LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Department of Economics

**Maria Rosa Borges & Raquel Machado**

# Modelling credit risk: evidence for EMV methodology on Portuguese mortgage data

**WP03/2020/DE/UECE**

WORKING PAPERS

ISSN 2183-1815

LISBOA
UNIVERSIDADE DE LISBOA

# Modelling credit risk: evidence for EMV methodology on Portuguese mortgage data

**Raquel Machado**
raquel.f.machado@gmail.com
ISEG (School of Economics and Management) of the
Universidade de Lisboa

**Maria Rosa Borges**
mrborges@iseg.ulisboa.pt
ISEG (School of Economics and Management) of the
Universidade de Lisboa
UECE (Research Unit on Complexity and Economics)
ORCID ID: 0000-0001-5340-471X

**Abstract**

Traditional credit risk models failed during the recent financial crisis and revealed weaknesses in forecasting and stress testing procedures. One of the main reasons for this failure was the fact that they did not include lifecycle and macroeconomic adverse selection effects. The Exogenous-Maturity-Vintage (EMV) models emerged in this context, in the credit risk literature. In this article, we assess the applicability of the EMV models to a dataset consisting of Portuguese mortgage data between 2007 and 2017, to study the determinants of default rates. We obtain and examine the exogenous, maturity and vintage curves from the dataset under analysis, plotting defaults rates through time, under each of the three component's logic (default rates by calendar period, by age and by vintage). We show that these curves follow the expected behavior. Finally, we identify a set of explanatory variables suitable to be incorporated in an EMV model specification, for forecasting purposes, and discuss the rationality for their inclusion in the model.

**Keywords**: credit risk; EMV models; mortgage loans; default rates; vintages

**JEL Codes**: G20; G21

**Funding**: UECE (Research Unit on Complexity and Economics) is financially supported by FCT (Fundação para a Ciência e a Tecnologia), Portugal. This article is part of the Strategic Project (UIDB/05069/2020).

# Modelling credit risk: evidence for EMV methodology on Portuguese mortgage data

## 1. Introduction

The traditional credit risk models failed during the recent financial crisis and revealed to be weak for forecasting and stress testing procedures. One of the main reasons for the failure of those models was the fact that they do not analyze jointly the lifecycle and the macroeconomic adverse selection effects. It is in this context that the Exogenous-Maturity-Vintage Models (EMV Models) emerge in the credit risk area, taking inspiration from the Age-Period-Cohort (APC) Model, used in the study of demography, epidemiology and sociology issues, among others.

The Exogenous-Maturity-Vintage (EMV) model aims to separate a time series into three components: exogenous, maturity and vintage. The exogenous component refers to the calendar date at which the event has occurred. Maturity refers to the time since the individual has entered the study. A vintage corresponds to a group of individuals entering in the study within a given time period, whom are expected to share common features. If, in the last three definitions, we replace "event" by "default" and "individual" by "loan/client", we get the appropriate concepts for credit risk context. In recent years, some authors implemented this model to the analysis of credit risk (Zhang 2009; Forster and Sudjianto 2013; Strydom 2017).

Previously to any application of the methodology, we must confirm its applicability to the dataset, such that the model fitting may translate the best description of data among the several possible templates. Particularly to this methodology, we must examine the portfolio dynamics by graphical representation of rates of default through time, under each of the three component's logic (default rates by calendar period, by age and by vintage). While default rates along the maturity component are expected to have a very well-shaped behavior, default rates by calendar period and vintage should be explainable by the macroeconomic scenario and the banks' internal operational policy, respectively. Additionally, this components analysis shall be helpful to identify the patterns to be fitted.

Assuming the EMV approach accomplishes the descriptive objective, the major challenge will be forecasting. Each component may provide insight into possible future scenarios, by relating the response variable to parametric functions of each component, based on predictable covariates. Under the well-defined and smooth shape assumptions, it seems reasonable to fit a curve to the maturity component that may be extrapolated for longer loans than the ones observed in the dataset. For the exogenous component, we may perform an historical analysis based on macroeconomic variables for which we have forecast values projected to the future. Moreover, the vintage behavior may be forecasted under a few simplifying assumptions. If it is hard to anticipate future underwriting standards, it may be plausible to assume that they will be similar to the recent vintages' ones.

In this paper we explore the applicability of EMV models to a sample consisting of mortage loans of a representative Portuguese financial institution, analyze a set of variables to be included in the model specification, and discuss the rationality behind their inclusion. We conclude that our data sample is suitable for EMV modeling purposes. Our paper contributes to a recent area of research that applies EMV models to the analysis of credit risk, where published studies are still scarse.

This paper is structured as follows. Section 1 introduces the subject. In section two we present the exogenous maturity vintage (EMV) and age period cohort (APC) models. Section three is dedicated to explain the EMV methodology, and in section four we present some previous implementations of the methodology. In section five we report the main results and finally, in section six, we conclude.

## 2. The Exogenous-Maturity-Vintage model

The origins of the EMV model are found in the demography, epidemiology and sociology fields under the Age-Period-Cohort (APC) model designation. We find multiple definitions in the literature for each component of the APC model, which we resume in the following: age refers to the time since a subject or entity has entered a study; period refers to the calendar date at which the outcome was measured. This component intends to capture the historical conditions (eventual environmental shocks) that affect everyone; a cohort is generally compounded by individuals that have some characteristic in common (Glenn 2005), it is identified by the moment when an individual or entity entered the study (due to his birth or a dated event). Therefore, APC Modeling focus on a dual-time domain: lifetime (age, maturity) and calendar time.

This model intends to measure its three components simultaneously, but identifying each effect separately from the other two (Keyes et al. 2010). In the credit risk area, researchers replaced age, period and cohort components with maturity (age of contract), exogenous (external environment influenced by macro-economic conditions) and vintage (heterogeneity introduced from the various origination groups) components, respectively, thus resulting in the naming EMV.

The first requirement to the application of this methodology is related with the nature of the data. This model is applied to a population changing through birth, death or aging, i.e., through the arriving of new individuals, the leaving of individuals and the "natural" maturation of the individuals. Data on credit loans meets these criteria as it is composed by different loans entering in different moments and followed through their lives (until default or liquidation). Additionally, a dataset adequate to this methodology will exhibit a differentiated behavior profile along the three components. This is the case for default events, for which we can specify for each component a prompt interpretation, presented in the following paragraphs.

Relatively to the exogenous component, it is straightforward to state that default rates depend crucially of macroeconomic conditions, summarized by factors such as confidence indexes, interest rates, housing prices or unemployment rates. In this component, we expect a dynamic and volatile behavior of the default rate.

The loan´s age (maturity component) should help explaining the default event, regardless of the vintage where it belongs or the period. Empirical evidence shows that the typical lifecycle pattern of a credit follows four distinct phases: at the very beginning of the credit there is low risk of default, then the risk of default starts to rise, after that the credit enters in a phase of systematically decreasing risk and finally, the likelihood of a default event to occur is expected to stabilize (at a very low level). The length and the severity of each phase depend strongly on the credit product. For instance, in households' mortgages, the peak occurs between two and four years after the credit has been granted. Initially, almost all clients are "good" clients, as they are getting familiar with the product (think for instance a person subscribing her first credit card) and even against a bad event (like an illness or unemployment) they may use savings to face their payment duties. However, intentionally or due to misfortune, some clients fall into arrears and after, eventually they default, and exit the credit portfolio. Therefore, as time evolves, the portfolio will retain the best clients and accordingly, the default rate will gradually decline, until only the best ones remain and the rate stabilizes at some positive value.

Loans originated within a specific period shall be defined as a vintage. Each vintage's loan shares initial characteristics that distinguish them from the other vintages' loans and ultimately, may represent a forward-looking measure of future performance. The differences between vintages result mainly from the credit market conditions at each moment, either on the demand or on the supply side of the market. Consumers, enterprises and financial institutions do not exhibit the same propensity for booking (for the first two) or lending (for the later one) credit at every moment. Nor the same characteristics, as the applicants' profiles and underwriting standards change over time. For instance, collateral valuations, collateral requirements, loan amounts required or operative spreads vary. The vintage curve is affected mainly by changes in banks' internal operational policies, namely concerning to credit approvals and marketing strategy. However, economic variables may also be useful to identify singularities specific to loans' granted at different business cycle phases.

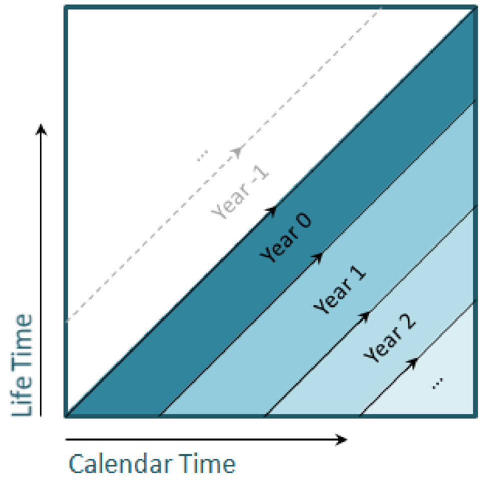
Every risk manager recognizes that this portfolio dynamics is present, but the majority think of each component as a separate issue. The EMV model allows modelling these components independently and simultaneously, in a tractable way. Moreover, unraveling exogenous, maturation and origination effects on historic information, is helpful for the goal of forecasting future trends.

## 3. EMV model

A database suitable for an EMV model consists of observed events (deaths, defaults, etc.) over a population composed by a restricted number of vintages (or birth cohorts), during a limited horizon and age span.

Conventionally, the data can be represented in a vintage diagram, as in Figure 1. The x-axis represents the period (calendar time), and the y-axis represents the age (maturity) of the entity (in this case, a loan). Each vintage starts in a point with coordinates (t, 0), then evolves to the right and upward at an angle of 45 degrees. Therefore, in a two-dimension graph, it is possible to identify simultaneously the period (x-axis), the age (y-axis) and vintage (diagonal bands) features.

**Fig. 1 Vintage diagram**




Generally, EMV analysis starts by the tabulation of the data, where rows and columns may refer to any two different components, and the table diagonals represent the third one. Suppose that we have a variable of interest, such as a default rate ($Y$), observed over time for a portfolio of loans. $Y$ is observed over time, for example, a series of months $(t = 1; ::: ; P)$ and is separately observed by age $(m = 1; ::: ; A)$ representing the maturity, or time on books, of the loan. The maximum maturity, $M$, is the largest value observed in the data. In addition to time $(t)$ and maturity $(m)$ the other factor which is initially used to explain variability in $Y$ is the vintage $(v)$ of the loan, which indexes the time of origination of a loan, is defined by

$$v = t - m$$

where the vintage that originates in the first period of observation can be labelled 1. For simplicity of notation, we define each observation of the average default rate, in a given period, maturity and vintage as,

$$Y_{mtv} = \frac{number\ of\ defaults}{total\ of\ loans\ at\ risk}$$

In Table 1, we present a possible tabulation of the data, where we can read all the components and rates in a single table. The maturity ($a$) is indexed to rows and the periods ($t$) to columns, such that the observations within a common vintage ($v$), i.e. loans originated in the same period, are read diagonally down to the right. Deliberately, vintages originated before the periods under analysis ($t < 1$) are excluded.

**Table 1**

**Contingency Table**

| | | Calendar time (period) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $t = 1$ | $t = 2$ | $t = 3$ | $\dots$ | $t = P - 2$ | $t = P - 1$ | $t = P$ |
| Maturity (age) | $m = 0$ | $Y_{0,1,1}$ | $Y_{0,2,2}$ | $Y_{0,3,3}$ | $\cdots$ | $Y_{0,P-2,P-2}$ | $Y_{0,P-1,P-1}$ | $Y_{0,P,P}$ |
| | $m = 1$ | | $Y_{1,2,1}$ | $Y_{1,3,2}$ | $\cdots$ | $Y_{1,P-2,P-3}$ | $Y_{1,P-1,P-2}$ | $Y_{1,P,P-1}$ |
| | $m = 2$ | | | $Y_{2,3,1}$ | $\cdots$ | $Y_{2,P-2,P-4}$ | $Y_{2,P-1,P-3}$ | $Y_{2,P,P-2}$ |
| | $\vdots$ | | | | $\ddots$ | $\vdots$ | $\vdots$ | $\vdots$ |

The easiest way to perceive whether age, period or vintage effects exist is to plot the data of the contingency table in a graph and then analyze the curves' patterns. The purpose of EMV modelling is to explain the variability in $Y$ as a function of the three components, age, maturity and vintage, and then to predict $Y$ into the future.

## 4. Previous implementations of the EMV model

In this section, we present some applications of the EMV decomposition approach to credit risk studies. Despite of its widespread use in demography, epidemiology and sociology studies, the use of the EMV tool in the credit risk context is still relatively scarce.

The dual-time dynamics model is a specific implementation of the EMV Model for consumer behavior modeling and delinquency forecasting, adopted by Strategic Analytics Inc. (Breeden 2007). Breeden et al. (2008) employed this technique to US mortgage data relating to historical default rates. Firstly, they decomposed nonparametrically the default rates into the three components. They considered 30-year of delinquency data for four segments. All exhibited maturation curves with profiles that corresponded to the expected. In a second step, they modelled the exogenous component with macroeconomic and management representative variables. In order to get a macroeconomic model for the exogenous component, they developed a multivariable model, considering lending rate, housing price index and unemployment rate, which captures quite well the evolution of the default rates. The vintage analysis identified three periods of higher default rates: 1994-1995, 2000-2001 and 2005-2006. The authors concluded that it was not the drop of the house prices that lead to the last peak, but instead, the maturation of the higher risk loans originated between 2004 and 2006 justified the 2007's peak of delinquency.

Zhang (2009) implemented the EMV methodology to credit risk modeling. The author started by applying the EMV methodology to Moody's annual corporate default rates, between 1970 and 2008, and found that the exogenous curve was more volatile than the other two and followed closely the macroeconomic cycles. Given the EMV framework, it is necessary to extrapolate each component function beyond the historical training data. The author adverts that this may be reliable for maturity curve, but not for exogenous or vintage curves due to the outside factors influencing the last two. Zhang (2009) further applied the EMV model to an American mortgage portfolio, collected between 2001 and 2007. He concluded that originations of 2005 and 2006 presented a higher default probability, confirming the previously mentioned effect of credit boom just before the beginning of the financial crisis. He found a strong relation between the increase of mortgage defaults and the declining of house prices and the rising of unemployment. Next, he added loan-level covariates in order to detect unobserved heterogeneity in the vintage effect. He found the expected relation with each covariate, namely relative to loan-to-value and to FICO scores. He concluded that the implementation of these methods to the credit risk area showed significant potentialities.

Forster and Sudjianto (2013) highlighted the relevance of the macroeconomic component for forecasting purposes. The authors applied the EMV decomposition to a credit card portfolio data of seven years. They also proposed a semiparametric approach, where the exogenous component is modeled with macroeconomic covariates, including the debt-to-income ratio, year-on-year increase in the log of the United Kingdom unemployment and year-on-year increase in the log of the debt-to-income ratio. The authors concluded that EMV models could be helpful in understanding the behavior of credit portfolios over time.

Strydom (2017) used the methodology set out by Zhang (2009) and Breeden (2007) to decompose the default rate for both a mortgage and personal loan portfolio of a South African bank. Over 500,000 individual accounts were tracked using the EMV decomposition, from 2007 to 2010. He found that the exogenous component accounted for a large share of the defaults from April 2008 to June 2009, indicating the deterioration of the credit quality of all loans during this period. The vintage impact on the mortgage default rates was above 50% at the start of the observation period, then reduced to 24% in October 2008 and increased again in 2009. The author found that the vintage and the economic cycles were the most important factors affecting the default rates of mortgage portfolios. Strydom (2017) then performed an out-of-sample study, using the EMV decomposition. He used a set of macroeconomic variables including interest rates, price indexes and other business cycle variables to estimate the exogenous component and combined this with the maturity and vintage components to forecast the portfolio default rates in 2013 and 2014. The author found that the estimated default rates using the EMV decomposition followed closely the observed default rates, in that out-of-sample period.

## 5. EMV data analysis

### 5.1. Contingency table

We begin by constructing the contingency table and examining it graphically. Our dataset includes monthly historic data of more than 100,000 mortgage loans in a time span of around 11 years from a representative Portuguese bank, from January 2007 to December 2017. In this section, we analyze the data aggregated annually, even though the raw data is monthly.

**Table 2**

**Contingency table for the dataset stratified by calendar year and maturity**

| | Calendar year | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Maturity | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| **]0,1]** | 0,2051% | 0,5491% | 0,3017% | 0,2078% | 0,3568% | 0,3826% | 0,0485% | 0,1266% | 0,1295% | 0,0805% | 0,1290% |
| **]1,2]** | | 0,9702% | 0,8419% | 0,3467% | 0,6822% | 0,8860% | 0,6236% | 0,2994% | 0,2115% | 0,2072% | 0,2349% |
| **]2,3]** | | | 0,7333% | 0,5713% | 0,4455% | 0,8950% | 0,7780% | 0,5053% | 0,2752% | 0,1222% | 0,1947% |
| **]3,4]** | | | | 0,5470% | 0,8160% | 0,7722% | 0,7995% | 0,7548% | 0,2958% | 0,2414% | 0,2314% |
| **]4,5]** | | | | | 0,6962% | 1,6660% | 0,5714% | 0,8644% | 0,7646% | 0,3994% | 0,2261% |
| **]5,6]** | | | | | | 0,8780% | 0,7699% | 0,5161% | 0,5783% | 0,4910% | 0,1693% |
| **]6,7]** | | | | | | | 0,5312% | 0,6938% | 0,4247% | 0,4629% | 0,4764% |
| **]7,8]** | | | | | | | | 0,5382% | 0,5880% | 0,3727% | 0,4054% |
| **]8,9]** | | | | | | | | | 0,3695% | 0,4414% | 0,3105% |
| **]9,10]** | | | | | | | | | | 0,1608% | 0,2943% |
| **]10,11]** | | | | | | | | | | | 0,2477% |

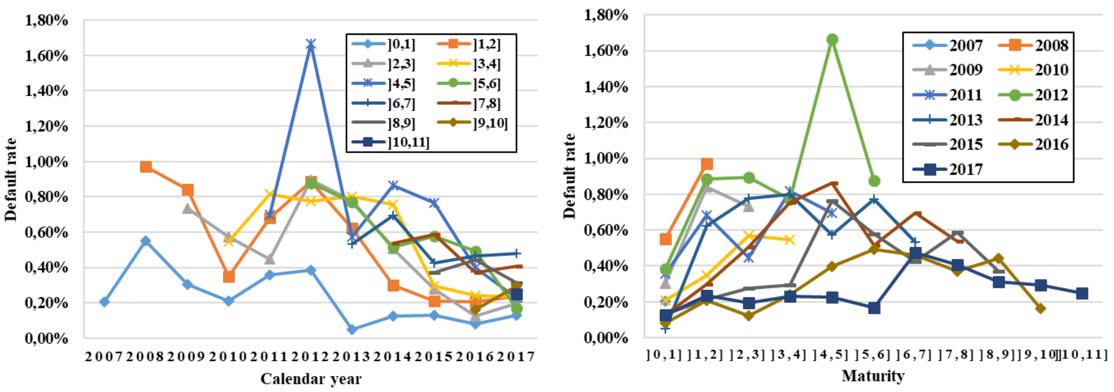
Table 2 displays the default rates for each calendar year (period) by maturity (years-on-book or age). Each cell of the table presents the percentage of loans with a certain maturity that defaulted in a specific year. Each row displays how the percentage of defaults changes across calendar year for loans within a given maturity. It is assumed that it is enough that a loan alive in a single month of a certain year within a specific maturity range to contribute to the calculation of the corresponding cell. Each column presents the maturity variation for each specific calendar year. Along the diagonals (from higher up towards the bottom right of the table), we may observe the behavior of each vintage over time.

The default rates range from 0.0805% in 2016 for loans with less than one year of maturity to 1.1666% in 2012 for loans within their fourth year of maturity.

The data of the table is plotted in the Figure 2, wherein each curve reveals the default rates for loans within a specific maturity, i.e. the values shown in each line of the contingency table.

Given the two chosen dimensions (calendar year and maturity), from Figure 2a we observe that the lowest default rates are observed on the loans with less than one year in maturity, and tend to increase until the maturity ]4,5] years. The highest default rate is observed in 2012, in this maturity. In fact, except for loans with maturity up to 2 years, the worst performance occurs in 2012. For maturities up to 2 years, the worst performance was in 2008 and 2009, at the beginning of the international financial crisis.

**Fig. 2 Contingency table stratified by calendar year and maturity**




*2a. Maturities* *2b.Calendar years*

In the years 2010 and 2011 default rates decreased and spiked in 2012. Note that Portugal faced a sovereign debt crisis during this specific period, which ultimately led to the application of a bailout program between April 2011 and May 2014. The graphs show that default rates tended to decrease every year, between 2013 and 2017.The recent years show also a worse performance for older loans, specifically the ones granted before 2012. After 2012, a general decrease of default rates is observed, signaling a vintage effect, i.e. loans granted in 2013 and in the following years performed much better. These results are similar to Leow and Crook (2016), who studied a large portfolio of credit card loans observed between 2002 and 2011 provided by a major UK bank, investigated the stability of the parameter estimates, and found that default rates were different for loans granted before and after the credit crisis of 2008. Strydom (2017) also found lower quality in the loans generated in South Africa prior to 2008 and claimed that this lower quality, together with the economic cycle effect, explained the increase in the default rates in that stressful period.

### 5.2. Exogenous, maturity and vintage curves

For a deeper analysis of the relevance of the EMV methodology to our dataset, we plot and examine the exogenous (E), maturity (M) and vintage (V) curves separately, using monthly data.

**Fig. 3 Exogenous curve (calendar time)**

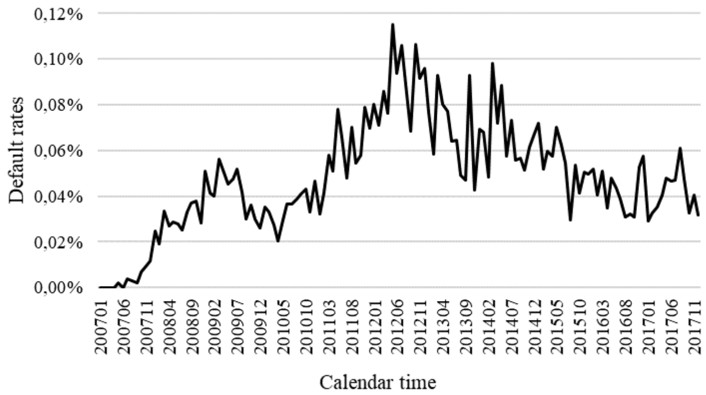



Figure 3 plots the default rates observed along the calendar time. As expected, the exogenous curve is dynamic and volatile. The major peak for default incidence occurs in May 2012, but all the period between January 2012 and June 2014 reveals high default rates. This period coincides with a deep sovereign debt crisis in Portugal, which ultimately lead to a bailout program.

Before 2012, we observe rising default rates until July 2009, then a modest improvement in the last quarter of 2009 and in 2010, and in the following years defaults started to rise, until the major peak. After June 2014, the graph shows a decline in default rates.

**Fig. 4 Maturity curve**

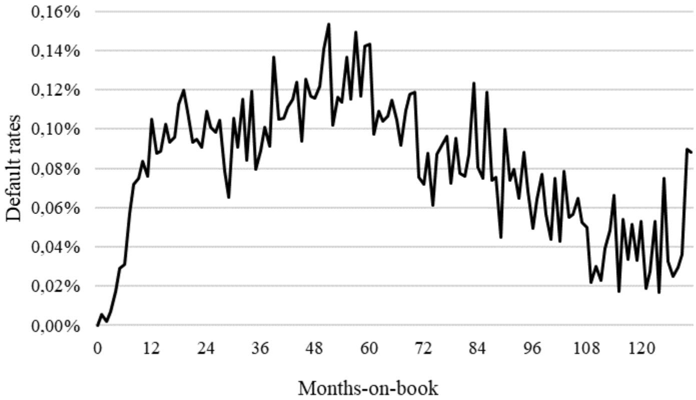



Figure 4 shows the defaults rates by maturity (i.e. by months-on-book). The plot is consistent with the typical maturity curve pattern: growing default rates until the loans get around 5 years of maturity and decreasing default rates in the following maturities. This declining phase may appear at first glance slighter than expected, but we shall notice that the time span of the dataset is around 11 years, which for a mortgage loan is less than the average time life. Nevertheless, there is evidence for the typical declining default occurrence beyond 5 years-on-book.

**Fig. 5 Vintage curve**

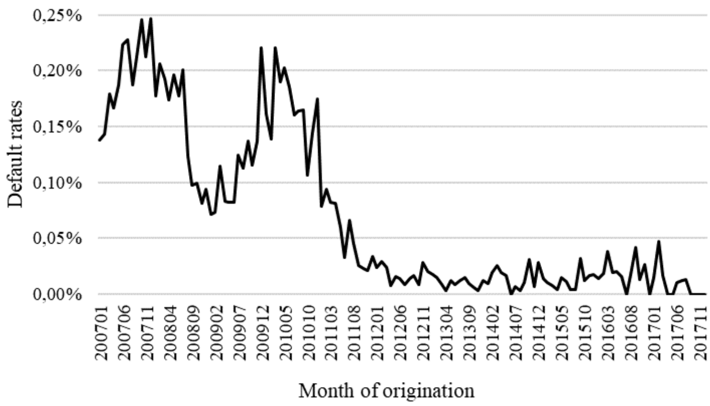



Figure 5 shows the default rates estimated for monthly vintages, where loans with a common origination month are aggregated and regarded as a vintage. There are clearly two major distinct periods in what vintage evolution concerns: the loans granted before September 2011 are characterized by higher delinquency rates, while loans granted after this date exhibit lower rates of default occurrence. We believe this fact results from changes on internal granting policies of banks.

Before September 2011, the loan vintages show greater heterogeneity. Along this period, there are three major peaks: one at the last quarter of 2007 and the others in December 2009 and March 2010, all reaching values above 0.20%. Note also the good performance of the vintages originated between October 2008 and October 2009, which may reflect an immediate reaction to the uncertain following the Lehman Brothers bankruptcy, in September 2008.

For vintages originated after September 2011, we observe less heterogeneity across vintages and the lowest default rates (always below 0.05%) of the sample. This behavior may be partly due to a camouflaged "maturity effect", because in the most recent vintages, the loans have lower maturities and may have not reached the maturity of around 5 years, where defaults rate peak. Nevertheless, the default rates in more recent vintage are clearly lower, and therefore likely reflect distinct vintage characteristics.

### 5.3. Explanatory variables

Previously to the modelling step, it is important to explore the variables that may be able to explain the exogenous and vintage curves. For this purpose, a set of variables is selected not only because they are likely to explain the default rates, but also because a subset of these are the most commonly used for designing stress scenarios (namely interest rates, gross domestic product, unemployment rate and house price index). Canais-Cerdá and Kerr (2014) showed that model specifications that incorporated interactions between macroeconomic variables and account characteristics performed better and generated projections that were more accurate.

However, the rationale for the inclusion of each explanatory variable in the EMV model needs to be validated.

#### 5.3.1. Explanatory variables for the exogenous curve

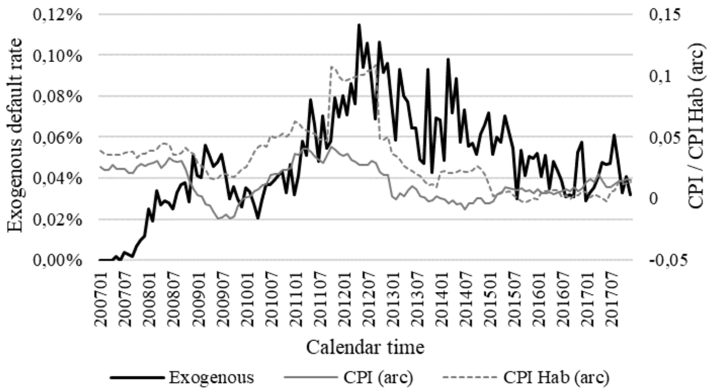
In this subsection, we explore a set of economic variables that may explain the behavior of the exogenous curve. We collected the historical series of the following variables: consumer price indexes (general and for housing, water, electricity, gas and other fuels); annual rate of change of gross domestic product (GDP); annual rate of change of private consumption; unemployment rate; 6 months Euribor rate; spread between 10 years' maturity of government bonds, for Portugal and Germany. These variables were selected not only because they are likely to explain credit defaults, but also because some of them are commonly used for designing stress scenarios, namely interest rates, GDP, unemployment rate and house price index. Similarly to Zhang (2009), Bellotti and Crook (2009), Forster and Sudjianto (2013), Strydom (2017) and Dirick et al. (2019), we analyze if these macroeconomic variables are correlated with the exogenous curve, and provide the rational for their inclusion as explanatory variables.

*Consumer price indexes*

We consider either the general consumer price index ("IPC") and the housing, water, electricity, gas and other fuels specific one ("IPC_Hab"), as the later may be higher correlated to the exogenous curve. We expect higher price levels to generate higher credit defaults.

In Figure 6, the exogenous curve is plotted together with the series of annual rates of change of the two consumer price indexes. It shows the expected behavior, for the housing expenses specific price index. Increasing annual rates of change are correlated with the rising default rates, which preceded the 2012's peak and lower rates afterwards. However, this relation is not observed prior 2010, and the overall correlation coefficient is only 0.293.

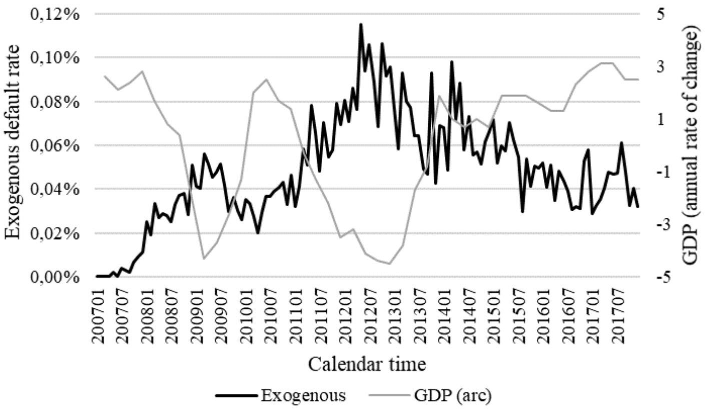
**Fig. 6 Exogenous curve and consumer price indexes (annual rates of change)**




*Change of GDP*

When the economy is growing, we expect lower credit delinquency, and the opposite is expected to occur, if the annual rates of change of GDP decrease. This pattern is visible throughout all the period under analysis (Figure 7) and the correlation coefficient is -0.564. The rise in the default events follows closely the fall of GDP and vice-versa, particularly between 2008 and 2014.

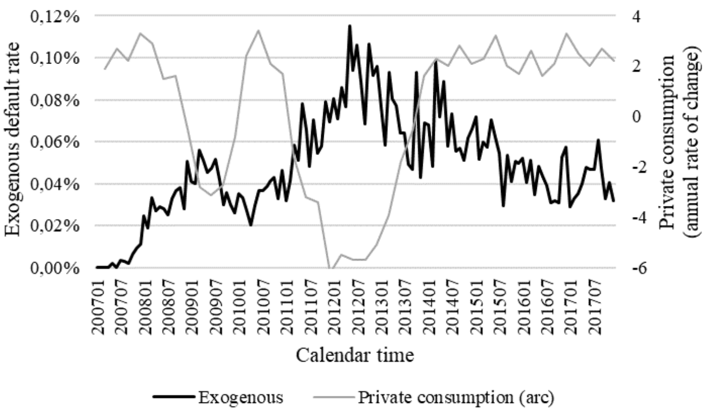
**Fig. 7 Exogenous curve and GDP (annual rate of change)**




*Change of private consumption*

Higher rates of change of private consumption mean a better economic environment and ultimately, less defaults. This pattern is seen throughout all the period under analysis (Figure 8), and the correlation coefficient is -0.541. The rise in the default events follows closely the fall of Private Consumption and vice-versa, particularly between 2008 and 2014.

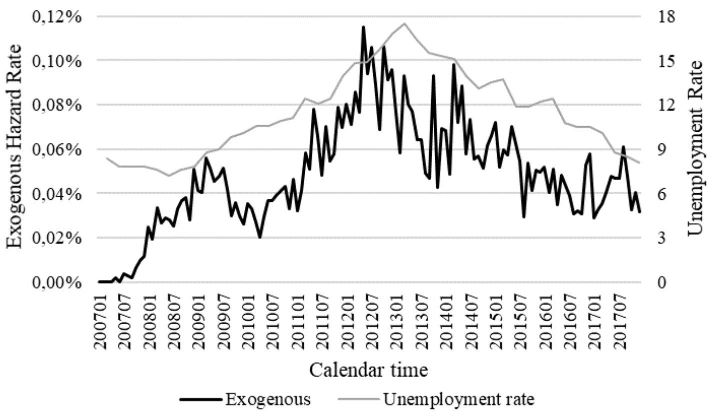
**Fig. 8 Exogenous curve and private consumption (annual rate of change)**



*Unemployment rate*

Employment is probably the most obvious driver for loan delinquency and generally reveals good correlation with default events, due to its direct impact on family's available resources to face loan's payments. We expect that the unemployment rate has a positive relation with credit delinquency. In Figure 9, the unemployment rate displays a triangle shape throughout the period under analysis, which accompanies the exogenous curve as expected. The correlation coefficient is 0.810.

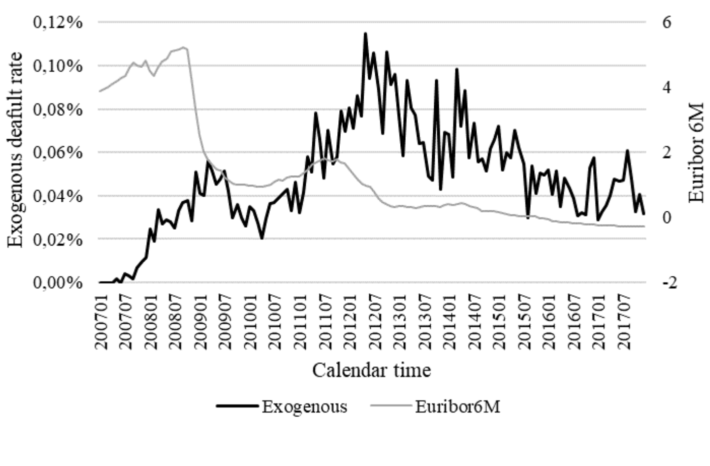
**Fig. 9 Exogenous curve and unemployment rate**




*6 months Euribor rate*

The 6 months Euribor rate series is the most commonly used interest rate for loans in Portugal. We expect, as the Euribor rate rises, the value of the loan payments to be made increase and ultimately, defaults may increase.

**Fig. 10 Exogenous curve and Euribor 6 months**

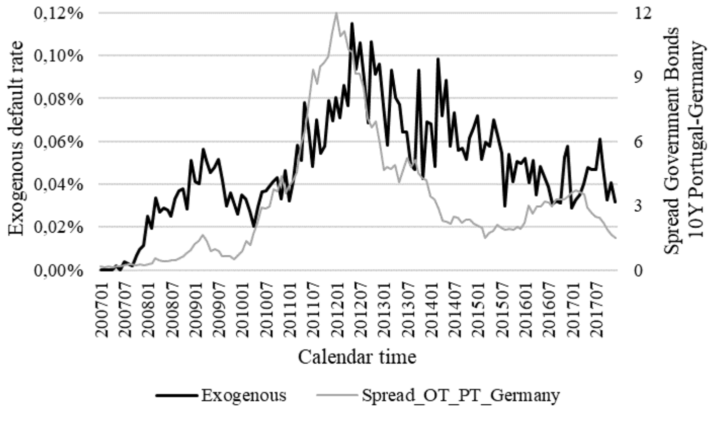


In Figure 10, the expected pattern is not visible throughout all the period under analysis. While along the neighborhood of the 2009's default peak the exogenous curve behavior appears to reflect the Euribor rate variation (particularly during the rising to the peak observed), afterwards the default rates do not respond as closely to the Euribor rate. The rising in Euribor rate until September 2008 may well justify the rise in defaults observed for that period. The following decline of the Euribor rate is accompanied by the expected decreasing of delinquency. However, the major peak of delinquency is not accompanied by a high level of the Euribor rate nor the declining of the delinquency occurred after appears to be due to a decrease of the Euribor rate level. The correlation coefficient also reveals the limited response of the exogenous curve to the Euribor rate variation. Contrary to previous expectations, its value is negative and equal to -0.487.

*Spread between 10 years' maturity government bonds for Portugal and Germany*

The recent financial crisis occurred in Portugal was, above all, a sovereign debt crisis. For this reason, the spread between 10 Years' maturity government bonds for Portugal and Germany may be a good covariate to consider for a macroeconomic explanation of defaults over the crisis' period. The spread reflects the perception of financial risk of the Portuguese economy comparatively with the German economy. Therefore, as the spread rises, default rates may upsurge.

**Fig. 11 Exogenous curve and spread between 10 years' maturity government bonds for Portugal and Germany**




This pattern is visible throughout the period under analysis (see Figure 11), and the correlation coefficient of 0.679. Particularly, the spread slightly rises during the 2009's delinquency peak and it severely increases and decreases along the 2012's delinquency peak.

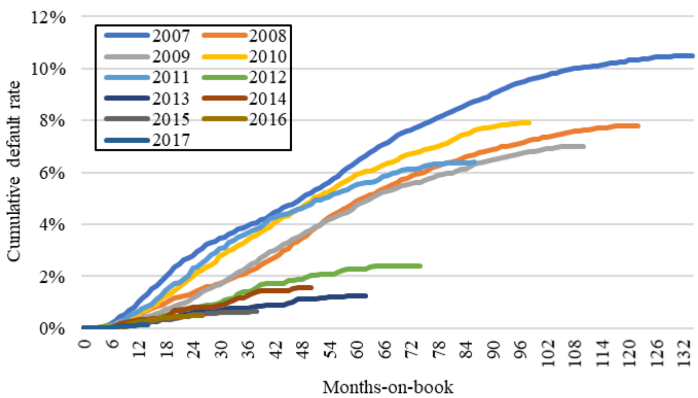
### 5.3.2. Explanatory variables for the vintage curve

*Cumulative default rate by vintage*

Before exploring the vintage specific characteristics, it is helpful to analyze the vintage's delinquency through an alternative perspective to the vintage curve. Therefore, we intend to characterize the vintages by their cumulative default rate, along age. Through this representation, we may not only assess the vintage's quality but also the evolution over the maturity for each vintage. This later feature shall be helpful to evaluate the best pattern to assign to the maturity curve.

In Figure 12 each curve displays, for a specific vintage, the percentage of defaults occurred since origination up to a given maturity (in months).

**Fig. 12 Cumulative default rate by vintage**



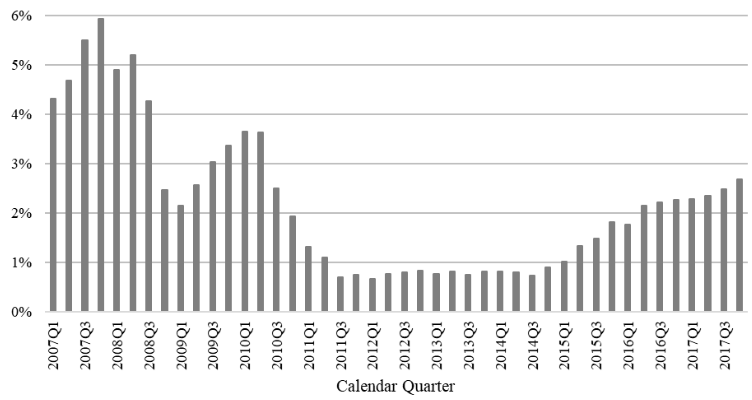

Clearly, there are two distinct periods in what credit quality concerns: the loans granted before 2012 and in 2012 or after. The first set of loans shows default rates much higher than the second set. The vintage of 2007 vintage is by far the lowest credit quality vintage throughout its time life. Then, from the worst to the best vintages we find the 2010 and 2011's granted loans and with a slightly better performance, the 2008 and 2009's vintages. The 2012 and 2014's vintage curves show a distinctive good behavior comparatively with the former and finally, the 2013, 2015 and 2016's loans denote the best vintage profile. For 2017 and 2018's vintage, it is early to conclude about their quality.

Relatively to the typical pattern of delinquency through maturity, there is some evidence for the older loans to sooth the increasing of the defaults beyond a certain age. For instance, for 2008 and 2009's vintages, the growth of defaults appears to have stabilized around 5 years of age, while for 2011's loans set defaults appear to have stabilized 1 year younger.

*Granting Evolution*

Figure 13 shows the distribution of the granted loans along time (by quarter).

**Fig. 13 Distribution of loans granted by quarter**




The highest percentages of loans granted are observed prior the third quarter of 2008. Then, a slight decrease in granting numbers is seen, which may well reflect an immediate reaction to the uncertain following the Lehman Brothers bankruptcy in September 2008. During 2009, a small rise in the percentage of loans granted is still observed, but the graph reveals undoubtedly that there was something dramatically affecting the granting evolution by 2010 (due to either supply's actions, demand's changes or even both). Indeed, half of the portfolio was granted until the end of 2009 and until the first quarter of 2010 the percentage of loans granted were much higher than afterwards. Between 2011 and 2014 there was the lowest percentage of loans being granted (less than 1% each quarter). After 2015, there was a successive rising proportion of loans being granted.

*Loan-to-value at the origination date*

For vintage distinction, the most obvious strategy is to use underwriting covariates upon origination, being the loan-to-value at the origination date the most commonly variable used for mortgages. The loan-to-value (LTV) corresponds to the loan amount over the collateral value.

**Fig. 14 Average loan-to-value**

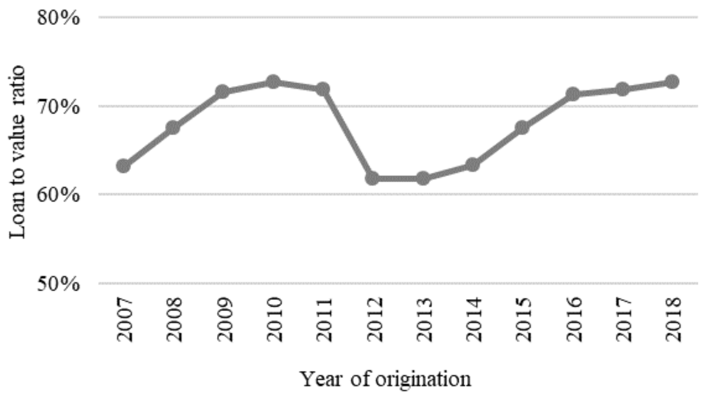



Figure 14 presents the average LTV of granted loans by year of origination. It shows that the loans granted in 2007 and 2012 to 2014 were granted supported in a lower average LTV (below 65%), while during the years of 2009 to 2011 and after 2016 the average LTV is above 70%.

## 6. Conclusions

Previously to any application of the EMV methodology, we must confirm its applicability to the dataset, such that the model fitting may translate the best description of data among the several possible templates.

To apply the EMV model, we examine the portfolio dynamics by graphical representation of defaults rates through time, decomposed by calendar time, age and vintage). While defaults rates along the maturity are expected to have a very well-shaped behavior, defaults rates by calendar period and vintage need to be confronted with the macroeconomic scenario and the internal operational policy, respectively. Additionally, this component analysis is helpful in identifying the patterns to be fitted.

In this study, we find strong evidence to support the application of the EMV model to a Portuguese mortgage loans dataset. We find that the exogenous, maturity and vintage curves follow the expected behavior, and we identify a set of explanatory variables suitable to be incorporated in a EMV model specification.

## References


Bellotti T, Crook J (2009) Credit scoring with macroeconomic variables using survival analysis. Journal of Operational Research Society 60, 1699–1707. https://www.jstor.org/stable/40295738

Breeden JL (2007) Modeling data with multiple time dimensions. Computational Statistics & Data Analysis 51(9):4761–4785. https://doi.org/10.1016/j.csda.2007.01.023

Breeden JL, Thomas L (2008) The relationship between default and economic cycles for retail portfolios across countries. Journal of Risk Model Validation 2:11–47. https://eprints.soton.ac.uk/156515/

Canais-Cerdá JJ, Kerr S (2014) Forecasting credit card portfolio losses in the Great Recession: a study in model risk. Journal of Credit Risk 11(1):29-57. https://doi.org/10.21314/JCR.2015.187

Dirick L, Bellotti T, Claeskens G, Baesens B (2019) Macro-economic factors in credit risk calculations: Including time-varying covariates in mixture cure models. Journal of Business & Economic Statistics 37(1):40-53. https://doi.org/10.1080/07350015.2016.1260471

Forster JJ, Sudjianto A (2013). Modelling time and vintage variability in retail credit portfolios: the decomposition approach. Lloyds Banking Group, Risk Division, Analytics and Modelling. https://arxiv.org/abs/1305.2815

Glenn ND (2005). Cohort analysis (2 ed.). London: Sage Publications. https://epdf.pub/cohort-analysis-2nd-ed-quantitative-applications-in-the-social-sciences.html

Keyes KM, Utz RL, Robinson W, Li G (2010) What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. Social Science Medicine 70(7):1100-1108. https://doi.org/10.1016/j.socscimed.2009.12.018

Leow M, Crook J (2016) The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. European Journal of Operational Research 249(2):457–464.

https://doi.org/10.1016/j.ejor.2014.09.005

Strydom P (2017) Macro economic cycle effect on mortgage and personal loan default rates. Journal of Applied Finance & Banking 7(6):1-27. https://www.scienpress.com/journal_focus.asp?main_id=56&Sub_id=IV&Issue=389663

Zhang A (2009). Statistical methods in credit risk modeling. Dissertation, University of Michigan. http://hdl.handle.net/2027.42/63707