Credit Risk Modelling Under Recessionary and Financial Distressed Conditions

Dendramis Y.[*] Tzavalis E.[†] Adraktas G. [‡]

January 18, 2016

Abstract

This papers provides clear cut evidence that recessionary and financial distressed conditions, as well as banning foreclosure laws, often introduced by governments to mitigate the effects of the economic and/or financial distressed conditions on mortgage loans, have adverse effects on the loan default probability. We argue that this may be attributed to long-term persistency of the above conditions, which can cause abrupt shifts in the probability of default of a loan. Our estimates indicate that these policies may also raise moral hazard incentives that borrowers will not maintain their payments in long run, even for loans with low LTV. Under these conditions, efforts of banks to restructure (or refinance) mortgage loans may not successfully affect future default probabilities. Our evidence is based on an extension of the discrete-time survival analysis model which allows for a structural break in its hazard rate function due to abrupt changes to exogenous events, like changes in political conditions. It is also robust to alternative specifications of the binary link function between default events and covariates. Asymmetric link functions are found to be more appropriate under financial distressed conditions.

JEL classification: G12, E21, E27, E43

Keywords: mortgage loans, survival analysis, structural breaks, financial distressed conditions, probability of default.

The authors would like to thank Jonathan Crook and the participants in the Credit Scoring and Credit Control XIV Conference held in Edinburgh 2015, for helpful comments and discussions. The views presented in the paper are those of the authors alone and do not present official views of their institutions.

[*]Department of Accounting & Finance, University of Cyprus, email: dendramis.yiannis@ucy.ac.cy.

[†]Department of Economics, Athens University of Economics and Business, email: etzavalis@aueb.gr, Corresponding author.

[‡]Retail Banking Credit Risk Division, email: georgios.adraktas@gmail.com.gr

1 Introduction

Since the subprime crisis, there is growing interest in modeling probability of default (PD) on residential mortgages (see, e.g., Gross and Souleles (2002), Elul et al (2010), Crook and Banasik (2012), Divino et al 2013, Campbell and Cocco (2015), among others). These mortgages constitute a large proportion of banks loan portfolios and household debt, and thus their default risk is of primary concern for both credit institutions and households. By Basell II and/or III accords, banks are required to hold a minimum amount of capital to absorb expected losses on mortgage defaults.

There are two prevailing views on what can explain the default probability of residential mortgages referred to in the literature as the negative equity and ability-to-pay hypotheses. According to the first hypothesis, a mortgage defaults if the value of the mortgaged property becomes less than the loan value. This can happen when house (property) prices decline. This hypothesis can be empirically examined by testing if the loan-to-value (LTV) ratio, where loan stands for the amount of the loan and value stands for the most recently estimated market value of its underlying collateral, is positively related to PD, especially for values of LTV bigger than unity.[1]

The ability-to-pay hypothesis claims that the PD of a loan depends on the ability of the obligor (borrower) to meet his/her periodic payments (installments). This ability may depend on changes in demographic characteristics (see, e.g., Folliant et al (1999)), liquidity constraints (see, e.g., Gross and Souleles (2002) and Elul et al (2010)) and changes in business cycle (macroeconomic) conditions, which are not expected at the time that the loan was granted. Key macroeconomic variables used to capture these conditions are the unemployment and/or growth rates of an economy (see, e.g., Bajari et al (2008), Gerardi et al 2013 and Gyourko and Tracy (2014)). In addition to above variables, one may also include changes of inflation and mortgage interest rates which affect obligors' real per capita income and their ability to pay interest rate and principal on their loans.

It is worthwhile to note at this point that the above hypotheses are not competitive in explaining PD, but they interact with each other. For instance, changes in business conditions also affect house prices and LTV, and thus they can be also thought of as being consistent with the negative equity hypothesis explanation of PD. Negative equity is not a sufficient condition for a borrower

[1] See Titman and Torous (1989), for a first empirical evaluation of this hypothesis.

to default, as also noted by Elul et al (2010). If the ability to pay of a borrower is high or his/her liquidity position is adequate, then she/he may prefer not to default on her/his loan, but to wait for house prices to restore.

Apart from the above variables, we can also consider institutional factors that influence PD. These can include lenders' policies (or practices) and/or government laws (or acts) helping borrowers with financial difficulties to prevent them from falling behind on payments and default. Under financial distressed conditions, to avoid massive losses and accommodate borrowers' financial conditions many banks tend to be actively involved in loan restructuring (or refinancing) of delinquent (or defaulted) loans. This is often done by extending the maturity term structure of existing loans and/or by reducing their installments (or payment) rates. To motivate such practices on behalf of credit units and also to protect the economy from massive collaterals liquidation, governments often introduce acts which ban foreclosures for a specific period of time. The banning period extends the time between foreclosure and collateral liquidation procedures. During this period, macroeconomic conditions and borrowers's financial and/or ability-to-pay conditions may improve. Thus, defaulted loans may be cured and lenders may avoid loses from massive property sales in auctions markets. Examples of such acts are the laws 3814/2010 and 4128/2013 of the greek government banning foreclosures on first residence. These laws do not allow banks to proceed with liquidations on their residential collaterals, in case of a defaulted borrower and given that all collections or legal actions have been exhausted. They can also raise moral hazard incentives that borrowers with ability to pay and without liquidity problems will not pay back their loans.

In this paper, we empirically examine how the above policies of banks and government affect PD. To our knowledge, this is the first paper which examines these questions. Ignoring these effects may obscure the true relationship between PD and behavioral (like the LTV), or macroeconomic, variables which are often assumed that determine this relationship. Our analysis relies on a unique panel data set consisting of loan-level residential data from the Greek economy. In particular, our data set consists of 85230 Greek individual mortgage accounts (loans) on monthly frequency, covering the period from 2008:01 to 2014:10. Using panel data, instead of cross-section data sets consisting of loan portfolios, enables us to reveal the dynamic effects of behavioral and macroeconomic variables on PD. Also, we can capture the qualitative effects of restructuring loan decisions and government acts banning foreclosures on PD. The Greek economy constitutes an interesting case to examine if the above institutional policies influence PD, under persistent recessionary and financial distressed conditions. During our sample, the Greek economy has experienced a severe and prolonged economic and financial crisis which led to a loss of its GDP by 24.6%, the unemployment rate increased from 7.8% in 2008 to 26.5% in 2014, whilst the residential real estate prices dropped cumulatively by 36.8% compared to their peak in 2008.

Our modelling approach of PD is based on a discrete-time survival analysis model, which allows for calculating PD of a borrower in a future period conditionally on the assumption that she/he has not been previously defaulted. Discrete-time survival models constitute reduced form econometric panel data models which are frequently used in the literature to examine if the equity and/or ability-to-pay hypotheses can explain mortgage defaults.[2] These models can estimate time-varying hazard rate of mortgages (reflecting the expected number of defaults), taking into account a number of application, behavioral and macroeconomic covariates, at a point of time, as well as the lenders' practices and foreclosures procedures discussed above that can influence default rates. These covariates or procedures control for the time variation of the baseline hazard rate function of a mortgage during its life. They can help us in understanding behavioral and/or exogenous sources of mortgage defaults. Discrete-time survival models can also show how long the mortgage survives before its default in a future period. They can be employed to provide out-of-sample probability and point forecasts of mortgage defaults for a future period of time, given current information on behavioral, macroeconomic variables and other events.

In our analysis, we have extended the discrete-time survival model into the following directions. First, we have allowed for a break in its baseline hazard rate function due to severe economic and/or credit crunch conditions of the economy after a point of time (see Leow and Crook (2014)). Instead of assuming this break point as known, we treat it as unknown and we estimate it from our data, endogenously, by adopting a search estimation procedure over a grid of possible breaks during our sample. This method can shed light on exogenous events that have played an important role in deteriorating the financial conditions of the economy and, hence, have caused abrupt shifts in the PD rate. The second extension of the model is that, apart from the logit function, we also

[2] See, e.g., Deng et al (2000), Calhoun and Deng (2002), Gross and Souleles (2002), Gerardi et al 2013, Crook and Bellotti (2010), and Bellotti and Crook (2013).

consider the cloglog function as a link function between PD and the different vectors of covariates (explanatory variables) assumed in our analysis. As it has heavier asymmetric tails compared to the logit, the cloglog function can better capture asymmetric responses of covariates on extreme events (see, e.g., Koenker and Yoon (2009)).

The paper provides a number of results which have important policy implications. First, they indicate that, during our sample, there is a clear cut structural break in the baseline hazard rate function after the end of year 2011, which has increased this rate upwards. This may be associated with the political uncertainty in year 2012 (double elections) and the severe deterioration of the financial and economic conditions in that year. Second, banning foreclosure laws tend to increase the future default probability. Our results also indicate that these procedures may raise moral hazard incentives that borrowers will not maintain their payments in long run, even for loans with lower than unity LTV. We argue that this may be attributed to the prolonged recessionary conditions held in the greek economy, during our sample. Under these conditions and incentives, efforts of banks to restructure (or refinance) mortgage loans may not successfully affect future default probabilities. Regarding the application, behavioral and macroeconomic variables used in our analysis, the paper finds that these variables have the predicted by the theory effects on PD. The above results are robust to different specifications of the binary link functions used in our analysis, namely the logit and cloglog functions. The paper shows that between these two functions, the cloglog exhibits better in- and out-of-sample forecasting performance.

The paper is organized as follows. Section 2 presents the discrete-time survival model and its extensions. Section 3 presents the estimation procedure of identifying the break point in the baseline hazard function rate. This is done under different specifications of the binary link function, as mentioned above. This section also presents the estimation results of the model and carries out an out-of-sample forecasting exercise to evaluate the performance of alternative specifications of the model considered. Section 4 concludes the paper.

2 Model setup and extensions

2.1 Model set up

Assume discrete time, denoted as $t = 0, 1, ..., T$, and that at $t + h$ a default event may happen, where $h = 0, 1, 2, ..., H$ are h periods (months) ahead. This default event can be captured by a binary variable $Y_{i,t+h}$, where $i = 1, 2, ..., N$ stands for an individual loan account. $Y_{i,t+h}$ takes the value 1 if a default occurs, and zero otherwise, i.e., $Y_{i,t+h} = 1$ for default and $Y_{i,t+h} = 0$ for non-default. A standard discrete-time survival model assumes that the probability of default (PD) of a loan i at time $t + h$, denoted as $P_{i,t+h}$ is given by the following reduced form equation:[3]

$$\begin{aligned} P_{i,t+h} &= \Pr(Y_{i,t+h}|Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \quad i = 1, 2, ..., N \text{ and } t = 1, 2, ..., T \\ &= \Phi(b_0 + b_1'\varphi_{t+h-d_i} + b_2'\xi_{i,t+h-d_i} + b_3'x_{i,t-m} + b_4'z_{t-p}), \ \forall\ h \end{aligned} \tag{1}$$

where $\Phi(.)$ is a function which links probability $P_{i,t+h}$ to different vectors of covariates. In particular, $\varphi_{t+h-d_i} = (d_i, d_i^2, \log(d_i),\ \log(d_i)^2)'$ is a vector of linear and nonlinear functions of the duration time of loan i, denoted as d_i, which reflects the number of periods (months) that loan account i has opened until $t + h$, i.e., $d_i = 1, 2, ...D_i$ months. The inner product of the vectors of slope coefficients b_1' and φ_{t+h-d_i}, i.e., $b_1'\varphi_{t+h-d_i}$, gives the baseline hazard rate of default, for different values of d_i. This is a smooth function of d_i, which captures the deterministic pattern of PD, for all loan accounts i. $\xi_{i,t+h-d_i}$ denotes a vector of application variables (AV) which is known only at the time of a loan application $t + h - d_i$, i.e., when a loan account was opened. Finally, $x_{i,t-m}$ is a vector of behavioral variables (BV) collected over the life of the loan and z_{t-p} is a vector of macroeconomic variables (MV), which are common for all individual loan accounts, i, in any period t of our sample. The lag structure of vectors $x_{i,t-m}$ and z_{t-p} (i.e., m and p) are appropriately chosen to provide conditional forecasts of $P_{i,t+h}$ based on the current information set I_t, which includes all the above explanatory variables.[4]

The application variables that we consider in our analysis include dummy variables capturing (i) *urban effects* on $P_{i,t+h}$, (ii) *age effects* associated with the age of the borrower at the time

[3] See, e.g., Bellotti and Crook (2013), and Hwang and Chu (2014).

[4] Note that m and p can take negative values, when they stand for lead orders.

of application (categorized by the following years of old intervals: 18-30, 30-40, 40-50), and (iii) housing and/or repair categories of a loan, denoted as *product codes.* If the purpose of a loan is for a house purchase, it is denoted as product code 1, for a house repair as product code 2 and, finally, for any other use as product code 0.

As behavioral variables (BV), we consider the ratio of delinquent amount to the contract amount, denoted as DTC, and the ratio of the total balance to the most recent collateral valuation, referred to as loan-to-value (LTV) in the literature. Since LTV may not capture, efficiently, negative equity effects (known also as strategic defaulters' effects) on $P_{i,t+h}$ for values of it close to unity, due to transaction costs of defaulting, moving and reputation costs etc (see Bhutta et al (2010)), we have also included a specification of this variable which considers, separately, the effects of LTV values bigger or equal than 120% on $P_{i,t+h}$. This variable is denoted as LTV≥120%. It equals to LTV, if LTV is 120% or more, and zero otherwise. The 120% value of the LTV is a threshold variable above which strategic defaulters may choose to default (see Goodstein et al 2011).

The variable DTC, defined above, constitutes a measure of delinquency to the total debt of the borrower. It is a measure of the consistency of the borrower to pay his/her loan. Another measure of this consistency is the number of times that a delinquent obligor has a positive amount in bucket 1 (i.e., it is past due for up to one month), in bucket 2 (past due up two months) and in bucket 3 (past due up to three months), over the history of a loan. This variable is defined as *delinquency buckets* (DB). This variable and DTC may be thought of as complementary variables in measuring the effects of borrower's delinquent attitude on $P_{i,t+h}$. DTC can be thought of as capturing the magnitude of this risk on $P_{i,t+h}$, while DB as reflecting the strength (degree) of this attitude per borrower.

In addition to the above behavioral variables, we also consider a dummy variable capturing the effects on $P_{i,t+h}$ of restructured loans by the banks. These are currently performing loans, assigned in new accounts, which have been defaulted before at least for one time. This dummy variable is defined as *redefaulted loans.* It takes the value 1, if a loan i has defaulted before time t, and zero otherwise. Including this variable into model (1) can indicate if loan restructuring practices of banks aiming to mitigate the effects of recessionary and financial distressed conditions on debt servicing have any significant effects on $P_{i,t+h}$.

Another dummy variable used in our analysis to capture behavioral effects on $P_{i,t+h}$ is denoted as *foreclosure moratorium (*FM*).* This variable equals to 1 for all the loan accounts protected by government laws (or acts) which ban foreclosures, and zero otherwise. Including this variable into analysis can indicate if government laws which do not allow banks to proceed with liquidations on their residential collaterals have important effects on $P_{i,t+h}$. As such laws may raise moral hazard incentives that borrowers with low LTV will not pay back their loans, we have also included a dichotomous variable, denoted as FM×LTV $\leq$ 90%, among the behavioral variables of model (1). This variable is equal to the LTV value of the loan accounts protected by the foreclosure moratorium law if LTV is less, or equal, to 90%, and zero otherwise. Since the value of the loan collateral is bigger than that of the loan, this variable may capture borrowers' motivation to avoid paying back their loans, by exploiting laws on foreclosures ban.

Finally, as macroeconomic variables, we consider the inflation and unemployment rates, and a weighted average of loan interest rates of the mortgage market. These variables constitute key macroeconomic variables in many models determining PD. They are also sampled on monthly basis, which is consistent with the frequency of our level-loan data. In our empirical analysis, we have also considered the growth rate of Gross Domestic Product (GDP). This was interpolated on a month to month basis, as it is given on quarterly frequency. But, this variance is not found to be significant at the 5% level, for all the alternative specifications of the model estimated. One can argue that its effects on $P_{i,t+h}$ can be better captured by the unemployment rate, given the monthly frequency of our data.

2.2 Extensions of model (1)

Since severe changes in economic and credit conditions may have caused abrupt shifts on PD, model (1) is extended to different directions. First, a vector of time(year)-specific dummy variables, denoted as v_t, which are common across all i, are included into its intercept to capture effects, like elections, government interventions or any policy announcements, on $P_{i,t+h}$. Second, to see if there is a point in our sample, say T_0, after which a structural break type of change occurs in baseline hazard rate function $b'_1\varphi_{t+h-d_i}$, the vector of slope coefficients of this function b_1 is broken into two components, denoted as $b_1^{(0)}$ and $b_1^{(1)}$, respectively, before and after break point T_0. By splitting vector b_1 in this way, we can investigate if the baseline hazard rate of the model remains stable, or change substantially, over our sample, reflecting changes in the credit conditions of the economy (see Leow and Crook (2014)). In our analysis, we will treat point T_0 as unknown and we will estimate it from the data. To this end, we will adopt a sequential maximum likelihood (ML) estimation procedure (see, Zivot and Andrews (1992), and Andrews (1993)). According to it, the selection of T_0 is viewed as the outcome of maximizing the conditional log-likelihood function of model (1) over a set of possible points T_0 in our sample. Based on the above extensions, model (1) can be written as follows:

$$\begin{aligned} P_{i,t+h} &= \Pr(Y_{i,t+h}|Y_{i,t}=0;\xi_{i,t+h-d_i},x_{i,t-m},z_{t-p}) \quad i=1,2,...,N \text{ and } t=1,2,...,T \\ &= \Phi(b_0+\gamma_1'v_t+b_1^{(0)\prime}\varphi_{t+h-d_i}+b_1^{(1)\prime}\varphi_{t+h-d_i}+b_2'\xi_{i,t+h-d_i}+b_3'x_{i,t-m}+b_4'z_{t-p}), \quad \forall\ h, \quad (2) \end{aligned}$$

where $b_1^{(0)\prime} \neq b_1^{(1)\prime}$, with $b_1^{(0)\prime}=0$ if $t \geq T_0$ and $b_1^{(1)\prime}=0$ if $t<T_0$.

Another extension of the model concerns the link function $\Phi(.)$. Apart from the logit function which is derived by the logistic distribution $F(w)=\frac{e^w}{1+e^w}$, we will employ in our analysis the cloglog function $\Phi(w)=\log(-\log(1-w))$, derived by the extreme value (or log-Weibull) distribution given as $F(w)=1-e^{-e^w}$ (see Lahiri and Yang (2013)). The cloglog link function has heavier and asymmetric tails than the logit, or probit, one, and thus it can better capture asymmetric responses of covariates on extreme events.

3 Empirical analysis

In this section, we present and discuss the results of our empirical analysis. Our analysis proceeds as follows. First, we present the estimates of model (1), which does not consider a break in its baseline hazard rate function. This is done for alternative specifications of the model which consider logit and cloglog link functions $\Phi(.)$, as well as year-specific dummies and qualitative dummy variables counting for *redefaulted loans* and *foreclosure moratorium* effects. Then, we estimate the above all specifications of the model allowing for a break in the baseline hazard function (see model (2)). Finally, we carry out an out-of-sample forecasting exercise to evaluate the relative performance of the above specifications of model (2) to forecast future average default probabilities.

3.1 Data

Our data set consists of a very rich set of a Greek bank individual loan accounts whose frequency is monthly. It covers the period from 2008:01 to 2014:10 and it consists of 85230 accounts (loans). The dummy variable capturing *urban effects* takes 1 for the loans around Athens (Attica), and zero otherwise. Regarding the collateral evaluation of loans, used to calculate LTV, we have recognized some special features of the Greek economy over the recent period. To this end, we use real estate indices provided by the bank of Greece to adjust the values of the collaterals, whose prices dropped cumulatively by 36.8% over our sample. To cover the different categories of collaterals, we have used five different real estate indices. These include: the Residential Real Estate, Warehouse /Storage, Building ground/Construction, Field for Utilization/Animals, Offices, Stores/Shops, Industrial and Agricultural fields.

Figure 1 graphically presents the key behavioral and macroeconomic variables employed in our analysis, namely the consumer price index (from which inflation is calculated), unemployment and average mortgage rates, and the behavioral variables LTV, DTC and DB. For exposition reasons, the behavioral variables were aggregated across all loan accounts i, at any point of time t. For the LTV and DB variables, we present the upper and lower 5% quantiles of them and their average value, over all i. Figure 2 presents graphs of the number of the defaulted and restructured by the bank loans, as well as the loans protected by government law banning foreclosures. Values of the correlation coefficients among the number of defaulted, restructured and protected loans, over time, and the above behavioral and macroeconomic explanatory variables are given in Table 1. Note that the table gives two sets of correlation coefficient values: the first is based on contemporaneous values of all variables, while the second is based on three-periods (months) lagged values of the behavioral and macroeconomic variables.

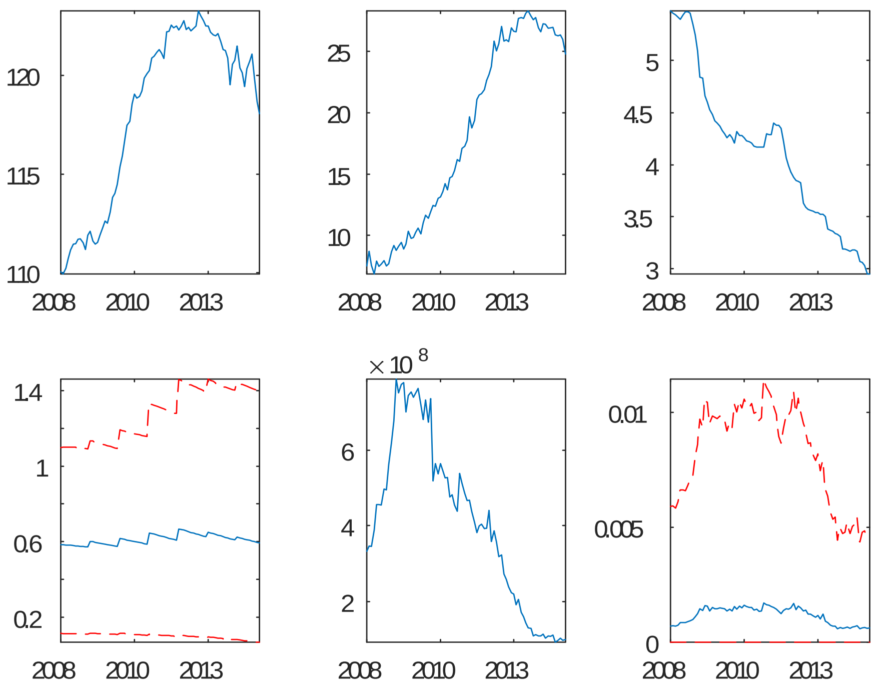

Figure 1: Graphs of the macroeconomic and behavioral variables; Top plots, left towards right side: CPI, unemployment and mortage interest rate) Bottom plots, left towards righ side: LTV(10% upper quantile, average values and lower 10% values), DTC and DB (10% upper quantile, average and lower 10% quantile).

Inspection of Figure 1 clearly indicates that the unemployment rate increases almost linearly, during our sample. It reaches its peak in the begging of year 2013 and, then, it starts declining. A similar picture appears for the consumer price index, used to calculate the inflation rate. Finally, the mortgage rate steadily declines during our sample, with the exception of a short period after April 7th 2011, where the ECB increased its lending rate to tighten inflation expectations. However, since year 2012 the ECB started lowering its interest rates to get banks to lend more to credit-starved customers, taking also into account that inflation expectations had been falling since the beginning of this year. Regarding the behavioral variables, the figure indicates that the upper quantile of the LTV, which presents the most likely to default loans, increases steadily since the beginning our sample, in year 2008, and it becomes stable towards the end of our sample, after year 2013. The two other behavioral variables, graphically presented in Figure 1 (namely, DTC

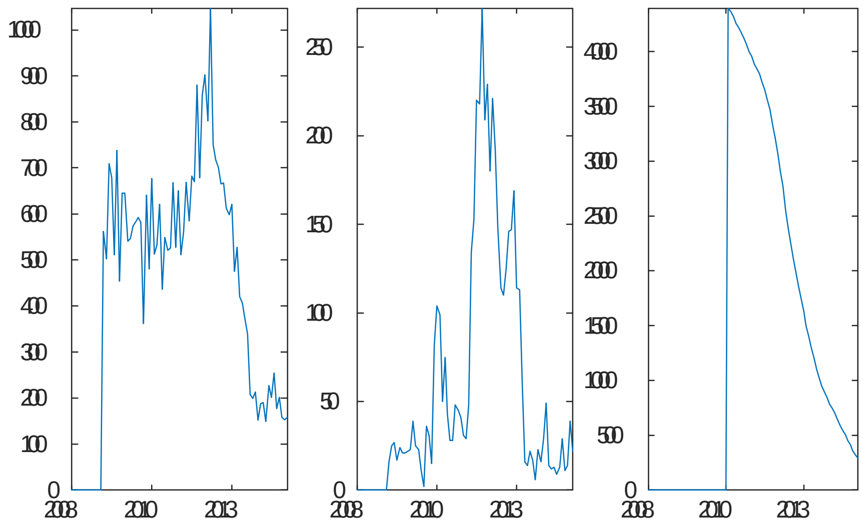

Figure 1: Figure 2: Defaulted, restructured and protected residential loans (from the left towards the right side)

and DB), exhibit similar patterns to that of LTV.

Figure 2 reveals that the number of defaulted loans during our sample is very volatile. This number reaches its peak in year 2012, and since then it starts declining until the end of the sample, where its becomes stable. A similar pattern is observed for the number of restructured loans. The number of protected loans picks up immediately after the implementation of the law on foreclosures ban, in year 2010. A comparison between Figures 1 and 2 indicates that the above changes (rises and falls) of the number of defaulted and restructured loans can be associated with those of the macroeconomic and behavioral variables, presented in Figure 1. As shown in Table 1, which presents values of the correlation coefficients among the above categories of variables, changes in the unemployment and inflation rates, as well as in LTV and DTC are positively and highly correlated to the number of defaulted, restructured and protected loans. This is true for both set of values of the correlation coefficients reported in the table, i.e., based on contemporaneous values of the behavioral and macroeconomic variables and their lagged values. Another interesting observation that can be made from the results of Table 1 is the very high positive correlation of unemployment and inflation rates with LTV. This may be taken as evidence that the deterioration in the macroeconomic and credit conditions, observed during our sample are closely linked.

Table 1: Correlation Coefficients

	Contemporaneous (with three-lags in parentheses)					
	LTV	DTC	DB	unemployment	inflation rate	mortgage rate
Defaulted loans	0.55 (0.33)	0.80 (0.82)	0.35 (0.47)	0.16 (0.009)	0.41 (0.30)	-0.20 (-0.07)
Restructured loans	0.57 (0.48)	0.40 (0.44)	-0.08 (-0.007)	0.37 (0.28)	0.59 (0.54)	-0.21 (-0.14)
Protected loans	0.54 (0.42)	0.44 (0.51)	-0.07 (-0.009)	0.29 (0.18)	0.65 (0.59)	-0.16 (-0.15)
LTV				0.69 (0.62)	0.81 (0.79)	-0.55 (-0.52)
DTC				-0.31 (-0.40)	0.07 (-0.10)	0.23 (0.23)
DB				-0.19 (-0.17)	-0.56 (-0.69)	0.63 (0.73)

Notes: This table presents values of the correlation coefficients of the number of defaulted, restructured and protected loans with key explanatory variables of the model, such as LTV, DTC, DB, and unemployment, inflation and mortgage interest rates. It presents correlation coefficients based on contemporaneous values of the above explanatory variables and, second, based on three-periods (months) lagged values of them (see values in parentheses).

3.2 Estimation Results

In this section, we present maximum likelihood (ML) estimates of the parameters of models (1) and (2), denoted by vector θ. These are given in Tables 2 and 3, respectively. Both of these tables present parameter estimates of different specifications of the models with the logit and cloglog link functions $\Phi(.)$, and for their versions considering also the qualitative dummy variables capturing the *redefaulted loans* and *foreclosure moratorium* effects on probability of default $P_{i,t+h}$. The lag order of both the behavioral and macroeconomic variables used in the estimation of the models are chosen to be m=p=3 months, which enables to forecast default probabilities $h = 3$ months ahead. The macroeconomic variables employed to estimate the models are quarter-on-quarter differences of them, obtained at any point (month) t of our sample.

The parameter estimates of Table 2, for model (1) ignoring a structural break in the baseline hazard function, are based on the maximization of the following likelihood function:

$$\log L(\theta) = \sum_{i=1}^{N} \log L_{i,h}(\theta),$$

where $\log L_{i,h}(\theta)$ is the log-likelihood function which corresponds to each individual loan account

i, i.e.,

$$\log L_{i,h}(\theta) = \sum_{t=1}^{T-h} Y_{i,t+h} \log P_{i,t+h} + (1 - Y_{i,t+h}) \log (1 - P_{i,t+h}) , \quad \forall\, i. \tag{3}$$

This estimation procedure assumes that the error terms of model (1) are independent across all accounts i, which constitutes the individual units of our panel data set.

Table 2: Estimates of alternative specifications of model (1)

	(I) logit		(II) cloglog		(III) logit		(IV) cloglog	
Intercept	-129.98	(1.13)	-41.52	(0.82)	-130.11	(3.12)	-45.75	(1.16)
2009	0.12	(0.04)	-0.02	(0.01)	0.28	(0.01)	0.01	(2e-3)
2010	0.18	(0.04)	0.004	(9e-3)	0.17	(0.05)	0.01	(0.09)
2011	0.40	(0.05)	0.05	(9e-3)	0.21	(0.03)	0.02	(3e-3)
2012	0.78	(0.03)	0.13	(9e-2)	0.57	(0.01)	0.09	(6e-2)
2013	0.25	(0.08)	-0.005	(8e-3)	0.13	(1e-3)	-0.04	(4e-3)
duration: d_i	2.84	(0.27)	0.94	(0.21)	2.86	(0.82)	1.04	(0.30)
d_i^2	-0.009	(0.001)	-0.003	(8e-4)	-0.009	(9e-4)	-0.003	(1e-4)
$\log(d_i)$	97.42	(1.04)	31.30	(0.65)	97.26	(2.49)	34.64	(0.62)
$\log(d_i^2)$	-24.62	(0.88)	-7.98	(0.18)	-24.59	(0.65)	-8.84	(0.16)
Application Variables								
age effects: (18,30]	0.28	(0.02)	0.07	(7e-3)	0.28	(0.03)	0.06	(7e-3)
(30,40]	0.36	(0.02)	0.09	(6e-3)	0.34	(0.03)	0.07	(6e-3)
(40,50]	0.39	(0.02)	0.09	(6e-3)	0.38	(0.02)	0.08	(7e-3)
product codes: 0	0.25	(0.02)	0.08	(9e-3)	0.29	(0.03)	0.09	(5e-3)
1	0.09	(0.02)	0.04	(7e-3)	0.10	(0.01)	0.03	(4e-3)
2	0.03	(0.001)	0.002	(8e-3)	0.04	(0.01)	0.04	(1e-3)
urban effects	-0.04	(0.01)	0.001	(3e-3)	-0.08	(0.01)	-0.01	(2e-3)

Table 2 (continued): Estimates of alternative specifications of model (1)								
Behavioral Variables	(I) logit		(II) cloglog		(III) logit		(IV) cloglog	
DTC	155.26	(1.00)	63.65	(0.33)	155.30	(0.85)	54.05	(0.32)
Deliquency buckets	0.03	(2e-3)	0.007	(1e-4)	0.02	(4e-3)	0.009	(1e-4)
LTV	0.27	(0.05)	0.004	(3e-3)	0.20	(0.01)	0.02	(0.01)
LTV> 120%					0.17	(0.02)	0.04	(0.02)
Redefaulted Loans					1.32	(0.04)	0.41	(0.01)
Foreclosure Moratorium (FM)					1.23	(0.03)	0.34	(0.08)
FM * LTV<0.90%					0.33	(0.04)	0.05	(0.01)
Macro Variables								
inflation	0.002	(2e-3)	0.002	(8e-4)	0.009	(1e-3)	3e-3	(8e-4)
mortgage rate	0.005	(0.01)	0.002	(7e-4)	1e-4	(1e-3)	3e-3	(0.001)
unemployment	0.010	(2e-3)	0.003	(2e-4)	0.013	(1e-3)	3e-3	(2e-4)
-loglik	165877.7		157128.9		163023.3		154992.9	
no. parameters	23		23		27		27	
R^2	0.19		0.22		0.21		0.24	
R^2-*adjusted*	0.19		0.22		0.21		0.24	
MAE	0.9211		0.939		0.9196		0.9188	
MSE	0.8512		0.8709		0.8576		0.8524	
MAE (Pr(.)>0.5)	0.007		0.003		0.0069		0.0012	
MSE (Pr(.)>0.5)	0.06		0.05		0.0690		0.0252	

Notes: The table presents maximum likelihood (ML) estimates of model (1), which does not assume a break in its baseline hazard rate function. This is done for the following specifications of it: With a logit and cloglog link functions (see Columns I and II, respectively), and with a logit and cloglog link functions including also the qualitative dummy variables in the set of behavioral variables (see Columns III and IV, respectively). The table presents values of a number of fit measures for the above alternative specifications of the model, i.e., the minus log-likelihood function value (denoted as - loglik), and Mcfaddens' coefficients of determination R^2 and R^2-adjusted, defined as $R^2 = 1 - \frac{\log L(\theta)}{\log L(\theta_0)}$ and R^2-adjusted=$1 - \frac{\log L(\theta)\text{-}K}{\log L(\theta_0)}$, where θ is the vector of the K-slope coefficients (parameters) of the models and θ_0 contains only the intercept as an explanatory variable (i.e., the remaining slope coefficient are all set to zero). It also includes the mean squared and absolute errors, denoted as MSE and MAE, respectively, evaluating the in-sample forecasting performance of the above all specifications. These are calculated based on the difference between the model-forecasted default rates for one-moth ahead, averaged over for all units i, and their corresponding observed default rates. They also include MSE and MAE errors of the point forecast of the default event, i.e., $Y_{i,t+1} = 1$. To calculate these forecasts, we assume $P_{i,t+1} = 1$, if $\Pr(Y_{i,t+1} = 1 | Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \geq 0.5$. These metrics are denoted as MSE(Pr(.)>0.5) and

MAE(Pr(.)>0.5).

The ML parameter estimates of model (2) (see Table 3) constitute conditional estimates on the break point T_0 of the baseline hazard function, implying that $b_1^{(0)\prime} \neq b_1^{(1)\prime}$. This point is found to be at the end of year 2011, based on the sequential estimation ML procedure described before. At this point, the ML reaches its maximum (supremum) value across all its values corresponding to different points of our sample which are searched for a break in b_1. More formally, the parameter estimates reported in Table 3 are based on the solution of the following optimization problem:

$$\underset{T_0 \in (1+l, T-h-l)}{sup} \log L(\theta | T_0),$$

where l gives the number of observations from the start and end of our sample which are trimmed out in order to implement our sequential searching procedure for a break point in b_1. For computational and economic (assumed lagged effects of explanatory variables), we have set up our grid search procedure over a period of three-months.[5]

To evaluate how well the alternative specifications of the models fit into the data, the tables report values of the following metrics: the minus log-likelihood function value (denoted as - loglik), and Mcfaddens' coefficients of determination R^2 and R^2-adjusted, defined as follows: $R^2 = 1 - \frac{\log L(\theta)}{\log L(\theta_0)}$ and R^2-adjusted$=1-\frac{\log L(\theta)\text{-}K}{\log L(\theta_0)}$, where θ is the vector of the K-slope coefficients (parameters) of the models and θ_0 is the intercept of the model (i.e., the remaining slope coefficient are all set to zero). In addition to the above, the tables also present in-sample measures of the forecasting ability of the models. These include the mean squared error (MSE) and mean absolute error (MAE) of the in-sample forecast errors of the probability forecasts. These errors are calculated

[5] For reasons of space, we do not report these results. To test more formally if there is a structural break in vector b_1, we can rely on the following likelihood ratio test statistic:

$$\underset{T_0 \in (1+l, T-h-l)}{sup} LR(T_0),$$

where $LR(T_0) = -2[\log L(\theta|T_0) - \log L(\theta)]$. Since the distribution of the above statistic under the null hypothesis of no break in b_1 is not standard, we have derived boorstrap critical values of it for different combinations of the cross-section units N and time-series T of our panel. As the dimension N of our panel is very large, we have kept fixed its time-dimension T and we have selected randomly different samples of cross-section units i from our entire data set, i.e., $N = \{100, 500, 1000, 2000\}$. Then, we have derived bootstrap critical values of test statistic $\underset{T_0 \in (1+l, T-h-l)}{sup} LR(T_0)$, for the above different values of N. As N increases, these converge to a critical value. Given this and our sample estimates of $\underset{T_0 \in (1+l, T-h-l)}{sup} LR(T_0)$, we can clearly reject the null hypothesis of no break against its alternative of a break point at point T_0.

based on the difference between the model-forecasted default rates for one-moth ahead, averaged over for all units i, and their corresponding observed default rates. They also include MSE and MAE errors of the point forecast of the default event, i.e., $Y_{i,t+1} = 1$. To calculate these forecasts, we transform the model-predicted probabilities of default events into point forecasts based on a threshold value 0.5. That is, we set $P_{i,t+1} = 1$, if $\Pr(Y_{i,t+1} = 1 | Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \geq 0.5$ (see, e.g., Lahiri and Yang (2013)). We henceforth denote these metrics as MSE(Pr(.)>0.5) and MAE(Pr(.)>0.5).

In terms of the log-likelihood and R^2 (or R^2-adjusted) values of models (1) and (2) reported in the tables, our results clearly support that the model with the break in its baseline hazard function fits better into the data than the other, which does not allow for a break. Note that this result is robust across all different specifications of the models estimated. Figures 3A-3D graphically present values of the baseline hazard rate function across the duration of a loan d_i (horizontal line). These values are implied by the different estimates of slope coefficients $b_1^{(0)\prime}$ and $b_1^{(1)\prime}$, before and after the end of year 2011, respectively. This is done for all the alternative specifications of model (2) presented in Table 3.

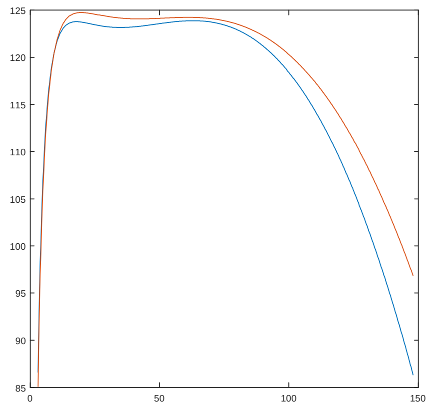

Figure 3A: Baseline hazard rate functions of model (2), with logit link function (lower graph before year 2012 and upper graph after this year)

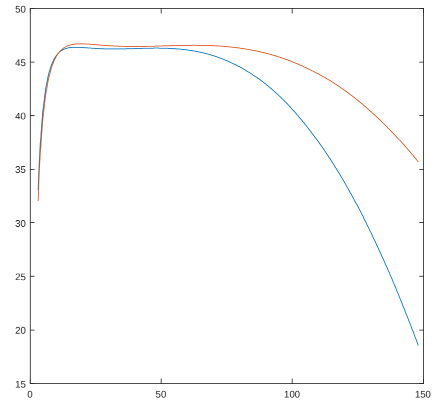

Figure 3B: Baseline hazard rate functions of model (2), with loglog link function (lower graph before year 2012 and upper graph after this year)

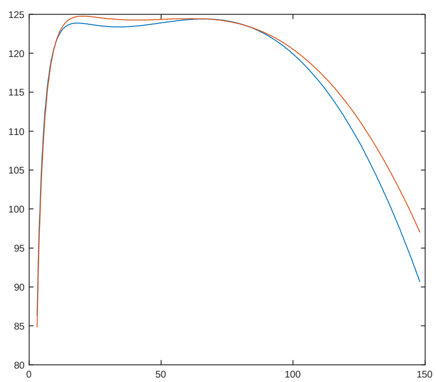

Figure 3C: Baseline hazard rate functions of model (2), with logit link function and qualitative dummies (lower graph before year 2012 and upper graph after this year)

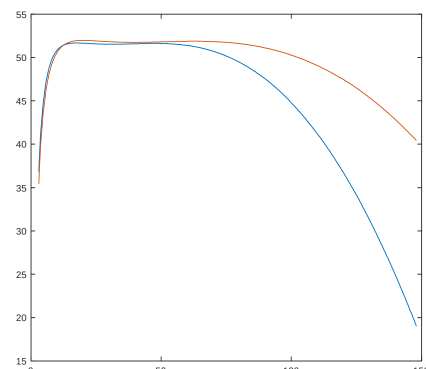

Figure 3D: Baseline hazard rate functions of model (2), with loglog link function and qualitative dumies (lower graph before year 2012 and upper graph after this year)

As can be seen from the above figures, the baseline hazard rate becomes more persistent after the break point T_0, identified by our data at the end of year 2011. After this point, the duration period of a high hazard rate (and, hence, default risk) is considerably extended, for all i, and this risk declines very slowly. This is more apparent for the cloglog link function. Although the hazard rate function has the usual hump-shaped form, with respect to d_i, observed in the literature, after point T_0 it reaches its peak (maximum) quite early (at about 12 months) and it stays at this level for a prolonged period of time, especially for the cloglog specification of link function $\Phi(.)$. Then, it starts declining very slowly. This pattern of the baseline hazard rate function may be obviously attributed to the deterioration of the credit risk conditions of the economy. These conditions become more severe after the end of year 2011, due to political uncertainty (the double elections in year 2012) and a possible exit of Greece from euro (referred to as GREXIT). As said before, the cloglog link function may have captured more efficiently than the logit one the responses of the above events on the probability of default.

Table 3: Estimates of alternative specifications of model (2)

	(I) logit		(II) cloglog		(III) logit		(IV) cloglog	
Intercept	-130.33	(0.73)	-48.42	(0.67)	-130.55	(0.68)	-53.75	(0.79)
2009	0.84	(0.05)	0.20	(0.01)	0.96	(0.06)	0.25	(0.01)
2010	1.00	(0.06)	0.23	(0.01)	0.93	(0.05)	0.24	(0.01)
2011	0.99	(0.05)	0.21	(0.01)	0.76	(0.05)	0.17	(0.01)
2012	0.72	(0.05)	0.13	(0.01)	0.54	(0.04)	0.11	(0.01)
2013	0.23	(0.04)	-0.01	(0.01)	0.15	(0.04)	-0.02	(8e-3)
duration: $d_i^{(0)}$	3.10	(0.09)	1.39	(0.03)	3.04	(0.08)	1.58	(0.04)
$d_i^{(0)^2}$	-0.01	(6e-4)	-0.006	(2e-4)	-0.01	(5e-3)	-0.006	(2e-4)
$\log(d_i^{(0)})$	98.53	(0.68)	37.60	(0.57)	98.19	(0.60)	41.92	(0.65)
$\log(d_i^{(0)^2})$	-25.56	(0.30)	-10.26	(0.18)	-25.30	(0.26)	-11.49	(0.20)
$d_i^{(1)}$	2.63	(0.05)	1.025	(0.02)	2.61	(0.04)	1.10	(0.02)
$d_i^{(1)^2}$	-0.09	(2e-4)	-0.0032	(1e-4)	-0.008	(4e-4)	-0.004	(1e-4)
$\log(d_i^{(1)})$	96.51	(0.66)	36.37	(0.51)	96.15	(0.60)	40.19	(0.60)
$\log(d_i^{(1)^2})$	-23.81	(0.24)	-9.08	(0.13)	-23.64	(0.21)	-9.93	(0.15)
Application Variables								
age effects: (18,30]	0.29	(0.03)	0.07	(7e-3)	0.28	(0.03)	0.07	(7e-3)
(30,40]	0.36	(0.03)	0.09	(6e-3)	0.34	(0.03)	0.08	(6e-3)
(40,50]	0.40	(0.02)	0.09	(6e-3)	0.39	(0.02)	0.09	(6e-3)
product codes: 0	0.25	(0.04)	0.08	(9e-3)	0.29	(0.04)	0.09	(9e-3)
1	0.09	(0.03)	0.03	(7e-3)	0.09	(0.03)	0.04	(7e-3)
2	0.007	(0.03)	-0.002	(8e-3)	0.03	(0.03)	0.007	(8e-3)
urban effects	-0.05	(0.02)	-0.0004	(3e-3)	-0.09	(0.02)	-0.01	(2e-3)

Table 3 (continued): Estimates of alternative specifications of model (1)								
Behavioral Variables	(I) logit		(II) cloglog		(III) logit		(IV) cloglog	
DTC	156.12	(0.9)	63.52	(0.33)	156.30	(0.86)	62.78	(0.34)
Deliquency buckets	0.03	(4e-4)	0.007	(1e-4)	0.02	(4e-4)	0.007	(1e-4)
LTV	0.27	(0.01)	0.004	(3e-3)	0.20	(0.01)	0.02	(3e-3)
LTV> 120%					0.16	(0.02)	0.04	(6e-3)
Redefaulted Loans					1.27	(0.04)	0.41	(0.01)
Foreclosure Moratorium (FM)					1.23	(0.03)	0.32	(9e-3)
FM * LTV<0.90%					0.33	(0.04)	0.06	(0.01)
Macro Variables								
inflation	-0.003	(3e-3)	0.0007	(8e-4)	0.005	(3e-3)	0.003	(9e-4)
mortgage rate	-0.001	(3e-3)	0.002	(7e-4)	0.005	(3e-3)	0.001	(8e-4)
unemployment	0.007	(1e-3)	0.002	(2e-4)	0.009	(1e-3)	0.002	(2e-4)
-loglik	165574.21		156766.58		162768.66		154304.91	
no. parameters	27		27		31		31	
R^2	0.18		0.23		0.20		0.25	
R^2-*adjusted*	0.18		0.23		0.20		0.25	
MAE	0.9133		0.938		0.9127		0.9050	
MSE	0.849		0.8657		0.8480		0.8410	
MAE (Pr(.)>0.5)	0.0073		0.0038		0.0081		0.0037	
MSE (Pr(.)>0.5)	0.0675		0.0495		0.0708		0.0487	

Notes: The table presents maximum likelihood (ML) estimates of model (2), which assumes a break in its baseline hazard rate function which is found to occur at the end of year 2011. This is done for the following specifications of the model: With a logit and cloglog link functions (see Columns I and II, respectively), and with a logit and cloglog link including also the qualitative dummy variables in the set of behavioral variables (see Columns III and IV, respectively). The table presents values of a number of fit measures of the above alternative specifications of the model, i.e., the minus log-likelihood function value (denoted as - loglik), and Mcfaddens' coefficients of determination R^2 and R^2-adjusted, defined as $R^2 = 1 - \frac{\log L(\theta)}{\log L(\theta_0)}$ and R^2-adjusted$=1 - \frac{\log L(\theta)\text{-}K}{\log L(\theta_0)}$, where θ is the vector of the K-slope coefficients (parameters) of the models and θ_0 is the intercept of the model (i.e., the remaining slope coefficient are set to zero). It also includes the mean squared and absolute errors, denoted as MSE and MAE, respectively, evaluating the in-sample forecasting performance of the above all specifications. These are calculated based on the difference between the model-forecasted default rates for one-moth ahead, averaged over for all units i, and their corresponding observed default rates. They also include MSE and MAE errors of the point forecast of the default event, i.e., $Y_{i,t+1} = 1$. To calculate these forecasts, we assume $P_{i,t+1} = 1$, if $\Pr(Y_{i,t+1} = 1|Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \geq 0.5$. These metrics are denoted as MSE(Pr(.)>0.5) and MAE(Pr(.)>0.5).

Regarding the parameter estimates and the alternative specifications of the model considered, the results of Table 3 reporting estimates of model (**2**) which fits better into the data, leads to a number of interesting conclusions which have important policy implications.[6] First, in terms of the -loglik and $R^{\mathbf{2}}$ (or $R^{\mathbf{2}}$-adjusted) values reported in the table, the best specification of model (2) is found to be that which assumes the cloglog link function $\Phi(.)$. As argued before, this function may capture better the asymmetric responses of covariates on extreme events on PD, $P_{i,t+h}$, compared to the logit one. This is in line with the superior performance of this function based on the MSE(Pr(.)>.5) and MAE(Pr(.)>.5) metrics, evaluating the point forecast performance of the model for default events. Both of these metrics take their lowest values for the cloglog specification of $\Phi(.)$.

Second, the versions of model (2) which include the qualitative dummy variables, capturing the *redefaulted loan* and *foreclosure moratorium effects* on PD, $P_{i,t+h}$, (see the estimates of columns III and IV of Table 3, respectively) perform better than those that they do not (see columns I and II). This can be clearly justified by the values of -loglik and R^2 (or R^2-adjusted), as well as all the forecasting performance metrics reported in the table. The effects of the above dummy variables on $P_{i,t+h}$ are positive and significant at 5%, or 1%, level. Note that these effects are also positive and significant for model (1), which does not allow for a break in its baseline hazard rate function (see Columns II and III of this table). The positive sign effects of the above variables on $P_{i,t+h}$ mean that a restructured loan, which had redefaulted before, or a loan which is protected by the foreclose moratorium law has higher probability to default in the future than a loan which is not subject to the above categories. These effects can be attributed to the persistency of the recessionary and financial distressed conditions held in the Greek economy, over our sample. These conditions may have affected the ability of borrowers to pay in long term, even for borrowers whose loans have been restructured. The above policies may be thus more successful under temporary

[6]Note that, with exception of the estimates of the slope coefficients of the baseline hazard rate function, most of the results of Table 3 are qualitatively consitent with those of Table 2, which provides estimates of the model without the break.

financial depressed conditions. The positive and significant effects of variable FM×LTV<90%, which captures foreclosure moratorium effects of loans with LTV<90%, on $P_{i,t+h}$ indicate that these policies may raise moral hazard incentives that borrowers will not maintain their payments in long run, even for loans with lower than one LTV values (i.e., LTV<0.90%).

Turning to the discussion about the effects of the other behavioral variables considered in our analysis on $P_{i,t+h}$, namely DTC, DB, LTV and LTV ≥120%, the results of Table 3 indicate that the sign of the slope coefficients of these variables are consistent with the theory and their estimates are significant at 5%, or 1%, level. This is true for all specifications of model (2) considered. Moreover, the positive sign and significant effects of LTV and LTV ≥120% on $P_{i,t+h}$ are consistent with the negative equity hypothesis which predicts that, if a loan has LTV>100%, then it will default. The positive sign and significant effects of DTC and DB on $P_{i,t+h}$ may capture the inability of borrowers to pay back their loans, since they tend to delay their loan repayments.

Fourth, for the specifications of the model that employ the cloglog link function (see columns II and IV of Table 3), the macroeconomic variables employed in the estimation of model (2), i.e., inflation, mortgage rate and unemployment rate, are all found to be significant at the 5% level. The signs of their slope coefficients are also consistent with the theory. The effects of a rise in mortgage interest rate and unemployment rate tend to increase PD, since they negatively affect the ability of borrowers to pay their loans. On the other hand, the positive effect of a rise in inflation on $P_{i,t+h}$ can be attributed to a fall in real per capita income, caused by this rise. The positive relationships between the above macroeconomic variables and $P_{i,t+h}$,as well as those between unemployment rate and behavioral variables LTV and DTC, shown in Table 1, can explain the fall of the number of the defaulted and restructured loans observed towards the end of our sample, shown in Figure 2. This fall may be attributed to the slight decrease of the unemployment rate, observed towards the end of our sample.

Fifth, regarding the application variables of the model, i.e., *age*, *production codes* and *urban variables*, the results of Table 3 reveal that these variables are all significant and economically meaningful. The effects of the age of a borrower at the time of application affect positively the probability of default. As was expected, our results indicate that borrowers of higher age affect more strongly $P_{i,t+h}$, compared to the younger ones. This happens because they are more vulnerable to service their debt, as they have shorter life expectancy horizon. From the three categories of mortgages considered, the results of the table indicate that the residential and home repair loans are the most risky ones. The negative sign of the slope coefficient of *urban* variable can be attributed to the fact that, compared to non-urban areas, most of the mortgages given to the urban ones are for first residence.

3.3 Out-of-sample forecasting performance

In this section, we evaluate the out-of-sample forecasting performance of model (2), which is found to better fit into the data. We consider both specifications of the model, with the logit and cloglog link functions $\Phi(.)$. These specifications also include the dummy variables capturing the qualitative effects of the *redefaulted loan* and *foreclosure moratorium* procedures on PD, $P_{i,t+k}$. That is, the specifications of the model used in our forecasting exercise correspond to those of columns III and IV of Table 3.

Since the in-sample results suggest that there is a break in the baseline hazard rate function of the model at the end of year 2011, in our out-of-sample exercise we forecast default probabilities of loan data starting from 2012:06 until the end of the sample, i.e., 2014:08. For this period, we obtain forecasts of $P_{i,t+h}$ over the next 3-months, for all loan accounts i. This is done for every time point of our out-of-sample interval.[7] Based on them, we can obtain point forecasts of default events (points), by setting $Y_{i,t+h} = 1$, if $\Pr(Y_{i,t+h} = 1|Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \geq 0.5$. To calculate the above forecasts, we rely on two different real-time recursive estimation approaches of the model, over the out-of-sample period. The first is based on an expanding window of data in which we estimate the model based on an initial sample, starting from 2008:01 until 2012:05. Based on these estimates, we obtain default probabilities over the next 6-months. Then, we add to our initial sample a window of 6 months observations, re-estimate the model and calculate the subsequent set of default probability forecasts. This procedure is sequentially repeated until the end of the sample. According to the second approach, known as rolling window, we do not use an

[7] Note that the expected probability of default, at time t, is calculated as follows. First, we estimate the survival probability of a loan account i, which is given as the product of the probability of not failing at each time period, conditional on not having failed before, i.e.,

$$S_i(t) = \prod_{s=1}^{t} (1 - P_{i,s})$$

The failure probability (probability of default) is given as $1 - S_i(t)$, for all i. See also Bellotti and Crook (2013).

expanding window to estimate the model and provide future default probabilities, but we consider a rolling window of six months. The first window starts from 2011:12 to 2012:05.

Table 4 presents the MSE and MAE metrics of the forecasting errors of the above approaches based on the actual default rates. Our forecasts constitute average, over all individual accounts i, estimates of default probabilities and default points. The results of the table lead to the following conclusions. First, as for the in-sample exercise, the out-of-sample forecasting performance of the model is very satisfactory. Our results clearly indicate that the specification of (2) with the cloglog link function outperforms that with the logit function. This is true for both the expanding and rolling window methods of forecasts. It is also true for both the probability default and default point forecasts. These results highlight the usefulness of the cloglog function in modelling PD. The better performance of this link function, compared to the logit one, can be attributed to its ability to capture asymmetric responses of covariates on the default events, as mentioned before. Third, between the expanding and rolling window forecast methods, the second method improves considerably the performance of the model. This can be obviously attributed to the fact that the rolling window based forecasts can better capture the effects of structural break type of changes on model parameters, or other type of instabilities, occurring during our sample.

Table 4: Out-of-sample forecasts

	logit	cloglog	logit	cloglog
	Expanding window		Rolling window	
	Probability default forecasts			
MAE	1.5185	1.1419	1.2419	0.9675
MSE	3.3039	2.2391	2.8941	2.0212
	Default event forecasts			
MAE (Pr(.)>.5)	1.1187	0.9027	0.9865	0.7957
MSE (Pr(.)>.5)	1.6925	1.3634	1.5507	1.2780

Notes: This table presents the mean absolute and squared error metrics of our out-of-sample forecasting exercise, denoted as MAE and MSE, respectively. This forecasting exercise covers the period from 2012:06 to 2014:08. The above metrics are calculated based on an expanding and a rolling window estimation approaches of model (2), over the out-of-sample interval. This is done for both the logit and cloglog specification of the model. To forecast the default events (points), we set $Y_{i,t+h} = 1$, if $\Pr(Y_{i,t+h} = 1 | Y_{i,t} = 0; \xi_{i,t+h-d_i}, x_{i,t-m}, z_{t-p}) \geq 0.5$. The MAE and MSE corresponding to the forecasts of these events are denoted as MAE (Pr(.)>.5) and MSE (Pr(.)>.5), respectively.

4 Conclusions

Based on a discrete-time survival model, this paper has modelled the probability of default (PD) of residential mortgages using an exclusive data set of the Greek economy, over period 2008-2014. During this period, this economy has experienced a severe economic crisis and extraordinary financial distressed conditions. Our results can be thus very useful in understanding sources of loan defaults, under the above conditions. The paper has extended the survival model to allow for a break in its baseline hazard rate function and it has tested a number of hypotheses of interest that may determine the default probability of a residential loan. Some of these hypotheses concern the effects of law banning foreclosures on PD. These procedures are often adopted to mitigate the effects of recessionary or financial distressed conditions on debt servicing and to avoid massive collaterals liquidation. Examining their effects on PD is of major interest in the literature of credit risk.

The paper has derived a number of interesting results, which have important policy implications. First, it shows that there is a common structural break in the baseline hazard rate function after the end of year 2011, which can be attributed to the political uncertainty and severe recessionary conditions held in the economy after that year, as well as the possibility of an exit of Greece from euro. After that year, the high levels of the baseline hazard rate function became more persistent than before, and they decreased very slowly. Second, the government law banning foreclosures introduced have significantly increased the future default probability. Our estimates indicate that such laws (or acts) may raise moral hazard incentives that borrowers will not maintain their payments in long run. We argue that this may be attributed to the prolonged recessionary conditions held in the economy. Under these conditions and government acts, efforts of banks to restructure (or refinance) mortgage loans may not successfully affect future default probabilities.

Third, the paper provides clear cut evidence that the probability of default also depends on behavioral variables, like the ratios of the loan-to-value and delinquent-to-contract amount, as well as on macroeconomic variables, like the unemployment and inflation rates. As was expected by the theory, the effects of the above all variables on PD are positive. A positive relationship between LTV and PD supports the negative equity hypothesis, while positive effects of the macroeconomic and the other behavioral variables on PD are in line with the ability-to-pay hypothesis.

References

Andrews, D.W.K. (1993), Tests for parameter instability and structural change with unknown change point, Econometrica, 61, 821 – 856.

Bajari, P., C.S. Chu and M. Park (2008), An empirical model of subprime mortgage default from 2000 to 2007, NBER WP 14625.

Bellotti, T. and J. Crook (2013), Forecasting and stress testing credit card default using dynamic models, International Journal of Forecasting, 563-574.

Bhutta, N, J.Dokko and H. Shan (2010), The depth of negative equity and mortgage default decisions, Finance and Economics Discussion Series, Federal Reserve Board, No 2010-35.

Calhoun, A. and Y. Deng (2002), A dynamic analysis of fixed-and adjustable-rate mortgages, Journal of Real Estate Finance & Economics, 2, 9-33.

Campbell, J.Y., and J.F. Cocco (2015), A model of mortgage default, Journal of Finance, forthcoming.

Crook,. J. and J. Banasik (2012), Forecasting and explaining aggregate consumer delinquency behaviour", International Journal of Forecasting,, 28, 145-160.

Crook, J. and T. Bellotti (2010), Time varying and dynamic models for default risk in consumer loans, Journal of Royal Statistics Society A, 173, 283-305.

Deng, Y., J.M. Quigley and R. van Order (2000), Mortgage terminations, heterogeneity and the exercise of mortgage options, Econometrica, 68, 275-307.

Divino, J.A., E.S. Lima and J. Orrillo (2013), Interest rates and default in unsecured loan markets, Quantitative Finance 13-12: 1925-1934.

Elul, R. N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt (2010), What “triggers” mortgage default?, The American Economic Review, 100, 490-494.

Folliant, J., W-V. Huang and J. Ondrich (1999) Stay, pay or walk away: A hazard rate analysis of FHA-insured mortgage terminations, mimeo, University of Syracuse.

Gerardi, K., K.F. herkenhoff, L.E. Ohanian and P.S. Willen (2013), Unemployment, negative equity and strategic default, Federal Reserve Bank of Atlanda, WP 2013-04.

Goodstein, R.M., P. Hanouna, C. D. Ramirez and C. Stahel (2011), Are foreclosures contagious?, FDIC Center for Financial Research, WP No 2011-4.

Gross, D.B., and N.S. Souleles (2002), An empirical analysis of personal bankruptcy and delinquency, The Review of Financial Studies, 15, 319-347.

Gyourko, J. and J. Tracy (2014), Reconciling theory and empirics on the role of unemployment in mortgage default, Journal of Urban Economics, 80, 87-96.

Hwang, R-C and C-K Chu (2014), Forecasting forward defaults with discrete-time hazard model, Journal of Forecasting, 33, 108-123.

Koenker, R. and J. Yoon (2009), Parametric links for binary choice models: A Fisherian-Bayesian colloquy, Journal of Econometrics, 152, 120-130.

Lahiri, K., and L. Yang (2013), Forecasting Binary Outcomes, Handbook of Economic Forecasting (Timmermann, A. and G. Elliott, eds), North-Holland, Vol. 2B, 1025-1106.

Leow, M. and J. Crook (2014), The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis, European Journal of Operational Research, 1-8.

McFadden, D. (1974), The measurement of urban travel demand, Journal of Public Economics, 3, 303-328.

Titman and Torous (1989), Valuing Commercial Mortgages: An empirical investigation of the contingent-claims approach to pricing risky debt, Journal of Finance, 44, 345-373.

Zivot, E., and D.W.K. Andrews (1992), Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis,Journal of Business & Economic Statistics 10, 251-270.