An Intensity Based Non-Parametric Default Model for Residential Mortgage Portfolios

Enrico De Giorgi, RiskLab, ETH Zurich

joint work with Vlatka Komaric, Credit Suisse Group

Risk Day 2001, Zurich
October 19, 2001

E-mail: degiorgi@math.ethz.ch

Homepage: http://www.math.ethz.ch/~degiorgi/

RiskLab: http://www.risklab.ch

An Intensity Based Non-Parametric Default Model for Residential Mortgage Portfolios

Part 1

Part 2

Introduction

- In April 2001 Swiss banks held over CHF 500 billion of debt in form of mortgages (about 63% of the total loan portfolios value held by Swiss banks).

- Estimated Swiss real estate market value is between 2300 and 2800 billion CHF (more than twice the market capitalization of all stocks included in Swiss Performance Index).

- About 86% of Swiss real estate are in the hands of private individuals.

- Not much research was done in this area so far:
  Lack of information, confidentiality, old and insufficient data, mortgages are regarded as a ”low risk” product in Switzerland.

Mortgage characteristics

A **mortgage** is a loan secured by a real estate property.
The purpose of the private clients is to finance the property.
We have the following traditional products:

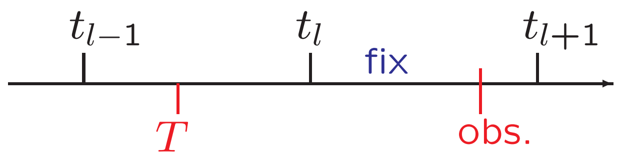
- *Adjusted-rate and term (ARM):*
  - no fixed maturity, interest rate follows market, but with a lag and is subject to politics;
  - prepayment is free for the clients (embedded option).
- *Fixed-rate and term:*
  - maturity and interest rate are fixed by the issue of the mortgage;
  - maturity usually of 2-5 years;
  - prepayment costs are charged to clients.

Default event

Definition. An obligor is said to default at time T if he loses the ability to make the next interest payment. Define the default indicator process for $t \geq 0$ by

$$X_t \stackrel{\text{def}}{=} 1_{\{T \leq t\}} = \begin{cases} 1 & \textit{default prior to time } t, \\ 0 & \textit{else.} \end{cases}$$

The observation of default is censored, it is observed only if the payment fails over a period of fixed length (usually 90 days) after it was due.



Common reasons for default

- unemployment;
- divorce;
- significantly interest rate increase.

Conditional intensity process

Let $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,p})$, $Y_{i,q} = (Y_{i,q}(t))_{t\geq 0}$ be a collection of predictors for obligor i, $i = 1, \ldots, n$.
$\mathcal{F}_{i,t} = \sigma(\mathbf{Y}_{i,s} : s \leq t)$ is the σ-algebra generated by $\widetilde{\mathbf{Y}}_{i,t} = (\mathbf{Y}_{i,s})_{0\leq s\leq t}$.
$\mathcal{D}_{i,t} = \sigma(X_{i,s} : s \leq t)$ is the σ-algebra generated by the default indicator $X_i = (X_{i,s})_{t\geq 0}$ of obligor i.
We define the enlarged filtration $\mathbb{G}_i = (\mathcal{G}_{i,t})_{t\geq 0}$ by $\mathcal{G}_{i,t} = \mathcal{F}_{i,t} \vee \mathcal{D}_{i,t}$.

Definition. Let $\mathbb{F}_i = (\mathcal{F}_{i,t})_{t\geq 0}$ be the flow of information from the predictors for obligor i. The conditional intensity process of the time to default T_i given $\mathbb{F}_i$, is the nonnegative, $\mathbb{F}_i$-predictable process $\lambda_i^{\mathbb{F}_i}$ such that the process $M_i = (M_{i,t})_{t\geq 0}$ defined by

$$M_{i,t} = X_{i,t} - \int_0^{t\wedge T_i} \lambda_{i,u}^{\mathbb{F}_i}\, du$$

is a $\mathbb{G}_i$-martingale.

Properties of the conditional intensity process $\lambda_i^{\mathbb{F}_i}$

Let $S_i(t\,|\,\mathcal{F}_{i,t}) = \mathbb{P}\Big[T_i > t\,|\,\mathcal{F}_{i,t}\Big]$ be the conditional survival probability and $f_i(t\,|\,\mathcal{F}_{i,t}) = \lim_{s\searrow 0} \frac{1}{s}\mathbb{P}\Big[T_i \in (t, t+s]\,|\,\mathcal{F}_{i,t}\Big]$ be the conditional density function of T_i. .

Under technical conditions, $\lambda_i^{\mathbb{F}_i}$ and f_i exist, and we have

- $\lim_{s\searrow 0} \frac{1}{s}\mathbb{P}\Big[T_i \in (t, t+s]\,|\,\mathcal{G}_{i,t}\Big] = 1_{\{T_i>t\}}\,\lambda_{i,t}^{\mathbb{F}_i}$.

- $\lambda_{i,t}^{\mathbb{F}_i} = \frac{f_i(t\,|\,\mathcal{F}_{i,t})}{S_i(t\,|\,\mathcal{F}_{i,t})}$.

- $S_i(t\,|\,\mathcal{F}_{i,t}) = \exp\Big(-\int_{d_i}^{t\vee d_i} \lambda_{i,u}^{\mathbb{F}_i}\,du\Big)$ where $d_i =$ time of issue.

On the set $\{T_i > t\}$ and for $\Delta t \ll 1$, $\lambda_{i,t}^{\mathbb{F}_i}\,\Delta t$ approximates the conditional probability that a default occurs during $(t, t+\Delta t]$.

Overview of the model

- Form homogeneous groups characterized by their credit rating.
- Model time-to-default as the first jump-time of an inhomogeneous Poisson process with stochastic intensity (doubly stochastic Poisson process or Cox process).
- Link the intensity to explaining factors (economic environment, mortgage characteristics, obligor characteristics).
- Given a realization of explaining factors, suppose individual defaults occur independently.
- Fit the model to a mortgage portfolio for determining the form of the linking functions.

The model

Let $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,p})$, $Y_{i,q} = (Y_{i,q}(t))_{t\geq 0}$ be a collection of predictors for the intensity process $\lambda_i^{\mathbb{F}_i}$ for obligor i. We model $\lambda_i^{\mathbb{F}_i}$ as a function of $\mathbf{Y}_i$.
We suppose that

$$\lambda_{i,t}^{\mathbb{F}_i} = \lambda_{i,0}\, h_{i,0}(t - d_i) \prod_{q=1}^{p} h_{i,q}(Y_{i,q}(t)).$$

We write $\lambda_{i,t}^{\mathbb{F}_i} = \lambda_{i,t}^{\mathbb{F}_i}(\theta_i; \mathbf{Y}_{i,t})$ where $\theta_i = (\log \lambda_{i,0}, \log h_{i,0}, \ldots, \log h_{i,p})$. Here $h_{i,0}, h_{i,1}, \ldots, h_{i,p}$ are the link functions to be estimated later.

Let $\eta_{i,t}^{\mathbb{F}_i}(\theta_i; \mathbf{Y}_{i,t}) = \log \lambda_{i,t}^{\mathbb{F}_i}(\theta_i; \mathbf{Y}_{i,t})$. Then we obtain

$$\eta_{i,t}^{\mathbb{F}_i}(\theta_i; \mathbf{Y}_{i,t}) = \log \lambda_{i,0} + \log h_{i,0}(t - d_i) + \sum_{q=1}^{p} \log h_{i,q}(Y_{i,q}(t)).$$

We suppose that $\mathbb{E}\Big[\log h_{i,q}(Y_{i,q}(t))\Big] = 0$ for $i = 1, \ldots, n$, $q = 1, \ldots, p$.

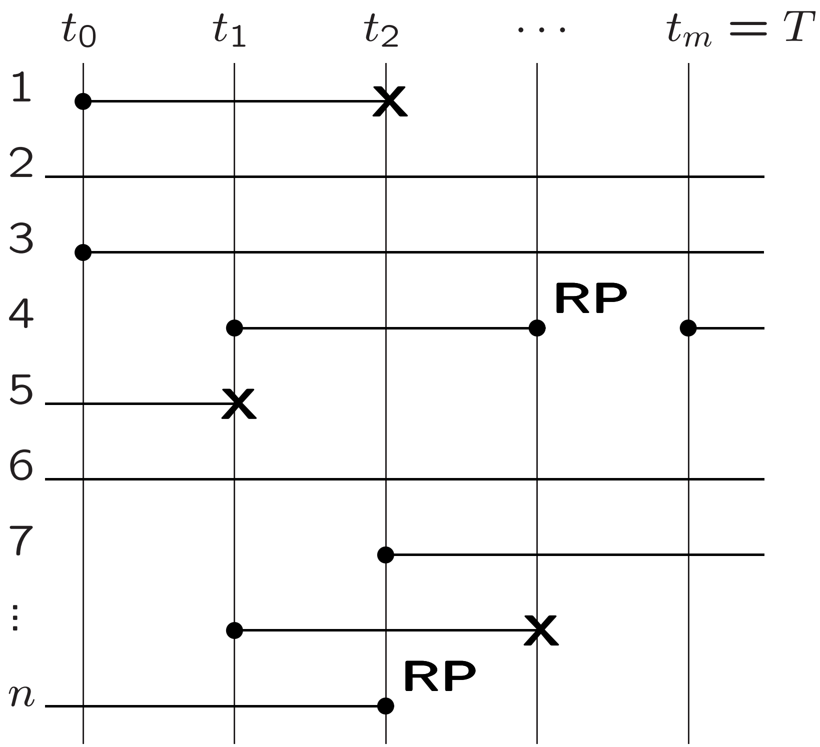
Assumptions

- The θ_i's are the same for all obligors in the same rating class.
  $\Rightarrow$ Functional form depends only on the rating class.

- Given $T_i > t$, the conditional probability that obligor i will survive time $t+s$ for $s > 0$ depends on the history only through the predictors at time t.
  $\Rightarrow$ Treats all the outstanding mortgages at time t in the same way.

- Given the predictors up to time t, defaults of obligors up to time t are conditionally independent.
  $\Rightarrow$ Dependence structure is totally described by the predictors.

Estimation of the model for one rating class



RP: repayment.
X: default.

- $\theta = (\log \lambda_0, \log h_0, \log h_1, \ldots, \log h_p)$ are the same for every obligor.
- Group obligors such that their predictors $\mathbf{Y}_i$ and their time of issue d_i are identical in every group (J groups).
- Let $0 = t_0 < t_1 < \cdots < t_m = T$.
- $O_{j,l}$ = number of outstanding mortgages during $(t_l, t_{l+1}]$ in group j.
- $D_{j,l}$ = number of mortgages defaulted during $(t_l, t_{l+1}]$ in group j.

Conditional likelihood function of the discretized model

Assuming that λ and the predictors are constant on $[t_l, t_{l+1})$, then on the set $\{T_i > t_l\}$ for obligor i in group j and $\mathbf{Y}_j = \mathbf{y}_j$ we have

$$\begin{aligned}
\mathbb{P}\Big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{G}_{i,t_l}\Big] &= \frac{\mathbb{P}\Big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{F}_{j,t_l}\Big]}{\mathbb{P}\Big[T_i > t_l \,|\, \mathcal{F}_{j,t_l}\Big]} \\
&= \frac{S(t_l \,|\, \mathcal{F}_{j,t_l}) - \mathbb{P}\Big[T_i > t_{l+1} \,|\, \mathcal{F}_{j,t_l}\Big]}{S(t_l \,|\, \mathcal{F}_{j,t_l})} \\
&= 1 - \exp\Big(-(t_{l+1} - t_l)\lambda_{t_l}(\theta; \mathbf{y}_{j,t_l})\Big) \\
&\overset{\text{def}}{=} u_{j,l}(\theta).
\end{aligned}$$

The likelihood function for the observation is thus given by

$$L(\theta) = \prod_{l=0}^{m-1} \prod_{j=1}^{J} \underbrace{\binom{O_{j,l}}{D_{j,l}} u_{j,l}(\theta)^{D_{j,l}} \Big(1 - u_{j,l}(\theta)\Big)^{O_{j,l} - D_{j,l}}}_{\text{binomial distribution}}.$$

Generalized additive model (GAM)

Let V be a real random variable. Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ be a set of predictors. Given $\mathbf{Y}$, V has the conditional distribution function $F_{\mathbf{Y}}$ with $\mu(\mathbf{Y}) = \mathbb{E}\Big[V \,\big|\, \mathbf{Y}\Big]$. We assume that for functions $f_1, \ldots, f_p$, we have

$$G(\mu(\mathbf{Y})) = \eta(\mathbf{Y}) = \alpha + \sum_{q=1}^{p} f_q(Y_q)$$

where G is the link function, $\mathbb{E}\Big[f_q(Y_q)\Big] = 0$ for $q = 1, \ldots, p$.
η is called an additive form, $\theta = (\alpha, f_1, \ldots, f_p)$ are the unknown parameters to be estimated. The triple $(\eta, G, F_{\mathbf{Y}})$ is called a GAM.

Remarks

- If all the f_q's are linear functions, then $(\eta, G, F_{\mathbf{Y}})$ is called a generalized linear model (GLM).
- For observations $(V_i)_{i=1,\ldots,M}$ we need $V_i \,|\, \mathbf{Y}_i \sim F_{\mathbf{Y}_i}$, independently.
- The GAM serves as a diagnostic tool for suggesting transformations of the predictors.

GAM estimation

If $V \,|\, \mathbf{Y} \sim F_{\mathbf{Y}}$ has an exponential family density

$$f_{\mathbf{Y}}(v;\xi,\phi) = \exp\left\{\frac{v\xi - b(\xi)}{a(\phi)} + c(v,\phi)\right\}, \qquad v \in \text{support}(F_{\mathbf{Y}})$$

where ξ is the natural parameter ($b'(\xi) = \mu$) depending on $\mathbf{Y}$, and ϕ is the dispersion parameter, then the local scoring algorithm with backfitting can be applied to solve the GAM (Hastie and Tibshirani, 1990).

Remarks

- $F_{\mathbf{Y}} = \text{binomial}(n, p(\mathbf{Y}))$ is an exponential family density with $\phi = 1$.
- The local scoring algorithm maximizes the likelihood function by a modified Newton-Raphson procedure.
- The local scoring algorithm converges for cubic smoothing splines.

Backfitting algorithm

Let $G = id$ and $(\eta, id, F_{\mathbf{Y}})$ the simple additive model, with $F_{\mathbf{Y}}$ an exponential family density. We have for $i = 1, \ldots, M$

$$V_i = \alpha + \sum_{q=1}^{p} f_q(Y_{i,q}) + \epsilon_i,$$

where $\epsilon_i = V_i - \mathbb{E}\Big[V_i \,|\, \mathbf{Y}\Big]$. The backfitting algorithm proceeds as follows:

- Initialization $r = 0$: $\widehat{f}_q^0 \equiv 0$ for $q = 1, \ldots, p$, $\widehat{\alpha}^0 = \frac{1}{M}\sum_{i=1}^{M} V_i$.
- Iteration $r \to r + 1$: cycle over $q = 1, \ldots, p$

$$\widehat{f}_q^{r+1} = S_q^{\lambda}\left(V_i - \widehat{\alpha}^r - \sum_{q'=1}^{q-1} \widehat{f}_{q'}^{r+1}(Y_{i,q'}) - \sum_{q'=q+1}^{p} \widehat{f}_{q'}^{r}(Y_{i,q'}) \,\Bigg|\, Y_q \right)_{i=1,\ldots,M}$$

until $\max_{i=1,\ldots,M} \left| \widehat{f}_q^{r+1}(Y_{i,q}) - \widehat{f}_q^{r}(Y_{i,q}) \right|$ is small enough.

S_q^{λ} denotes a smoothing operator (linear) with smoothing factor λ.

Stepwise selection technique

- Choice of the smoothing method (smoothing spline, local regression, kernel regression,...).

- For each function f_q define a set Θ_q of alternatives of increasing complexity for the corresponding smoothing operator S_q^λ, in terms of the number of degrees of freedom df ($df = 0$ for one S_q^λ means that $f_q \equiv 0$, $df = 1$ means f_q linear).

- Let $\widehat{\theta}_1 \in \Theta = \mathbb{R} \times \Theta_1 \times \cdots \times \Theta_p$. Define $\widehat{\theta}_2$ by increasing the complexity one step forward in $\Theta_{q'}$ for exactly one $q' = 1, \ldots, p$ in $\widehat{\theta}_1$ ($\widehat{\theta}_1$, $\widehat{\theta}_2$ are nested models).

- Compare the two models by testing the null hypothesis $H_0 : \theta = \widehat{\theta}_1$ against the alternative $H_A : \theta = \widehat{\theta}_2$ using a χ^2-test.

Akaike information criterion

Let $(\eta, G, F_{\mathbf{Y}})$ be a GAM and let $F_{\mathbf{Y}}$ be an exponential family density with dispersion parameter ϕ.

We define the Akaike information criterion for the model $\widehat{\theta} \in \Theta$ by

$$AIC_{\widehat{\theta}} = D(\widehat{\theta}; \mathbf{v}) + 2\,\phi\, df_{\widehat{\theta}},$$

where $df_{\widehat{\theta}}$ is the number of degrees of freedom of the model.

- AIC is a penalized version of the deviance D.
- AIC accounts for the number of degrees of freedom used by the smoothers.
- Usually a lower AIC implies that the model fits better then another.
- AIC offers a criterion for comparing two models $\widehat{\theta}_1, \widehat{\theta}_2 \in \Theta$, nested or non-nested.
- No specific statistical test is associated with comparing AIC's.

Reformulation of the default model as GAM

Let

$$
\begin{aligned}
V_{j,l} &= \frac{D_{j,l}}{O_{j,l}} \\
u_{j,l}(\theta) &= 1 - \exp\Big(-(t_{l+1} - t_l)\,\lambda(t_l, \theta \,|\, \mathbf{y}_{j,t_l})\Big)
\end{aligned}
$$

then

$$
\begin{aligned}
V_{j,l} &\sim \frac{1}{O_{j,l}} \text{binomial}(O_{j,l}, u_{j,l}(\theta)) \\
G(u_{j,l}(\theta)) &= \log \lambda_0 + \log h_0(t - d_j) + \sum_{q=1}^{p} \log h_q(y_{j,q}(t_l))
\end{aligned}
$$

where $u_{j,l}(\theta) = \mathbb{E}_\theta\Big[V_{j,l} \,|\, \mathbf{y}_{j,t_l}\Big]$ and $G : (0,1) \longrightarrow \mathbb{R}, \mu \mapsto \log(-\log(1-\mu))$ is the link function (the complementary log log-function).

$\Rightarrow$ Generalized additive model.

Data set

- Sub-portfolio $\mathcal{P}$ with 73683 Swiss residential mortgages.
- $t_0 =$ 1st quarter 1994, $t_m =$ 4th quarter 2000.
- Observation of $\mathcal{P}$ follows at the *end* of each quarter (March 31, June 30, September 30, December 31).
- The mortgage product and the mortgage interest rate r_{i,t_l} applied during the quarter $[t_{l-1}, t_l)$ are available for $i = 1, \ldots, 73683$ and $l = 1, \ldots, m$.
- Obligors belongs to 26 different economic and political regions (26 cantons).
- Two rating classes are considered:
  A=higher rating and B=lower rating.

Predictors

For obligor i ($i = 1, \ldots, 73683$) we use the following predictors.

- Quarter of the year $Y_{i,0}$: $Y_{i,0}(t_l) = k$, if t_l is the k-th quarter.
- Quarterly regional unemployment rate $Y_{w,1}$, if obligor i lives in region $w = 1, \ldots, 26$.
- Lags of $1, \ldots, 16$ quarters for the regional unemployment rate are considered (notation: $Y_{w,1}^{(r)}$, $w = 1, \ldots, 26$, $r = 1, \ldots, 16$).
- Indicator variable $Y_{i,3}$ for mortgage product: adjusted-rate ($Y_{i,3} = 1$), fixed-rate mortgage ($Y_{i,3} = 2$).
- Levels $Y_{i,4}$ for the relative interest rate change over last quarter:

$$Y_{i,4}(t_l) = \begin{cases} 1 & \text{if } x_{i,t_l} < 0, \\ 2 & \text{if } x_{i,t_l} = 0, \\ k+1 & \text{if } x_{i,t_l} \in (a_{k-1}, a_k], \;\; k = 2, 3, \\ 5 & \text{if } x_{i,t} > 0.5, \end{cases}$$

where $x_{i,t_l} = \frac{r_{i,t_l}}{r_{i,t_{l-1}}} - 1$, $a_1 = 0$, $a_2 = 0.25$ and $a_3 = 0.5$.

Selected models

- We have $J = 260$ groups of obligors characterized by the predictor realizations (6500 observations of $O_{j,l}$ and $D_{j,l}$ for each rating class).
- 3265 non-zero observations of $O_{j,l}$ for rating A, and 2713 non-zero observations of $O_{j,l}$ for rating B.

The following models has been selected by our criterion:

- Rating A

$$G(u_{j,l}(\widehat{\theta}_{\mathrm{A}})) = \widehat{\alpha}_{\mathrm{A}} + \widehat{f}^{(11)}_{1,\mathrm{A}}(y^{(11)}_{j,1}(t_l)) + \\ + \left(\widehat{\beta}_{3,\mathrm{A}}\, 1_{\{y_{j,3}(t_l)=1\}} + \widehat{\gamma}_{3,\mathrm{A}}\right) + \widehat{f}_{4,\mathrm{A}}(y_{j,4}(t_l)).$$

- Rating B

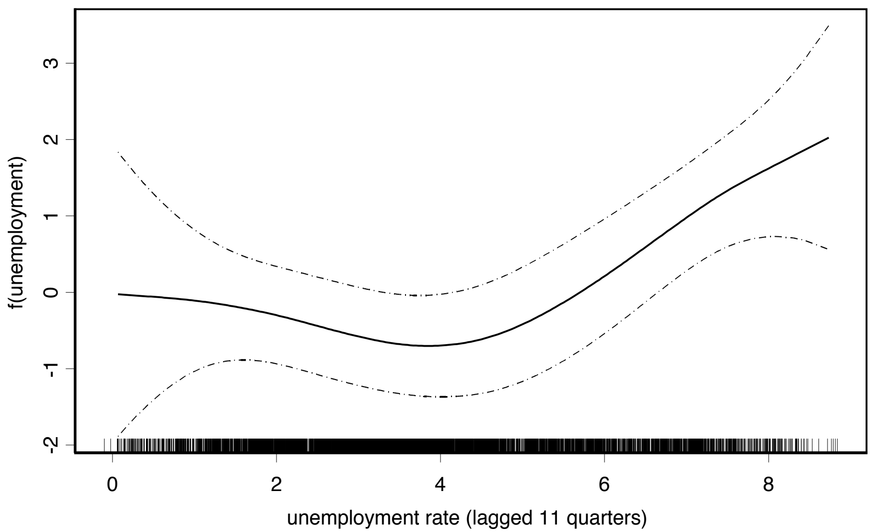
$$G(u_{j,l}(\widehat{\theta}_{\mathrm{B}})) = \widehat{\alpha}_{\mathrm{B}} + \widehat{f}^{(q)}_{0,\mathrm{B}}(y_0(t_l)) + \widehat{f}^{(8)}_{1,\mathrm{B}}(y^{(8)}_{j,1}(t_l)) + \\ + \left(\widehat{\beta}_{3,\mathrm{B}}\, 1_{\{y_{j,3}(t_l)=1\}} + \widehat{\gamma}_{3,\mathrm{B}}\right) + \widehat{f}_{4,\mathrm{B}}(y_{j,4}(t_l)).$$

Parametric estimates

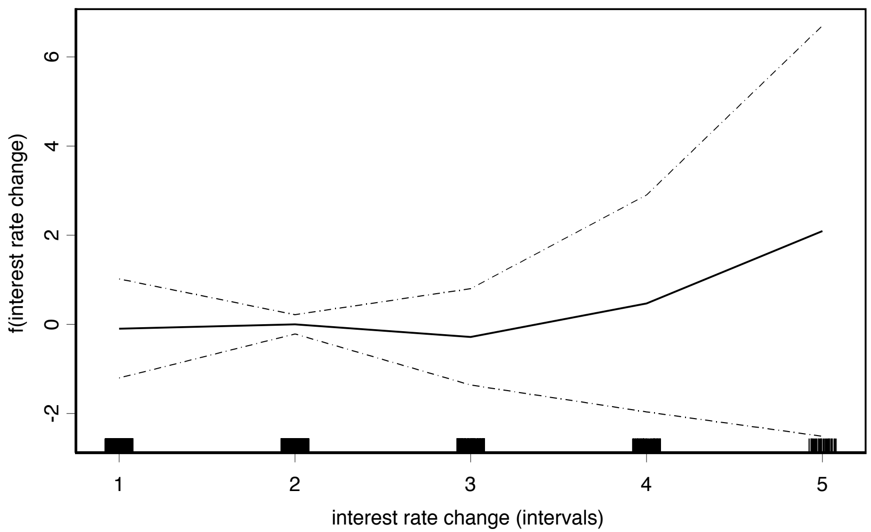
Rating		$\widehat{\alpha}$	$\widehat{\beta}_3$	$\widehat{\gamma}_3$
A	Estimate	-9.9108	-1.3568	0.6740
	Standard error	0.7752	0.4443	0.2207
	Approx. 95% CI	-11.4612 -8.3604	-2.2454 -0.4682	0.2326 1.1154
B	Estimate	-6.8644	-1.7893	0.8462
	Standard error	0.3636	0.1690	0.0799
	Approx. 95% CI	-6.1372 -7.5916	-2.1273 -1.4513	0.6864 1.006

Parametric estimates for the two models (Rating A and Rating B), with standard errors and approximated 95% confidence intervals.

Non-parametric estimates: rating A

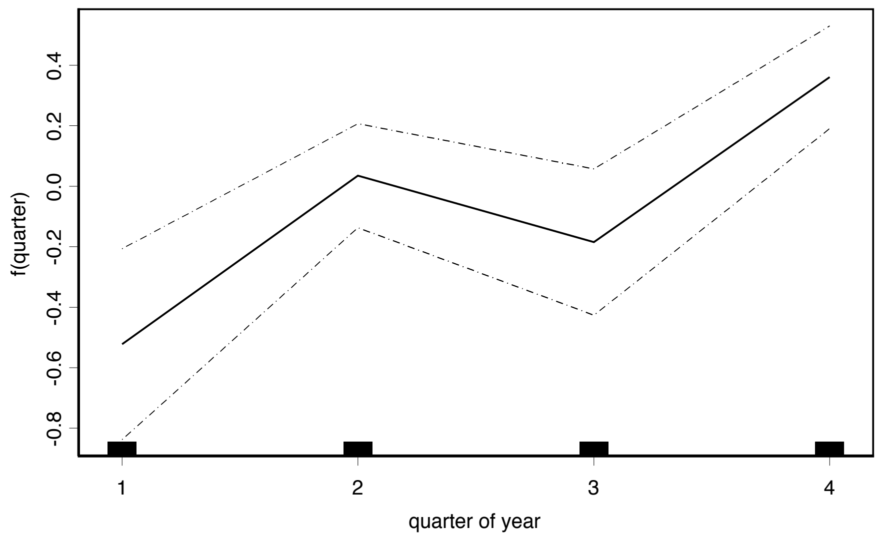


Spline estimation $\widehat{f}^{(11)}_{1,\mathrm{A}}$ *with 1.2 degrees of freedom. Dotted lines give the approximated 95% confidence interval.*

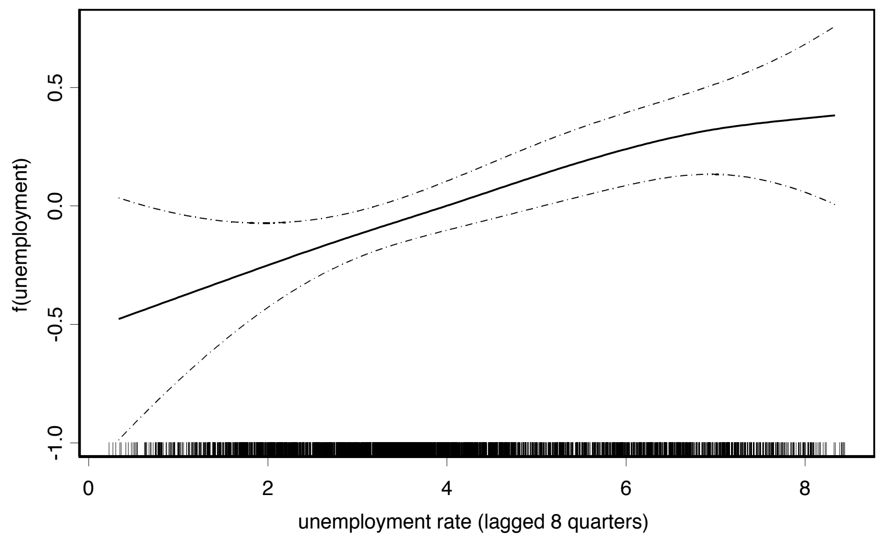


Spline estimation $\widehat{f_{4,\mathrm{A}}}$ *with 2 degrees of freedom. Dotted lines give the approximated 95% confidence interval.*

Non-parametric estimates: rating B

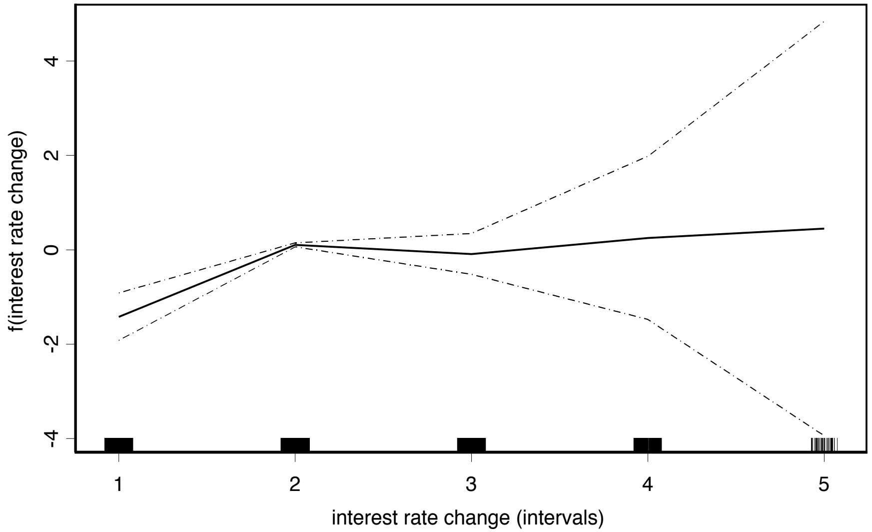


Spline estimation $\widehat{f}^{(q)}_{0,\mathrm{B}}$ *with 2 degrees of freedom. Dotted lines give the approximated 95% confidence interval.*



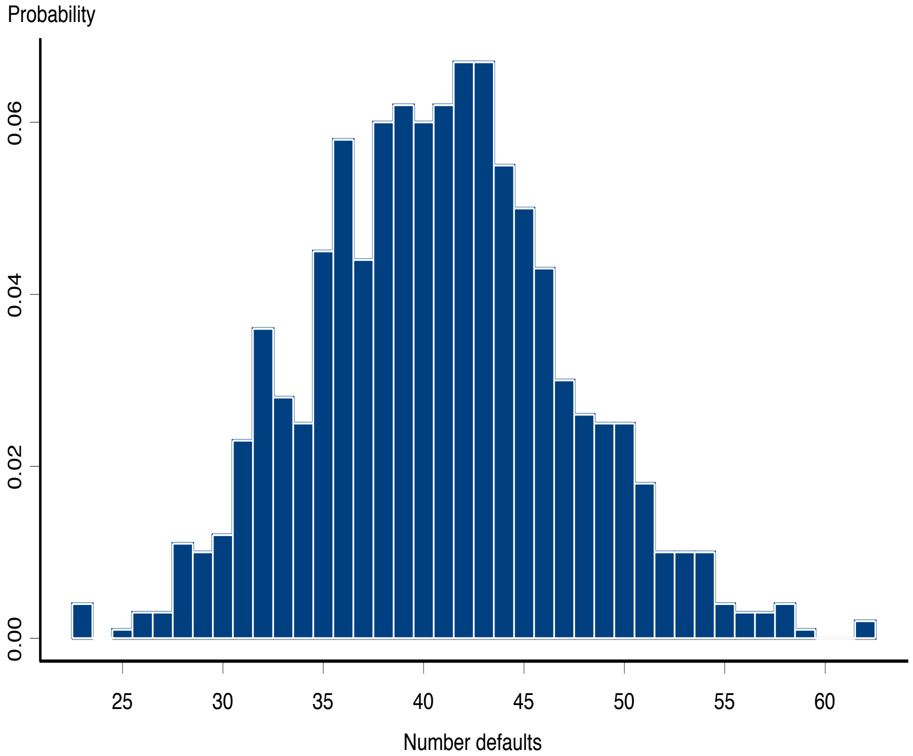
Spline estimation $\widehat{f}^{(8)}_{1,\mathrm{B}}$ *with 1.1 degrees of freedom. Dotted lines give the approximated 95% confidence interval.*

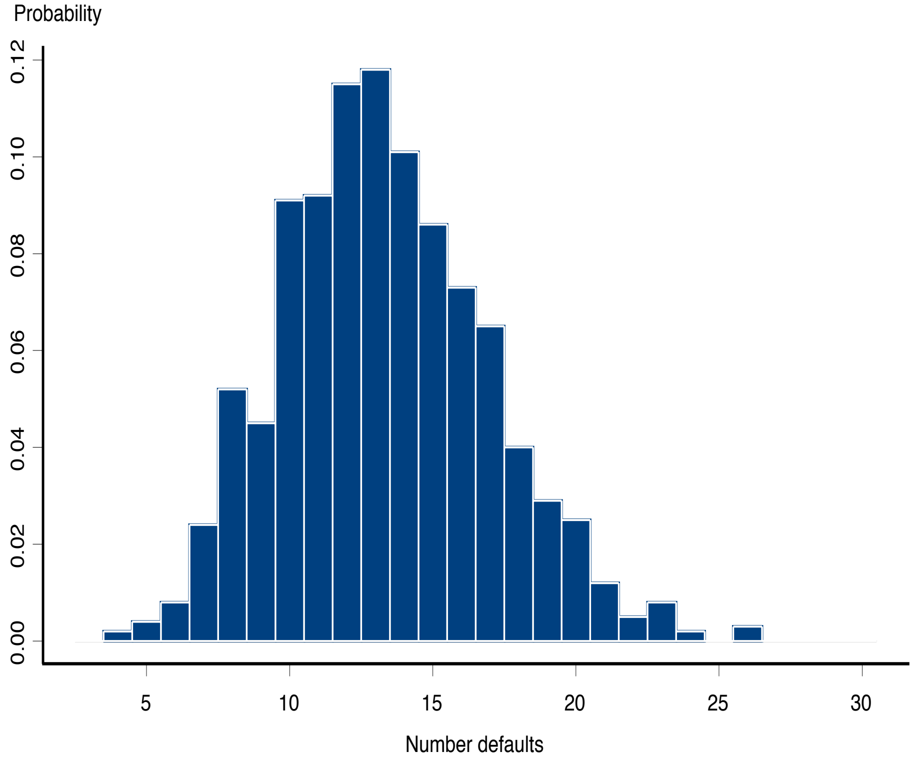
Non-parametric estimates: rating B (2)



Spline estimation $\widehat{f}_{4,\mathrm{B}}$ *with 1.9 degrees of freedom. Dotted lines give the approximated 95% confidence interval.*

Simulation under different scenarios





1000 simulations of the total number of defaults during the first quarter 2001 in a portfolio $\mathcal{P}'$ with 100000 obligors. Obligors in $\mathcal{P}'$ are distributed among the 26 regions, the 2 mortgage products and the 2 rating classes as in portfolio $\mathcal{P}$ at the end of the last quarter 2000. Two scenario for the interest rate are considered: increase of 0.75% (left histogram), decrease of 0.5% (right histogram).

Conclusion

Advantages of the model:

- Dynamical model.
- Choice of the predictors very flexible.
  (The model suggests how data has to be collected.)
- Link the default process to the macro-economical environment.
- Dependence structure given by the common predictors.
- Applicable to available data.

Further research:

- Stochastic modeling of recoverables.
- Stochastic modeling of predictors.