PROBABILITY OF DEFAULT ON A RETAIL LOANS PORTFOLIO, UNDER BASEL 3 REQUIREMENTS

Author : Coman Iulian-Ştefan

Supervisor : Professor Moisă Altăr

Doctoral School of Finance and Banking

June 2014

Overview

1. Motivation
2. Dataset
3. Survival Analysis – Proportional Hazard Model(with and w/o macro variables)
4. Logistic regression (with and w/o macro variables)
5. Validation – out of sample and out of time
6. Conclusions
7. Bibliography

Motivation

- In credit risk management, banks witness to estimate their risk parameters under Internal Rate Base approach and to diminish cost of capital allocated to unexpected losses (Basel III capital requirements are **increasing**).

- Based on customer's behavioral characteristics, their goal is to evaluate probability of default at *each* point in time(or within a specific time period), in order to strengthen their credit-scoring models (Stepanova, Thomas - 2000).

- Macroeconomic environment influences customers' reimbursement capacity and lending business direction (Bellotti, Crook – 2007).

- Evaluating time-to-default will also give them a more appropriate level of business **profitability** and portfolio management.

Dataset (1)

- Period : **April 2010 – March 2013** (follow-up time for each client was 12 months rolling window since loan disbursement).
- Aggregated portfolio contains **26,229** private individual customers with unsecured personal loans .
- Default threshold is based on more than **90** past due days and a past due amount of 100 EUR (Basel 2 and 3 methodology).
- Categorical variables are selected based on weight of evidence and information value criteria.
- Macroeconomic variables were introduced with **3** months lag for every moment of observation.

Dataset (2)

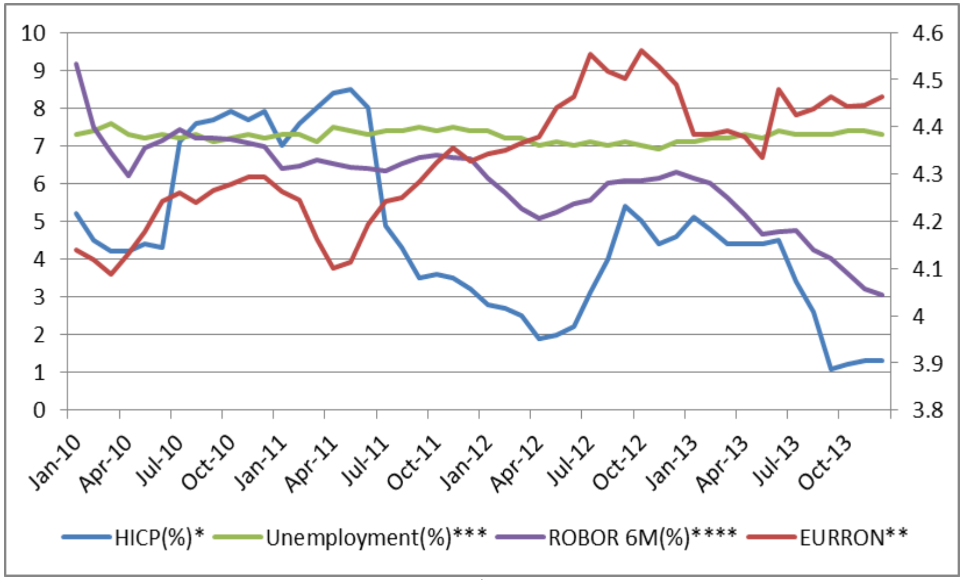
Weight of Evidence / IV

Variable	IV
risk_class	1.781569
net_rev_class	0.247856
seniority_cat	0.162486
age_cat	0.108771
currency	0.00033
gender	0.019547
Employment_status	0.021622
insurance	0.003228
loan_term_cat	0.058786
marital_sts	0.030841
reg_buc	0.000395

Main evidence

- Risk class at the moment of loan disbursement is most powerful variable.
- Net revenue class, seniority category and age group were included into estimation.

Macroeconomic indicators



Main evidence

- Public sector wages cut (25%)
- Hike of VAT from 19 to 24%
- Key rate from 6.25 to 4%

Survival analysis – Proportional hazard model (1)

- Client is followed-up on a monthly basis, using censoring mechanism .
- Hazard function is estimated at moment T , given the fact that client didn't default at time t (T>t), using Breslow approximation.

$$h(t) = \lim_{\delta t \to 0} \{\frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t}\}$$

- Cox model allows to measure the effect of time dependent covariates (X) on time-to-failure of client.

$$h(t, X) = e^{(\beta X)} h_0(t)$$

No need to estimate !

Survival analysis – Proportional hazard model (2)

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
313299	1467	311832	99.53

- At portfolio level, there were 5.593% default events, regardless their month of observation
- Most of customers default in the first 6 months of reimbursement

Parameter	DF	Parameter estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Net_rev_class	1	-0.00327	0.0006186	28.0131	<.0001	0.997
Risk_class	1	-0.00977	0.0002546	1472.1734	<.0001	0.990
Amount	1	0.00001	0.0000149	45.0662	<.0001	1.000
Rata_dob	1	0.01314	0.00481	7.4769	0.0062	1.013
Seniority_cat	1	-0.00247	0.0007328	11.3386	0.0008	0.998
Age	1	-0.02139	0.00233	83.9619	<.0001	0.979

- An increase of 1% of interest rate could lead to a 1.3% increase of HR*
- An increase of net revenue category drop HR by 0.3%
- HR for Risk class drops by 1 percentage point

* HR – Hazard Ratio

Survival analysis – Proportional hazard model (3)

Parameter	DF	Parameter estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Net_rev_class	1	-0.00329	0.0006186	28.2047	<.0001	0.997
Risk_class	1	-0.01085	0.000374	842.0032	<.0001	0.989
Amount	1	0.00001	0.0000149	45.0595	<.0001	1.000
Rata_dob	1	0.01314	0.00481	7.4673	0.0063	1.013
Seniority_cat	1	-0.00246	0.0007327	11.2943	0.0008	0.998
Risk_class_time	1	0.0004321	0.0001009	18.3559	<.0001	1.000
Age	1	-0.02139	0.00233	83.9618	<.0001	0.979

! Testing HR at different levels, it increases as risk class is weaker and amount level is higher

Accuracy Ratio is 30%.

- Based on Schoenfeld residuals and correlation with time, PH model assumption that effect of variables *do not vary* with time is violated for Risk_class.
- Product between variable and time (Risk_class_time) was introduced => HR is **constant** for variable over observation period.
- HR are similar to those obtained in the initial model.

Logistic regression (1)

- Logistic regression was performed using continuous time-dependent variables, based on complementary log -log function.

$$\log\left[\frac{S(t)}{1-S(t)}\right] = \beta_0^* + \beta_1^* x_1 + .. + \beta_k^* x_k - \gamma \log(t)$$

- Basic idea : Default is treated as a **continuous** event because we don’t know the exact moment when it happened (only interval between t-1 and t moments), supposing client didn’t default at previous time interval.

S(t) – survival function (probability that event does not happen up to time *t*

Logistic regression (2)

Parameter	DF	Parameter estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	-2.6647	0.1224	473.6806	<.0001
Amount	1	0.000092	0.0000149	37.0788	<.0001
Rata_dob	1	0.0139	0.00485	8.2006	0.0042
Seniority_cat	1	-0.00284	0.000727	15.2869	<.0001
Age	1	-0.0205	0.00232	78.0875	<.0001
Risk_class	1	-0.00984	0.000254	1499.5422	<.0001
Time	1	-0.5526	0.0151	1335.1927	<.0001

Variable	Estimate	Hazard ratio
Amount	0.0000092	1.000
Rata_dob	0.0139	1.014
Seniority_cat	-0.00284	0.997
Age	-0.0205	0.980
Risk_class	-0.00984	0.990
Time	-0.5526	0.575

- Variables with most discriminatory power are *Risk class* and *Age* group.
- *Net revenue category* is not significant, but its effect could be added into intercept.
- A difference between risk categories tend to have 1% decrease of HR (reevaluation is done yearly or at moment of credit limit renewal).
- Accuracy Ratio is 84.1%.

Survival analysis – PH model with macro variables (1)

- Client is followed-up on a monthly basis, using censoring mechanism, and its hazard function is estimated based on specific covariates .

- Macroeconomic variables are introduced into Cox model as time dependent covariates (X) – 3 months lagged, in order to catch influences on default event.

- Adding effects into equation is done based on stepwise selection, like in models without economic indicators.

Survival analysis – PH model with macro variables (2)

Parameter	DF	Parameter estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Amount	1	0.0000462	0.0000156	8.7809	0.003	1.000
Seniority_cat	1	-0.00987	0.0007371	179.1555	<.0001	0.990
Age	1	-0.02637	0.00241	119.9262	<.0001	0.974
Hicp	1	-0.28918	0.01819	252.8128	<.0001	0.749
Eurron	1	-5.3909	0.34928	238.2192	<.0001	0.005
Unemployment	1	-2.49076	0.21494	134.2799	<.0001	0.083
Robor6m	1	0.64606	0.04675	190.9518	<.0001	1.908

- Likelihood Ratio based on Breslow approximation show that model fits better the data.
- A 1% decrease of HICP leads to a decrease of 25.1% of HR.
- HR for Amount and Seniority category remain stable, as in PH model w/o macroeconomic indicators.
- Accuracy Ratio is 42.62%, an improvement from initial model.

Logistic regression with macro variables

Parameter	DF	Parameter estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	3.7918	1.1102	11.6649	0.0006
Amount	1	0.0000467	0.0000156	8.9214	0.0028
Rata_dob	1	0.0178	0.00555	10.2753	0.0013
Seniority_cat	1	-0.00955	0.000731	170.8026	<.0001
Age	1	-0.0243	0.0024	102.3617	<.0001
Hicp	1	-0.0904	0.0153	34.7033	<.0001
Eurron	1	-1.3516	0.2427	31.0042	<.0001
Time	1	-0.5447	0.0153	1269.0568	<.0001

Variable	Estimate	Hazard ratio
Amount	0.0000487	1.000
Rata_dob	0.0178	1.018
Seniority_cat	-0.00955	0.990
Age	-0.0243	0.976
HICP	-0.0904	0.914
EURRON	-1.3516	0.259
Time	-0.5447	0.580

- Based on stepwise selection, macroeconomic variables with most discriminatory power are *HICP* and *exchange rate*.
- Most powerful idiosyncratic variable is seniority category.
- A 1% increase in interest rate leads to a 1.8% increase of HR and 2.4% for age category.
- Accuracy Ratio is 72.1%, while ROC curve is 86.1% (close to ideal model)

Validation

- Basel II and III recommends that banks should evaluate their models on test samples, in order to check robustness and discriminatory power.

- Validation out of sample is done on 33% of customers, selected random uniform, rest of observations being kept into training sample.

- Out of time validation requires estimating coefficients of first year sample (customers with loans disbursed between April 2010 and March 2011) and cross checking how good second and third year models classify bad customers.

Out of sample validation

- PH model without economic indicators register an accuracy ratio of 31.35% on training and 27.38% on validate samples, while logistic regression has more discriminatory power : 83.6% - training, 85% - validate.

- Introducing variables about economy position **improves** discriminatory power for PH model (41.89 % - training, 44% - validate), but **lowers** it for logistic regression (72.3% and 71.8% respectively).

- Also, ROC curve for logistic regression stays close to the ideal model on both models.

Out of time validation

Model	Accuracy Ratio
PH w/o macro variables (2nd year)	41.84%
PH w/o macro variables (3rd year)	42.73%
LR w/o macro variables (2nd year)	88.60%
LR w/o macro variables (3rd year)	80.40%
PH with macro variables (2nd year)	28.28%
PH with macro variables (3rd year)	30.41%
LR with macro variables (2nd year)	78.20%
LR with macro variables (3rd year)	67.10%

- In terms of accuracy ratio, all the models register close indicators to the ones resulted in first year.
- LR with macro variables performs better in second year, but weakens in the third, which can be a sign that model need to be re-estimated.
- Economic conditions also have changed in 2011-2012, in comparison with 2010.

Conclusions (1)

- In terms of accuracy ratio, logistic regression with time-dependent covariates performs **better** on the entire sample (also on both analysis and validate).

- Most important variables are net revenues and risk class, but age group or seniority can add predictive power to the models.

- Macroeconomic factors improved estimates for PH model, although some coefficients must be treated carefully; HICP influenced the most hazard ratio and exchange rate can affect customer's capacity of reimbursement.

- Since unsecured personal loans are granted up to 5 years, bank could assess probability of survival at different points of time (24, 36, 48, 60 months)

Conclusions (2)

- Also, bank should be aware of default and set defaults acceptance rated for each segment of clients and loans.

- A further research can be extending analysis period for an entire business cycle (6-7 years – Koopmann) and portfolio stress testing in a growing economy.

- Also, survival analysis could be done in geographical areas with high concentration, because bank needs to know customers who default and **when** the event happens.

Bibliography

Bellotti T. , Crook J. (2007), “Credit Scoring with macroeconomic variables using survival analysis”, Credit Research Center, University of Edinburgh

Cao R., Vilar J. and Devia A. (2009), “Modelling consumer credit risk via survival analysis”, Universidad de Coruna

Lyn T. and Stepanova M. (2002), “Survival Analysis methods for personal loan data”, Operations Research, vol. 50, no.2, 277-289

Caivano G. and Bonini S. (2013), “The survival analysis approach in Basel II credit risk management: modeling danger rates in the loss given default parameter”, Journal of Credit Risk, volume 9, no.1, 101-118

Im J-K., Apley DW., Qi C. and Shan X (2012), “A time-dependent proportional hazards survival model for credit risk analysis”, Journal of the Operational Research Society, no.63, 306-321

Witzany J., Rychnovsky M. and Charamza P. (2012), “Survival Analysis in LGD Modeling”, European Financial and Accounting Journal, vol. 7, no.1, 6-27

Aas K. (2005), “The Basel II IRB approach for credit portfolios”, Norwegian Computing center, SAMBA/33/05

Thank you for your attention !