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

Enrico De Giorgi*

RiskLab
Swiss Federal Institute of Technology
ETH Zentrum
CH-8092 Zürich
Tel.: +41 1 632 07 99
Fax.: +41 1 632 10 85
degiorgi@math.ethz.ch

22nd November 2001

Abstract

In April 2001 Swiss banks held over CHF 500 billion in mortgages. This important segment accounts for about 63% of all the loan portfolios of Swiss banks. In this paper we restrict our attention to residential mortgages held by private clients, i.e. borrowers who finance their property by the loan and we model the probability distribution of the number of defaults using a non-parametric intensity based approach. We consider the time-to-default and, by conditioning on a set of predictors for the default event, we obtain a log-additive model for the conditional intensity process of the time-to-default, where the contribution of each predictor is described by a smooth function. We estimate the model by using a local scoring algorithm coming from the generalized additive model.

1 Introduction

A mortgage is a lien or claim against a real estate property given by the borrower to the lender (usually the bank) as a security for money borrowed. In other words a mortgage is a loan collateralized by real estate, which obligate the borrower to make a series of payments of interest and principal.
In April 2001 Swiss banks held over CHF 500 billion in mortgages. This important segment accounts for about 22% of all the assets and about 63% of all loan portfolios of Swiss banks[1].

*The author wishes to thank Vlatka Komaric of Credit Suisse Group for providing the data set used in this work and Jürg Burkhard and Urs Wolf of Credit Suisse Group for the support and the interesting discussions. I'm grateful to Uwe Schmock and to seminar participants at RiskLab, UBS AG and University of Lausanne for suggestions and remarks. Research of the author was financially supported by Credit Suisse Group, UBS AG and Swiss Re through RiskLab, Switzerland.

[1]Source: Swiss National Bank, "Statistisches Monatsheft", May 2001, available from the website of the Swiss National Bank, http://www.snb.ch.

Moreover, the market value of all real estate in Switzerland has been estimated at between CHF 2,300 billion and CHF 2,800 billion (see Credit Suisse Group (2000)), more than twice as much as the market capitalization of the companies listed in the Swiss Performance Index. Nevertheless, not much research has been done in this area so far and the estate market is characterized by the limited amount of information available to investors.
In 2000 47% of the overall Swiss property estate was purely residential buildings, while the remaining 53% is dived between commercial buildings (12%), offices, administrative and public buildings (14%), industrial, retail and commercial premises (14%) and other properties (agricultural, hotel, catering) (13%). Moreover, about 86.5% of Swiss properties are owned by private individuals and only 13.5% by insurance companies, pension funds, real estate companies or real estate funds (see Credit Suisse Group (2000), Chapter 1, for more details on the Swiss real estate market).
The percentages reported above give an idea about the importance of the market of private residential mortgages, i.e. residential mortgages held by private individuals, whose purpose is to finance their own real estate by the loan. We want to address our attention to this particular sector, for which some traditional approaches for modeling credit risk, as for example the firm's value approach of Merton, are not appropriate. Moreover the credit quality of private clients is usually determined by the issue of the mortgage and a new evaluation is made only when interest payments are delinquent for some period, and thus when the credit quality of the borrower has already deteriorated, even if a default has not occurred. The lack of specific information about each single counterpart in a residential mortgage portfolio, has to be considered for modeling the credit risk.

Portfolio managers dealing with big portfolios of residential mortgages are responsible for maintaining adequate reserves to cover future losses that may occur on outstanding mortgages. The credit risk associated with mortgage portfolios is essentially the risk that borrowers will default and fail to meet interest rate payments on the outstanding balance (the *default risk*) and the risk that given a default, the collateral value of the defaulted mortgage (i.e. the minimum between the current house value and the face value of the mortgage's note) is less then the outstanding balance plus unpaid interest.
In the literature different explanatory variables for a default in a mortgage contract has been identified. Smith, Sanchez, and Lawrence (1999) and Deng (1995) select mortgage and economic characteristics for predicting default and for calculating the probability of incurring a loss on a defaulted loan. Deng (1995), Santos Silva and Murteira (2000) use borrower characteristics, such as the payment-income ratio, which is usually only observable by the issue of the mortgage. Follian, Huang, and Ondrich (1999) include in their model duration, location, demographic and economic variables as covariates to explain default. The contingent claim approach, which is considered by Kau and Keenan (1998), treats default as a rational decision, such that a default occurs if the house value (equity value) falls below the value of the mortgage. This approach considers a "strategic" mortgage default as a put option on the mortgage itself, i.e. the option of selling back the house to the lender in exchange for eliminating the mortgage obligation. The behaviour of private individuals, whose purpose is to finance their property with the loan, is, however, not always rational in the sense of the economic theory. Deng and Quigley (2000) propose combining the financial value of the put option in the contingent claim framework, with non-option related variables, such as unemployment or divorce rates.

For our study, we consider quarterly observation of a big Swiss residential mortgage portfolio from 1994 to 2000, selecting a sub-portfolio of only private clients with a one-family house financed by the loan, divided in 26 economic and geographical regions across Switzerland (26 cantons). Analyzing our data set, which has the advantage of containing observations during the economic recession of the 90's, we deduce that a mortgage default is usually triggered by numerous personal "non-financial" reasons, more than by a rational economic decision (see also Deng and Quigley (2000)). One common cause for default is unemployment; another is divorce. In the case of unemployment the income of the borrower can dramatically decrease and the consequence will be the inability to pay the interest on the outstanding balance. Therefore, considering a big portfolio, we conclude that the number of defaults, as well as the loss amount, is correlated to the economic environment, i.e. to economic factors such as unemployment and interest rates, or to social and demographic developments, such as the increase of the number of divorces observed in Switzerland in the last 10 years. Normally, there is a lag of one, two or more years between a macro-economic or social event and an increase in the number of defaults. This time lag can be explained by the fact that borrowers with economic difficulties will often use own savings to pay the interest before defaulting: this behaviour is quite typical for private clients, in contrast to corporate clients.

Before introducing our approach, we give some insight into the problem we treat in this work. We consider a simple fixed rate and fixed maturity mortgage contract between two parties, B and C. Party B (the bank) lends money to party C (the client), with a fixed interest rate r and maturity τ. In return party C gives B a guarantee covered by a collateral, i.e. the property financed by the loan. The value of the coverage has an upper bound v, deterministic and independent of the house value over the time, as stipulated in the mortgage contract by the issue of the mortgage. The bank fixes maximal coverage by v and this amount is usually greater than the initial outstanding balance. As already stated above, B is faced with essentially two risks: the risk that C will not be able to pay periodical interests (i.e. he defaults) and the risk that the collateral given as a guarantee will fall in value and be worth less than the outstanding balance at the time of default plus additional unpaid interests. Suppose that $(B_t)_{t\in\{0,\cdots,\tau\}}$ is a process denoting the outstanding balance of party B at time $t = 0, \ldots, \tau$. B_t denotes the outstanding principal plus unpaid interest; B_0 represents the money lent to party C at the beginning of the contract. $(V_t)_{t\in\{0,\cdots,\tau\}}$ is a stochastic process giving the value of the collateral at time t (selling value, realized value).
Given a default at time T, the loss function L depends essentially on B_T and V_T. According to the mortgage contract, the bank can realize $\min(V_T, v)$ and the loss at time T is thus given by

$$L_T = [B_T - \min(V_T, v)]^+ . \tag{1}$$

This last equation suggests that L_T is positive if and only if $B_T > \min(V_T, v)$. This means that a default causes a loss if and only if the collateral represents insufficient coverage, and in a big portfolio this is only a small percentage of all mortgages defaulted on.
For a big portfolio is thus the total number of defaults and the dependence between default events and the value of the collateral which can be relevant for risk management purposes. Moreover, as stated above, the number of defaults is related to the economic environment, as well as to the house price, and specific economic scenarios can thus imply major losses.

Looking at equation (1), we conclude that the loss function depends on three random variables: the default time, the outstanding balance and the value of the collateral at default. A static model, characterized by a fixed time horizon, will not capture the behaviour of the default process under different scenarios. The observation that default is usually triggered by specific macro-economical, social or personal conditions, makes a static approach inappropriate, since the behaviour of the obligors under different conditions will not be taken into consideration.
We propose an intensity based approach for modeling the time to default, which we take to be the first-jump-time of an inhomogenous Poisson process with stochastic intensity, also called doubly stochastic Poisson or Cox Process. The main idea consists in conditioning on a set of explaining variables, which affect the borrowers' behaviour, and to consider borrower defaults as independent given this set of information about the economic environment. This hypothesis has some empirical

relevance for private individuals. The intensity process is directly related to the explaining variable, as in a proportional hazard rate (PHR) model of Cox and Oakes (1984), but in our model the relationship is not assumed to be log-linear as in a PHR model. We simply assume some smooth, perhaps non-linear, relationship.
The goal is to model the dependence structure of a mortgage credit portfolio through a common scenario and, by applying the conditioning technique, to obtain computational advantages. In Section 2 the default event is defined. Section 3 introduces the model. Section 4 is devoted to the estimation methodology. Section 5 presents our results and Section 6 concludes our proposal. Technical results are given in the appendix.

2 Mortgages Default

Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space. Let $\mathcal{P} = \{(d_i, v_i, B_i, V_i, r_i), i = 1, \ldots, n\}$ be a portfolio of n mortgages outstanding during some period after time t_0 (we just eliminate from $\mathcal{P}$ mortgages defaulting in t_0). For mortgage i, d_i denotes the time of issue, v_i the maximal coverage stipulated in the mortgage contract, $B_i = (B_{i,t})_{t \geq d_i}$ is a process giving the outstanding balance at time t, $V_i = (V_{i,t})_{t \geq d_i}$ is a stochastic process giving the house value at time t and $r_i = (r_{i,t})_{t \geq d_i}$ is a process (stochastic or deterministic) giving the interest rate applied to mortgage i at each time t. We suppose that a mortgage portfolio is totally characterized by $\mathcal{P}$. Each mortgage in $\mathcal{P}$ either has been issued before t_0 ($d_i < t_0$) or after t_0 ($d_i \geq t_0$). In the first case, the mortgage is not observed from the start of the contract but only after t_0. In the sequel we use alternately and with the same meaning the words "mortgage" and "obligor", the latter indicating private individuals borrowing money under a mortgage agreement.

We divide the time interval $[t_0, \infty)$ in sub-intervals $(t_l, t_{l+1}]$, $l \in \mathbb{N}$ of the same length. Each interval $(t_l, t_{l+1}]$ represents one year, one quarter or one month respectively and each time t_l, $l \in \mathbb{N}$ represents the *discrete points in time where an interest payment on the mortgage is due.* The length $|t_{l+1} - t_l|$ is set to be the time unit, meaning that $|t_{l+1} - t_l| = 1$. Let $\mathcal{T} = \{t_l \mid l \in \mathbb{N}\}$. At each time $t_l \in \mathcal{T}$ the obligor may be punctual and make an interest and/or principal payment or he may repay the loan in its entirety. A default is observed when the interest payment due at time $t_l \in \mathcal{T}$ is delinquent over a period of fixed length (usually 90 days) after t_l and the loan has not be repaid at time t_l (see Figure 1). We suppose that the observation of a default is right-censored and we define the default time as follows.

Definition 2.1 (Default). *An obligor is said to* default *at time $T \in [t_0, \infty)$ if he loses the ability to make the next interest payment at time $\bar{t} = \min\{t \in \mathcal{T} : t \geq T\}$ and to repay the outstanding balance before $\bar{t}$ or at time $\bar{t}$. $T = \infty$ means that a default will not occur, i.e. the observation of the default time is left-censored by a repayment.*

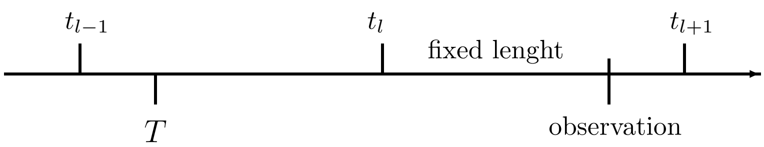

Figure 1: Time to default T and observation of T.

This definition seems to be quite complex, but it will be very useful for modeling the time-to-default, since we can assume that time-to-default is a continuous variable and a default can thus occur at any time t and not only at discreet points in time where the interest payment is due.

Let $T_i : \Omega \longrightarrow [t_0, \infty)$ be a positive random variable giving the time of default for obligor i, $i = 1, \ldots, n$. Under our assumptions on $\mathcal{P}$ we have $\mathbb{P}\big[T_i = t_0\big] = 0$. Moreover $\mathbb{P}\big[T_i > t\big] > 0$, $\forall t > t_0$. The default indicator is a stochastic process defined by $X_i = 1_{\{T_i \leq t\}}$ for $t \geq t_0$. We define the default indicator process for each $t \geq t_0$ independently from d_i. If $d_i > t_0$ then X_i will be identical to zero on $[t_0, d_i]$. We denote the loss function of the portfolio $\mathcal{P}$ by $L = (L_t)_{t \geq t_0}$, where $L_t : \Omega \longrightarrow \mathbb{R}^+$ is given thought

$$L_t = \sum_{i=1}^{n} X_{i,t} \left[B_{i,t} - \min(V_{i,t}, v_i)\right]^+ . \tag{2}$$

Naturally, the portfolio manager is interested in the difference between $L_{t_{l+1}} - L_{t_l}$ giving the specific reserve to be accumulated for period $(t_l, t_{l+1}]$ to face to new losses. This difference can be negative, meaning that some specific reserve can be relaxed. A defaulted mortgage will remain in the portfolio for some period after default, meaning that a defaulted mortgage will not be treated as a definitively closed position. During the period a defaulted mortgage remains outstanding, the credit recovery department of the bank tries to recover the outstanding balance, by liquidating the property given as a guarantee.

3 Model for the default probability

We restrict our attention to the problem of modeling the default probability. We consider the portfolio $\mathcal{P}$ introduced in the previous section, as well as the times of default $T_1, \ldots, T_n$.
We denote by $\mathbb{F}_i = (\mathcal{F}_{i,t})_{t \geq t_0}$ the flow of information available for obligor i, $i = 1, \ldots, n$, at each time $t \geq t_0$. With $\mathbb{D}_i = (\mathcal{D}_{i,t})_{t \geq t_0}$ we denote the natural filtration of the default indicator process X_i for obligor i, i.e. $\mathcal{D}_{i,t} = \sigma(X_{i,s} : s \leq t)$. Moreover, $\mathbb{G}_i = (\mathcal{G}_{i,t})_{t \geq t_0}$, where $\mathcal{G}_{i,t} = \mathcal{D}_{i,t} \vee \mathcal{F}_{i,t} \equiv \sigma\left(\mathcal{D}_{i,t} \cup \mathcal{F}_{i,t}\right)$ denotes the enlarged σ-algebra, giving all the information about the predictors and default indicator at each time $t \geq t_0$, for obligor i.

Following Jeanblanc and Rutkowski (2000), for $t \geq t_0$ we define the conditional intensity process of the time to default T_i given $\mathbb{F}_i$ as follows.

Definition 3.1 (Conditional intensity process). *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 stochastic process $M_i = (M_{i,t})_{t \geq t_0}$ defined by*

$$M_{i,t} = X_{i,t} - \int_{t_0}^{t \wedge T_i} \lambda_{i,u}^{\mathbb{F}_i} \, du \tag{3}$$

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

This is the martingale characterization of the intensity process, sometimes called the pre-intensity process (Duffie and Gàrlenau 1999). We suppose that for all $i = 1, \ldots, n$ the conditional intensity process $\lambda_i^{\mathbb{F}_i}$ exists. Appendix A gives more details on the technical conditions we need for the existence. By application of the martingale property, of the Lebesque averaging theorem and the Lebesque dominated-convergence theorem successively, we obtain

$$\lim_{s \searrow 0} \frac{\mathbb{P}\big[T_i \in (t, t+s] \,|\, \mathcal{G}_{i,t}\big]}{s} = 1_{\{T_i > t\}} \lambda_{i,t}^{\mathbb{F}_i} . \tag{4}$$

This last equation suggests that the conditional intensity process at time t, $\lambda_{i,t}^{\mathbb{F}_i}$, corresponds to the "classical" intensity on $\{T_i > t\}$. On the set $\{T_i \leq t\}$ the "classical" intensity will be just identical to zero, since the probability for a default will be zero if a default has already occurred (default is considered as an absorbing state). Moreover, this equivalence gives us a useful and practical interpretation of the conditional intensity process $\lambda_i^{\mathbb{F}_i}$. On the set $\{T_i > t\}$ and for $\Delta t \ll 1$, $\lambda_{i,t}^{\mathbb{F}_i}\Delta t$ gives the conditional probability that a default occurs during $(t, t+\Delta t]$ conditioned on all the information $\mathcal{G}_{i,t}$ available up to time t.

Some technical results used in the sequel are given in Appendix A. The Corollary of Lemma A.1 applied on $T = T_i$, $\mathcal{G}_t = \mathcal{G}_{i,t}$ and $\mathcal{F}_t = \mathcal{F}_{i,t}$ for $i = 1, \dots, n$, implies that

$$\mathbb{P}\big[T_i \in (t, t+s] \,|\, \mathcal{G}_{i,t}\big] = 1_{\{T_i > t\}} \frac{\mathbb{P}\big[T_i \in (t, t+s] \,|\, \mathcal{F}_{i,t}\big]}{\mathbb{P}\big[T_i > t \,|\, \mathcal{F}_{i,t}\big]}. \tag{5}$$

Dividing by s and taking the limit $s \searrow 0$ it follows that on $\{T_i > t\}$

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

where $f_i(t \,|\, \mathcal{F}_{i,t}) = \lim_{s \searrow 0} \frac{1}{s} \mathbb{P}\big[T \in (t, t+s] \,|\, \mathcal{F}_{i,t}\big]$ is the conditional density function of T_i and $S_i(t \,|\, \mathcal{F}_{i,t})$ is the conditional survival function. The conditional density function $f_i(. \,|\, \mathcal{F}_{i,t})$ exists if the conditional intensity process $\lambda_i^{\mathbb{F}_i}$ exists.

The goal is to model the conditional intensity. Before proceeding we note that the survival function can be written as a function of the intensity process, as follows for $t \geq t_0$

$$S_i(t \,|\, \mathcal{F}_{i,t}) = \exp\left(- \int_{t_0 \vee d_i}^{t \vee d_i} \lambda_{i,u}^{\mathbb{F}^i} du \right)$$

which implies that $\lambda_{i,t}^{\mathbb{F}_i} = \frac{f_i(t \,|\, \mathcal{F}_{i,t})}{S_i(t \,|\, \mathcal{F}_{i,t})}$ also holds on $\{T_i \leq t\}$.

We now consider the portfolio $\mathcal{P}$. Mortgage defaults are triggered by some obligors specific or by mortgage-specific or by external, environment-specific factors. We suppose that we find a set of predictors for the default event of obligor i. Mathematically, we have a multi-dimensional stochastic process $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,p})$, such that each component $Y_{i,q}$ ($q = 1, \dots, p$) represents an explaining factor for the event of default of obligor i, as for example the regional unemployment rate or the regional divorce rate. The history up to time t of the predictors gives the additional flow of information available at time t. Thus we can assume that the filtration $\mathbb{F}_i$ introduced before, is the natural filtration of $\mathbf{Y}_i$, i.e. $\mathcal{F}_{i,t} = \sigma(\mathbf{Y}_{i,s} : s \leq t)$.
We model the intensity process as follows. We suppose that we can find real valued strictly positive, measurable functions $h_{i,0}, \dots, h_{i,p}$, and a strictly positive constant $\lambda_{i,0}$ such that for $t \geq d_i$

$$\lambda_{i,t}^{\mathbb{F}_i} = \lambda_{i,0} h_{i,0}(t - d_i) \prod_{1=1}^{p} h_{i,q}(Y_{i,q}(t)). \tag{6}$$

We assume that $h_{i,0}(0) \neq 0$. This assumption prevents the conditional intensity from being identical to zero at time d_i. Without loss of generality we suppose that for all i, $h_{i,0}(0) = 1$.[2] We write

[2]We can still define $\widetilde{h}_{i,0}$ by $\widetilde{h}_{i,0} = \frac{h_{i,0}}{h_{i,0}(0)}$ and $\widetilde{\lambda}_{i,0} = \lambda_{i,0} h_{i,0}(0)$

$\lambda^{\mathbb{F}_i}_{i,t} = \lambda^{\mathbb{F}_i}_{i,t}(\widetilde{\theta}_i; \mathbf{Y}_{i,t})$ where $\widetilde{\theta}_i = (\lambda_{i,0}, h_{i,0}, h_{i,1}, \ldots, h_{i,q})$ is the parameter to be estimated. The use of the *tilde* will be clear later, when we apply the logarithm to equation (6) to obtain an additive form. We introduce additional conditions on $\widetilde{\theta}_i$ in the sequel.

Equation (6) suggests the following interpretation. The functions $h_{i,1}, \ldots, h_{i,p}$ give the sensitivity of obligor i to the predictors $Y_{i,1}, \ldots, Y_{i,q}$. $h_{i,0}$ is the so-called base-line intensity function and gives the contribution of the duration $t - d_i$ (life time of the mortgage) to the conditional intensity process; $\lambda_{i,0}$ is a "time-invariant intensity".
The model explains the following behaviour of the conditional intensity. We suppose that, at the beginning of the mortage agreement, an expected intensity $\lambda_{i,0}$ can be associated to obligor i. If the obligor's behaviour is not affected by any predictors $Y_{i,1}, \ldots, Y_{i,p}$, then we expect no contribution to the intensity process to be given by $\mathbf{Y}_i$, meaning that $h_{i,q} \equiv 1$, for $q = 1, \ldots, p$. If moreover the duration contribute nothing to the intensity process, then $h_{i,0} \equiv 1$ and the conditional intensity process is reduced to a constant $\lambda_{i,0}$ (this would imply a Poisson process with constant intensity). That the reason why we call $\lambda_{i,0}$ a "time-invariant intensity".
Usually, we observe that obligor's behaviour changes during the life of the mortgage, meaning that the probability of incurring a default increases or decreases. Some factor $Y_{i,q}$ (mortgage specific, obligor's specific or depending on the macro-economic environment) affects the ability of obligor i to pay the interest rate on the mortgage, changing stochastically the default intensity. Equation (6) says that predictors $\mathbf{Y}_i$ and the duration $t - d_i$ affect the realization of $\lambda^{\mathbb{F}_i}_i$ in a multiplicative way, as explained by the functions $h_{i,q}$, for $q = 0, \ldots, p$.

When a rating system exists, we can assume that the credit quality of an obligor is captured by his rating: credit quality in our framework means exactly the ability to pay the interest rate, as well as the ability to pay back the outstanding balance. The credit quality of an obligor will, in addition, capture the ability to react to specific scenarios, i.e. to the specific realization of predictors $\mathbf{Y}_i$. In our model this is described by the functions $h_{i,q}$, for $q = 0, \ldots, p$ as well as by the constant $\lambda_{i,0}$, which gives an expected intensity at time d_i. Therefore, if a rating system exists, the parameter $\widetilde{\theta}_i$ will be identical for each obligor with the same rating. We will use this intuition for the estimation of our model.
If no rating system exists, then equation (6) suggests a methodology to create one. We can select a family of predictors, which are significant for describing the credit quality of the obligor and can then estimate the model under the assumption that obligors are identical, i.e. all the $\widetilde{\theta}_i$ are the same. For the time d_i the realization of the conditional intensity can be computed for each obligor, given realization $\mathbf{y}_{i,d_i}$ of the predictors. Following the realized values for $\lambda^{\mathbb{F}_i}_{i,d_i}$ obligors can then be grouped and a rating system can be defined.

For the estimation of the model it is useful to consider the logarithm of the conditional intensity, which is given by an additive form, as follows

$$\eta^{\mathbb{F}_i}_{i,t}(\widetilde{\theta}_i; \mathbf{Y}_{i,t}) = \log \lambda^{\mathbb{F}_i}_{i,t}(\widetilde{\theta}_i; \mathbf{Y}_{i,t}) = \log \lambda_{i,0} + \log h_{i,0}(t - d_i) + \sum_{q=1}^{p} \log h_{i,q}(Y_{i,q}(t)). \tag{7}$$

For the sake of simplicity, we introduce the parameter $\theta_i = (\log \lambda_{i,0}, \log h_{i,0}, \ldots, h_{i,p})$ which is obtained by a log-transformation of the components of $\widetilde{\theta}_i$. In the sequel we only consider the parameter θ_i and thus write $\lambda^{\mathbb{F}_i}_{i,t}(\theta_i; \mathbf{Y}_{i,t})$ instead of $\lambda^{\mathbb{F}_i}_{i,t}(\widetilde{\theta}_i; \mathbf{Y}_{i,t})$. The same for $\eta^{\mathbb{F}_i}_{i,t}(\widetilde{\theta}_i; \mathbf{Y}_{i,t})$. Without loss of generality we suppose that $\mathbb{E}\big[\log h_q(Y_{i,q}(t)\big] = 0$ for each $t \geq t_0$, each $q = 1, \ldots p$, $i = 1, \ldots n$ (we can still define $\widetilde{h}_q$ such that $\log \widetilde{h}_q(Y_{i,q}(t)) = \log h_q(Y_{i,q}(t)) - \mathbb{E}\big[\log h_q(Y_{i,q}(t))\big]$ for $q = 1, \ldots, p$ and $\log \widetilde{h}_0(t - d_i) = \log \widetilde{h}_0(t - d_i) + \sum_{q=1}^{p} \mathbb{E}\big[\log h_q(Y_{i,q}(t))\big]$).

Moreover, we assume that obligors who default up to time t are conditionally independent given the history of the predictors up to time t. This assumption seems reasonable for the kind of portfolio we are considering in this work, i.e. a portfolio of private individuals (for companies this assumptions would not be realistic). In fact the conditional independence implies that, given a scenario through the predictors, obligor defaults occur independently, meaning that the dependence structure is fully described by the common scenario.

For estimation purposes we now want to introduce a homogeneous group of obligors. We define K classes of obligors, where the functional form of the model is identical for all obligors of a given class. In other words, we suppose that obligors can be grouped in K classes, such that the θ_i's are the same for all obligors in the same class. If two obligors belong to the same class, then the sensitivity of the conditional intensity to the explaining factors will be the same and the only reason for their exertion of two different realization of the conditional intensity, is because the predictors realization are different. As remarked above, if a rating system exists, then we can identify the K classes with the rating categories. For each class $k = 1, \dots, K$ only one parameter θ_i has to be estimate.
Moreover, we form J groups of obligors where the predictors $\mathbf{Y}_i$ and the duration $t - d_i$ are identical. If we use only macro-economic predictors, the groups represent regional groups of obligors with same duration. We can also group the obligors following other criteria, as intervals where the ratio $\frac{B_{i,t}}{L_{i,t}}$, called the loan-to-value ratio, or the interest rate $r_{i,t}$, have some specific values (Kau and Keenan 1998).

4 Estimation methodology

We fix a time horizon $\tau = t_m$, $t_m \in \mathcal{T}$. We have $[t_0, \tau] = \{t_0\} \cup \left(\bigcup_{l=0}^{m-1} (t_l, t_{l+1}]\right)$, $t_l \in \mathcal{T}$, $l = 0, \dots, m$. Our observation is made over $[0, \tau]$. Defaults times are denoted by $T_1, \dots, T_n$ as before. If a default is observed, i.e. $T_i \leq \tau$, then the contribution of the observation to the conditional likelihood function is $f_i(T_i; \theta_i \,|\, \mathcal{F}_{i,T_i})$. Otherwise the observation is right-censored by τ, or by a repayment of the mortgage at time $c_i \in [0, \tau]$ (in this case $T_i = \infty$ by definition): the contribution to the conditional likelihood function is given by $S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau})$. Since obligor defaults are conditionally independent, the conditional likelihood function is thus given by

$$\begin{aligned} \mathcal{L} &= \prod_{T_i \leq \tau} f_i(T_i; \theta_i \,|\, \mathcal{F}_{i,T_i}) \prod_{T_i > \tau} S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau}) && (8) \\ &= \prod_{T_i \leq \tau} \left(\lambda^{\mathbb{F}_i}_{i,T_i}(\theta_i; \mathbf{Y}_{i,T_i}) S_i(T_i; \theta_i \,|\, \mathcal{F}_{i,T_i})\right) \prod_{T_i > \tau} S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau}). && (9) \end{aligned}$$

Our observations are both right- and left-censored, in the sense that the exact time T_i, where the inability to pay the interest appears, is unobserved. We only know that $T_i \in (t_l, t_{l+1}]$ for some l, meaning that a payment has been made at time t_l but not at time t_{l+1} for some l. We rewrite the likelihood function taking this characteristic of our observation into consideration. If obligor i defaults during the time interval $(t_l, t_{l+1}]$ then the contribution to the likelihood function will be $S_i(t_l; \theta_i \,|\, \mathcal{F}_{i,t_l}) - S_i(t_{l+1}; \theta_i \,|\, \mathcal{F}_{i,t_{l+1}})$, meaning that he survives time t_l but not time t_{l+1}. If obligor i survives the time horizon τ or a repayment occurs at time $c_i \in [t_0, \tau]$, then the observation of the time-to-default is right-censored and the contribution to the conditional likelihood function will be $S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau})$. The conditional likelihood is given as follows

$$\mathcal{L} = \prod_{l=0}^{m-1} \prod_{T_i \in (t_l, t_{l+1}]} \left(S_i(t_l; \theta_i \,|\, \mathcal{F}_{i,t_l}) - S_i(t_{l+1}; \theta_i \,|\, \mathcal{F}_{i,t_{l+1}})\right) \prod_{T_i > \tau} S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau}). \quad (10)$$

This likelihood function is very hard to handle analytically since each factor is given by an exponential function or by a difference of exponential functions of the integral of the conditional intensity process. Moreover, we lose some information about the number of obligors surviving time t_l for each $t_l \in \mathcal{T}$. This kind of information would be very useful for estimating the "sensitivity" of obligors' behaviour with respect to some specific realized predictors. In fact, not only defaults but also survivals will provide information on the behaviour of obligors under specific conditions or specific realization of the predictors. If we just consider obligors who survive time τ, or repay the mortgages before τ, through $S_i(c_i \wedge \tau; \theta_i \,|\, \mathcal{F}_{i,c_i \wedge \tau})$, then information about single periods are lost, which should be avoided since we already have a dearth of information.

Looking at equation (6) we observe that the intensity at time t depends on the history $\mathcal{F}_{i,t}$ of the predictors only through the last realized value $\mathbf{Y}_{i,t}$. It follows that we can simply consider each time period $(t_l, t_{l+1}]$, $l = 0, \dots, m-1$ separately. Moreover, obligor defaults up to time t are assumed to be conditionally independent given $\mathcal{F}_{i,t}$, and thus we suggest that our consideration of the conditional likelihood function can be restricted to each rating class $k = 1, \dots, K$ separately. In fact the conditional independence implies that one rating class will not contribute to the maximum-likelihood estimation of the parameters for another rating class. Note that the θ_i's are the same for all obligors in the given class under our assumptions; for the sake of simplicity, in the sequel we consider one class $k = 1, \dots, K$ and thus drop the index i: we write $\theta = (\log \lambda_0, \log h_0, \log h_1, \dots, \log h_p)$ and $\lambda^{\mathbb{F}_i}$ instead of $\lambda_i^{\mathbb{F}_i}$.

We divide the obligors in J groups as introduced in the previous section. Each group is homogeneous with respect to the predictors, as well as the duration $t - d_i$. By $D_{j,l}$ we denote the number of mortgages in group j defaulting during $(t_l, t_{l+1}]$ and by $O_{j,l}$ the number of mortgages outstanding in group j (at risk) during this period ($j = 1, \dots, J$). The probability that obligor i in group j defaults during $(t_l, t_{l+1}]$, given that he survive time t_l and given the predictors up to time t_l, is given by equation (5) for $t = t_l$, $s = t_{l+1} - t_l$. On the set $\{T_i > t_l\}$ we have

$$\begin{aligned} \mathbb{P}\big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{G}_{i,t_l}\big] &= \frac{\mathbb{P}\big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{F}_{j,t_l}\big]}{\mathbb{P}\big[T_i > t_l \,|\, \mathcal{F}_{j,t_l}\big]} \\ &= \frac{S(t_l \,|\, \mathcal{F}_{j,t_l}) - \mathbb{P}\big[T_i > t_{l+1} \,|\, \mathcal{F}_{j,t_l}\big]}{S(t_l \,|\, \mathcal{F}_{j,t_l})}. \end{aligned}$$

Under our assumptions, $\mathbb{F}_i = \mathbb{F}_j$ for obligor i in group j. Using

$$\mathbb{P}\big[T_i > t_{l+1} \,|\, \mathcal{F}_{j,t_l}\big] = S(t_l \,|\, \mathcal{F}_{j,t_l})\, \mathbb{E}\Big[\exp\left(-\int_{t_l \vee d_j}^{t_{l+1} \vee d_j} \lambda_u^{\mathbb{F}_j}(\theta; \mathbf{Y}_{j,u}) du\right) \,|\, \mathcal{F}_{j,t_l}\Big]$$

we obtain that, on $\{T_i > t_l\}$, for mortgage i in group j, it follows

$$\mathbb{P}\big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{G}_{i,t_l}\big] = 1 - \mathbb{E}\Big[\exp\left(-\int_{t_l}^{t_{l+1}} \lambda_u^{\mathbb{F}_j}(\theta; \mathbf{Y}_{j,u}) du\right) \,|\, \mathcal{F}_{j,t_l}\Big]. \tag{11}$$

The typology of our data set makes it necessary to discretize the intensity process, as well the process describing the set of predictors. With $\mathbf{y}_j = (\mathbf{y}_{j,t})_{t \in [t_0, \tau]}$ we denote the realized vector of predictors for obligors in group $j = 1, \dots, J$. We suppose that the predictors are constant on each interval $[t_l, t_{l+1})$ and that the function h_0 is piecewise constant on $[t_l, t_{l+1})$. We have that, on $\{T_i > t\}$, the conditional probability is given by

$$\mathbb{P}\big[T_i \in (t_l, t_{l+1}] \,|\, \mathcal{G}_{i,t_l}\big] = 1 - \exp\left(-\frac{1}{t_{l+1} - t_l} \lambda_{t_l}^{\mathbb{F}_j}(\theta; \mathbf{y}_{j,t_l})\right) = u_{j,l}(\theta). \tag{12}$$

Therefore, $u_{j,l}(\theta)$ denotes the conditional probability that a default occurs during $(t_l, t_{l+1}]$ given that the mortgage survives time t_l and given the realization at time t_l of the predictors for obligors in group j. The number of defaults in a group j is thus binomial distributed with conditional probability $u_{j,l}(\theta)$. The contribution of period $(t_l, t_{l+1}]$ to the conditional discretized likelihood function is thus given by

$$\mathcal{L}_l(\theta) = \prod_{j=1}^{J} \binom{O_{j,l}}{D_{j,l}} u_{j,l}(\theta)^{D_{j,l}} \left(1 - u_{j,l}(\theta)\right)^{O_{j,l}}.$$

The total conditional discretized likelihood function follows directly using the independence between successive periods, we have:

$$\mathcal{L}(\theta) = \prod_{l=0}^{m-1} \mathcal{L}_l(\theta) = \prod_{l=0}^{m-1} \prod_{j=1}^{J} \binom{O_{j,l}}{D_{j,l}} u_{j,l}(\theta)^{D_{j,l}} \left(1 - u_{j,l}(\theta)\right)^{O_{j,l}}. \tag{13}$$

The conditional total likelihood function given by equation (13) is simply the likelihood function of mJ independent observations, which are all binomial but not identically distributed, since the probability $u_{j,l}(\theta)$ changes, as well as the parameter $O_{j,l}$ of the distribution. Moreover if we consider the definition of $u_{j,l}(\theta)$ we can express $u_{j,l}(\theta)$ as a function of the additive form $\eta_{t_l}^{\mathbb{F}_j}(\theta; \mathbf{y}_{j,t_l})$ given by equation (7), i.e.

$$\eta_{t_l}^{\mathbb{F}_j}(\theta; \mathbf{y}_{j,t_l}) = \log \lambda_0 + \log h_0(t_l - d_j) + \sum_{q=1}^{p} \log h_q(y_q(t_l)).$$

For this purpose we define the function $G : (0, 1) \to \mathbb{R}, x \mapsto \log(-\log(1 - x))$, the so called complementary log log-function; we have:

$$G(u_{j,l}(\theta)) = \eta_{t_l}^{\mathbb{F}_j}(\theta; \mathbf{y}_{j,t_l}) = \log \lambda_0 + \log h_0(t_l - d_j) + \sum_{q=1}^{p} \log h_q(y_q(t_l)). \tag{14}$$

In this last equation the convention introduced before, namely that $t_{l+1} - t_l = 1$, for $l = 0, \ldots, m-1$ is used explicitly.

4.1 Reformulation of the model as GAM

Combining equations (13) and (14) we suggest a reformulation of the model for the number of defaults as a generalized additive model (GAM). An overview of GAM and some technical results are given in Appendix B. More details can be found in Hastie and Tibshirani (1990).

For $l = 0, \ldots, m-1$ and $j = 1, \ldots, J$ we define the conditionally independent random variables $v_{j,l}$ as the ratio $\frac{D_{j,l}}{O_{j,l}}$. We suppose that $v_{j,l} \sim \frac{1}{O_{j,l}}\text{binomial}(O_{j,l}, u_{j,l}(\theta))$ conditionally on $u_{j,l}(\theta)$, where $u_{j,l}(\theta)$ is defined by equation (12). The observation $(v_{j,l})_{j,l}$ has the conditional likelihood function $\mathcal{L}(\theta)$ as given by equation (13). Moreover, $\mathbb{E}\left[v_{j,l} \,|\, \mathbf{Y}_{j,t_l} = \mathbf{y}_{j,t_l}\right] = u_{j,l}(\theta)$ and $u_{j,l}(\theta)$ is related to an additive form as shown above (equation (14)).

Let $\alpha = \log \lambda_0$, $f_q = \log h_q$ for $q = 0, \ldots, p$. Resuming, we obtain the following problem. Estimates $\theta = (\alpha, f_0, \ldots, f_p)$, given conditionally independent observations $v_{j,l}$ and observed predictors $\mathbf{y}_{j,t_l}$ for $l = 0, \ldots, m-1$, $j = 1, \ldots, J$ such that

$$v_{j,l} \sim \frac{1}{O_{j,l}}\text{binomial}(O_{j,l}, u_{j,l}(\theta)), \tag{15}$$

$$G(u_{j,l}(\theta)) = \alpha + f_0(t_l - d_j) + \sum_{q=1}^{p} f_q(y_{j,t_l}). \tag{16}$$

For the sake of simplicity, we drop the index j, l and write $\widetilde{\mathbf{v}} = (v_{1,1}, v_{1,2}, \ldots, v_{J,m-1}, v_{J,m})' \in \mathbb{R}^M$ and $\widetilde{\mathbf{y}}_{(l-1)\,J+j} = (t_l - d_j, \mathbf{y}'_{j,l})' \in \mathbb{R}^{p+1}$, for $l = 1, \ldots, m$, $j = 1, \ldots, J$, such that $(\widetilde{v}_i, \widetilde{\mathbf{y}}_i)$ represents a pair of observations $(i = 1, \ldots, M)$.
By reformulating of the problem as a GAM, we see that it is possible to maximize the likelihood function $\mathcal{L}(\theta)$ and to find the maximum likelihood estimation of θ using the technique developed for estimating a GAM, i.e. using a local scoring algorithm with backfitting (see Appendix B). This procedure is implemented in standard software packages, as, for example, S-Plus. Details on the S-Plus implementation can be found in Chambers and Hastie (1992).

4.2 Model selection

We consider the GAM introduced in the previous section (equations (15) and (16)). The main goal is to identify the functional form of the model by a nonparametric technique. The fitted function will serve as a *diagnostic tool* to inspire parsimonious reparametrizations of some variables, using log transformation, inverse transformation or polynomials. An accurate deviance test can be performed on the transformed model, but it is nevertheless very useful to find a model selection procedure within the GAM framework.
We have to estimate the parameter α as well as the functions $f_0, \ldots, f_p$. For this purpose the local scoring algorithm (see Appendix B) approximates in the backfitting loop each function f_q, $q = 0, \ldots, p$ by using a weighted smoothing operator $\mathbf{S}_q^{\lambda_q}$, where λ_q denotes the smoothing factor (see Appendix B and C for technical details) for the qth term. In this work we use smoothing splines, which are also introduced in Appendix C. Other classes of smooth operators are available, such as weighted locally regression, Gaussian kernel regression or B-splines.

Restricting the choice of the weighted smoothing operators to the class of weighted smoothing splines $\mathcal{C}_{\text{spline}} = \{\widehat{f}_\lambda \mid \widehat{f}_\lambda$ smoothing spline with smoothing factor $\lambda \geq 0\}$ we can address the question of which term has to be included in the model and also how smooth it has to be.
We follow the model selection technique proposed by Hastie and Tibshirani (1990, Section 9.4.1). A discussion on smoothing factor selection for smoothing splines is given in Appendix C, as well as the definition of effective degrees of freedom for a smoothing spline. Here, we extend the smoothing factor selection procedure to the GAM.

We refer to the local scoring algorithm introduced in Appendix B. By $\mathbf{S}_q^{\lambda_q}$ we denote the smoothing operator applied to the qth term ($q = 0, \ldots, p$) and by $\mathbf{W}$ the weighted least square operator for the constant α, at *convergence* of the local scoring algorithm (the last step). Since we are dealing with smoothing spline estimations (which are linear), we need only to consider $\mathbf{S}_q^{\lambda_q}$ as the smooting matrix for the qth term. Let $\widehat{f}_q$ (we drop the index λ for the moment, see Appendix C) be the smoothing spline estimation of f_q and $\widehat{\mathbf{f}}_q = (\widehat{f}_q(\widetilde{y}_{1,q}), \ldots, \widehat{f}_q(\widetilde{y}_{M,q}))'$. We have

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\mathbf{f}}_0 \\ \vdots \\ \widehat{\mathbf{f}}_p \end{pmatrix} = \begin{pmatrix} \mathbf{W}(\mathbf{z} - \sum_{q=0}^p \widehat{\mathbf{f}}_q) \\ \mathbf{S}_0^{\lambda_0}(\mathbf{z} - \sum_{q\neq 0} \widehat{\mathbf{f}}_q) \\ \vdots \\ \mathbf{S}_p^{\lambda_p}(\mathbf{z} - \sum_{q\neq p} \widehat{\mathbf{f}}_q) \end{pmatrix}. \tag{17}$$

At convergence, we can thus write the estimation of the additive form $\widehat{\eta} = (\widehat{\eta}(\widetilde{\mathbf{y}}_1), \ldots, \widehat{\eta}(\widetilde{\mathbf{y}}_M))'$, $\widehat{\eta}(\widetilde{\mathbf{y}}_i) = \widehat{\alpha} + \sum_{q=0}^p \widehat{\mathbf{f}}_{q,i}$ $(i = 1, \ldots, M)$, as $\widehat{\eta} = \mathbf{R}^\lambda \mathbf{z}$, where $\mathbf{R}^\lambda$ is the weighted additive-fit operator and $\mathbf{z}$ is the adjusted variable by the final step of the local scoring algorithm. $\lambda = (\lambda_0, \ldots, \lambda_p)'$ is the vector of smoothing factors. Analogously, as with smoothing splines, the total number of effective degrees of freedom of the model is defined by

$$df_\lambda = \operatorname{tr}(\mathbf{R}^\lambda). \tag{18}$$

The number of degrees of freedom for the error term is defined by

$$df_\lambda^{\text{err}} = M - \text{tr}\left[2\mathbf{R}^\lambda - (\mathbf{R}^\lambda)'\mathbf{A}\mathbf{R}^\lambda\mathbf{A}^{-1}\right], \tag{19}$$

where $\mathbf{A}$ is the weight given by the expected information matrix in the last step (at convergence) of the local scoring algorithm (see Appendix B for more details).
In the backfitting loop of the local scoring algorithm each function f_q, $q = 0, \ldots, p$ is fitted by a smoothing spline. As we discuss in Appendix C the number of effective degrees of freedom for the spline estimation of the qth term has to be selected. In Appendix C we show that the number of effective degrees of freedom, defined by the trace of the smoothing matrix, uniquely determines the smoothing factor λ, since the relationship between smoothing factor and the effective number of degrees of freedom is strictly monotone.

In the GAM framework we define the effective number of degrees of freedom for the spline estimation of the qth term by

$$df_q^{\lambda_q} = \text{tr}(\mathbf{S}_q^{\lambda_q}) - 1. \tag{20}$$

We subtract the 1 since in the GAM the constant is isolated.

We define the set $\Theta_q \subset \mathcal{C}_{\text{spline}}$ giving the different value for the effective number of degrees of freedom $df_q^{\lambda_q}$ which are allowed for the spline estimation of f_q ($q = 0, \ldots, p$): Θ_q contains alternatives of increasing complexity for the choice of the number of degrees of freedom for the q-th term. One can, for example, give the alternatives $df_q^{\lambda_q} = 0$, $df_q^{\lambda_q} = 1$, $df_q^{\lambda_q} = 3$ and $df_q^{\lambda_q} = 5$ for the smoothing spline estimation of the qth term: degree of freedom $df_q^{\lambda_q} = 0$ for example means that the q-th term is deleted from the model, $df_q^{\lambda_q} = 1$ means that f_q is forced to be linear (the least square line usually has two degrees of freedom, but for the GAM we subtract 1 since the constant is isolated). We assume that, for each q, the qth term can be deleted from the model, i.e. $0 \in \Theta_q$ for $q = 0, \ldots, p$. We set the *model space* $\Theta = \mathbb{R}^+ \times \left(\bigotimes_{q=0}^p \Theta_q\right)$ and we define the *null model* or initial model by $\widehat{\theta}^0 = (\widehat{\alpha}^0, 0, \ldots, 0) \in \mathcal{M}$. This is to say that the conditional intensity process at the beginning is simple assumed to be constant and $\widehat{\alpha}^0$ is the least square estimation of α. Before we introduce the model selection technique, we have to define a criterion for comparing the fit of different models. Following Hastie and Tibshirani (1990) we use the χ^2-test and the *AIC* criterion.

Let $\widetilde{\mathbf{v}} \in \mathbb{R}^M$ and $\widetilde{\mathbf{y}}_i \in \mathbb{R}^{p+1}$, $i = 1, \ldots, M$ as defined in the previous section. $\widehat{\theta} \in \Theta$ denotes the estimation of θ obtained by the local scoring procedure, where the smoothing spline estimation of the qth term is an element of Θ_q for each $q = 0, \ldots, p$.
$\widehat{\mu} \in \mathbb{R}^M$ is defined by $\widehat{\mu}_{(l-1)J+j} = G^{-1}(\eta_{j,l}^{\mathbb{F}_j}(\widehat{\theta}; \mathbf{y}_{j,t_l}))$, $l = 1, \ldots, m$, $j = 1, \ldots, J$, where $\eta_{t_l}^{\mathbb{F}_j}(\widehat{\theta}; \mathbf{y}_{j,t_l})$ is given by equation (7) for $\widehat{\theta}$ and G is the link function as defined in the previous section. With $l(\theta) = \log \mathcal{L}(\theta)$ we denote the log-likelihood function. The *likelihood-ratio statistic* for $\widehat{\theta}$, also called deviance, is defined by

$$D(\theta; \widetilde{\mathbf{v}}) = 2\left\{l(\theta_{\max}; \widetilde{\mathbf{v}}) - l(\widehat{\theta}; \widetilde{\mathbf{v}})\right\} \tag{21}$$

where $\theta_{\max} = \arg\max_\theta l(\theta; \widehat{\mathbf{v}})$. Considering the product form of the likelihood function $\mathcal{L}(\theta)$ (see equation (13) and Appendix B for the general case where the observations follows an exponential family density), one can rewrite the deviance as

$$D(\theta; \widetilde{\mathbf{v}}) = \sum_{i=1}^M D(\widehat{\mu}_i; \widetilde{v}_i), \tag{22}$$

where $\widehat{\mu}_i$ is defined for $i = 1, \dots, M$ as above. According to Hastie and Tibshirani (1990) the χ^2-distribution is a quite useful approximation of the distribution of the deviance, although the deviance is not χ^2-distributed. In fact bootstrap simulations suggest that this approximation is informative under some assumptions (see McCullagh and Nelder (1995), Chapter 4.4.3 for a discussion on generalized linear model with binomial distribution). The deviance function is more directly useful for comparing two nested models.
Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be two models in Θ, which differ only by a single term (not the constant): we suppose that $\widehat{\theta}_1$ is the smaller model, meaning that for one specific $q = 0, \dots, p$ the function f_q is simply forced to be identical to 0 in the model $\widehat{\theta}_1$ but not in the model $\widehat{\theta}_2$ ($\widehat{\theta}_1$ and $\widehat{\theta}_2$ are called *nested models*). We want to test the null hypothesis $H_0 : \theta = \widehat{\theta}_1$, with respect to the alternative $H_A : \theta = \widehat{\theta}_2$. For the binomial distribution (for which the dispersion parameter ϕ is equal to 1, see Appendix B), the asymptotic distribution of the difference

$$D(\widehat{\theta}_2; \widehat{\theta}_1, \widetilde{\mathbf{v}}) = D(\widehat{\theta}_1; \widetilde{\mathbf{v}}) - D(\widehat{\theta}_2; \widetilde{\mathbf{v}}) \tag{23}$$

is approximated by a χ^2-distribution with $df_q^{\text{err}} = df^{\text{err}}(\widehat{\theta}_1) - df^{\text{err}}(\widehat{\theta}_2)$ degrees of freedom, under the null hypothesis H_0 (i.e. the effect f_q of the q-the variable is absent). $df^{\text{err}}(\widehat{\theta}_1)$, $df^{\text{err}}(\widehat{\theta}_2)$, are the numbers of degrees of freedom for the error in model 1 and model 2, respectively (see equation (19)). The exact values for $df^{\text{err}}(\widehat{\theta}_i)$ $(i = 1, 2)$ are quite difficult to obtain. A useful approximation for the difference df_q^{err} is given by $df_q^{\lambda_q} = \text{tr}(\mathbf{S}_q^{\lambda_q}) - 1$, where $\mathbf{S}_q^{\lambda_q}$, as above, is the matrix representing the linear smoothing operator (smoothing spline) for f_q (see Hastie and Tibshirani (1990, Chapters 5 and 6) and Hastie and Tibshirani (1987)). The same idea can be applied to test linearity in one term - say the q-th term - or to test an increase of complexity in exactly one term. For additive models, Cantoni and Hastie (2000) have proposed exact statistics for testing the null hypothesis H_0 that the number of degrees of freedom for the q-th term in the model is equal to df_q, against the alternative H_A that it is bigger, while the other terms are forced to be identical in both models.

In addition to the analysis of deviance presented above, we use another criterion for comparing two models. Let V_i be a random variable with the same distribution as the realization $\widetilde{v}_i$, $i = 1, \dots, M$ (it is a binomial distribution, see equation (15)). The prediction error (PE) for the model $\widehat{\theta} \in \Theta$ is given by

$$PE = \frac{1}{M} \mathbb{E}\big[\sum_{i=1}^{M} D(\widehat{\mu}_i; V_i)\big] \tag{24}$$

where $\widehat{\mu}_i$ $(i = 1, \dots, M)$ is defined above. We introduce the AIC statistic

$$AIC = \frac{1}{M} D(\widehat{\theta}; \widetilde{\mathbf{v}}) + 2 df_{\widehat{\theta}} \frac{1}{M}, \tag{25}$$

where $df_{\widehat{\theta}}$ is the effective number of degrees of freedom for the model $\widehat{\theta}$, as defined by equation (18), and there $\mathbf{R}^\lambda$ denotes the weighted smoothing operator corresponding to the model $\widehat{\theta}$. The AIC statistic is asymptotically unbiased for the prediction error PE (Ishiguro, Kitagawa, and Sakamoto 1986).
The smoothing factor selection also involves the AIC statistics. As for the smoothing factor selection technique implemented by smoothing spline (see Appendix C), we consider the cross-validated deviance

$$CV = \frac{1}{M} \sum_{i=1}^{M} D(\widehat{\mu}_i^{-i}, \widetilde{v}_i), \tag{26}$$

where $\widehat{\mu}_i^{-i}$ denotes the fitted value for μ_i, leaving the ith data point out of the sample. Minimizing CV with respect to smoothing factors $\lambda = (\lambda_0, \ldots, \lambda_p)'$ is computationally expensive, since each trial λ needs the calculation of $D(\widehat{\mu}_i^{-i}, \widetilde{v}_i)$, for $i = 1, \ldots, M$, and thus the local scoring algorithm has to be run M times. We extend the consideration done in Appendix C for smoothing splines and define the generalized cross-validation GCV by

$$GCV = \frac{1}{M} \frac{\sum_{i=1}^{M} D(\widehat{\mu}_i; \widetilde{v}_i)}{(1 - \frac{1}{M} df_{\widehat{\theta}})^2}. \tag{27}$$

Using that for $x \approx 0$ we have $(1-x)^{-2} \approx 1 + 2x$, and that $\frac{1}{M} df_{\widehat{\theta}}$ is small for M big enough, we obtain

$$GCV \approx \frac{1}{M} \sum_{i=1}^{M} D(\widehat{\mu}_i; \widetilde{v}_i) + 2df_{\widehat{\theta}} \frac{1}{M} \frac{1}{M} \sum_{i=1}^{M} D(\widehat{\mu}_i; \widetilde{v}_i) \tag{28}$$

and thus GCV and AIC are equal to first order.

Model selection technique

To compare the fit of two models we use the AIC statistic defined by equation (25). We start with the null model $\widehat{\theta}^0$. The AIC statistic for the null-model is computed using equation (25). The first step of the model selection is the following. For $q = 0, \ldots, p$, the model $\widehat{\theta}^{1,q} \in \Theta$ is obtained by increasing the complexity of the q-th term in $\widehat{\theta}^0$ one step forward in Θ_q, while the other terms $q' \neq q$ are kept fixed equal to zero. The estimation of the model $\widehat{\theta}^{1,q}$ is given by the local scoring algorithm, and $AIC_{\widehat{\theta}^{1,q}}$ is computed. Between the $p+1$ models generated by this first step, the model $\widehat{\theta}^1 \in \Theta$, which is defined by

$$\widehat{\theta}^1 = \arg\min \left\{ AIC_{\widehat{\theta}^{1,q}} \mid \widehat{\theta}^{1,q} : q = 0, \ldots, p \right\} \tag{29}$$

is selected if and only if $AIC_{\widehat{\theta}^1} < AIC_{\widehat{\theta}^0}$. In the case where $AIC_{\widehat{\theta}^1} \geq AIC_{\widehat{\theta}^0}$, the model $AIC_{\widehat{\theta}^0}$ is the best model (with respect to the selection criterion) and the model selection procedure stops. The general steps from one model $\widehat{\theta}^r \in \Theta$, $r = 1, \ldots$ to the next model $\widehat{\theta}^{r+1} \in \Theta$ is the following.

(i) For $q = 0, 1, \ldots, p$ define the model $\widehat{\theta}_+^{r+1,q}$ by increasing the complexity of the qth term in $\widehat{\theta}^r$ one step forward in Θ_q , while the other $q' \neq q$ are kept fix.

(ii) For $q = 0, 1, \ldots, p$, if the qth term in not identical to 0 in $\widehat{\theta}^r$, define the model $\widehat{\theta}_-^{r+1,q}$ by decreasing the complexity of the qth term in $\widehat{\theta}^r$ one step backward in Θ_q, while the other $q' \neq q$ are kept fix. If the qth term in $\widehat{\theta}^r$ is identical to 0, $\widehat{\theta}_-^{r+1,q} = \widehat{\theta}^r$.

(iii) Define

$$\widehat{\theta}^{r+1} = \arg\min \left\{ AIC_{\widehat{\theta}_+^{1,q}}, AIC_{\widehat{\theta}_-^{1,q}} \mid \widehat{\theta}_+^{r+1,q}, \widehat{\theta}_-^{r+1,q} : q = 0, \ldots, p \right\}. \tag{30}$$

(iv) If $AIC_{\widehat{\theta}^{r+1}} < AIC_{\widehat{\theta}^r}$, select the model $\widehat{\theta}^{r+1}$ and continue the selection procedure, otherwise select the model $\widehat{\theta}^r$ and stop the selection procedure.

5 The data

We estimate the default model (6) by applying the GAM reformulation given in Section 4 in equations (15) and (16).

Our Swiss portfolio $\mathcal{P}$ contains 73683 mortgages held by private clients for some period between January 1994 and December 2000[3]. We consider only first-mortgages, which we selected form the available data. The portfolio is observed at the end of each quarter (March 31, June 30, September 30, December 31), as well as the default process for each obligor. Only mortgages which we were able to observe at least two subsequent points in time were selected (this means that selected mortgages were outstanding for more then three months). This restriction is necessary to identify a default which occurred during a given quarter, since one has to make sure that the defaulted mortgage was at risk at beginning of the quarter (and not already defaulted) and thus observed at end of the last quarter. We denote by Q1.94, Q2.94,..., Q3.00, Q4.00 the *end* of each quarter between 1994 and 2000. The portfolio $\mathcal{P}$ is thus observed at time Q1.94, Q2.94,..., Q3.00, Q4.00: the observations of $\mathcal{P}$ at time Q3.94 and Q3.95 fail. Using the notation introduced in the previous sections, we define t_0=Q1.94, t_1=Q2.94, t_2=Q3.94, etc. t_{28}=Q4.00. No information about the time of issue d_i of a mortgage i is available for our analysis ($i = 1, \ldots, 73683$). Instead we assume that $d_i = t_0$ for each mortgage and thus h_0 in equation (6) is independent of d_i.

In addition, our data set contains information about the mortgage product (adjusted-rate mortgage or fixed-rate mortgage) and the mortgage interest rate valid for the last quarter, for each quarter and for each obligor. Obligor belong to 26 different political and economic regions across Switzerland. Each region has its own regional government and regional laws, apart from the federal, central government, so that a difference in the political environment can be observed (see Credit Suisse Group (2000) on the real estate market characteristics in the different economic regions). In the existing rating system obligors are divided either in higher credit quality, “Rating A” or a lower credit quality, “Rating B”.

Figure 2 gives the structure of our data set. Besides the observation of defaults, the data set also contains observations of repayments, even if they cannot be identified with certainty, since a mortgage can disappear from the portfolio for reasons other than repayment.

In the sequel we enumerate the predictors $Y_{i,q}$, $q = 1, \ldots, p$, which we select for our model.

Time

The function h_0 depends on the time $t - d_i$, since the issue of the mortgage. We assume that $d_i = t_0$ for all the obligors ($i = 1, \ldots, 73683$). The time $t - d_i$ should capture seasonality in the data set (Jegadeesh and Ju 2000) and thus we also verify whether there is some contribution to the conditional intensity of that quarter of the year to which the observation corresponds, i.e. Q1,...,Q4. We define the variable Y_0 independent of obligor i by $Y_0(t_l) = k$, if t_l corresponds to the k-th quarter of the year (i.e. t_l has the form Qk.xx).

The exact age of the mortgage would have more impact on the value of the recovery rate, as shown for example by Smith, Sanchez, and Lawrence (1999).

Regional quarterly unemployment rate

One of the common reasons for private clients failing to pay the interest rates on the mortgage is unemployment. Naturally, it is not possible to define an indicator of unemployment for each obligor, so we use a regional variable, the *regional quarterly unemployment rate*, provided by the Swiss Federal Statistical Office. For each region we test lags of one up to 16 quarters (4 years). We define the quarterly unemployment rate in region w (for $w = 1, \ldots, 26$) by $Y_{w,1}$ and the lagged rate by $Y_{w,1}^{(r)}$, i.e. $Y_{w,1}^{(r)}(t) = Y_{w,1}(t - r)$ for $r = 1, \ldots, 16$ (time units is a quarter). We expect the

[3] Our data set represents a sub-portfolio provided by Credit Suisse Group.

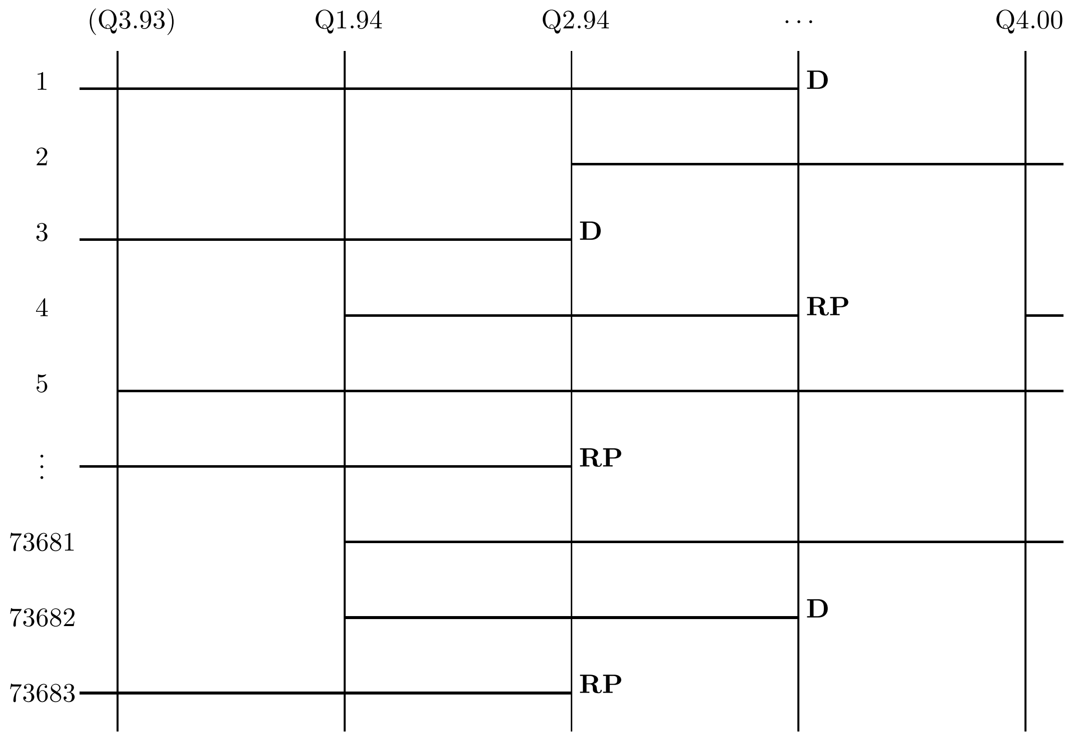

Figure 2: Data set with mortgage defaults (**D**) and mortgage repayments (**RP**).

lagged variables to be more significant for explaining the default process and that the function h_1 in the model increases, meaning that a high quarterly unemployment rate (1 up to 16 quarters before the quarter we are looking at) will imply a higher default rate.

Regional yearly divorce rate
Another quite significant cause for default seems to be divorce. As an indicator for divorce, we use the *yearly regional divorce rate* provided by the Swiss Federal Statistical Office. It gives the percentage of divorces during one year. The rate is assumed to be constant over the year. Instead of considering the rate itself we take the yearly absolute change of the rate for each region and lags of 1 to 4 years. If by $div_{w,2}(t)$ we denote the regional yearly divorce rate in region $w = 1, \ldots, 26$ at time t and we define $Y_{w,2}$ by

$$Y_{w,2}(t) = div_{w,2}(t) - div_{w,2}(t-4).$$

By $Y_{w,1}^{(r)}$ we denote the lagged predictor, i.e. $Y_{w,1}^{(r)}(t) = Y_{w,1}(t-4r)$ for $r = 1, \ldots, 4$ (time unit is a quarter). We also expect the lagged rate for divorce to be more significant. Since changes of the divorce rates over the last 10 years are not extreme, we expect the contribution of this variable to the default intensity to be small or absent.

Mortgage product
Our portfolio $\mathcal{P}$ contains adjusted-rate mortgages, as well as fixed-rate mortgages. The first type is characterized by a variable interest rate and maturity: neither is fixed in the mortgage contract.

The mortgage interest rate follows a reference market interest rate, although with a time lag, and is subject to politics. The obligor can repay the outstanding balance at each time and prepayment is free from additional transition costs. The second type of mortgage product is characterized by a fixed interest rate and maturity: both are fixed in the mortgage contract. An obligor is protected from increases in the interest rate but cannot profit from a future cut in interest rate. A prepayment of the mortgage is costly.
We code the mortgage product of obligor i ($i = 1, \ldots, 73683$) by a 0-1 variable $Y_{i,3}$ where

$$Y_{i,3}(t) = \begin{cases} 0 & \text{if adjusted rate mortgage,} \\ 1 & \text{if fixed rate mortgage.} \end{cases} \tag{31}$$

We expect the contribution of mortgage product to the intensity process to be higher for adjusted-rate mortgages.

Mortgage interest rate
Default is defined an an obligor being unable to meet the interest payment on his mortgage (see the definition in Section 2). For a fixed-rate mortgage the interest rate does not change during the life of the mortgage and the obligor does not have to face with an increase of the quarterly payment. This is not true for adjusted-rate mortgages. If the quarterly mortgage interest rate increases, then the quarterly charge for the obligor increases too. $(r_{i,t})_{t \geq d_i}$ denotes the interest rate applied at time t on the outstanding balance of obligor i ($i = 1, \ldots, 73683$). We consider the relative change $x_{i,t}$ of the interest rate with respect to the previous quarter, i.e. for each quarter $t = t_l \in \mathcal{T}$ we have

$$x_{i,t} = \frac{r_{i,t} - r_{i,t-1}}{r_{i,t}}. \tag{32}$$

Since the value of $x_{i,t}$ differs from zero for only a low percentage of the mortgages in our data set, we group obligors with the criterion that $x_{i,t}$ belongs to given intervals. This means that our predictor $Y_{i,4}$ based on the interest rate is defined by levels, as follows

$$Y_{i,4}(t) = \begin{cases} 1 & \text{if } x_{i,t} < 0, \\ 2 & \text{if } x_{i,t} = 0, \\ k+1 & \text{if } x_{i,t} \in (a_{k-1}, a_k],\ k = 2, \ldots, K, \\ K+2 & \text{if } x_{i,t} > a_K, \end{cases} \tag{33}$$

where $a_1 = 0$ and $(a_k)_{k=1,\ldots,K}$ is a strictly increasing sequence. We use $K = 3$, $a_2 = 0.25$, $a_3 = 0.5$. Moreover this definition is very useful for grouping obligors if their predictors are identical.

5.1 Results

Considering our predictors Y_q (we now drop the index i), $q = 0, 1, \ldots, 4$, $Y_1^{(r)}$, $r = 1, \ldots, 16$, $Y_2^{(r)}$, $r = 1, \ldots, 4$ and their realizations, we obtain, for each rating class, $J = 260$ groups of obligors (26 regions, 2 mortgage types, 5 intervals for the interest rate), where their predictors are identical on each time interval $(t_l, t_{l+1}]$. Since we are considering 25 time intervals $(t_l, t_{l+1}]$, each one quarter long, the total number of realizations of $D_{j,l}$ and $O_{j,l}$ that we observe is 6500. For A-rated obligors, only $M = 3265$ observations of $O_{j,l}$ are different from zero; for B-rated obligors we have $M = 2713$ non-zero observations of $O_{j,l}$. We estimate the models for the intensity process of each rating class separately, as stated in the previous sections. The general model has the form

$$\begin{aligned} G(u_{j,l}(\theta)) = \alpha + f_0(t_l) + f_0^{(q)}(y_0(t_l)) \quad &+ \quad \sum_{r=1}^{16} f_1^{(r)}(y_{j,1}^{(r)}(t_l)) + \\ &+ \quad \sum_{i=1}^{4} f_2^{(r)}(y_{j,2}^{(r)}(t_l)) + f_3(y_{j,3}(t_l)) + f_4(y_{j,4}(t_l)), \end{aligned}$$

where $y_{j,q}(t_l)$ and $y_{j,q}^{(r)}(t_l)$ again denote as the realizations of the predictors at time t_l, in group j (y_0 depends only on the time t_l and thus we drop the index j).

We apply the model selection technique introduced in the previous section to our data set, starting with the null model for both rating classes. Table 1 shows the estimated values for the constant α, the residual deviance, effective number of degrees of freedom and the AIC statistic (multiplied with the number of observations M) for the two models. For both rating classes we select the

Rating	$\widehat{\alpha}^0$	Residual deviance	$df_{\widehat{\theta}^0}$	$M * AIC_{\widehat{\theta}^0}$
A	-9.379	322.023	1	324.023
B	-6.298	1380.706	1	1382.706

Table 1: Null model estimates.

variables to be included in the model following the criterion on the χ^2-test introduced in the previous section, with a confidence level of 1%. The variable $Y_1^{(r)}$ for $r = 8, 9, 11, 14, 15$, Y_3 and Y_4 has been selected for A-rated obligors; and Y_0, $Y_1^{(r)}$ for $r = 7, 8, 9, 11$, Y_3 and Y_4 for B-rated obligors.

We define the model set Θ by allowing the smoothing spline estimation of the function $f_1^{(r)}$ (for $r = 1, \ldots, 16$) to have up to 20 effective degrees of freedom, which implies a lot of variability. The choice of the maximal effective number of degrees of freedom is arbitrary: we want to ensure that the complexity of the model can be increased as long as the AIC criterion allows it, since the number of observations is big enough. The smoothing spline estimation of the function f_3 has maximal 1 degree of freedom, since the variable Y_3 takes only two values (it is a (0,1)-variable). For the spline estimation of the function f_4 we restrict the choice to maximal 3 degrees of freedom (only 5 distinct values). Finally, we allow for the smoothing spline estimation of the function f_0 to have up to 15 effective degrees of freedom and the smoothing spline estimation of $f_0^{(q)}$ to have maximal 3 degrees.

Starting with $\widehat{\theta}_0$ we apply the model selection procedure based on the AIC-statistic, which we introduced in the previous section. The following model has been selected for the higher rated class , A:

$$G(u_{j,l}(\widehat{\theta}_{\text{A}})) = \widehat{\alpha}_{\text{A}} + \widehat{f}_{1,\text{A}}^{(11)}(y_{j,1}^{(11)}(t_l)) + \left(\widehat{\beta}_{3,\text{A}}\, 1_{\{y_{j,3}(t_l)=1\}} + \widehat{\gamma}_{3,\text{A}}\right) + \widehat{f}_{4,\text{A}}(y_{j,4}(t_l)). \quad (34)$$

For the lower rated class, B, we have:

$$G(u_{j,l}(\widehat{\theta}_{\text{B}})) = \widehat{\alpha}_{\text{B}} + \widehat{f}_{0,\text{B}}^{(q)}(y_0(t_l)) + \widehat{f}_{1,\text{B}}^{(8)}(y_{j,1}^{(8)}(t_l)) + \left(\widehat{\beta}_{3,\text{B}}\, 1_{\{y_{j,3}(t_l)=1\}} + \widehat{\gamma}_{3,\text{B}}\right) + \widehat{f}_{4,\text{B}}(y_{j,4}(t_l)). \quad (35)$$

Table 2 gives the residual deviance, the effective number of degrees of freedom and the AIC-statistics for both models. The estimated parametric components of the two models are presented in Table 3, together with standard error and approximated 95%-confidence intervals, which correspond to $\pm 2\times$ times the standard error. The point estimates of α indicate that the expected probability of default for A-rated obligors is less then for B-rated obligors. This suggests that the rating system is consistent and has predictor power. Looking at the point estimates of $\widehat{\beta}_{3,\cdot}$ and $\widehat{\gamma}_{3,\cdot}$ we deduce that the default rate of A-rated and B-rated obligors is smaller - 74% and %84 respectively - than the default rate of adjusted-rate mortgages if we keep other factors constant. For A-rated obligors the point estimates are less significant. It follows that a default is more likely to occur on adjusted-rate mortgages. The same behaviour (but with smaller differences in percentage) has been found in Smith, Sanchez, and Lawrence (1999), Table 1. This can be intuitively explained by the fact that fixed-rate mortgages are not influenced by an increase of the interest rate. Moreover, defaults are more likely to be observed for adjusted-rate mortgages, since the term of a fixed-rate mortgage is usually 2-5 years in Switzerland and a default will generally not occur immediately following the issue of the mortgage contract, or during the first 12-18 months. In fact the bank tends to sell adjusted-rate mortgages to obligors with lower credit quality, since the mortgage contract is easier to set up.

Rating	Residual deviance	$df_{\widehat{\theta}}$	$M * AIC_{\widehat{\theta}}$
A	273.635	6.98	287.595
B	1082.055	9.89	1101.851

Table 2: Model estimation

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	-2.2454	0.2326
		-8.3604	-0.4682	1.1154
B	Estimate	-6.8644	-1.7893	0.8462
	Standard error	0.3636	0.1690	0.0799
	Approx. 95% CI	-6.1372	-2.1273	0.6864
		-7.5916	-1.4513	1.006

Table 3: Model parameter estimates, with standard error and approximated 95% CI.

Figures 3-7 give the smoothing spline estimates $\widehat{f}^{(11)}_{1,\mathrm{A}}$, $\widehat{f}_{4,\mathrm{A}}$, $\widehat{f}^{(q)}_{0,\mathrm{B}}$, $\widehat{f}^{(8)}_{1,\mathrm{B}}$ and $\widehat{f}_{4,\mathrm{B}}$. The model selection procedure selects a small number of degrees of freedom for each function, meaning that more complexity in the model will not significantly decrease the deviance (see the model selection procedure). The approximated 95% confidence intervals are also given. We observe that the confidence band is larger near the bounds of our observation sets. This can be explained by the small number of extreme realization of the predictors.

The quarter of the year to which the observation corresponds, contributes only to the intensity process of B-rated obligors. The second and specially the last quarter of the year present the most important contributions. These results reflect in part the common experience within the bank: in fact a high number of defaults is usually observed at the end of the year.

Unemployment affects the ability to pay the interest rate for both rating classes. The default rate of A-rated obligors increases significantly for extreme realization of the unemployment rate: a positive contribution is observed for a rate higher than 6%. For B-obligors an increase is almost

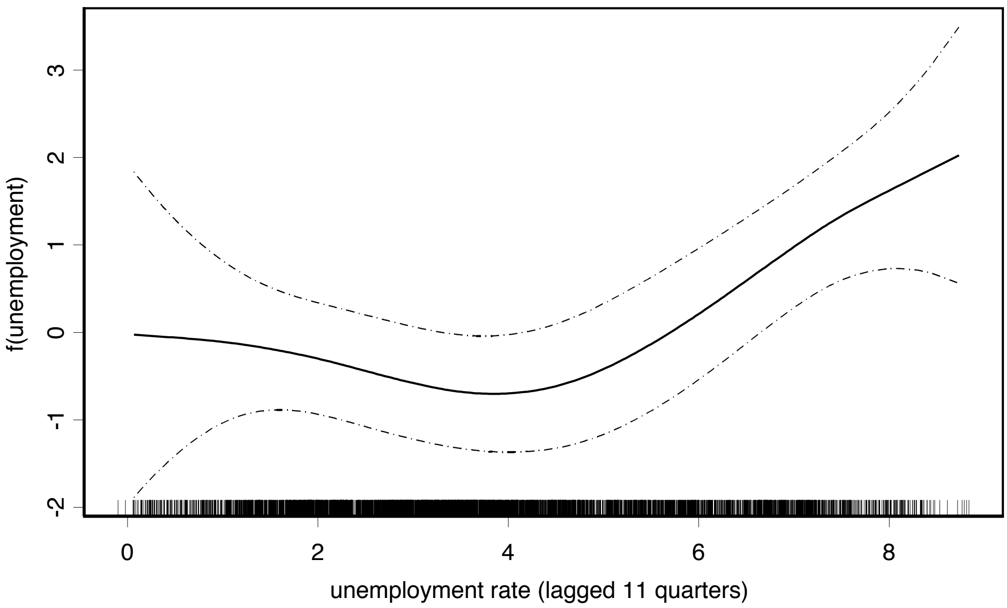

Figure 3: Spline estimation $\widehat{f}^{(11)}_{1,\mathrm{A}}$ with 1 degree of freedom. Dotted lines give the 95% approximated confidence interval.

linear and a positive contribution is already observed for a rate higher than 4%. A-rated obligors seem to be less affected by unemployment: nevertheless, the default rate changes dramatically if the unemployment rate reaches high levels. B-rated obligors are also more sensitive to lower unemployment rates. The model suggests that higher rated obligors are more likely to remain current under a "normal" scenario then B-rated obligors, but they will also be affected by stress scenarios. Moreover, we observe that the rate is lagged three quarters more for A-rated obligors than for B-rated obligors.
The percentage change of the interest rate also contributes significantly to the default intensity. For A-rated obligors the contribution becomes important if the relative change in the interest rate over one quarter is higher than 25% (level 3), which implies an increase of the interest payment of the same amount. If changes are lower, the contribution seems to be less important. This suggests that the ability to pay the interest rate is strong and will be affected only under stress scenarios. This behaviour is expected for obligors with higher credit quality. For B-rated obligors a big difference of the contribution of this predictor to the conditional intensity process is observed between level 1 (decrease) and level 2 or higher (interest rate remains unchanged or increases).

The estimated models serve to generate the distribution function of the number of defaults for each rating class and the whole portfolio. Under the assumption that defaults are conditionally independent and that, given the specific realization of the predictors, the number of defaults follows a binomial distribution, we simulate the probability distribution function for each rating class under given scenarios, as the sum of conditionally independent binomial distributions, with probability $u_{j,l}(\widehat{\theta})$, given by the inverse G^{-1} of the link function, applied to the additive forms (34) and (35).
To illustrate this point we present a simple simulation result. We consider a portfolio $\mathcal{P}'$ with 100000 obligors. We assume that obligors in $\mathcal{P}'$ are distributed among the 26 regions and the 2

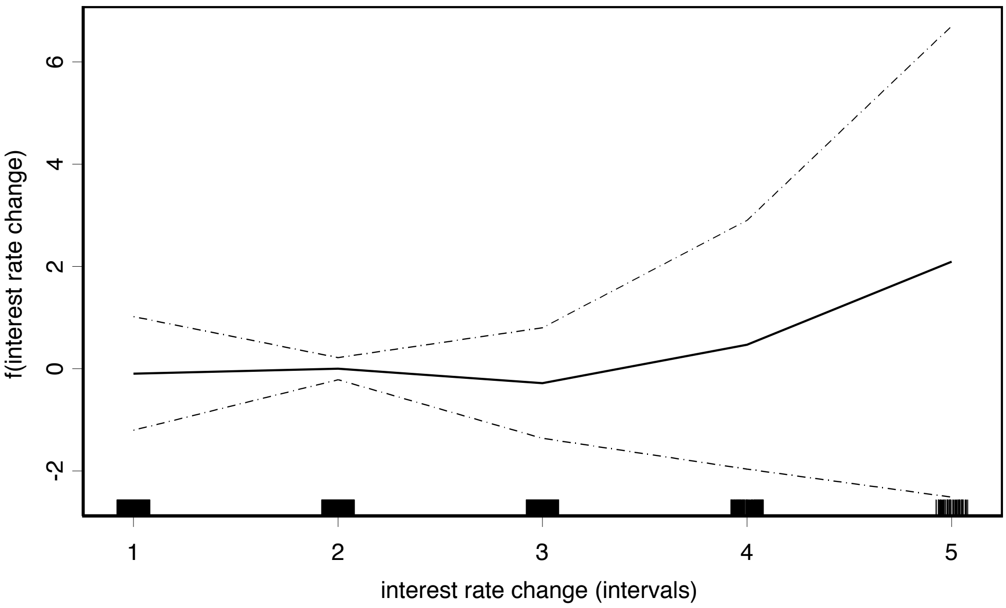

Figure 4: Spline estimation $\widehat{f}_{4,\mathrm{A}}$ with 2 degrees of freedom. Dotted lines give the approximated 95% confidence interval.

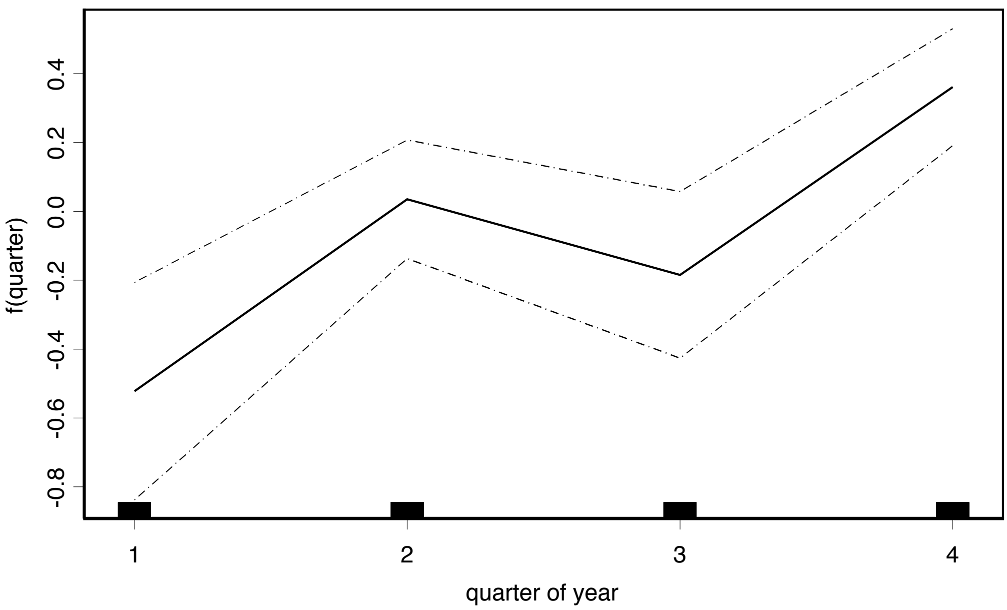

Figure 5: Spline estimation $\widehat{f}^{(q)}_{0,\mathrm{B}}$ with 2 degrees of freedom. Dotted lines give the approximated 95% confidence interval.

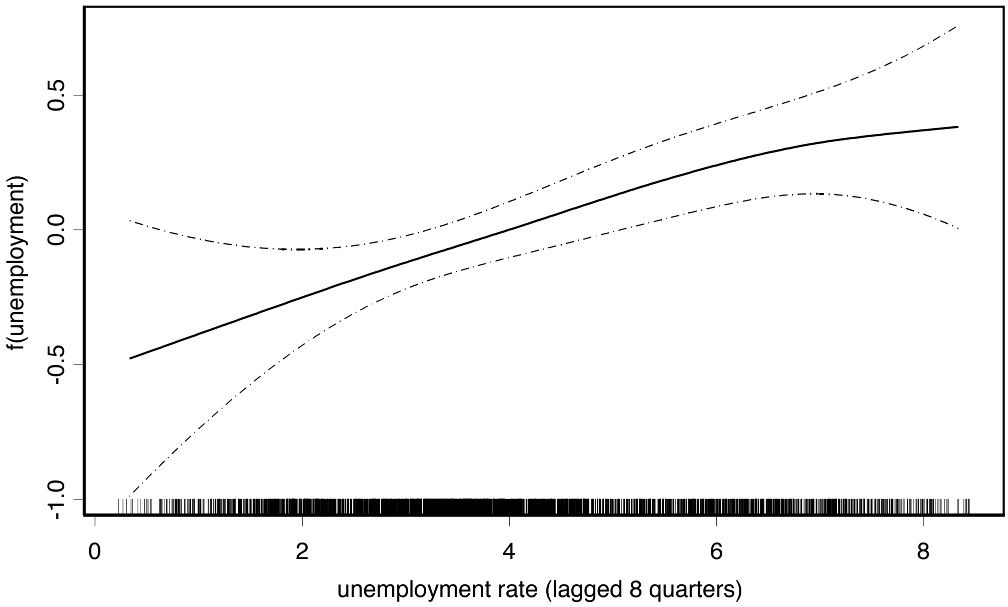

Figure 6: Spline estimation $\widehat{f}^{(8)}_{1,\mathrm{B}}$ with 1 degree of freedom. Dotted lines give the approximated 95% confidence interval.

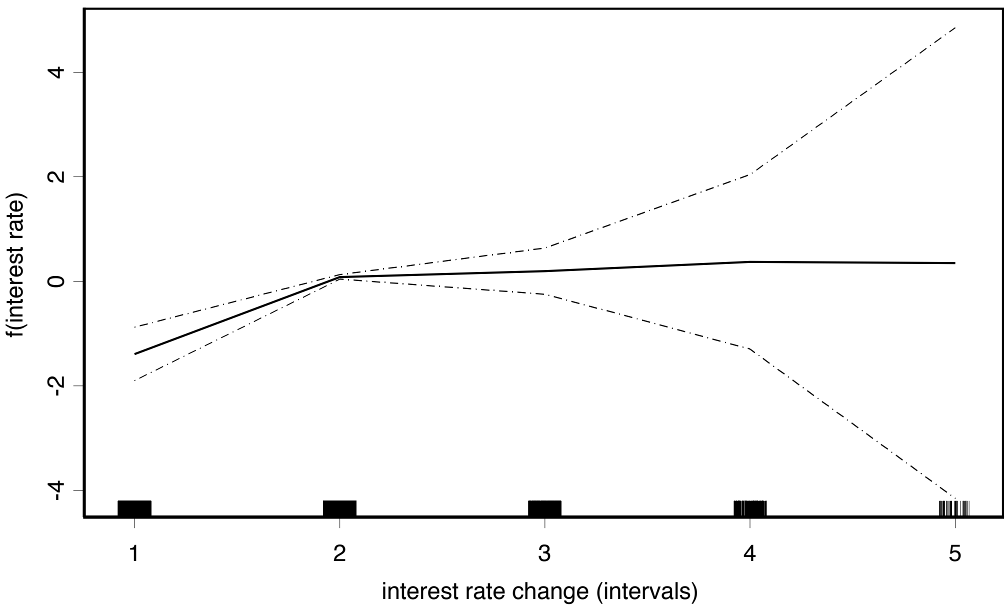

Figure 7: Spline estimation $\widehat{f}_{4,\mathrm{B}}$ with 1.9 degrees of freedom. Dotted lines give the approximated 95% confidence interval.

two mortgage products, in the same way as in portfolio $\mathcal{P}$ at the end of the last quarter 2000. We simulate the distribution function of the number of defaults in $\mathcal{P}'$ for the first quarter of 2001 under two different scenarios regarding the interest rate. The first scenario (scenario 1) supposes an interest rate increases of 0.75 percentage points for all outstanding adjusted-rate mortgages at the end of the last quarter 2000. The second scenario (scenario 2) supposes a decrease of the interest rate of 0.5 percentage points for all the outstanding adjusted-rate mortgages at the end of the last quarter 2000. The lag in the unemployment rate used in our model (8 and 11 quarters, respectively) are taken from our observation of $Y_{w,1}^{(r)}$ ($r = 8, 11,\ w = 1, \ldots, 26$) up to December 2000.

Our assumptions on the interest rate are simplistic. Other scenarios for the mortgage interest rate can be generated, following, for example, the model proposed by Burger (1998, Chapter 4). Moreover, one can naturally suppose that an increase or a decrease of the interest rate will not involve all the outstanding adjusted-rate mortgages, but only a few or in different ways. Under complex scenarios the implementation technique of our default model will, in any case, not change. Figure 8 gives the conditional distribution functions for the total number of defaults in portfolio $\mathcal{P}'$ during the first quarter 2001, under the two different scenarios for the interest rate given above.

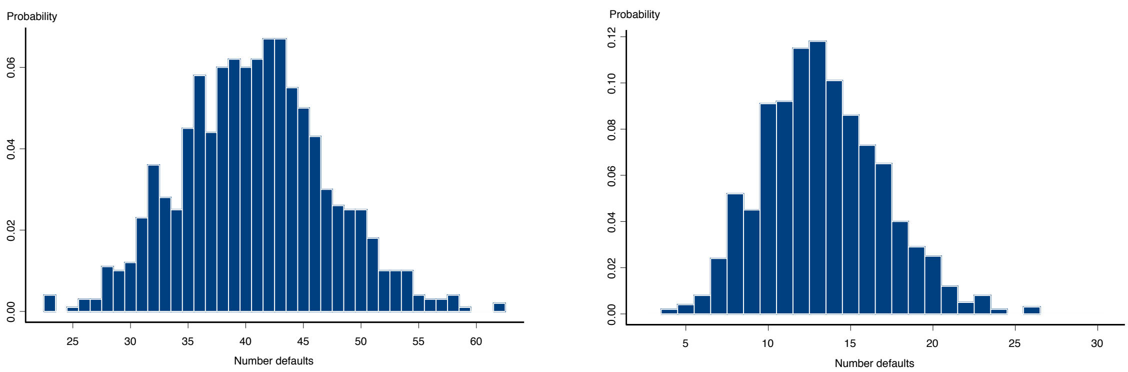

Figure 8: Conditional distribution function for the total number of defaults in portfolio $\mathcal{P}'$ under scenario 1 (left histogram) and scenario 2 (right histogram). We have computed 1000 simulations of the total number of defaults.

6 Conclusion

In this work we have presented an approach for modeling the distribution function of the number of defaults for a residential mortgage portfolio. We have considered the time-to-default and the associated conditional intensity process, given a set of predictors for the default event. The model is very flexible with respect to the choice of the predictors. We have used macro-economic variables, such as the unemployment rate, mortgage specific variables, interest rates and the mortgage product and obligor specific characteristics, such as the region where he lives. We estimate the model by using a non-parametric technique, stemming from generalized additive models and we use smoothing splines to estimate the relationship between predictors and the conditional default intensity.

Our approach offers a dynamical framework to model the default risk and therefore to better capture the sensitivity of private individuals to different scenarios. The model allows other predictors

which are relevant for the default process to be used and suggests that other information about single obligor is needed, to better capture the sensitivity of each obligor to some scenario.
The result of this paper should contribute to the problem of modeling the default risk for retail portfolios, which are often characterized by a lack of information about the credit quality of each party during the life of the loan.
Further research has to be done in the direction of modeling the recovery rate for defaulted mortgages.

A Appendix: Conditional intensity process

Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space and $T : \Omega \longrightarrow \mathbb{R}^+$ be a positive random variable. T denotes the time where a mortgage is defaulted on. We assume that $\mathbb{P}\big[T = 0\big] = 0$ and $\mathbb{P}\big[T > t\big] > 0$ for each $t \in \mathbb{R}^+$. We define the default indicator process $X = (X_t)_{\{t \geq 0\}}$ by

$$X_t = 1_{\{T \leq t\}} = \begin{cases} 1 & \text{if} \quad T \leq t \\ 0 & \text{otherwise} \end{cases} .$$

The σ-algebra $\mathbb{D} = (\mathcal{D}_t)_{\{t \geq 0\}}$ is defined by $\mathcal{D}_t := \sigma(X_s : 0 \leq s \leq t)$ and denotes the smallest σ-algebra such that X is adapted. Let $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}^+}$ be an additional flow of information available at time t and $\mathbb{G} = (\mathcal{G}_t)_{t \in \mathbb{R}^+}$ where $\mathcal{G}_t = \mathcal{D}_t \vee \mathcal{F}_t \equiv \sigma\left(\mathcal{D}_t \cup \mathcal{F}_t\right)$ is the enlarged filtration. For $t \in \mathbb{R}^+$ we define the stochastic process F by

$$F_t = \mathbb{P}\big[T \leq t \,|\, \mathcal{F}_t\big], \tag{36}$$

i.e. F gives the conditional distribution function of T given $\mathbb{F}$. F is a bounded, nonnegative $\mathbb{F}$-submartingale. Moreover the unconditional distribution function $t \mapsto \mathbb{P}\big[T \leq t\big]$ is right-continuous, which implies that F has a unique right-continuous modification (Protter 1990, Chapter 1).

Definition A.1 ($\mathbb{F}$-hazard process). *We suppose that $F_t < 1$ for all $t \in \mathbb{R}^+$. Then the* $\mathbb{F}$-hazard process $\Gamma = (\Gamma_t)_{t \in \mathbb{R}^+}$ *is defined by*

$$\Gamma_t = -\ln(1 - F_t) \tag{37}$$

for $t \in \mathbb{R}^+$. We have $S_t = 1 - F_t = \exp(-\Gamma_t)$.

In the sequel we prove that Γ_t satisfies a martingale characterization and, under some assumptions it is the only process with this property.
We introduce the filtration (it is one) $\mathbb{G}^* = (\mathcal{G}_t^*)_{\{t \geq 0\}}$ defined by

$$\mathcal{G}_t^* = \{A \in \mathcal{G} \mid \exists B \in \mathcal{F}_{i,t} : A \cap \{T > t\} = B \cap \{T > t\}\} .$$

We can easily check that $\mathcal{G}_t \subset \mathcal{G}_t^*$. It follows that on the set $\{T > t\}$ each $\mathcal{G}_t$-measurable random variable coincides with a $\mathcal{F}_t$-measurable random variable. We prove the following Lemma.

Lemma A.1. *Let Y be a $\mathcal{G}$-measurable random variable, then for each $t \in \mathbb{R}^+$*

$$\mathbb{E}\big[1_{\{T>t\}} Y \,|\, \mathcal{G}_t\big] = 1_{\{T>t\}} \frac{\mathbb{E}\big[1_{\{T>t\}} Y \,|\, \mathcal{F}_t\big]}{\mathbb{P}\big[T > t \,|\, \mathcal{F}_t\big]} = 1_{\{T>t\}} \exp(\Gamma_t) \mathbb{E}\big[1_{\{T>t\}} Y \,|\, \mathcal{F}_t\big]. \tag{38}$$

Proof. Using that, on the set $\{T > t\}$, any $\mathcal{G}_t$-measurable random variable is an $\mathcal{F}_t$-measurable random variable, we find an $\mathcal{F}_t$-measurable random variable Z such that

$$\mathbb{E}\big[1_{\{T>t\}} Y \,|\, \mathcal{G}_t\big] = 1_{\{T>t\}} \mathbb{E}\big[Y \,|\, \mathcal{G}_t\big] = 1_{\{T>t\}} Z.$$

Taking the conditional expectation with respect to $\mathcal{F}_t$ we obtain

$$\mathbb{E}\big[1_{\{T>t\}}Y\,|\,\mathcal{F}_t\big] = \mathbb{P}\big[T>t\,|\,\mathcal{F}_t\big]Z$$

and thus a formula for Z, which can be inserted in to the previous equation. □

Corollary. *Let $0 \le t \le s$, then we have*

(i) $\mathbb{P}\big[t<T\le s\,|\,\mathcal{G}_t\big] = 1_{\{T>t\}}\frac{\mathbb{P}\big[t<T\le s\,|\,\mathcal{F}_t\big]}{\mathbb{P}\big[T>t\,|\,\mathcal{F}_t\big]}$.

(ii) $\mathbb{P}\big[T>s\,|\,\mathcal{F}_t\big] = \mathbb{P}\big[T>t\,|\,\mathcal{F}_t\big]\exp(\Gamma_t-\Gamma_s)$.

Proof. (i) Apply Lemma A.1 on $Y = 1_{\{T\le s\}}$.

(ii) We have

$$\begin{aligned}
\mathbb{P}\big[T>s\,|\,\mathcal{F}_t\big] &= \mathbb{E}\big[\mathbb{E}\big[1_{\{T>s\}}1_{\{T>t\}}\,|\,\mathcal{G}_t\big]\,|\,\mathcal{F}_t\big]\\
&\overset{\text{Lemma}}{=} \mathbb{E}\big[1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[1_{\{T>t\}}1_{\{T>s\}}\,|\,\mathcal{F}_t\big]\,|\,\mathcal{F}_t\big]\\
&= \mathbb{E}\big[1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[\mathbb{E}\big[1_{\{T>s\}}\,|\,\mathcal{F}_s\big]\,|\,\mathcal{F}_t\big]\,|\,\mathcal{F}_t\big]\\
&= \mathbb{E}\big[1_{\{T>t\}}\mathbb{E}\big[\exp(\Gamma_t)\exp(-\Gamma_s)\,|\,\mathcal{F}_t\big]\,|\,\mathcal{F}_t\big]\\
&= \mathbb{P}\big[T>t\,|\,\mathcal{F}_t\big]\mathbb{E}\big[\exp(\Gamma_t-\Gamma_s)\,|\,\mathcal{F}_t\big].
\end{aligned}$$

□

Using Lemma A.1 we prove the following Proposition.

Proposition A.1. *The process $\widetilde{M}$ defined by the formula*

$$\widetilde{M}_t := 1_{\{T>t\}}\exp(\Gamma_t) = (1-X_t)\exp(\Gamma_t) = \frac{1-X_t}{1-F_t} \tag{39}$$

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

Proof. Let $s \ge t$, we have

$$\begin{aligned}
\mathbb{E}\big[1_{\{T>s\}}\exp(\Gamma_s)\,|\,\mathcal{G}_t\big] &= \mathbb{E}\big[1_{\{T>t\}}1_{\{T>s\}}\exp(\Gamma_s)\,|\,\mathcal{G}_t\big]\\
&\overset{\text{Lemma}}{=} 1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[1_{\{T>t\}}1_{\{T>s\}}\exp(\Gamma_s)\,|\,\mathcal{F}_t\big]\\
&= 1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[1_{\{T>s\}}\exp(\Gamma_s)\,|\,\mathcal{F}_t\big]\\
&= 1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[\mathbb{E}\big[1_{\{T>s\}}\exp(\Gamma_s)\,|\,\mathcal{F}_s\big]\,|\,\mathcal{F}_t\big]\\
&= 1_{\{T>t\}}\exp(\Gamma_t)\mathbb{E}\big[\exp(\Gamma_s)\underbrace{\mathbb{E}\big[1_{\{T>s\}}\,|\,\mathcal{F}_s\big]}_{=1-F_s=\exp(-\Gamma_s)}\,|\,\mathcal{F}_t\big]\\
&= 1_{\{T>t\}}\exp(\Gamma_t).
\end{aligned}$$

□

We now introduce the definition of the $\mathbb{F}$-martingale hazard process Λ.

Definition A.2 ($\mathbb{F}$-martingale hazard process). *Let $(\Omega,\mathcal{G},\mathbb{P})$ be a probability space and $T:\Omega\longrightarrow\mathbb{R}^+$ a random time. An $\mathbb{F}$-predictable right-continuous increasing process Λ is called an* $\mathbb{F}$-martingale hazard process *of the random time T if and only if the process $M=(M_t)_{t\in\mathbb{R}^+}$ defined by*

$$M_t = X_t - \Lambda_{t\wedge T} \tag{40}$$

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

The $\mathbb{F}$-martingale hazard process of the random time T exists under some technical conditions, as stated by the following theorem. Moreover, it is unique up to time T, meaning that Λ_t is uniquely defined on $\{T > t\}$ (Jeanblanc and Rutkowski 2000, Section 4).

Theorem A.1 (Existence of the $\mathbb{F}$-martingale hazard process). *Suppose that the process $(F_t)_{t\in\mathbb{R}^+}$ defined by equation (36) is an increasing continuous process. Then the $\mathbb{F}$-martingale hazard process exists and is given by*

$$\Lambda_t = \int_0^t \frac{dF_u}{1 - F_u}. \tag{41}$$

Proof. See Jeanblanc and Rutkowski (2000, Proposition 4.7) with the additional assumption that $(F_t)_{t\in\mathbb{R}^+}$ is continuous. □

Remark A.1. *If $(F_t)_{t\in R^+}$ is absolutely continuous, then $dF_t = f_t\,dt$ for a process $(f_t)_{t\in\mathbb{R}^+}$. $(f_t)_{t\in\mathbb{R}^+}$ gives the conditional density function of T given $\mathbb{F}$. Looking at the definition of the $\mathbb{F}$-martingale hazard process, we see that the nonnegative, $\mathbb{F}$-predictable process $\lambda^{\mathbb{F}} = (\lambda_t^{\mathbb{F}})_{t\in\mathbb{R}^+}$ defined by*

$$\lambda_t^{\mathbb{F}} = \frac{f_t}{1 - F_t} = \frac{f_t}{S_t} \tag{42}$$

satisfies the property that $\left(X_t - \int_0^{t\wedge T} \lambda_u^{\mathbb{F}}\,du\right)_{t\in\mathbb{R}^+}$ is a $\mathbb{G}$-martingale. A process with this property is called conditional intensity process *of T given $\mathbb{F}$, as formulated in the following definition.*

Definition A.3 (Conditional intensity process). *The* conditional intensity process *of the random time T given $\mathbb{F}$ is the nonnegative, $\mathbb{F}$-predictable process $\lambda^{\mathbb{F}}$ such that the process $M = (M_t)_{t\in\mathbb{R}^+}$ defined by*

$$M_t = X_t - \int_0^{t\wedge T} \lambda_u^{\mathbb{F}}\,du \tag{43}$$

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

The next goal is to prove that, under some restriction on the process F, the $\mathbb{F}$-hazard process Γ satisfies the martingale characterization introduced by the definition of the $\mathbb{F}$-martingale hazard process Λ. The following Theorem gives the condition for which the process $(X_t - \Gamma_{t\wedge T})_{\{t\geq 0\}}$ is a $\mathbb{G}$-martingale.

Theorem A.2. *Suppose that the process F defined by equation (36) is an increasing and continuous process. Then the process N defined by $N_t = (X_t - \Gamma_{t\wedge T})_{\{t\geq 0\}}$ is a $\mathbb{G}$-martingale.*

Proof. N is $\mathbb{G}$-adapted. Moreover, using the integration by part formula for function with finite variation (Protter 1990), we have

$$\widetilde{N}_t := (1 - X_t)\exp(\Gamma_t) = 1 + \int_0^t \exp(\Gamma_u)\left((1 - X_u)\,d\Gamma_u - dX_u\right)$$

since Γ is an increasing continuous process, under the same assumptions for F. We have

$$N_t = X_t - \Gamma_{t\wedge T} = X_t - \int_0^t (1 - X_u)\,d\Gamma_u = \int_0^t \left(dX_u - (1 - X_u)\,d\Gamma_u\right) = -\int_0^t \exp(-\Gamma_u)\,d\widetilde{N}_u$$

and thus N is $\mathbb{G}$-martingale, since $\widetilde{N}$ it is one and Γ is $\mathbb{G}$-adapted. □

We have proved that, if F is continuous and increasing, then the $\mathbb{F}$-hazard process satisfies the martingale characterization. A consequence of this Theorem is that, under these assumptions on F, we have $\Gamma_t = \Lambda_t$, $\forall t \in \mathbb{R}^+$. We formulate these results in the following Proposition.

Proposition A.2. *Suppose that the process F is an increasing and continuous process, then $\Gamma_t = \Lambda_t$.*

In the following Proposition we pull together some results of this section.

Proposition A.3. *Let $T : \Omega \longrightarrow \mathbb{R}^+$ be a random time. $\mathbb{F}$ and $\mathbb{G}$ are filtrations on $(\Omega, \mathcal{G}, \mathbb{P})$ as defined before. We suppose that T is absolutely continuous and admits a conditional intensity process $\lambda^{\mathbb{F}}$, given the filtration $\mathbb{F}$. We have*

(i) $\mathbb{P}\big[T > s \,|\, \mathcal{F}_t\big] = \mathbb{P}\big[T > t \,|\, \mathcal{F}_t\big]\mathbb{E}\big[\exp\big(-\int_t^s \lambda_u^{\mathbb{F}}\, du\big) \,|\, \mathcal{F}_t\big]$.

(ii) $S_t = 1 - F_t = \mathbb{P}\big[T > t \,|\, \mathcal{F}_t\big] = \exp\Big(-\int_0^t \lambda_u^{\mathbb{F}}\, du\Big)$.

(iii) $\lambda_t^{\mathbb{F}} = \frac{f_t}{S_t} = \frac{f_t}{1-F_t}$, *where* $f_t\, dt = dF_t$.

B Appendix: Generalized additive models

Let V be a random variable. Suppose that, given $\mathbf{Y} = (Y_1, \cdots, Y_p)$, the random variable V has a conditional distribution $F_{\mathbf{Y}}$ with expected value $\mu = \mathbb{E}\big[V \mid \mathbf{Y}\big]$, where $\mu = \mu(\mathbf{Y})$ is determined by the following equation

$$G(\mu(\mathbf{Y})) = \eta := \alpha + \sum_{q=1}^{p} f_q(Y_q), \tag{44}$$

for a constant α and functions (usually smooth functions) $f_1, \ldots, f_p$. The right side of equation (44) is called an *additive form* and represents the systematic component of the model. The function G is the so-called *link function*, linking the conditional expected value to the predictors $(Y_1, \cdots, Y_p)$. In (44) the assumption $\mathbb{E}\big[f_q(Y_q)\big] = 0$ is implicit, since one can define $\widetilde{f}_q = f_q - \mathbb{E}\big[f_(Y_q)\big]$ and $\widetilde{\alpha} = \alpha + \sum_{q=1}^{p} \mathbb{E}\big[f_q(Y_q)\big]$. The triple $(\eta, G, F_{\mathbf{Y}})$ defines a standard generalized additive model (GAM).
Given a GAM, one can impose some restrictions on the additive form η. We can suppose that $f_q(Y_q) = \beta_q Y_q$ for some parameter β_q, $q = 1, \ldots, p$. In this case, the relationship will be linear and we call $(\eta, G, F_{\mathbf{Y}})$ a generalized linear model (GLM).

V is said to follow an *exponential family density*, if the conditional cumulative distribution function $F_{\mathbf{Y}}$ has a density function given by

$$\rho_{\mathbf{Y}}(v; \xi, \phi) = \exp\left(\frac{v\xi - b(\xi)}{a(\phi)} + c(v, \phi)\right), \quad v \in \operatorname{supp}(F_{\mathbf{Y}}), \tag{45}$$

where $\xi = \xi(\mathbf{Y})$ is called natural parameter and ϕ is the dispersion parameter. The function $a(\phi)$ is often of the form $a(\phi) = \frac{\phi}{w}$ where ϕ is constant over the observations and w is a known, a priori given, weight which varies from observation to observation.
From $0 = \int_{\mathbb{R}} \frac{\partial}{\partial \xi} \rho_{\mathbf{Y}}(v; \xi, \phi)\, dv$ it follows $b'(\xi) = \mu(\mathbf{Y})$. Moreover, from $0 = \int_{\mathbb{R}} \frac{\partial^2}{\partial^2 \xi} \rho_{\mathbf{Y}}(v; \xi, \phi)\, dv$ we obtain $\mathrm{Var}\,[V \,|\, \mathbf{Y}] = a(\phi) b^{''}(\xi)$. In the case of an exponential family density we denote the GAM by the triple $(\eta, G, \rho_{\mathbf{Y}})$.

Example
Suppose $V \sim \frac{1}{n}\text{binomial}(n,p)$ where $p = p(\mathbf{Y})$. Then for $v \in [0,1]$ such that $nv \in \{0,\dots,n\}$ we have

$$\mathbb{P}\big[V = v \,|\, \mathbf{Y}\big] = \binom{n}{nv} p^{nv}(1-p)^n = \exp\left[n\left(v\log\left(\frac{p}{1-p}\right) + \log(1-p)\right) + \log\binom{n}{nv}\right]$$

Let $\xi = \log\left(\frac{p}{1-p}\right)$, $b(\xi) = -\log(1-\mathrm{e}^{\xi})$, $\phi = n$, $a(\phi) = \frac{1}{\phi}$, $c(v,\phi) = \log\binom{\phi}{\phi v}$. We have

$$\mathbb{P}\big[V = v \,|\, \mathbf{Y}\big] = \exp\left(\frac{v\xi - b(\xi)}{a(\phi)} + c(v,\phi)\right).$$

We can easily check that $p = \frac{\mathrm{e}^{\xi}}{1+\mathrm{e}^{\xi}}$, thus $b'(\xi) = p = \mathbb{E}\big[V \,|\, \mathbf{Y}\big]$; $a(\phi)b''(\xi) = \frac{1}{n}p(1-p) = \mathrm{Var}\left[V \,|\, \mathbf{Y}\right]$. Moreover, the canonical link is given by $G(p) = \log\left(\frac{p}{1-p}\right)$, which is the well-known logit link function. Following the remark on $a(\phi)$ given above, we consider the parameter n of the binomial distribution as a known, a priori given, weight for the observation, and the dispersion parameter ϕ is identical to 1 for the binomial case.

We first consider a GLM to explain the estimation technique. Let $(\eta, G, \rho_{\mathbf{Y}})$ be a GLM with an exponential family density $\rho_{\mathbf{Y}}$. We have $G(\mu(\mathbf{Y})) = \eta$ where $\eta = \alpha + \beta_1 Y_1 + \dots + \beta_p Y_p$. Let $\mathbf{y}_i = (y_{i,1},\dots,y_{i,p})'$, $i = 1,\dots,M$ be subsequent observations of the predictors, $\mu_i = \mu(\mathbf{y}_i)$, $i = 1,\dots,M$ are the corresponding conditional expected values and v_i the observed conditionally independent realization of V. Moreover, $\eta = (\eta_1,\dots,\eta_M)$, where $\eta_i = G(\mu_i)$. The goal is to estimate the parameter $\theta = (\alpha, \beta_1, \dots, \beta_p)'$ defining the GLM. The likelihood function of the observations is the following

$$\mathcal{L}(\theta) = \prod_{i=1}^{M} \rho_{\mathbf{Y}}(v_i; \xi_i(\theta), \phi_i) = \prod_{i=1}^{M} \exp\left(\frac{v_i\xi_i(\theta) - b(\xi_i(\theta))}{a(\phi_i)} + c(v_i,\phi_i)\right)$$

where $\xi_i(\theta) := \xi(\theta, \mathbf{y}_i)$.
For maximizing the log-likelihood function $l(\theta) = \log\mathcal{L}(\theta)$ we consider the partial derivative with respect to θ; we obtain the score equations

$$0 = \frac{\partial l}{\partial\theta_{q+1}}(\theta) = \sum_{i=1}^{M}(v_i - \mu_i)\frac{\partial G^{-1}}{\partial\eta}\bigg|_{\eta_i = G(\mu_i)} Var_i^{-1} y_{i,q}, \quad \text{for} \quad q = 0,\dots,p, \tag{46}$$

where $y_{i,0} = 1, \forall\, i = 1,\dots,M$ and $Var_i = \mathrm{Var}\left[V \,|\, \mathbf{y}_i\right]$. We have used the formulae $\mathbb{E}\big[V \,|\, \mathbf{Y}\big] = b'(\xi)$ and $\mathrm{Var}\left[V \,|\, \mathbf{Y}\right] = a(\phi)b''(\xi)$ derived above for $\xi = \xi(\mathbf{Y})$.
The maximum likelihood estimation θ_{MLE} of θ is obtained by solving numerically the non-linear equations (46) with respect to θ. A standard method to solve these equations is the Fisher-scoring algorithm, which is a Newton-Raphson algorithm using the expected infomation matrix instead of the observed one (details are given in Seber and Wild (1989, Chapter 2.2) and McCullagh and Nelder (1995)). Hastie and Tibshirani (1990) propose solving (46) by a form of the iteratively-reweighted-least-squares (IRLS) called *adjusted dependent variable regression*, which is equivalent to the Fisher-scoring procedure for the case where V follows an exponential family density. The advantage of the adjusted dependent variable regression is that no special optimization software is required. We introduce the IRLS algorithm in the next section with the extension for the GAM case.

B.1 Adjusted dependent variable regression

The suggestion for the following derivation of the adjusted dependent variable regression procedure comes from Hastie and Tibshirani (1990, Exercise 6.1). We show that, under the assumption of an exponential family density, the adjusted dependent variable regression is simply the Fisher-scoring algorithm, which is a modification of the Newton-Raphson algorithm for solving non-linear equations.

Using the same notation as above, we define the score function and the expected Fisher-information matrix for θ by $\mathcal{S}_\theta = \frac{\partial l}{\partial \theta}$ and $\mathcal{I}_\theta = \mathbb{E}\big[-\frac{\partial^2 l}{\partial\theta\partial\theta'}\big]$ respectively. The score function and the expected Fisher-information matrix for η are denoted by $\mathcal{S}_\eta = \frac{\partial l}{\partial \eta}$ and $\mathcal{I}_\eta = \mathbb{E}\big[-\frac{\partial^2 l}{\partial\eta\partial\eta'}\big]$ respectively. We solve the equation $\frac{\partial l}{\partial \theta} = 0$ using the Fisher-algorithm. By $\mathbb{Y}$ we denote the design matrix, i.e.

$$\mathbb{Y} = \begin{pmatrix} 1 & y_{1,1} & \dots & y_{1,p} \\ \vdots & \vdots & & \vdots \\ 1 & y_{M,1} & \dots & y_{M,p} \end{pmatrix}.$$

We have $\eta = \mathbb{Y}\theta$. Let θ^0 be the start value by the Fisher-scoring algorithm, the Newton-Raphson step with $\mathcal{I}_\theta$ is the following

$$\theta^1 = \theta^0 + \mathcal{I}_{\theta^0}^{-1}\mathcal{S}_{\theta^0}.$$

Rewriting this equation using $\mathcal{S}_\eta$ and $\mathcal{I}_\eta$, for $\eta^0 = \mathbb{Y}\theta^0$ we obtain

$$\theta^1 = (\mathbb{Y}' I_{\eta^0}\mathbb{Y})^{-1}\mathbb{Y}'\mathcal{I}_{\eta^0}(\eta^0 + \mathcal{I}_{\eta^0}^{-1}\mathcal{S}_{\eta^0})$$

which is exactly the LS solution of a weighted regression of $\mathbf{z} = \eta^0 + \mathcal{I}_{\eta^0}^{-1}\mathcal{S}_{\eta^0}$ on $\mathbb{Y}$ with weight $\mathcal{I}_{\eta^0}$. Up to now no restriction on the distribution family of V was considered. The last equation can be viewed as a generalization of the adjusted dependent variable regression. We now suppose that V follows an exponential family density (45): for this cas we compute $\mathcal{S}_\eta$ and $\mathcal{I}_\eta$ explicitly. We use that $\mu_i = b'(\xi_i(\theta))$, $b''(\xi_i(\theta)) = a(\phi_i)^{-1}Var_i$ and $\eta_i = G(\mu_i)$, it follows for $i, j = 1, \dots, M$

$$\frac{\partial \xi_i(\theta)}{\partial \eta_j} = 1_{\{i=j\}} \frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_i} \left(b''(\xi)\bigg|_{\xi=\xi_i(\theta)} \right)^{-1} = 1_{\{i=j\}} \frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_i} a(\phi_i) Var_i^{-1}$$

Using the formula for $\frac{\partial \xi_i(\theta)}{\partial \eta_j}$ we obtain

$$\begin{aligned} \mathcal{S}_{\eta,j} &= \frac{\partial l}{\partial \eta_j} = \frac{\partial}{\partial \eta_j} \sum_{i=1}^{M} \frac{v_i \xi_i(\theta) - b(\xi_i(\theta))}{a(\phi_i)} + c(v_i \phi_i) \\ &= \sum_{i=1}^{M} \frac{1}{a(\phi_i)} \left(v_i \frac{\partial}{\partial \eta_j} \xi_i(\theta) - \frac{\partial}{\partial \eta_j} b(\xi_i(\theta)) \right) \\ &= \sum_{i=1}^{M} \frac{1}{a(\phi_i)} \left(v_i - \frac{\partial}{\partial \xi} b(\xi)\bigg|_{\xi=\xi_i(\theta)} \right) \frac{\partial}{\partial \eta_j} \xi_i(\theta) \\ &= (v_j - \mu_j) \frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j} Var_j^{-1}. \end{aligned}$$

Analogously

$$\begin{aligned}
\frac{\partial^2 l}{\partial \eta_i \partial \eta_j} &= \frac{\partial}{\partial \eta_i} S_{\eta,j} \\
&= -\frac{\partial G^{-1}}{\partial \eta_i}\bigg|_{\eta_j} Var_j^{-1} \frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j} + (y_j - \mu_j)\left(\frac{\partial Var_j^{-1}}{\partial \eta_i}\frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j} + Var_j^{-1}\frac{\partial}{\partial \eta_i}\frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j}\right) \\
&= -1_{\{i=j\}}\left(\frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j}\right)^2 Var_j^{-1} + (y_j - \mu_j)1_{\{i=j\}}\left(\frac{\partial Var_j^{-1}}{\partial \eta_j}\frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j} + Var_j^{-1}\frac{\partial^2 G^{-1}}{\partial^2 \eta_j}\bigg|_{\eta_j}\right)
\end{aligned}$$

and by taking the expectation we have

$$\mathcal{I}_{\eta,ij} = \mathbb{E}\big[-\frac{\partial^2 l}{\partial \eta_i \partial \eta_j}\big] = 1_{\{i=j\}}\left(\frac{\partial G^{-1}}{\partial \eta_j}\bigg|_{\eta_j}\right)^2 Var_j^{-1}$$

Hence in the case of an exponential family density, the weights $\mathbf{w} = (w_1, \ldots, w_M)'$ and the adjusted variable $\mathbf{z} = \eta^0 + \mathcal{I}_{\eta^0}^{-1}\mathcal{S}_{\eta^0}$ for the weighted regression $(B.1)$ are simply given by

$$w_i = \left(\frac{\partial G^{-1}}{\partial \eta}\bigg|_{\eta_i^0}\right)^2 (Var_i^0)^{-1} \tag{47}$$

and

$$z_i = \eta_i^0 + (v_i - \mu_i^0)\left(\frac{\partial G^{-1}}{\partial \eta}\bigg|_{\eta_i^0}\right)^{-1} \tag{48}$$

where $\mu_i^0 = G^{-1}(\eta_i^0)$.
The adjusted dependent variable regression procedure therefore consists in a repeated weighted regression of the adjusted dependent variable z_i with respect to $\mathbf{y}_i$ and weight w_i $(i = 1, \ldots, M)$. The algorithm is formally given in the sequel.

Algorithm B.1 (Adjusted dependent variable regression)
Initialization r=0: Initial value $\alpha^0 = G(\frac{1}{M}\sum_{i=1}^M v_i)$, $\beta_1^0 = \cdots = \beta_p^0 = 0$.
Iteration $r \rightarrow r+1$:
Compute for $i = 1, \ldots, M$:

$$\eta_i^r = \alpha^r + \sum_{q=1}^p \beta_q y_{i,q}, \quad \mu_i^r = G^{-1}(\eta_i^r), \quad Var_i^r = \mathrm{Var}\,(V\,|\,\eta_i^r)\,. \tag{49}$$

Moreover, for $i = 1, \ldots, M$ we define:

$$\begin{aligned}
z_i &= \eta_i^r + (v_i - \mu_i^r)\left(\frac{\partial G}{\partial \mu}\bigg|_{\mu_i^r}\right), \\
w_i &= \left(\frac{\partial G}{\partial \mu}\bigg|_{\mu_i^r}\right)^2 (Var_i^r)^{-1}.
\end{aligned}$$

Estimate the parameters $\alpha^{r+1}, \beta_1^{r+1}, \ldots, \beta_p^{r+1}$ by regressing $\mathbf{z} = (z_1, \ldots, z_M)'$ on $(\mathbf{y}_1, \ldots, \mathbf{y}_M)'$ with weights $\mathbf{w} = (w_1, \ldots, w_M)'$.
Repeat until the change of the log-likelihood function $l(\theta)$ is sufficiently small.

B.2 Extension to the GAM

The idea of the adjusted dependent variable regression can be extended to the case of the GAM. This is done by the local scoring algorithm of Hastie and Tibshirani (1990, Chapters 6.3 and 6.5) by replacing the linear regression step with a non-parametric additive regression step. We restrict the choice of the functions $f_1, \ldots, f_p$ in (44) to the class of smooth functions following Hastie and Tibshirani (1987). In our work we have considered smoothing splines. Here we want briefly to resume the main idea of the local scoring algorithm, given an intuitive development. Details can be found in Hastie and Tibshirani (1986, 1990). A recent reference on GAM is moreover Schimek and Turlach (2000).
The goal is the estimation of $\alpha, f_1, \ldots, f_p$ in the generalized additive model (44). We first assume that the link function corresponds to the identity, i.e.

$$\mathbb{E}\big[V \,|\, \mathbf{Y}\big] = \alpha + \sum_{q=1}^{p} f_q(X_q) \tag{50}$$

where $\mathbb{E}\big[f_q(Y_q)\big] = 0$ for every q. As a suggestion for the *backfitting algorithm* we consider the simple case where the model $V = \alpha + \sum_{q=1}^{p} f_q(Y_q) + \epsilon$ is correct for some ϵ independent of $\mathbf{Y}$, with $\mathbb{E}\big[\epsilon\big] = 0$. Moreover, we assume that α and $f_{q'}$ are known for $q' \neq q$. We have for $R_q = V - \alpha - \sum_{q' \neq q} f_{q'}(Y_{q'})$,

$$\mathbb{E}\big[R_q \,|\, Y_q\big] = f_q(Y_q) \tag{51}$$

which moreover minimizes $\mathbb{E}\Big[\Big(V - \alpha - \sum_{q' \neq q} f_{q'}(Y_{q'})\Big)^2\Big]$. This suggests the following algorithm for estimating each f_q given estimates for $f_{q'}$, $q' \neq q$.

Algorithm B.2 (Backfitting algorithm)
Initialization $r = 0$: $\alpha^0 = \mathbb{E}\big[V\big]$, $f_1^0 \equiv \cdots \equiv f_p^0 \equiv 0$.
Iteration $r \to r + 1$:
$\alpha^r = \mathbb{E}\big[V - \sum_{q=1}^{p} f_q^r(Y_q)\big]$
Cycle over $q = 1, \ldots, p$:

$$R_q = V - \alpha^r - \sum_{k=1}^{q-1} f_k^{r+1}(Y_k) - \sum_{k=q+1}^{p} f_k^r$$
$$f_q^{r+1}(Y_q) = \mathbb{E}\big[R_q \,|\, Y_q\big]$$

Until: $\mathbb{E}\Big[\Big(V - \alpha^r - \sum_{q=1}^{p} f_q^{r+1}(Y_q)\Big)^2\Big]$ fails to decrease.

In the world of the finite sample the conditional expectation in the backfitting algorithm is replaced by a scatterplot smoother $\mathbf{S}_q[.]$. Let $\mathbf{v} \in \mathbb{R}^M$ denote the vector of observations for V and $\mathbf{y}_q \in \mathbb{R}^M$, for $q = 1, \ldots, p$ the corresponding observed predictors, then we have

$$\mathbf{f}_q^{r+1} = \mathbf{S}_q[\mathbf{r}_q \,|\, \mathbf{y}_q],$$

where $\mathbf{f}_q^{r+1} = (f_q^{r+1}(y_{q,1}, \ldots, f_q^{r+1}(y_{q,M}))'$ and $\mathbf{r_q} = \mathbf{v} - \alpha - \sum_{k=1}^{q-1} \mathbf{f}_k^{r+1} - \sum_{k=q+1}^{p} \mathbf{f}_k^r$. More details smoothing operator are given in Härdle (1990), Hastie and Tibshirani (1990). An overview is given in Appendix C.
The parameter α is initialized by $\alpha^0 = \frac{1}{M} \sum_{i=1}^{M} v_i$ and α^r for $r \geq 1$ can be estimated by using a least square estimator, which corresponds to $\frac{1}{M} \sum_{i=1}^{M} (v_i - r_{q,i})$.

Hastie and Tibshirani (1990) derive the local scoring algorithm by considering the expected log-likelihood function $\mathbb{E}\big[l(\eta(\mathbf{Y}, V)\big]$, where $\eta = \alpha + \sum_{q=1}^{p} f_q$. They choose the $\widehat{\eta}$ that maximizes the expected log-likelihood, i.e. $\mathbb{E}\big[l(\widehat{\eta}(\mathbf{Y},V)\big] = \max_\eta \mathbb{E}\big[l(\eta(\mathbf{Y},V)\big]$. For the maximizer $\widehat{\eta}$ one can show that $\mathbb{E}\big[\frac{\partial l}{\partial \eta}\,|\,Y_q\big]\big|_{\widehat{\eta}} = 0$ for every q (Hastie and Tibshirani 1990, Chapter 6.5.1). By linearizing this equation with a first order Taylor expansion about a current guess η^0, we obtain

$$\eta^1 = \mathbb{E}\big[\eta^0 - \frac{\partial l}{\partial \eta}\big|_{\eta^0}\left(\mathbb{E}\big[\frac{\partial^2 l}{\partial^2 \eta}\,|\,\mathbf{Y}\big]\big|_{\eta^0}\right)^{-1}\,|\,\mathbf{Y}\big]$$

and thus, with the same notation as in the previous section, we have

$$\eta^1(\mathbf{Y}) = \mathbb{E}\big[\eta^0 + \mathcal{I}_{\eta^0}^{-1}\mathcal{S}_{\eta^0}\,|\,\mathbf{Y}\big]$$

We define the adjusted dependent variable by $Z = \eta^0 + \mathcal{I}_{\eta^0}^{-1}\mathcal{S}_{\eta^0}$, then we have $\eta^1(\mathbf{Y}) = \mathbb{E}\big[Z\,|\,\mathbf{Y}\big]$, which has the same form as equation (50) with Z instead of V. This suggests the implementation of the backfitting procedure for estimating $\alpha, f_1, \dots, f_p$. The local scoring algorithm in the case of exponential family density is the following.

Algorithm B.3 (Local scoring algorithm)
Define $\epsilon_1 > 0$ and $\epsilon_2 > 0$ very small numbers.
Initialization $r = 0$: $\alpha^0 = G\left(\frac{1}{M}\sum_{i=1}^M v_i\right)$, $\mathbf{f}_1^0 = \dots = \mathbf{f}_p^0 = 0$.
Iteration $r \to r+1$:
For $i = 1, \dots, M$ compute

$$\eta_i^r = \alpha^r + \sum_{q=1}^p f_{q,i}, \quad \mu_i^r = G^{-1}(\eta_i^r), \quad Var_i^r = \mathrm{Var}\,(V\,|\,\eta_i^r)\,.$$

Moreover, for $i = 1, \dots, M$ we define

$$\begin{aligned} z_i &= \eta_i^r + (v_i - \mu_i^r)\left(\frac{\partial G}{\partial \mu}\bigg|_{\mu_i^r}\right), \\ w_i &= \left(\frac{\partial G}{\partial \mu}\bigg|_{\mu_i^r}\right)^2 (V_i^r)^{-1}. \end{aligned}$$

$\mathbf{z} = (z_1, \dots, z_M)'$, $\mathbf{w} = (w_1, \dots, w_M)'$.

Backfitting Loop:
Let $\mathbf{f}_q^{(0)} = \mathbf{f}_q^r$ for $q = 1, \dots p$.
For $q = 1, \dots, p$ (backfitting iteration):

$\mathbf{r}_q = \mathbf{z} - \alpha^r - \sum_{k=1}^{q-1}\mathbf{f}_k^{(1)} - \sum_{k=q+1}^{p}\mathbf{f}_k^{(0)}$,
$\mathbf{f}_q^{(1)} = \mathbf{S}_q\,[\mathbf{r}_q\,|\,\mathbf{y}_q, \text{weight} = \mathbf{w}]$.

If $\|\mathbf{f}_q^{(0)} - \mathbf{f}_q^{(1)}\| < \epsilon_1$ for $q = 1, \dots, p$ stop the backfitting loop.
Else set $\mathbf{f}_q^{(0)} = \mathbf{f}_q^{(1)}$ for $q = 1, \dots, p$ and repeat the backfitting iteration oncemore.

Set

$$\begin{aligned} \mathbf{f}_q^{r+1} &= \mathbf{f}_q^{(1)} \text{ for } q = 1, \dots, p \text{ and} \\ \alpha^{r+1} &= \text{WLSE}\left[\mathbf{z} - \alpha - \sum_{q=1}^p \mathbf{f}_q^{r+1}\,|\,1, \text{weight} = \mathbf{w}\right], \end{aligned}$$

where WLSE denotes the weighted least square estimator of α.
Repeat the r-iteration until

$$\frac{\sum_{q=1}^{p} \|\mathbf{f}_q^{r+1} - \mathbf{f}_q^{r}\|}{\sum_{q=1}^{p} \|\mathbf{f}_q^{r}\|} < \epsilon_2.$$

$\mathbf{S}_q[.]$ now represents a *weighted* scatterplot smoothing procedure: it can be chosen from a variety of candidates, such as cubic-spline, B-spline, local regression or kernel regression. We have used cubic splines.

Example
Let $V \sim \frac{1}{n} binomial(n,p)$ conditional on a vector of predictors $\mathbf{Y}$.
Suppose that $p = p(\mathbf{Y}) = \mathbb{E}\big[V \mid \mathbf{Y}\big]$ is given by

$$G(p(\mathbf{Y})) = \alpha + \sum_{q=1}^{m} f_q(Y_q), \tag{52}$$

where $G : (0,1) \longrightarrow \mathbb{R}, \mu \mapsto \log(-\log(1-\mu))$.
Let $\mathbf{y}_i$ $(i = 1,\ldots,M)$ be subsequent observations of $\mathbf{Y}$ and $v_i \sim \frac{1}{n_i} binomial(n_i, p_i)$ the observed realization of V, where $p_i = p(\mathbf{y}_i)$ satisfies (52). We have

$$\begin{aligned} \frac{\partial G}{\partial \mu} &= -\left[(1-\mu)\log(1-\mu)\right]^{-1}, \\ Var_i &= \mathrm{Var}\left[v_i \,|\, \mathbf{y}_i\right] = \frac{1}{n_i} p_i(1-p_1). \end{aligned}$$

The adjusted dependent variable $\mathbf{z}$ and the weights are thus given by

$$\begin{aligned} z_i &= \eta_i - (v_i - p_i)\left[(1-p_i)\log(1-p_i)\right]^{-1}, \\ w_i &= \left[(1-p_i)\log(1-p_i)\right]^{-2} \left[\frac{1}{n_i} p_i(1-p_i)\right]^{-1}. \end{aligned}$$

C Appendix: Scatterplot smoothing

In the previous section we introduced the local scoring algorithm for solving the GAM. In the backfitting loop, the conditional expectation is estimated by using a scatterplot smoothing, denoted by $\mathbf{S}[.]$. In this section we consider some general characteristics of smooth operators, details can be found in Härdle (1990) and Hastie and Tibshirani (1990).
We consider a simpler version of equation (50), i.e.

$$\mathbb{E}\big[V \,|\, Y\big] = f(Y) \tag{53}$$

for some unknown function f. The goal is the estimation of the function f given pairs of observations $\{(v_i, y_i) : i = 1,\ldots,M\}$, where v_i, y_i are the realization of random variables V_i and Y_i respectively, following (53). Rewriting equation (53) we have

$$V_i = f(Y_i) + (V_i - \mathbb{E}\big[V_i \,|\, Y_i\big]). \tag{54}$$

Let $\epsilon_i := V_i - \mathbb{E}\big[V_i \,|\, Y_i\big]$. We have $\mathbb{E}\big[\epsilon_i\big] = 0$ and $\mathrm{Var}\left[\epsilon_i\right] = \mathbb{E}\big[V_i^2\big] - \mathbb{E}\big[f(Y_i)^2\big] =: \sigma_i^2$. We suppose that ϵ_i, $i = 1,\ldots,M$ are independent. For $\mathbf{v} = (v_1,\ldots,v_M)'$ and $\mathbf{y} = (y_1,\ldots,y_M)'$ we denote by $\widehat{f} = \mathbf{S}[\mathbf{v} \,|\, \mathbf{y}]$ the estimation of f by a scatterplot smoother S. S is called a *linear smoother* if $\mathbf{S}[a_1\mathbf{v}^1 + a_2\mathbf{v}^2 \,|\, \mathbf{y}] = a_1\mathbf{S}[\mathbf{v}^1 \,|\, \mathbf{y}] + a_2\mathbf{S}[\mathbf{v}^2 \,|\, \mathbf{y}]$, i.e. it is linear in the observation $\mathbf{v}$, given observation

$\mathbf{y}$. In this case we can simply write $\widehat{\mathbf{f}} = \mathbf{Sv}$, where $\mathbf{S} = (S_{ij})$ is a $M \times M$ matrix depending on $\mathbf{y}$, called *smoother matrix*, $\widehat{\mathbf{f}} = (\widehat{f}(y_1), \dots, \widehat{f}(y_M))$.

In the sequel we introduce a class of linear smoothing operators, i.e. the *smoothing splines*. We assume that the variance of the error term ϵ_i is a constant σ^2 independent of i. This assumption has to be relaxed by the smoothing operator used in the local scoring algorithm: we refer to Hastie and Tibshirani (1990, Sections 3.11 and 5.4.1), to Green and Silverman (1994, Secion 3.5) and to Opsomer and Kauermann (2000) for some extension to the case where weights are to be considered.

C.1 Smoothing spline

We consider equation (54). Our goal is to estimate the function f. This section is devoted to the smoothing spline technique, which is one possible choice for the smoothing operator $\mathbf{S}$ introduced in the previous section.
Given observations $\{(v_i, y_i) : i = 1, \dots, M\}$, $\mathbf{v} = (v_1, \dots, v_M)'$, a common measure of goodness of fit for a function g is the residual sum of square $\sum_{i=1}^{M} (v_i - g(y_i))^2$. If we minimize this quantity allowing g to be any curve, then we can easily find a g with $g(y_i) = v_i$ for all i and thus the residual sum of square will be identical to 0. This "solution" will not be very useful: first because it will not be a unique solution, second because no structure in the data can be found, since a very high local variation exists. The spline smoothing approach solves this problem by "penalizing" the local variation with a so called *roughness penalty* $\int (g''(y))^2 \, dy$. The smoothing spline approach minimizes the weighted sum

$$\sum_{i=1}^{M} (v_i - g(y_i))^2 + \lambda \int_a^b (g''(y))^2 \, dy \tag{55}$$

over the set of all the twice differentiable functions on $[a, b]$, where λ is the *smoothing factor*, representing the rate of exchange between residual error and roughness of the curve g, $a := \min_i\{y_i\}$, $b := \max_i\{y_i\}$. The smoothing factor λ governs the tradeoff between the goodness of fit to the data and the "wigglyness" of the function: larger values of force f to be smoother. If $\lambda \nearrow \infty$ then the penalty term dominates, forcing $g'' = 0$ on $[a, b]$: the solution tends to the least square line. If $\lambda \searrow 0$ then the roughness penalty becomes unimportant and the solution will tend to a twice differentiable interpolating function. For each constant $\lambda \geq 0$ the optimization problem has a unique solution $\widehat{f}_\lambda$ on the set of all the twice differentiable functions on $[a, b]$. The solution has the following properties:

(i) $\widehat{f}_\lambda$ is a cubic polynomial between two successive y-values;

(ii) at each y_i, $\widehat{f}_\lambda$ and its first two derivatives are continuous;

(iii) in a and b the second derivative of $\widehat{f}_\lambda$ is equal to zero.

and is thus a cubic spline with knots $y_1, \dots, y_M$ (Green and Silverman 1994, Chapter 2).

Smoothing splines are linear smooth operators. In fact the fits $\widehat{\mathbf{f}}_\lambda = (\widehat{f}_\lambda(y_1), \dots, \widehat{f}_\lambda(y_M))$ of the vector $\mathbf{f} = (f(y_1), \dots, f(y_M))'$ can be written as a linear transformation of the vector of observations $\mathbf{v}$, i.e. $\widehat{\mathbf{f}}_\lambda = (I + \lambda K)^{-1}\mathbf{v}$ (Green and Silverman 1994, Chapter 2). Here $I + \lambda K$ is a strictly positive matrix, K gives the roughness penalty, λ controls the smoothness of the fit. We denote the smoothing matrix by $\mathbf{S}^\lambda = (I + \lambda K)^{-1}$.

The choice of the smoothing factor λ is a crucial decision for the estimation of the smoothing spline $\widehat{f}_\lambda$. Hastie and Tibshirani (1986) affirm that it is not convenient to express the desired smoothness of the $\widehat{f}$ in terms of λ, since the meaning of the smoothing parameter depends on the units of the prognostic factor y. Instead, it is possible to define an *effective number of parameters* or *degrees of freedom* of a cubic spline smoother, and then to use a numerical search to determine the value of λ to yield this number.
Following Hastie and Tibshirani (1990) we define the effective number of degrees of freedom by $df_\lambda = \mathrm{tr}(\mathbf{S}^\lambda)$, where tr(.) denotes the "trace". It can be shown that

$$df_\lambda = 2 + \sum_{i=3}^{M} \frac{1}{1 + \lambda\, w_i} \tag{56}$$

where w_i are the eigenvalues of the matrix K, assuming that $w_1 = w_2 = 0$ (two such eigenvalues exist). Therefore, there is a strictly monotone relationship between df_λ and λ, which implies that λ is uniquely determined by df_λ . If $\lambda \searrow 0$ (interpolating twice differentiable functions), $df_\lambda \nearrow M$; if $\lambda \nearrow \infty$ (least square line) then $df_\lambda \searrow 2$. The definition of the effective degree of freedom through $\mathrm{tr}(S_\lambda)$ can, moreover, be motivated by an analogy with the classical parametric regression (Green and Silverman 1994, Section 3.4). One can show for example that

$$\mathbb{E}\Big[\sum_{i}^{M} \Big(v_i - \widehat{f}_\lambda(y_i)\Big)^2\Big] = \big\{M - \mathrm{tr}\left[2\mathbf{S}^\lambda - \mathbf{S}^\lambda(\mathbf{S}^\lambda)'\right]\big\}\,\sigma^2 + (\mathbf{b}^\lambda)'\mathbf{b}^\lambda \tag{57}$$

where $\mathbf{b}^\lambda = \mathbf{f} - \mathbb{E}\big[\mathbf{S}^\lambda \mathbf{y}\big]$. In the case of the linear regression $df_\lambda^{\mathrm{err}} = M - \mathrm{tr}\left[2\mathbf{S}_\lambda - \mathbf{S}^\lambda(\mathbf{S}^\lambda)'\right]$ would be exactly equal to $M - p$, where p is the number of parameters to be estimate (and $2\mathbf{S}_\lambda - \mathbf{S}^\lambda(\mathbf{S}^\lambda)' = \mathbf{S}^\lambda$ if $\mathbf{S}^\lambda$ is the least square operator) .
Thus we define df_λ as the effective number of degrees of freedom, also for the general case where $\mathbf{S}^\lambda$ is the smoothing spline operator. Moreover, we refer to van der Linde (2000, Section 2.3.1) for a detailed discussion on the effective number of degrees of freedom.

The classical approach for estimating the smoothing factor λ considers the cross-validation sum of squares CV. We suppose that the smoothing spline estimation $\widehat{f}_\lambda^{-i}$ is performed for each $i = 1, \ldots, M$ by leaving out the ith data point (v_i, y_i). The λ that minimizes the cross-validation sum of squares

$$CV(\lambda) = \frac{1}{M}\sum_{i=1}^{M}\Big(v_i - \widehat{f}_\lambda^{-i}(y_i)\Big)^2 \tag{58}$$

is selected. The idea is to test the quality of the predictor $\widehat{f}_\lambda^{-i}$ on the "new" observation (v_i, y_i) for each $i = 1, \ldots, M$. For linear smoothing splines the estimation $\widehat{f}_\lambda^{-i}(y_i)$ can be written as a function of v_i, $\widehat{f}_\lambda(y_i)$ and the diagonal element S_{ii}^λ of $\mathbf{S}^\lambda$ (Green and Silverman 1994, Theorem 3.1). We have

$$v_i - \widehat{f}_\lambda^{-i} = \frac{v_i - \widehat{f}_\lambda(y_i)}{1 - S_{ii}^\lambda}. \tag{59}$$

It follows that

$$CV(\lambda) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{v_i - \widehat{f}_\lambda(y_i)}{1 - S_{ii}^\lambda}\right)^2. \tag{60}$$

The diagonal elements of $\mathbf{S}^\lambda$ are not easy to obtain. This motivates the approximation of S^λ_{ii} by $\frac{1}{M}\mathrm{tr}(\mathbf{S}^\lambda)$ and the definition of the *generalized cross-validation GCV* by

$$GCV(\lambda) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{v_i - \widehat{f}_\lambda(y_i)}{1 - \frac{\mathrm{tr}(\mathbf{S}^\lambda)}{M}}\right)^2 . \tag{61}$$

Rewriting GCV with df_λ we obtain

$$GCV(df_\lambda) = M\sum_{i=1}^{M}\left(\frac{v_i - \widehat{f}_\lambda(y_i)}{M - df_\lambda}\right)^2 . \tag{62}$$

Using that, for $x \approx 0$, $(1-x)^{-2} \approx 1 + 2x$ and that, $\frac{1}{M}\mathrm{tr}(\mathbf{S}^\lambda)$ is small enough for M and λ big enough, we can approximate $GCV(df_\lambda)$ by

$$GCV(df_\lambda) \approx \frac{1}{M}\sum_{i=1}^{M}\left(v_i - \widehat{f}_\lambda(y_i)\right)^2 + 2\frac{df_\lambda}{M}\frac{1}{M}\sum_{i=1}^{M}\left(v_i - \widehat{f}_\lambda(y_i)\right)^2 . \tag{63}$$

References

Barlow, R., and F. Proschan (1975): *Statistical theory of reability and life testing.* Holt, Rinehart and Winston, New York.

Burger, W. (1998): "Das Zinsänderungsrisiko variabler Bankgeschäfte : die Risikoanalyse und Bewertung der variablen Hypotheken und Spargelder," Ph.D. thesis, University of Zurich.

Cantoni, E., and T. Hastie (2000): "Degree of freedom tests for smoothing splines," Statistics Department, Stanford University.

Chambers, J., and T. Hastie (1992): *Statistical models in S.* Wadswoth & Brooks, Cole Advanced Books & Software, Pacific Grove, California.

Cox, D., and D. Oakes (1984): *Analysis of survival data.* Chapman and Hall, London.

Credit Suisse Group (2000): "The Swiss real estate market, facts and trends," *Ecomomic Research*, April 2000.

Deng, Y. (1995): "Mortgage termination: an empirical hazard model with stochastic term structure," Working Paper, University of California at Berkeley.

Deng, Y., and J. Quigley (2000): "Woodhead behaviour and the pricing of residential mortgages," Working Paper No. W00-004, Institute of Business and Economic Research, University of California at Berkeley.

Duffie, D., and N. Gàrlenau (1999): "Risk valuation of collateralized debt obligations," Working Paper, Graduate School of Business, Stanford University.

Finger, C. (2000): "A comparison of stochastic default rate models," The RiskMetrics Group, New York.

Follian, J., W.-V. Huang, and J. Ondrich (1999): "Stay, pay or walk away: a hazard rate analysis of FHA-insured mortgage terminations," Draft paper, Freddie Mac and University of Syracuse.

Grandell, J. (1976): *Doubly stochastic Poisson processes.* Springer, Berlin-New York, Lectures Notes in Mathematics 529.

Green, P., and B. Silverman (1994): *Nonparametric regression and generalized additive models - A roughness penalty approach.* Chapman and Hall, London, Monographs on Statistics and Applied Probability 58.

Härdle, W. (1990): *Applied nonparametric regression.* Cambridge University Press, Cambridge, Econometric Society Monographs No.9.

Hastie, T., and R. Tibshirani (1986): "Generalized additive model," *Encyclopedia of Statistical Science*, 1(6), 297–318.

——— (1987): "Local likelihood estimation," *Journal of the American Statistical Association*, 82(398), 559–567.

——— (1990): *Generalized additive models.* Chapman and Hall, London, Monographs on Statistics and Applied Probability 43.

Ishiguro, M., G. Kitagawa, and Y. Sakamoto (1986): *Akaike information criterion statistics.* KTK Scientific Publishers, Tokyo.

Jeanblanc, M., and M. Rutkowski (2000): "Modelling of default risk: mathematical tool," Working Paper, Equipe d'Analyse et Probabilité, Université d'Evry Val d'Essonne, available from http://www.univ-evry.fr/labor/lami/maths/prepubli/.

Jegadeesh, N., and X. Ju (2000): "A non-parametric prepayment model and valuation of mortgage-backed-securities," *Journal of Fixed Income*, 10(1), 50–67.

Kalbfleisch, J., and R. Prentice (1990): *Akaike information criterion statistics.* John Wiley and Sons, New York, Wiley Series in Probability and Mathematical Statistics.

Kau, J., and D. Keenan (1998): "Patterns of rational default," Working Paper, University of Georgia.

Li, D. (2000): "On default correlation: a copula function approach," *Journal of Fixed Income*, 8(1), 43–54.

McCullagh, P., and J. Nelder (1995): *Generalized linear model - Second Edition.* Chapmann and Hall, New York, Monographs on Statistics and Probability 37.

Opsomer, J., and G. Kauermann (2000): "Weighted local polynomial regression, weighted additive model and local scoring," Working paper, Iowa State University.

Protter, P. (1990): *Stochastic integration and differential equations.* Springer, New York.

Santos Silva, J., and J. Murteira (2000): "Estimation of default probabilities using incomplete contracts data," Draft paper, Universidade de Coimbra.

Schimek, M., and B. Turlach (2000): *Additive and generalized additive models.* John Wiley and Sons, New York.

Seber, G., and C. Wild (1989): *Nonlinear regression.* John Wiley and Sons, New York, Wiley Series in Probability and Mathematical Statistics.

SHUMWAY, T. (1999): "Forecasting bankruptcy more accurately: a simple hazard model," Working paper, University of Michigan.

SMITH, L., S. SANCHEZ, AND E. LAWRENCE (1999): "A comprehensive model for managing credit risk on home mortgage portfolios," *Decision Sciences*, 27(2), 291–317.

SNYDER, D. (1975): *Random point processes*. John Wiley and Sons, New York.

VAN DER LINDE, A. (2000): "Variance estimation and smoothing-parameter selection for spline regression," in *Smoothing and regression*, ed. by M. Schimek, chap. 2. John Wiley and Sons, New York.