Credit Risk Modelling Under Distressed Conditions

Dendramis Y.* Tzavalis E.† Adraktas G. ‡ Papanikolaou A.

July 20, 2015

Abstract

Using survival analysis, this paper estimates the probability of default of residential mortgages based on a data set from the Greek economy, which has experienced a severe economic crisis over the last five years, after the global financial crisis of year 2008. This paper provides clear cut evidence that macroeconomic variables and behavioural variables, capturing the economic crises effects, constitute crucial variables in interpreting the probabilities of defaults and their deterioration. In addition to these variables, the paper also indicates that the new auctions' law banning foreclosures for the first residence, introduced by the government to mitigate the effects of the economic crisis on a mortgage loan default, had positively impact the default probability. In addition, this paper indicates that restructuring of mortgage loans, which have previously defaulted, decreases the default probability from 100%, that used to be, as defaulted loan (increasing in other words the curing rate).

JEL classification: G12, E21, E27, E43

Keywords: mortgages, survival analysis, financial stress, probabilty of default

*School of Economics and Finance, Queen Mary University of London. email: y.dendramis@qmul.ac.uk.
†Department of Economics, Athens University of Economics and Business. email: etzavalis@aueb.gr, Corresponding author.
‡Alpha Bank, Greece, Retail Banking Credit Risk Division. email: georgios.adraktas@alpha.gr.

1 Introduction

Since the subprime crisis, there is growing interest in modeling default probability on residential mortgages (see, e.g., Gross and Suleles (2002), Elul et al (2010), Crook and Basik (2012), Divino et al (2013), Campbell and Cocco (2014), etc). These mortgages constitute a large proportion of banks loan portfolios and of household debt. Furthermore, they critically depend on changes in business cycle and/or credit crunch conditions, as well as protection consumer laws (or acts) concerning liquidation and foreclosure proceedings. Most of the recent studies working on this area (see, e.g., above) rely on aggregate or portfolio loan data as they examine the effects of changes in macroeconomic and liquidation conditions (e.g., interest rates) and equity prices on the probability of default on a mortgage loan.

In this study, in addition to macroeconomic factors we examine how demographic and behavioural variables affect the probability that an obligor will default in a future period. In so doing, we rely on a panel data set consisting of 85230 individual mortgage accounts (loans) with monthly frequency, covering the period from 2008:01 to 2014:10. Using disaggregated data, like our panel data set, than aggregated data (e.g., loan portfolios) in answering the above question can lead to more robust and accurate inference about the effects of both application and time varying behavioural variables on the probability of default. During the above period, the Greek economy has experienced a severe 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 the peak in 2008. At the same time, there is a foreclosures ban on first residence, not allowing Banks, by this way, 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.

Our analysis is based on discrete survival model which allows for calculating readily the probability of a default of an obligor (referred to as hazard rate) in a future period, related to the age of the loan. This is done conditional on no prior default and the current and/or future macroeconomic conditions in the economy, taking also into account a number of application and behavioural explanatory variables, along with collateral information. Thus, the model can show how long the mortgage survives before its default in a future period. It can also provide estimates of the hazard rates over different future horizons, conditional on the current state of a loan and values of the explanatory variables.[1] The performance of the model is evaluated by conducting an out of sample forecasting exercise of the probability of default of an individual loan. Concerning the macroeconomic covariates of the probability of default considered by the model, these are assessed based on forecasts of them based on a VAR model, estimated recursively, over our sample, to capture possible business cycles effects.

The results of the paper provide a number of interesting conclusions on modelling probability of defaults. First, they indicate standard demographic, macroeconomic and application variables are consistent with the theory effects on the probability of default of an individual loan. In particular, from the behavioral variables we have found that the ratio of delinquent amount to the contract amount and the ratio of the total balance to collateral value (LTV) seem to have the highest effects on the probability of default, while from the macroeconomic variables the unemployment and loan rates. Regarding the application variables, the paper provides clear cut evidence of sizeable and lasting effects of the restructuring process of mortgage loans, which has previously defaulted, and the new auctions' law banning foreclosures on first residence. Both of these procedures have introduced by the government and the banking sector to mitigate the effects of the economic crisis on a mortgage loan default.

The paper is organized as follows. Section 2 presents the discrete survival model, it discuss its estimation procedure and presents a procedure of predicting probabilities of default through the model. Section 3 presents the empirical results of the paper and conducts the forecasting exercise. Section 4 concludes the paper.

2 Model setup, Estimation and Forecasting

A survival analysis model is an appropriate model for time-to-event data. To present a discrete version of the model, let us denote as t the number of months since an account (loan) i was opened (i.e., the duration, or passage time) and the date that it was opened as

[1] For a more recent survey on the use of survival models for consumer credit risk models, see Crook and Bellotti (2010) and Hwang and Chu (2014).

l_i. Variable d_{it} is a binary variable which takes values at t 1 if account i defaults and 0, if it does not default. Then, define the following explanatory variables: ξ_i denotes a vectors of application variables (AV) known only at the time of application only (i.e., $t - c??$), x_{it} is a vector of behavioural variables (BV) collected over the life of the loan and z_t is a vector of macroeconomic variables (MV), which are common for all accounts on a date of our sample. In addition to these variables, we can also consider a number of time specific, or intervention, dummy variables, v_t, which are common across all i, which can capture exogenous events, such as government interventions concerning liquidation and foreclose proceedings of the mortgage loan market.

Based on the above definitions, the discrete survival model predicts that the probability of default (PD) for each loan i at time t is given as follows:

$$\begin{aligned} P_{it} &= \Pr(d_{it=1} | d_{is} = 0; s < t, c_i, v_{t-k}, x_{i,t-m}, z_{t-p}) \\ &= \Phi(b_0 + b_1' \varphi_t + b_2' \xi_i + b_3' v_{t-k} + b_4' x_{i,t-m} + b_5' z_{t-p}), \end{aligned} \tag{1}$$

where $\Phi(w) = \frac{1}{(1+e^{-w})}$ is a logistic function, k, m and p denote lag orders, $\varphi_t = (t, t^2, \log(t), \log(t)^2)'$ of functions of the duration time t of loan i, b_0 is an intercept, and b_1, b_2, b_3, b_4 and b_5 are vectors of slope coefficients. The terms of vector φ_t are functions of t which enable us to capture a smooth pattern of the hazard rate over t.

For the purposes of this paper, an account is considered as defaulted, i.e., $P_{it} = 1$, if it is past due more than 90 days on any material credit obligation, or if it is distressed restructured (at the restructuring date more than 90dpd). The vector of the application variables (AV) that we will employ in the estimation of the model consist of loan terms and conditions captured by dummy variables or by linear, quadratic and logarithmic trends for the loan duration effects (see definition of vector φ_t). In particular, the dummies employed reflect urban effects (here, Attica), age effects (18-30, 30-40, 40-end), housing and/or repair effects (product code dummies). If the purpose of a loan is for a house purchase, it is denoted as product code 1, for repair as product code 2 and for other use as product code 0.

In addition to the above application dummies, we also include a dummy capturing re-defaulted events. Even though standard survival analysis does not consider accounts that have defaulted before, the greek legal framework and the practice of the greek banks obliged us to account for obligors which have defaulted before, and, after their default, they have restructured their loan in order to start repaying it. To adapt this idea to the survival model, we consider these accounts, as new account, with a dummy variable (denoted as $redefaulted$) taking a unity if these accounts have re-defaulted before, and zero otherwise. Finally, another application dummy variable considered in our analysis stands for capturing the effects of a government law, introduced in August 2010, which does not allow banks to proceed with liquidations on their residential collaterals for first residence in the case of a defaulted borrower, given that all collections or legal actions have been exhausted. To see if this law has affected probability of default P_{it}, we have defined a dummy variable (denoted as $auctions'\ law$) which take unity for all the loan accounts protected by this law, and zero otherwise.

In model (1), as behavioural variables (BV) we consider the installment amount, the ratio of delinquent amount to the contract amount, as a measure of delinquency to the total debt of the obligor, and the ratio of the total balance to the most recent collateral valuation (referred to as LTV in the literature). We also introduce a measure of the consistency of the obligor to pay his/her loan, by defining the variable "sum of buckets 1-3". This variable measures the number of times that the obligor has a positive amount in bucket 1 plus the number of times that the obligor has a positive amount in bucket 2 plus the number of times that the obligor has a positive amount in bucket 3 over the history of the obligor. The buckets 1 to 3 reflect if the obligor has any installment in delinquency which is rolled over at the next month.

Regarding the collateral evaluation, we have recognized some special features of the Greek economy over the recent period. To this end, we assume that the Greek real estate sector has seen a more than 30% decline in their prices, depending on the category of the collateral. To account for this huge drop of the collateral value, we use data on real estate sector indices for different categories of collaterals. Based on these, we construct a time series of collateral values given the date of the collateral value estimates, their values, the categories of the collaterals, and the time series of the real estate index. We have used five real estate indices, given by the Bank of Greece. These are as follows: the residential Real Estate, Warehouse /Storage, Building ground/Construction, Field for utilization/animals, Offices, Stores/Shops, Industrial and Agricultural field.

Finally, as macroeconomic variables (MV) in model (1) we consider the inflation and unemployment rates, a weighted average of loans rates of the mortgage market, and the Gross Domestic Product (GDP) growth rate. The latter is interpolated on a month to month base. In our empirical results, we present estimates of the model with the macroeconomic variables that have been found to be significant at the 5% significance level.

2.1 Estimation and forecasting results

In this subsection we present how to estimate the model and to evaluate its predictive performance based on backtesting procedures. To estimate the model (1), we will rely on the maximum likelihood procedure (ML) based on the following likelihood for each individual (borrower) i, for whom we observe T_i observations:

$$\log L_i = \sum_{j=1}^{T_i} d_{ij} \log P_{ij} + (1 - d_{ij}) \log (1 - P_{ij}) \tag{2}$$

Taking logarithms of the above function, the log-likelihood function of (2), over all individual loans i, is given as follows:

$$\log L = \sum_{i=1}^{N} \log L_i$$

Maximization of this function produces estimates of the slope coefficients of model (1), with well asymptotic properties.

To examine the forecasting performance of model (1), we will conduct an out of sample forecasting exercise which will evaluate the performance of the model relative to other models by calculating rolling average default probability with the observed ratio of defaulted loans, over the total number of loans, that is the observed default probability. Our forecasting exercise will be conducted, recursively. First, we will estimate all the models compared using an initial window of data up to a given date of the sample its end date. Using these estimates, we then produce recursive forecasts of default in the next h-months ahead and compute average default probabilities. We repeat this exercise until the end of the sample.

To compute the default probabilities over the next h-months, we use the model with h-lags back on the behavioral variables. Since this is a proper out of sample exercise, which tries to replicate real world situation, we use a VAR model to forecast the macro variables of the model. The probability of default over the next h-months is calculated as follows. The estimated survival probability of an individual account i at some time t is given as the product of the probability of not failing at each time period, conditional on not having failed previously. That is, it is calculated as follows:

$$S_i(t) = \prod_{s=1}^{t} (1 - P_{is}) \tag{3}$$

The failure probability is give as $1 - S_i(t)$. This gives the probability of default P_{it}. This probability can be used to compute credit scores and capital requirements.

3 Empirical analysis

In this section, we present and analyse the results of our empirical analysis. First, we presents the estimates of model (1), based on the ML estimation procedure described before and, then, we evaluate the performance of the model by conducting an out of sample forecasting exercise. The data set employed in our analysis consists of a very large set of Greek household load data whose frequency is monthly. It covers the period from 2008:01 to 2014:10 and it consists of 85230 accounts (loans).

3.1 Estimation Results

The estimation results of model (1) are presented in Tables 1A and 1B. Table 1A presents results where the model does not include the macroeconomic variables, while Table 1B it includes these variables. Note that in estimating the model, we have used 12 lags back for the behavioural (BV) variables and 3 for the macroeconomic ones (MV). The choice of these lags was based on information criteria, such as the Akaike criterion, and on our need to provide forecasts of probability of default P_{it} up to 12-months ahead.

Table 1A Estimates of Model (1), without macroeconomic variables		
BEH 12 lags		
-loglik=203538.48	estimate	std error
constant	-68.4778	2.9506
Installment Amount	0.0001	1e-5
$redefaulted\ accounts$	1.7354	0.0333
Attiki	-0.1882	0.0107
age>=18 and age<=30	0.0424	0.0278
age>30 and age<=40	0.1089	0.0238
age>40 and age<=60	0.174	0.0229
$Auctions'\ law$	1.3972	0.0166
product code 0	0.1915	0.0367
product code 1	0.1414	0.0312
product code 2	0.4843	0.0319
nums of pos. bucket1	0.0349	0.0004
delinquent amount/contract amount	81.1893	1.0212
t (=time since the account has opened)	1.9327	0.0687
t^2	-0.0072	0.0015
log(t)	52.1319	2.31
$\log(t)^2$	-14.3289	0.5836
total balance/ts of col. valuation	0.0967	0.0092

The results of both tables lead to a number of interesting conclusion, which have important policy implications. First, the sign of all BV is found to be consistent with the theory. As was expected, an increase in the installment amount, the ratio of delinquent amount to the contract amount and the ratio of the total balance to collateral value will lead to an increase in the probability of P_{it}. The same is true for the inconsistency of the obligor to repay his/her loan, defined by variable "sum of buckets 1-3". From the above all variables, the ratio of delinquent amount to the contract amount and the ratio of the total balance to collateral value seem to have the highest effects on P_{it}.

Regarding the effects of MV on P_{it}, our results are also consistent with the theory. They show that an increase in unemployment and loan rates will tend to increase the probability of default. This happens because under these conditions there will be a deterioration of the economic conditions and the payment ability of loan obligors. The positive effect of inflation on P_{it} can be interpreted by the indirect effect that an increase of this variable will have on rising interest rates, due to the monetary policy of the central bank.

Table 1B: Estimates of model (1), with the macroeconomic variables		
BEH 12 lags, Macro 3 Lags		
-loglik=206014.54	estimate	std error
constant	-77.9295	2.8995
Installment Amount	0.0001	1e-5
$redefaulted\ accounts$	1.879	0.0347
Attiki	-0.2123	0.0103
age>=18 and age<=30	0.0783	0.0245
age>30 and age<=40	0.1285	0.0213
age>40 and age<=60	0.1815	0.0204
$Auctions'Law$	1.4211	0.0178
product code 0	0.2209	0.0338
product code 1	0.1752	0.0285
product code 2	0.5612	0.0299
inflation	0.0834	0.0148
unemployment	0.0068	0.0012
loan interest rate	0.2168	0.0206
delinquent amount/contract amount	118.2799	1.0083
t (=time since the account has opened)	2.0668	0.0653
t^2	-0.0075	0.0002
$log(t)$	58.0743	2.2372
$log(t)^2$	-15.6633	0.5608
total balance/ts of col. valuation	0.1301	0.0098

Turning into the discussion of the effects of the application variables on P_{it}, the results of the table indicate that the variables capturing the redefaulted events and the new auctions' law banning foreclosures have very important positive effects on P_{it}. More specifically, a loan which has defaulted before and has been restructured will have higher probability of default than a loan whose is current state is no default. The higher than unity estimate of coefficient of variable $redefaulted\ accounts$ indicates that this procedure of loan restructuring may have permanent effects (if not explosive) on P_{it}. Thus, as a policy device one may consider for a constant number of loans restructuring. In addition to $redefaulted$, lasting effects on the probability of default P_{it} has the legislation about the auctions for the first residence, introduced in August 2010. The coefficient estimate of variable $auctions'\ law$ is bigger than unity, which implies that for the loans protected by this law the probability of default rises substantially and permanently.

Regarding the remaining application variables, it is interesting to note that the results of our table indicate that loans to urban areas (see variable Attiki) reduces the probability of default, compared to those in non-urban areas.

To see if our model produces estimates of default probability over time t often presented in the literature, in Figure1 we present estimates of the implied by our model hazard rate calculated based on the following function $1.98 \cdot t - 0.0073 \cdot t^2 + 54.58 \cdot \log(t) - 14.89 \cdot (\log(t))^2$. This pick us at 15 months and then it declines slowly over time, since those who are likely to default drop out. But, note that the rate the the hazard probability decreases after its pick is very slow and it appears to have another pick after 55 months. This pattern of this hazard rate function may be attributed to the effects of the loan restructuring procedure, reducing the possibility of a loan to drop out.

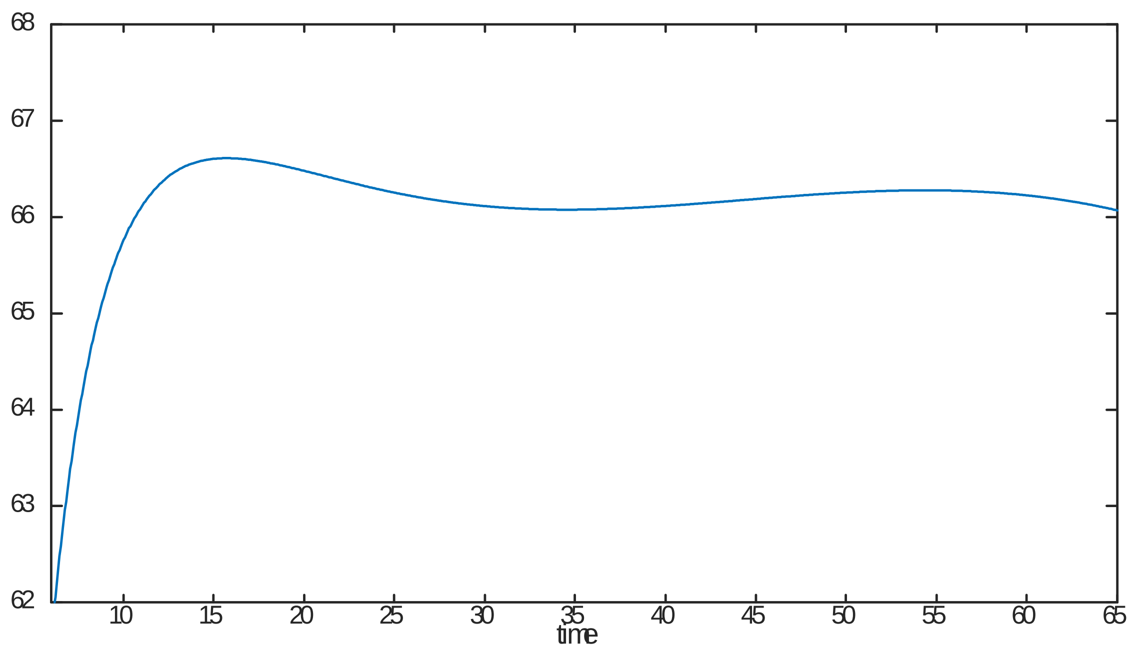

Figure 1: The graph presents values of the hazard function.

The results of our forecasting exercise are given in Tables 2A and 2B. Table 2A provides forecasts of default probabilities 3 months ahead, while Table 2B of 12 months ahead. As said before, the forecasting performance of model (1) is based on an out of sample forecasting exercise calculating rolling average default probability with the observed ratio of defaulted loans, over the total number of loans, that is the observed default probability. First, we estimate the model using data up to date 31-Oct-2012. Using these estimates, we then produce recursive forecasts of default in the next 12 months (or 3 months) and compute average default probabilities. We repeat this exercise for each month up to date 31-Oct-2013, where at this date we re estimate our model. Again, produce recursive forecasts of default in the next 12 months (or 3 months) and compute average default probabilities, while we re estimate it using the all the available sample.

Table 2A: Out of sample forecasts of 3 months ahead default probabilities

forecast default over period		observed default rate	average default probability
30-Nov-2012	31-Jan-2013	0.039	0.037
31-Dec-2012	28-Feb-2013	0.037	0.035
31-Jan-2013	31-Mar-2013	0.036	0.035
28-Feb-2013	30-Apr-2013	0.032	0.035
31-Mar-2013	31-May-2013	0.031	0.034
30-Apr-2013	30-Jun-2013	0.027	0.037
31-May-2013	31-Jul-2013	0.026	0.035
30-Jun-2013	31-Aug-2013	0.022	0.035
31-Jul-2013	30-Sep-2013	0.018	0.032
31-Aug-2013	31-Oct-2013	0.015	0.031
30-Sep-2013	30-Nov-2013	0.014	0.031
31-Oct-2013	31-Dec-2013	0.013	0.031
30-Nov-2013	31-Jan-2014	0.013	0.022
31-Dec-2013	28-Feb-2014	0.013	0.022
31-Jan-2014	31-Mar-2014	0.014	0.022
28-Feb-2014	30-Apr-2014	0.014	0.024
31-Mar-2014	31-May-2014	0.017	0.023
30-Apr-2014	30-Jun-2014	0.016	0.023
31-May-2014	31-Jul-2014	0.016	0.023
30-Jun-2014	31-Aug-2014	0.014	0.024
31-Jul-2014	30-Sep-2014	0.013	0.023
31-Aug-2014	31-Oct-2014	0.012	0.023
30-Sep-2014	30-Nov-2014	0.008	0.022
31-Oct-2014	31-Dec-2014	0.004	0.022
30-Nov-2014	31-Jan-2015	NA	0.012

To compute the default probability for account i over the next 12 months, we set $t = 12$

in $1 - S_i(t)$ where $S_i(t)$ is given by (3), while for an account over 3 months we set $t = 3$. The observed default rates are calculated as follows. For a given time period (say from $d1$ to $d2$), we observe the defaulted loans over the total number of loans. The results of Tables 2A and 2B indicate that our model produces accurate estimates of observable default rates, for most of the out-of-sample points considered. As was expected, for $h = 3$ periods ahead the model exhibits its highest forecasting ability. This may be attributed to the best forecasts of the macroeconomic variables, included in the model.

Table 2B: Out of sample forecasts of 12 months ahead default probabilities

forecast default over period		observed default rate	average default probability
30-Nov-2012	31-Oct-2013	0.106	0.157
31-Dec-2012	30-Nov-2013	0.098	0.153
31-Jan-2013	31-Dec-2013	0.091	0.149
28-Feb-2013	31-Jan-2014	0.082	0.148
31-Mar-2013	28-Feb-2014	0.076	0.153
30-Apr-2013	31-Mar-2014	0.07	0.151
31-May-2013	30-Apr-2014	0.065	0.15
30-Jun-2013	31-May-2014	0.062	0.151
31-Jul-2013	30-Jun-2014	0.058	0.15
31-Aug-2013	31-Jul-2014	0.056	0.148
30-Sep-2013	31-Aug-2014	0.055	0.148
31-Oct-2013	30-Sep-2014	0.054	0.148
30-Nov-2013	31-Oct-2014	0.053	0.131
31-Dec-2013	30-Nov-2014	0.05	0.131
31-Jan-2014	31-Dec-2014	0.046	0.132
28-Feb-2014	31-Jan-2015	0.041	0.134
31-Mar-2014	28-Feb-2015	0.038	0.132
30-Apr-2014	31-Mar-2015	0.032	0.13
31-May-2014	30-Apr-2015	0.028	0.129
30-Jun-2014	31-May-2015	0.022	0.133
31-Jul-2014	30-Jun-2015	0.017	0.133
31-Aug-2014	31-Jul-2015	0.012	0.132
30-Sep-2014	31-Aug-2015	0.008	0.135
31-Oct-2014	30-Sep-2015	0.004	0.131
30-Nov-2014	31-Oct-2015	NA	0.085

4 Conclusions

Using a discrete survival model, this paper has estimated the default probabilities of residential mortgages based on a data set from the Greek economy, which has experience a severe economic crisis over the last five years, after the global financial crisis of year 2008. To assess the performance of the model to fit into the data, the paper has conducted an out-of-sample forecasting exercise.

The paper provides a number of interesting results which have important policy implications. In particular, it shows that macroeconomic variables and behavioural variables, capturing the economic crises effects, constitute very important variables in interpreting the default probabilities of mortgage loans. From the set of behavioural variables, these variables include the ratio of delinquent amount to the contract amount and the ratio of the total balance to collateral value, while from the macroeconomic variables they include the unemployment and loan rates. Regarding the effects of application variables, the paper provides clear cut evidence that, apart from the standard application variables (e.g., age and duration), very important effects on the default rate of mortgage loans have the restructuring procedure mortgage loans, which have previously defaulted, and the new auctions' law banning foreclosures on first residence. Both of these procedures have introduced by the government and the banking sector to mitigate the effects of the economic crisis on a mortgage loan default, but they have the adverse effects on the probability of default.

References:

Campbell, J.Y., and J.F. Cocco, 2014, A model of mortgage default, American Economic Review,

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

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 100-2, 490-494.

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

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