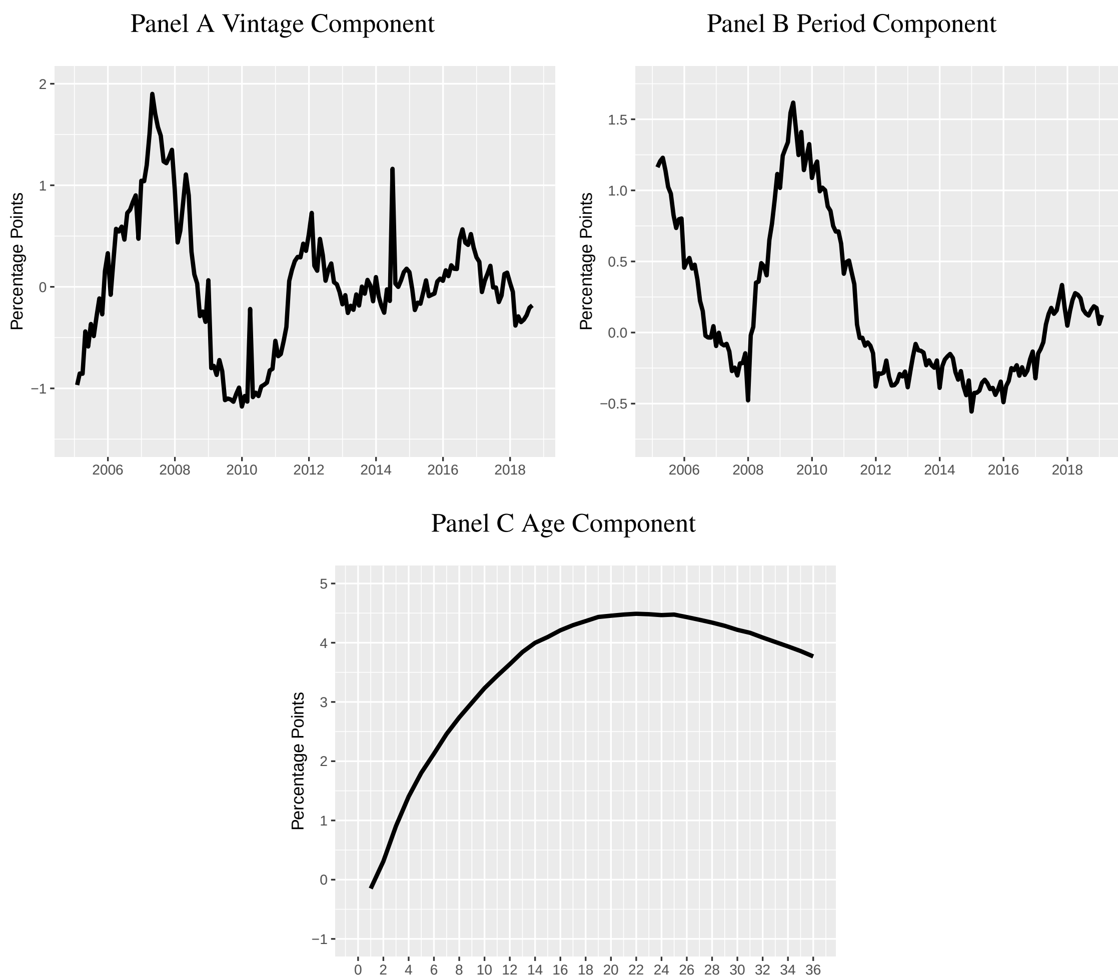


Table 3: Unit root tests

	Levels		First difference	
	KPSS	**PP**	**KPSS**	**PP**
Period Component	1.123***	−2.302	0.189	−15.02***
Vintage Component	0.185***	−2.620	0.122	−16.62***
log-ISE	3.457***	−1.871	0.302	−18.19***
Unemployment Rate	2.786***	−2.004	0.139	−23.804***
Inter-bank Interest Rate	0.815***	−1.916	0.097	−7.188***
Interest Rate Spread	1.431***	−2.848*	0.105	−12.784***
log-Aggregate Credit	3.334***	−3.013	0.117	−3.125**

This table presents the statistic values for the KPSS test (H_0 : the series is stationary) and Phillips and Perron (PP, H_0 : the series is non-stationary). *, ** and *** indicate if the null hypothesis is rejected at 10%, 5%, and 1% significance level, respectively.

Table 4: Shin cointegration test

Variables	Statistic
Macroeconomic Variables and Period Component	0.103^{***}
Macroeconomic Variables and Vintage Component	0.051^{**}

This table presents the statistic values for the Shin cointegration test (Shin, 1994), where the null hypothesis indicates that the variables are cointegrated. *, **, and *** denote rejecting the null hypothesis at 10%, 5% and 1% significance level, respectively. The set of macroeconomic variables corresponds to: ISE, unemployment, inter-bank interest rate, interest rate spread, aggregate credit, Component.

Table 5: VAR residual tests

Model	Multivariate Jarque-Bera	Multivariate ARCH-LM	Q-Stat
Macroeconomic Variables and Period Component	0.00	0.06	0.07
Macroeconomic Variables and Vintage Component	0.01	1.00	0.52

This table presents the $p-$values of some tests for the residual of the VAR models specified in (12). The null hypotheses of Jarque Bera, ARCH-LM and Q-stat tests are normality, homoscedasticity and no-autocorrelation of the model errors, respectively. The set of macroeconomic variables corresponds to: ISE, unemployment, inter-bank interest rate, interest rate spread, aggregate credit, Component.


BANCO DE LA REPÚBLICA
COLOMBIA