Understanding Default Risk Through Nonparametric Intensity Estimation

Fabien COUDERC
University of Geneva and FAME

Research Paper N°141
May 2005

FAME - International Center for Financial Asset Management and Engineering

UNIVERSITÉ DE GENÈVE

Understanding Default Risk Through Nonparametric Intensity Estimation

F. Couderc*

First Version: 10/02/2003

This Version†: 03/30/2005

Abstract

This paper investigates instantaneous probabilities of default implied by rating and default events. We propose and apply an alternative measurement approach to standard cohort and homogenous hazard estimators. Our estimator is a smooth nonparametric estimator of intensities, free of bias and unambiguously more accurate. It also avoids the Markovian framework and takes care of censoring. Using Standard & Poor's ratings database we then show that intensities vary both with respect to calendar time and ageing time. We deeper investigate the behaviour of through-the-cycle default probabilities, update and complement knowledge on documented non Markovian patterns. Results do not support associated timeliness problems but indicate a low reactivity of ratings in terms of magnitude. Because of their target horizon, they indeed integrate the mean reverting feature of default intensities.

JEL Classification: C14, C41, G20, G33.

Keywords: default intensity, hazard estimation, censored durations, non Markovian framework, through-the-cycle ratings.

---

* FAME & University of Geneva.

† I gratefully acknowledge data support by Standard & Poor's. In addition, I thanks financial support from the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management (NCCR FINRISK Project 5) and from the Geneva Research Collaboration Foundation (GRC). Olivier Renault, Arnaud de Servigny, Olivier Scaillet, René Stulz and Jean-David Fermanian are particularly thanked for helpful comments, as well as FAME colleagues. The views expressed therein are mine but not necessarily those of Standard & Poor's or any other institution. All remaining errors are mine. Contact information: FC@fame.ch.

Introduction

Credit research has been extremely active for proposing pricing and risk management models, but empirical studies of their inputs, and more precisely of default probabilities and correlations, were still rare events up to the last two years. Recent literature brushes away the cobwebs from migration matrices estimation techniques, searching for accurate and tractable methodologies. These works first comply with difficulties we have to face on data: confidentiality, short time series and sometimes reduced cross-sectional observations. Second, they rely on a priori, or conveniencies on the default process and ratings: rating systems are represented as Markov chain, issuers with the highest quality should not be zero probability defaulters, and so on. Yet it is well known that the data does not necessarily support such assumptions, or even violates the Markovian property. Among others, Kavvathas (2000) or Lando & Skodeberg (2001) documented non Markovian patterns due to durations spent in a previous rating class or up to the measurement point. Bahar & Nagpal (1999) exhibit dependency with respect to the previous rating change, referred to as the rating momentum.

As current pricing and risk management models are highly sensitive to default probabilities and to default correlations, researchers and practitioners now try to provide cautious signals on their estimates. Christensen, Hansen & Lando (2004) examine confidence levels on one year homogenous hazard probability estimates. Gordy & Heitfield (2004) study efficient estimation using cohorts. Over the same one year horizon Jafry & Schuermann (2004) compare these standard estimation techniques through various matrix metrics and present implied capital requirements. But all these papers concentrate on estimation techniques ground on the Markov property. Surprisingly, even if this hypothesis has been shown to be misspecified, alternative approaches are still missing. There is no real attempt to encompass the two time dimensions of the default risk: firms are exposed to business conditions (referred to as calendar time effects) but their reaction capacity or implications of the economy may differ according to the firm's various ages (in the whole rating process or in particular rating classes, referred to as ageing effects). Yet the time at which the firm operates (calendar time) and from which it started to operate (ageing time) may have distinct implications. The latter effects are usually assumed to be constant. As a consequence, little is known about profiles of default probabilities through the rating process and through time. Documented Markovian violations might be only due to a spurious interpretation of the rating system which is not built by agencies as a Markov chain, but which rather try to provide a cross-sectional ranking of firms. In addition, the traditional cohort and homogenous hazard estimators cannot be used to correctly investigate the temporal span of these non Markovian patterns.

Meanwhile, several discussions around the so-called through-the-cycle default probabilities implied by rating agencies' outputs have been opened. Differences in estimates along with the business cycle (e.g. see Nickell, Peraudin & Varotto (2000)) cast doubt on this through-the-cycle (TTC) label. With the Basel II banking regulation rules and the fear to amplify the business cycle, a good understanding of these through-the-cycle default probabilities becomes critical. Some researchers (e.g. see Loffler (2004)) advocate for their use to compute capital requirements because these probabilities are more stable than point-in-time default (PIT) probabilities (usually internally computed by banks or from the KMV approach). Others are more prudent because of their reported poor capacity to predict the default events or to explain credit spreads (e.g. see Altman & Kao (1992)). We claim that it is necessary to rely on an other kind of estimators, so as to infer default probabilities from agencies' outputs not flawed by structural misspecification and to understand what they are capturing before using them in pricing or risk management models.

In this paper we choose to take the point of view of durations in order to derive default probabilities. It allows to recover default intensities from rating changes and default events. Intensities can be interpreted in terms of probability of default on the very short run. The continuous-time homogenous hazard technique is a parametric example of intensity estimation (it assumes that intensities are constant). Worthwhile instruments, intensities provide clear treatments of censoring and truncation mechanisms which affect credit data. But first and foremost, durations do not necessarily call upon the Markovian assumption and can embrace several point of views over the rating process. We can for instance either look at durations within rating classes (i.e. conditional on "no rating change") or along with the whole rating process. Close to this study, Fledelius, Lando & Nielsen (2004) propose a way to separate ageing and calendar time effects. However, they use a biased smoothing technique which requires in addition discrete time adjustments and which do not clearly handle censoring.

Contributions of this study are both methodological and empirical. From empirical prospects, we first deliver new insights on the behaviour of default probabilities in the rating process under the objective measure. We point out that rating agencies success in preserving cross-sectional ranking through time. Second, we show that both ageing and calendar time effects coexist generating large variation in instantaneous probabilities for all rating classes. Our results justify conditional modelling of default probabilities using time series of covariates. Third, we update and complement knowledge on deviations from the Markovian hypothesis. In particular we show the temporal span of the rating momentum and its mean reversion. We finally evidence that ratings do not lag underlying probability changes but rather under react in terms of magnitude because their target horizon matches the mean reverting time of instantaneous default probabilities. In other words, our results suggest that the TTC label would be more appropriated when referring to the credit cycle.
From methodological prospects, we first rely on a non Markovian setup thanks to durations. Second, we introduce a new smooth nonparametric estimator of default intensities and probabilities, and establish its consistency and asymptotic normality. The estimator is based on Gamma kernels which do not suffer from the boundary bias present in symmetric kernels. We show that the impact of this bias can lead to unacceptably large errors in one year default probability estimates when using smooth kernel estimators (thus allowing intensities to vary with ageing time). In a bivariate version, we provide a way to derive the whole dependence structure (the bivariate survival function and its copula function) taking into account censoring.

The first section of the paper defines the intensity estimator and gives its asymptotic properties. We check robustness of our estimator and draw comparisons with nonparametric estimators proposed earlier in the literature using simulations. The second section describes the

Standard & Poor's database and discusses its characteristics. We investigate default intensities in the third section. We exhibit the shape of intensities, discuss the role of ratings and figure out calendar time effects. Then, we explore and comment known non Markovian patterns. The fourth section extends the estimator to the bivariate case so as to obtain the dependence structure. A final section concludes.

I. A Smooth Nonparametric Intensity Estimator

The technology we employ to estimate default probabilities relies on kernel estimators of intensities as introduced by Ramlau-Hansen (1983) (see also Yandell (1983) and Tanner & Wong (1983)). We modify their estimator in order to accommodate for an asymmetric Gamma kernel. The Gamma kernel has a major advantage as it allows for the estimation on the short run (i.e. on the boundary at zero[1]), whereas symmetric kernel estimators can only provide estimations on the interior of the support. In the following, we define the Gamma kernel intensity estimator, give its asymptotic properties and assess its performance over its competitors. Its advantages will constitute essential elements for a correct evaluation of credit risk since as we will see other smooth estimators are not reliable. In addition, homogeneity and Markov assumptions used by industry standard cohort and hazard rate estimators are relaxed. Further technical details, proofs and comments are gathered in appendix.

A. The GRHE estimator

Assumptions

As we have already underlined we look at distributions of durations, or times-to-default. Subsequent developments are based on standard assumptions and techniques used in the counting process literature[2]. To make the reading easier, we keep the time-to-default prospect without loss of generality concerning this intensity estimator, and focus on a given risk class. First, we put forward a setup allowing for a correct definition of intensities.

Assumption I.1 *All firms in the studied risk class are assumed to be homogenous and independent.*

Assumption I.2 *The censoring mechanisms are random and independent from the default process. They will be reported through a process $Y(u)$.*

[1] Symmetric kernels such as the standard Gaussian kernel, allocate weight over both negative and positive spectra. This leads to a bias at zero which is the lower bound for duration data. However this bias has subtantial repercussions over a domain which includes at least $[0, b[$, b being equal to the smoothing parameter. For further details on asymmetric kernels and their properties, see Chen (2000) or Bouezmarni & Scaillet (2001).

[2] See Andersen & al. (1997) for a complete presentation of counting processes estimation methods. They also deliver a complete survey on the multiplicative intensity specification and on the Nelson-Aalen estimator of the cumulative hazard rate used hereafter.

$Y(u)$ is left unspecified at this stage and will be clarified later on. Let D_i denote the uncensored duration up to default of firm i and C_i the censored duration. For a pool of n firms, Assumptions I.1 and I.2 imply respectively that $D_1, .., D_n$ (resp. $C_1, .., C_n$) are identically and independently distributed, and that censoring is uninformative.

Let N_u^i be the counting process associated with a given firm i in the risk class, and $\mathcal{F}_u^i$ its natural filtration augmented by its censoring indicator. This process is initially null when the firm enters the class and jumps to one when it defaults. We also set $U_i = \min(C_i, D_i)$ the time at which the firm leaves the class either because of censoring (C_i) or because it defaults (D_i). The U_i's, jointly with the indicators of censoring, are the true observations. In this context, provided that N_u^i is a non-explosive point process adapted to $\mathcal{F}_u^i$, we assume further that there exists a predictable non-negative process $\lambda^i(u)$ with $\int_0^u \lambda^i(s)\, ds < \infty$ such that

$$M_u^i = N_u^i - \int_0^u \lambda^i(s)\, ds$$

is a martingale adapted to the filtration $\mathcal{F}_u^i$. $\lambda^i(u)$ is the $\mathcal{F}_u^i$-intensity[3] of the process N_u^i, or equivalently the intensity of default associated with the firm i. Intensities can be interpreted as probabilities of default on the very short run (let us say over one day for durations around 10 years).

For estimation purposes, we then consider the aggregate process $N_u = \sum_i N_u^i$ which counts the total number of defaults in the class. We define a general filtered probability space $(\Omega, (\mathcal{F}_u), P)$ whose filtration $\mathcal{F}_u$ is the σ-algebra generated by the individual counting processes N_u^i and censoring indicators: $\mathcal{F}_u = \sigma\left(\mathcal{F}_u^1 \vee .. \vee \mathcal{F}_u^n\right)$[4] (when necessary, this filtration generated over all individuals from class I will be denoted by $\mathcal{F}_u^I \equiv \mathcal{F}_u$). Additional assumptions are required :

Assumption I.3 *All firms default eventually, i.e.* $D_i < \infty, \ \forall i$.

Assumption I.4 *The intensity of individual counting processes satisfies the Multiplicative Intensity Model:*

$$\lambda^i(u) = \alpha(u)\, Y^i(u)$$

where $\alpha(u)$ *is deterministic and called the hazard rate, whereas* $Y^i(u)$ *is a predictable and observable process.*

$Y^i(u)$ is then typically given by[5] $Y^i(u) = \mathbb{I}_{(u^- < C_i)} \mathbb{I}_{(u \leq D_i)}$, that is, denotes the presence of the firm in the observed sample up to u days into the class. Along with this formulation we also state (through u^-) that default occurs before censoring.

[3] From a probabilistic point of view the intensity accounts for the expectation of the jump in the counting process at time u given information at u^- : $\mathbb{E}\left[dN_u^i \middle| \mathcal{F}_{u^-}^i \right] = \lambda^i(u)\, du$. See Bremaud (1981) for further technical details.

[4] We recall that $\sigma(\mathcal{A} \vee \mathcal{B})$ is the smallest σ-algebra containing both $\mathcal{A}$ and $\mathcal{B}$.

[5] In the following, $\mathbb{I}(A)$ stands for the indicator function on set A.

As a consequence of Assumptions I.1 to I.4, N_u is a non-explosive counting process adapted to $\mathcal{F}_u$ with intensity $\lambda(u)$ defined by

$$M_u = N_u - \int_0^u \lambda(s)\ ds = N_u - \int_0^u \alpha(u)\, Y(u)\ ds \tag{1}$$

where M_u is a $\mathcal{F}_u$-martingale and $Y(u) = \sum_i Y^i(u)$.

Definition

In this framework, we are interested in the estimation of the baseline hazard function $\alpha(u)$. Calling $\Lambda(u)$ the cumulative hazard function which integrates $\alpha(u)$, denoting further $\widehat{\Lambda}(u)$ the Nelson-Aalen estimator of $\Lambda(u)$, the Ramlau-Hansen estimator $\overline{\alpha}(u)$ is then defined by

$$\overline{\alpha}(u) = \frac{1}{b}\int_0^\infty K\left(\frac{u-s}{b}\right) d\widehat{\Lambda}(s) = \frac{1}{b}\int_0^\infty K\left(\frac{u-s}{b}\right)\frac{J(s)\ dN_s}{Y(s)} \tag{2}$$

where K is a symmetric kernel with support on $[-1,1]$, b being the bandwidth (or equivalently the smoothing parameter) and $J(s) = \mathbb{I}_{(Y(s)>0)}$ guaranteeing existence of the Nelson-Aalen estimator.

We modify the estimator by substituting an asymmetric Gamma kernel for the symmetric kernel $K(.)$. The main reason lies in the fact that the Ramlau-Hansen estimator is not properly defined on $[0, b[$. For instance, considering rating classes above CCC, a meaningful b (in the sense of cross-validation - see the appendix) is about 2 years implying that we cannot evaluate consistenly short term default probabilities. We correct for this unacceptable drawback.

Definition I.1 *Under Assumptions I.1 to I.4, the Gamma kernel estimator $\widehat{\alpha}(u)$ of the hazard rate (Gamma Ramlau-Hansen Estimator, GRHE) is defined by*

$$\widehat{\alpha}(u) = \int_0^\infty \frac{s^{u/b}\, e^{-s/b}}{b^{u/b+1}\,\Gamma\left(\frac{u}{b}+1\right)} d\widehat{\Lambda}(s) = \int_0^\infty \frac{J(s)}{Y(s)} \frac{s^{u/b}\, e^{-s/b}}{b^{u/b+1}\,\Gamma\left(\frac{u}{b}+1\right)} dN_s. \tag{3}$$

The empirical counterpart of the estimator replaces the integral by a sum over observed default durations and $Y(u)$ is computed as the number of firms for which the last time of observation U is greater than u. $Y(u)$ is usually described as the risk set and handles censoring. Intuition behind this nonparametric estimator is as follows: probabilities of default on the very short run are estimated as a weighted average of past (if any), current and subsequent observed instantaneous probabilities of default. The weights are determined by the kernel, the bandwidth, as well as by the durations between the event and other default events. We recall that survival probabilities can then be retrieved as $P(U > u) = \exp\left(-\int_0^u \widehat{\alpha}(u)\ du\right)$. In addition, this estimator can be used for nonparametric specification tests in the framework of Fernandes & Grammig (2002).

Asymptotic Properties

The following propositions state that the GRHE is asymptotically unbiased both on the interior and at the zero boundary. Technical assumptions are precised in appendix so as to ligthten the exposition.

Proposition I.1 *Under assumptions (A1), the GRHE is pointwise consistent and asymptotically unbiased for all $u \in [0, \infty[$.*

Proposition I.2 *Under assumptions (A2), we get in the interior $\left(\frac{u}{b} \longrightarrow \infty\right)$*

$$\sqrt{n\sqrt{b}}\left(\widehat{\alpha}(u) - \alpha(u)\right) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}\left(0, \frac{\alpha(u)}{\tau(u)\sqrt{u}2\sqrt{\pi}}\right)$$

and at the boundary $\left(\frac{u}{b} \longrightarrow \kappa\right)$

$$\sqrt{nb}\left(\widehat{\alpha}(u) - \alpha(u)\right) \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}\left(0, \frac{\alpha(u)\ \Gamma(2\kappa+1)}{\tau(u)\ 2^{1+2\kappa}\Gamma^2(\kappa+1)}\right)$$

$\tau(u)$ is explicitly defined in appendix and represents the survival probability just before time u, taking into account censoring. In other words, it corresponds to the distribution of the U_i's[6]. In our case κ is equal to zero. Proposition I.2 implies that in most of situations[7] accuracy declines when one considers increasingly long durations as $\tau(u)$ tends to zero. This loss of power is more severe as riskiness increases, although we emphasize that contrary to other smooth estimators, the Gamma kernel partially limits this global decline because of the presence of the square root of the duration at the denominator of the variance. In order to present realistic confidence intervals, through this paper we will generally rely on small sample bootstrapped intervals rather than on asymptotic confidence intervals.

Small sample confidence intervals can be obtained using Efron (1981) bootstrap procedure. This approach consists of rebuilding samples from both the observed defaults and the censoring indicators. Hence, both durations and censoring are jointly drawn with replacement. The highly censored nature of default data makes this joint drawing crucial (see Strawderman and Wells (1997)).

[6] Remark that in our case of independent censoring, if $S_D(u)$ is the survival function of the D_i's and $S_C(u)$ the survival function of the C_i's, then the survival function of the U_i's is given by $\tau(u) = S_D(u)\ S_C(u)$.

[7] More precisely, it will not be the case if the intensity has a finite support (for instance if we are sure that all firms default before ten years) and if intensities decrease faster than the survival function of the U_i's. These are not realistic conditions for credit risk.

Remark that kernel estimators are known to be quite sensitive to the choice of bandwidth, i.e. to the degree of smoothness of the estimator. In our empirical applications, we compute the optimal bandwidth by cross-validation using a grid-search algorithm. More specifically, we used the well-tried leave-one-out method (see Wand & Jones (1995) for a discussion of bandwidth selection methods) to determine the bandwidth which minimises the mean of squared in-sample prediction errors.

B. Comparison with Standard Estimation Techniques

Cohort and Homogenous Hazard Methods

There are two popular approaches to evaluate default probabilities. The industry standard is the well know cohort method. Probabilities of default at a given horizon are computed as the ratio of defaulting firms over firms in the sample at the begining of the estimation window. Typically, rating agencies produce one year probabilities defining a calendar year as the estimation window. Notice that the cohort technique ignores censoring and assumes that the homogenous Markovian property is fulfilled: whatever the time spent in the class before the 1st of January of each year, all firms in a class are exposed to the same default probability over the year. From a duration prospect, it corresponds to an exponentiality assumption (i.e. constant intensity). Evidences against such a behaviour have largely been documented (e.g. Kavvathas (2000) exhibits dependence of such probabilities with respect to past duration). The so-called homogenous hazard rate is another popular methodology, widespread among academics. Here, the term homogenous is synonymous with constant. Keeping the same basis, the technique focuses on intensities and provides more efficient estimations as it takes into account left censoring. It follows partly the Nelson-Aalen insights in that way.

The GRHE relaxes both the constant hazard hypothesis and the Markovian setup, allowing the intensity to vary as duration increases. As standard estimators do not rely on the same framework, they cannot be used to assess the performance of the GRHE. We thus need to focus on other smooth estimators of intensities, but considering exponential durations we compare its outputs with the ones of cohort and homogenous hazard approaches.

Smooth Estimators of Intensities

Estimation of smooth hazard rate functions involves two challenges. The first one lies in the boundary biais, the second one in the fact that the variance of the estimator widens as duration increases. In order to tackle with these issues, Müller & Wang (1994) proposed an estimation procedure with adaptative kernels and bandwidths. The kernel family used in their approach is still with finite suppport, but the support reduces near the boundary and becomes non symmetrical, so as to decrease this boundary biais. Adaptative or local bandwidths also allow for diminishing the growth rate of the variance using higher bandwidths when the local number of observations is low. However the estimation process of such estimators is very costly, and the final accuracy highly depends on some ad-hoc inputs[8]. In addition their complexity prevents from the specification of multivariate estimators. The GRHE acts along with the same dimensions but keeping simplicity of evaluation.

In our simulations, we choose four benchmarks from kernel estimators of intensities: the Epanechnikov kernel, the biquadratic kernel, and their adaptative counterparts as defined by Müller & Wang (1994).

Simulation Results

Table I presents mean squared errors (MSE) and their decomposition for a constant unit intensity (i.e. exponential distribution for times-to-default). We also ran simulations over a decreasing intensity (Weibull distribution) and over a bell-shape intensity (log-normal distribution)[9]. Results are displayed for different uniform censoring schemes (10%, 20% and 40%) which cover the range found in our rating data.

[INSERT TABLE I HERE]

Results are clearly in favour of the GRHE. As pointed up, the Epanechnikov and biquadratic kernels constitute the more basic solution. They offer the same level of MSE with a little advantage to the Epanechnikov kernel as the censoring proportion increases. The use of adaptative kernels and bandwidths enhances the performance of the estimator by 30% to 20% but the GRHE goes beyond these estimators reducing the average MSE of the basic kernels by 50% to 30%. Most of the efficiency comes from a huge cutting-down of the bias. However, remark that enhancements decrease as the censoring increases. In terms of variance, the GRHE is similar to the adaptative Epanechnikov kernel. Let us observe in more details the behaviour of the different estimators from Figure 1, concentrating on the unit exponential distribution under 10% uniform censoring.

[INSERT FIGURE 1 HERE]

Picture 1(a) shows as expected that the GRHE operates the MSE reduction on the short term and in particular at the boundary. Additional lowering is realized on the long term whereas all estimators are equivalent around the mean of the distribution. Concerning the variance (Figure 1(b)) the GRHE offers higher variations on the short run but decreases long term variations as expected. Picture 1(c) displays the squared bias and picture 1(d) targets long term squared bias (which is far lower than at the boundary). From these simulations it is clear that the GRHE is unbiased on the short run but it slightly suffers from higher bias on the long run. This is due to the lower convergence rate of the GRHE with respect to its competitors and constitutes its main weakness.

[8] See the detailed estimation procedure in appendix.

[9] Results not reported here: findings do not differ from the exponential case. Moreover, different sample sizes do not modify the relative performance between estimators.

Remark that multiplying the duration axes by 1000 (and thus the squared bias by 1e-6) gives the magnitude of errors made by the other estimators for firms with a B- or CCC+ rating: with an average life of 34 months, the boundary bias has large impacts over the first 17 months in the class. Credit risk management cannot afford such errors over a so long period. In terms of probabilities, if we consider 1000 simulated samples of 100 individuals with 40% of censoring (37% of CCC observations are censored) and keeping this average default duration of 1000 days, the true one year probability is equal to 30.58%. Notice that the exponential case is the most favourable situation for both cohort and homogenous hazard techniques. Estimations from the Epanechnikov kernel displays a probability of 23.79%, its adaptative version of 27.84% and the GRHE of 30.86%. On the same samples, the cohort and the homogenous hazard estimator exhibit respectively an average default probability of 27.72% and 28.55%. It indicates that operating standard kernel smoothers can lead to disastrous results, and more interestingly that they are less accurate than techniques used by practitionners. However, the GRHE outperforms all other methods. From the one hand, it implies that the use of boundary unbiased estimators is crucial. From the other hand, it shows that both censoring and hazard non homogeneity matter.

II. Default Data Description

Ratings Data

Our ratings data was extracted from Standard & Poor's Credit Pro 6.6 database. This database contains S&P's rating histories for 10439 companies over the period January 1981 to December 2003. Overall 33044 rating migrations are recorded in Credit Pro as well as 1386 defaults and default rate ranges from 3% to 29% across industries. Remark that credit reviews concluding to no changes in rating are not reported in the database. Such data could indicate whether agencies revise their credit risk assessment more frequently during specific part of the business cycle, of the firm life cycle, or not. This could induce a bias in our analysis conditional on the business cycle as it would lead to sharper decrease/increase in estimated default intensities during peaks and throughs. This behaviour would also accentuate non Markovian patterns such as the rating drift, namely the fact that a firm which has been recently downgraded is more likely to be downgraded than another one. However, S&P argues that ratings are reviewed on a regular basis, and more frequently if substantial new information arises. A rating is only modified when the likelihood of default changes significantly and if this change is not purely transitory. The Credit Pro database has already been used and extensively described by Bangia & al. (2002) over the period January 1981 to December 1998. From that date the database has been strongly improved and in particular with respect to the industrial classification.

Within our sample, firms are classified by industrial groups and each of them has been refined by subindustry criteria. Practically, we have at our disposal 13 industries or 526 distinct subindustries distributed among 93 countries. But 6897 firms (66 %) are US ones. Moreover, S&P attributes 25 distinct ratings plus the NR one, but we aggregate the data coming from a grade and its plus/minus modifiers because of minimal population requirements. Besides, all grades below B- have been put in the CCC class. An important point to interpret following results lies in the fact that these ratings are mid term assessments of the credit riskiness of firms and should ideally not be altered by business cycle changes on this period. We expect that probabilities of default are quite stable over a 3 to 5 year horizon. That kind of ratings are called through-the-cycle (TTC) and assess the capacity to face with a given stress event. These ratings are usually opposed to bank internal rating systems which evaluate directly the marginal probability of default (PIT probabilities).

Censoring and Truncation

Rating events require careful treatment as three sources of censoring are present in the database. Figure 2 shows the number of firms entering into the rating process year by year, as well as the number of observed non investment grade firms (i.e. $Y(u)$) as duration increases. Left truncation arises from the fact that 1371 issuers had already received a rating before they were included in the database (i.e. before January 1981). We do not have information about the attribution date of their first rating and therefore for robustness checks we run all estimations both on the full sample and on the reduced sample excluding left-censored data (the reduced sample contains 9068 companies and 25993 rating migrations). Obviously, a first type of right censoring is also an inherent feature of any ratings database as most companies survive after the end of the recordings. Another specific type of right censoring requires specific consideration. Some companies leave the rating process and fall into the not-rated (NR) category. Several reasons may explain this fact: the rated company may be acquired by another firm or may simply decide no longer to be rated by S&P. The database has the nice feature to identify firms that migrated to NR and subsequently defaulted. The NR class is not a complete loss of information: although there is no more indications of credit quality, a NR firm is a non defaulter.

[INSERT FIGURE 2 HERE]

There is no obvious concensus on how to handle NR events, in particular for the purpose of calculating transition matrices. Accordingly it requires some discussion. Some authors assume that the migration to NR is a neutral event. They then redistribute the probability of migration to NR on all other rating classes by dividing the probabilities of migration to non-NR classes by (1-Pr(migration to NR)). Other authors take NR events as potential signs of weakness for a corporate (the withdrawal of information may be an attempt by the company to conceal its difficulties). They then reallocate the NR probability to downgrade states only. In fact, looking carefully at the database, one can observe that many companies falling into NR leave this class direcly for the default state. This represents about 15% of default observations. Thus we choose to apply a more efficient treatment to NR events, as allowed by the database. We erase NR events when we know that firms will go out from that state later, and we consider as right censored firms which stay into that state up to the 12/31/2003. This strategy keeps full information on default, even if it needs to assume that between corresponding NR events and defaults companies conserved their previous rating.

Times-to-Default

Let us notice that such a database allows to consider two types of durations, implying two different approaches to the behaviour of default probabilities. On the one hand, we can look at times-to-default from entry in a risk class up to the last available observation. This perspective constitutes the primary goal of ratings: ratings rank firms in cross-section with respect to their expected default probability over three to five years. Thus splitting the sample of firms on the basis of ratings should lead to homogenous classes of risk at these horizons. In this setting, default can then be considered as a binary process but with distinct distributions for each cross-sectional (rating) band. On the other hand, we can examine times-to-default conditional on staying in a given risk class up to the default time. By doing so, the emphasis is put on rating migrations giving stonger relevance to rating changes, and the default process is represented as a multistate process. Actually the difference lies in the way default probabilities are assumed to evolve. In the first case firms can change continuously their default probability and the evolution of corresponding intensities can be interpreted as debt life cycle. In the second case the default intensity should remain constant within a rating class and jump when the rating is revised. The latter corresponds to the standard Markov chain assumption. Remark finally that time profiles of intensities, if non constant, would reflect impacts of time spent within a rating class, a well know deviation from the Markov property.

III. Behaviour of Times-to-Default

From our knowledge, little is known about shape and evolution of default probabilities. Fledelius & al. (2004) made first analyses but the estimation technique they used suffers from boundary bias and oversmoothing, and lack of efficiency because of discrete time adjustments. Yet knowledge of the probabilities' patterns or equivalently intensity dynamics under the physical measure is of prime interest. First, risk managers require to model the evolution of realized default probability. Second, empirical intensities become a major input of pricing models, the form of risk premia being fixed (e.g. see Jarrow & al. (1997), Kijima & Komoribayashi (1998) or Acharya & al. (2002)), and in specific cases objective intensities being directly the critical quantity (e.g. see Bielecki & al. (2004) (2004) for an application to hedging). Remark that intensities features are also interesting for reduced-form modelling which assumes specific dynamics. These dynamics are most often chosen so as to provide convenient formulas but do not rest on a careful analysis of empirical behaviours.

Our framework lets us study these default intensities on a continuous time basis and to exhibit their movements through time, without relying on a spurious Markovian assumption or on biased inference techniques. Using unconditional durations to default and durations conditional on staying in a given risk class, we exhibit default intensity dynamics and provide analysis of the role of ratings. We then investigate the time span and potential fading out of some documented deviations from the homogenous Markov property.

A. Default Intensities and the Role of Ratings

Dynamics along with the Rating Process

Figure 3 considers durations up to the end of the firms' life. It reveals default intensities and their 95% bootstrap confidence intervals for investment grades (AAA to BBB, noted IG) and non investment grades (BB to C, noted NIG) risk classes. We also include corresponding survival probabilities.

[INSERT FIGURE 3 HERE]

As expected IG and NIG intensity curves have markedly different levels but overall exhibit large variations through time. Both intensities present a shape with two humps (or possibly three) that is recurring throughout our analyses as well as over single rating classes. A first explanation is economic or equivalently lies in calendar time effects. During the period covered by our sample, three US recessions have been observed, one at the beginning of each decade. We can conjecture that the observed humps show increases in intensities due to macro-economic conditions. The magnitude of this humps may reflect the numbers of firms present in the sample during each of these recessions. Figure 2(a) indeed shows that left truncated observed firms represented most of the companies which faced the first recession. The first hump may then be amplified by firms which entered the process after 1996 (they constitute almost 35 percent of the sample). The same remark applies to the other hump.

Alternatively this double bell shape may be explained by pure measurement issues and more specifically by the impact of censoring events on the estimator. Dramatic drops in the size of the risk set may create jumps in intensities[10]. Indeed, the first decrease in NIG intensity occurs roughly at the mean of censored observations. We could suspect that humps represent average life of firms or average times before a rating change, which both strongly lower the risk set size and highly raise instantaneous probabilities at subsequent default times. However, Figure 2(b) does not exhibit any sharp decrease in the size of the risk set which would provide support for this second hypothesis. Hence, we favor the first explanation.

[INSERT FIGURE 4 HERE]

Figure 4 shows survival probabilities over industrial sectors. If the global trend is similar for most of industries, some huge differences can be observed. At the top of secure sectors, we can find the financial institutions whereas the telecommunications can be found at the bottom. These graphs highly advocate the necessity to specialize modelling sector by sector.

[INSERT FIGURE 5 HERE]

[10] Remember that the number of firms at risk enters into the measure through the $\frac{1}{Y(s)}$ term.

Finally, Figure 5 shows that the shape of low speculative grade (B and CCC) intensity curves is globally downward sloping, while intensities for higher quality corporates are upward sloping with humps. These results indicate that usual distributions can not reproduce intensity patterns: exponential distributions (or equivalently the homegenous Markov assumption) imply flat intensities whereas Weibull, log-normal or log-logistic can only provide one bump. If some more regular functions could be used to fit survival probabilities (for instance, by analogy with interest rates, one could fit polynomial functions), it seems highly challenging to fully parametrize intensities in an unconditional setting.

Impact of Vintages

We have just put forward macro-economic shocks as a potential explanation for the two humps in intensity curves. Note that Fledelius & al. (2004) isolate ageing and calendar time effects on intensities, using a bivariate version of their estimator and integrating out marginal effects. In fact, they assume either a multiplicative structure for these effects or an additive one. They obtain almost flat intensities with respect to firm's age and very smooth evolution through calendar time, but our simulations manifest that these results are doubtful because of the smoothing technique they use[11]. To avoid marginal integration as they do, we choose to split our sample by year in which the firms entered the NIG class, i.e. by "vintages". These vintages additionally offer a first informal test of the necessity to model intensity conditioning on calendar time information, or equivalently to time series of covariates. Figure 6 presents results on quaterly sliding vintages from 1985 to 1989 and proves that it exists a term structure of default probabilities with strong variations through time, but also on the short and mid term durations. Default intensity variations are consequently the result of the joint impact of calendar time and ageing time effects, the former dominating on the long run, the latter on the short run.

[INSERT FIGURE 6 HERE]

The results are indeed strongly supportive of the hypothesis that humps are caused by macro-economic shocks. We can observe that the humps shift to the left as the year of reference increases. This implies that the peaks are observed at the same calendar time period. For instance, the first peak of the 1989 vintage occurs at a duration of about 2 years, while the first peak of the 1987 vintage is located at the 4 year horizon. Both intensity peaks therefore reflect the 1991 recession. The second hump also clearly corresponds to the 2000-2001 recession. The mid 1985 to mid 1986 vintages are peculiar as they exhibit a sharp drop in intensity over the first year before increasing progressively to reach a maximum at about five years (again corresponding to 1991). The initial fall in intensity may be attributable to the end of the default crisis in the energy sector which took place from 1987 onwards. Furthermore, the vintages clearly figure out a distinct short-mid term behaviour with respect to the age of the firm (time-to-default). Indeed, as we pointed out earlier, the intensity increases over the first

[11] Remark that their estimator has not been tested in our simulations. However it involves a double smoothing with Epanechnikov kernels and as a consequence should lead to a higher bias.

one to three years in the NIG class before reaching a somewhat average intensity level. This is also true for the vintages covering 1986, right after a first adjustment (i.e. a fast decrease). We find evidences of a specific short-mid term feature for all classes. It suggests that conditional on calendar time information, intensities should be monotonous and increasing (or decreasing for the B and CCC classes) as duration raises up. Clear consequences for potential parametric modelling are the following. Weibull distribution should be adapted and the constant hazard assumption might be valid on the long run. We also observe that peaks vary in magnitude and that crises also differ in length. Simple regime-switching models will struggle to capture these features.

Ratings: Preserving Risk Ranking

Rating agencies claim that the purpose of ratings consists of ordering firms with respect to the default risk they face. The difference in magnitudes between IG and NIG intensities from Figure 3 shows that initial ratings successfully rank firms. Figures 7(c) and 7(d) confirm this fact on rating classes. It additionnally suggests a global stability of the risk faced by a firm over its life along with the rating process as probability curves do not intersect even on the long run. Numerous researchers in corporate finance believe in the existence of a target leverage. Thinking about leverage as a rough measure of default risk, our results are in accordance with such a perspective.

[INSERT FIGURE 7 HERE]

Rating changes should then reflect changes in, or deviations from, the initial level of default risk through the life of the firm. Comparing intensities of unconditional durations with durations conditional on no rating change, proves that agencies succeed in preserving such a cross-sectional ranking. Figures 7(a), 7(b) and 8 display empirical conditional intensity and survival probability curves, as well as 95% bootstrap confidence intervals. Confidence intervals widen compared to Figure 3 as a smaller part of the corresponding default distribution can be observed. The two intensity curves still present two humps but differences can be observed in location and magnitude.

[INSERT FIGURE 8 HERE]

In particular, looking carefully at Figure 8(a), we observe that the IG intensity drops slightly at extremely short time horizons before increasing up to about 10 years. However it is noticeable that the IG intensity curve is almost flat for the first five years, significant changes arising at longer horizons. On the contrary the IG intensity is no longer stable over these first five years on Figure 3(a). Recall that the former tracks firms until either default or exit from IG or NIG classes, while the latter considers durations to default, irrespective of any intermediate downgrade or upgrade[12]. Hence, it is intuitive and in line with agencies claims

[12] Here we only consider the broad risk categories (IG and NIG), not strictly speaking ratings. Results are the same over individual rating classes.

to find that the IG (resp. NIG) intensity is lower (resp. higher) on Figure 8. On the same way, Figure 8(b) presents a higher first peak and the decrease between 5 and 15 years is less sharp. This can be attributed to firms surviving the first crisis (which materialises at around 5 years) or simply improving their solvency and being upgraded to IG. Their default intensity then becomes similar to other IG firms, thereby decreasing the intensity on Figure 3(b) after 5 years. That kind of seasoning effect is well-known in prepayment analysis of mortgage backed securities. Such findings still hold over rating classes. Intensities are more stable over the first five years within rating classes. But if it shows that agencies achieve to preserve risk ranking changing assigned ratings, we first have to underline that this stability spans closer to five than three years. This horizon confirms the "long term" label of through-the-cycle ratings (e.g. see Altman & Rijken (2004)) in comparison with the short term assessment given by point-in-time ratings. In addition, there are irregularities on the short run which constitutes a recurring pattern over all rating classes: IG intensity decreases during the first six months (Figure 8(a)) and NIG intensity strongly increases up to two years (Figure 8(b)). We already underline that these irregularities reflect firms' local (i.e. within a given class) ageing effects. Loosely speaking there are due to an informational adjustment during the first months in a new rating class.

Looking further at survival probabilities, humps coincide with local maxima of the slope. Comparing Figures 8(c) and 3(c), as well as Figures 8(d) and 3(d) demonstrates that the rating process successes to identify major changes in default risk. Indeed, if after five years 2% of firms which started rated as investment company died, up to 99.8% of firms which stayed as investment companies were still alive. The reverse is true for non investment company but at larger horizons manifesting difficulties to upgrade to IG. In particular, it is striking to observe that after 10 years around 50% of non investment companies died. From risk management purposes, the real 10 years default probability that one should consider for a firm which enter the market in the investment class should be 4.5% (given by Figure 3(c)) instead of the 0.5% observed from Figure 8(c).

B. Non Markovian Patterns

Some deviations to the Markovian assumption have been emphasized in the credit literature. Seminals papers from Nickell, Peraudin & Varotto (2000) and Bangia & al. (2002) show that unconditional on the economic cycle transition matrices can not be considered as Markovian whereas conditioning on the NBER classification of business cycles is meaningful. Bahar & Nagpal (1999) as well as Lando & Skodebeg (2001) argue that other non Markovian effects exist, like the so-called well documented rating momentum. But usual frameworks can only provide partial results on these effects. As our methodology does not ground on the Markovian property, we give additional information on these effects and in particular exhibit their persistency or reverting behaviour.

Rating Momentum, Rating Reactivity and Risk Classes Homogeneity

The homogeneity of risk classes is a tricky issue as it is fully correlated with the confidence that we are willing to put in the ratings. Yet documented non Markovian features cast doubt on this homogeneity. That is the reason why agencies have refined their initial range of ratings, adding 1-3 variants for Moody's or +/- for Standard & Poor's. In addition, the extended ratings were perceived as a partial answer to the debated low reactivity of ratings with respect to changes in firms' likelihood of default. Following results point out the links between these questions.

[INSERT FIGURE 9 HERE]

Figure 9(a) shows intensities conditioning with respect to downgrade/upgrade as well as the whole BB class intensities, looking at times-to-default over the whole rating process. A, BBB and B classes also exhibit this pattern but the difference in the first year is slightly lower than for the BB class. Recently downgraded firms face a higher risk, and consequently are more likely to be downgraded again than other firms in the same risk class. Conditioning on a downgrade indeed leads to higher intensity over the first two years. It proves that the rating momentum phenomenon is not an artefact due to agencies' rating methodology but is truly called by credit risk behaviour. Our estimates show that the rating momentum should also apply to upgraded firms, although to a lesser extent, as found by Bahar & Nagpal (1999). However, our methodology allows to investigate the time-span of these effects. We observe readjustments which take from one to three years: the light and normal curves converge. It confirms the existence of a mean-reverting pattern in the intensity of default. Remark that the difference between downgraded and upgraded firms is statistically significant[13] (Figure 11(b) displays the difference and its bootstrap 95% confidence interval). Interestingly original ratings of the firms still have an impact on the long term, after readjustment has taken place. From 4 years, the relation is reversed, indicating that after a while downgraded firms return to lower risk states. This is a clear evidence of non Markovian behaviour.

Nevertheless, if the rating momentum can now be justified by changes in default intensities, the issue lies in the way agencies incorporate the phenomenon. Looking at durations conditional on staying into the rating class, Figure 9(c) shows (for the BB case) that ratings are correctly changed in accordance with changes in default risk. Focusing on the first six years within the BB class we cannot observe sharp decreases or increases in intensities, nor convergence between upgraded and downgraded issuers. In particular, a low reactivity of ratings would have been indicated by variations over the first year in the class. Some researchers still conclude to a low reactivity of ratings. Because of capital requirements, it would be an undesirable feature. Results from previous studies are not clear. Altman & Saunders (2001) claim that ratings lag the business cycle and D'Amato & Furfine (2003) differentiate new ratings from existing ones. But such a behaviour would be a problem only if the business cycle and the evolution of default

[13] For comparative purposes between different intensity curves, we compute pointwise confidence interval on distances by bootstraping both curves and looking at all combinations of bootstrap samples. The interval is given by the upper and lower 5% among all computed differences.

probabilities were co-cyclical. There is no evidence of such a linkage (e.g. see Koopman & Lucas (2004)). The issue is consequently more closely related to the reactivity of ratings to changes in underlying default probabilities than to timeliness. Our results suggest that the rating reactivity problem resides in the magnitude of rating changes rather than in the time at which the event occurs as Figure 9(c) clearly shows that the standard classes are not homogenous with respect to previous ratings. From bootstrap samples, the difference between the green and red curves is statistically different: downgraded firms have always higher intensities than upgraded ones during the period when they share the same rating. One may believe that the extended rating range correct for this lack of granularity. The database does not allow us to consider more disaggregated classes to check for such an improvement. Notice that given the disastrous consequences that a rating change can induce on firms asset, a straight explanation for this "step by step" phenomenon would say that agencies are frightened of large changes in ratings. Figure 9(c) still advocates for another explanation. As the situation does not reverse even after 10 years, as it was the case on Figure 9(a), it is more likely that agencies anticipate future reversals or wait for further information on the firms (the recently launched Watchlists and Outlooks of Moody's confirm surch a perspective, see Hamilton & Cantor (2004)). Corroborating this phenomenon, Löffler (2004) shows that ratings react on permanent shocks in default probabilities rather than to transitory shocks. Hence, issues concerning potential lagging phenomena between the credit cycle and capital requirements should not be due to a lack of granularity (e.g. see Altman & Saunders (2001)) but rather to an inadequacy between the horizon of ratings and the one of capital requirement rules.

The Macro-Economic Conditions

We now analyse further how intensities behave when conditioned on macro-economic conditions. We used NBER classification to pool firms in "expansion" or "contraction" periods when they enter into the rating process. Figure 10(a) presents intensities and Figure 11(a) plots differences in intensities between expansion and contraction, for non investment grades.

[INSERT FIGURE 10 HERE]

[INSERT FIGURE 11 HERE]

As expected in recessions firms follow countercyclical intensities, showing conservatism of analysts whereas a peak is reached between 4 to 6 years after the crisis. This lag may correspond to the length of the default procedure as advocated by Altman (1989). Moreover we can see that survivors to crisis are less risky than others on the long run, which constitutes a very intuitive result. Statistics on these differences confirm these findings. Notice that the same study on durations up to the last day in the sample delivers the same patterns but differences are only significant between 10 to 13 years. Indeed in that case the sample includes firms which have been quickly upgraded to IG on the short run and firms which have been downgraded on the long run, smoothing the differences. The data used in Figure 10 also indicates that the "speed of ratings" increases in recession. The average time-to-default from a recession period is 1237 days compared to 1654 in a growth period. Similarly durations to exit (which include migrations to IG and NR) decrease to 1421 days from 1656 in an expansion period.

These results are consistent with findings of Nickell, Peraudin & Varotto (2000). Over one year, the probability of default is indeed higher in expansion than in contraction. However, considering the target horizon of ratings, integrating intensities shows that three to six years default probabilities do not exhibit any statistical difference. It justifies the through-the-cycle label of these ratings: the business cycle can induce variations on the short term but rating classes, or equivalently cross-sectional ranks, are stable on the mid term. Moreover, it matches the results of Altman & Rijken (2004) who find an adequacy between point-in-time default probabilities and through-the-cycle default probabilities at a six years horizon.

IV. Nonparametric Estimation of the Dependence Structure

Up to now we focused on univariate estimation. A major concern in pricing of credit portfolios and complex derivatives consists of modelling the dependence between the different identified risk classes. One needs in particular to estimate correlations and maybe the full dependence structure through copula functions. For that purpose, we extend our univariate setting to the bivariate case keeping the main features of the GRHE, the absence of boundary bias and the treatment of censoring. In two dimensions, bias effects and censoring become more pronounced. Then, we briefly described estimation of the bivariate survival and copula functions.

A. Multivariate Hazard Estimation

We concentrate on two risk classes. Consider a firm i which belongs to the risk class I and a firm l which belongs to the risk class L. Assumption I.1 does not preclude possible dependence among firms in the same class. Moreover, we claim that the independence hypothesis contained in this assumption can be relaxed by an equidependence hypothesis[14] without implying any modifications on the consistency and the unbiased property of the estimator (it modifies the asymptotic variance). It then allows for estimation of the dependence within rating class. Let us define the bivariate default counting process by $N^{i,l}(u_i, u_l)$. This process equals one if and only if both default of company i has been observed before u_i and default of company l has been observed before u_l. Therefore, as soon as one marginal time-to-default is censored, the process can no more jumps to one. This process allows us to study the joint intensity of default. Once again we rely on the multiplicative intensity model:

Assumption IV.1 *The bivariate intensity of the counting processes $N^{i,l}$ satisfies the Multiplicative Intensity Model:*

$$\lambda^{i,l}(u_i, u_l) = \alpha(u_i, u_l)\, Y^{i,l}(u_i, u_l)$$

[14] See Gourieroux & Monfort (2002) for details on the equidependence assumption.

where $\alpha(u_i, u_l)$ is deterministic and called the bivariate hazard rate, whereas $Y^{i,l}(u_i, u_l)$ is a predictable and observable process.

As before $Y^{i,l}(u_i, u_l)$ handles censoring and ensures correct definition of the intensity up to the joint default event, that is $Y^{i,l}(u_i, u_l) = \mathbb{I}_{\left(\left(u_i^- < C_i \,\cap\, u_i \leq D_i\right) \,\cap\, \left(u_l^- < C_l \,\cap\, u_l \leq D_l\right)\right)}$. Under Assumption I.2, the indicator function can be split and rewritten

$$Y^{i,l}(u_i, u_l) = \mathbb{I}_{(u_i \leq D_i \,\cap\, u_l \leq D_l)} \mathbb{I}_{\left(u_i^- < C_i\right)} \mathbb{I}_{\left(u_l^- < C_l\right)}.$$

Note that involved processes are scalar-valued. Besides, $N^{i,l}$ corresponds to the product of individual counting processes $N^{i,l}(u_i, u_l) = N^i(u_i) N^l(u_l)$, but facilitates the definition of a so-called bivariate intensities. Pedroso, De Lima & Pranab (1997) studied a vector-valued counting process $\left(N^i(u_i), N^l(u_l)\right)$ but were unable to exhibit a proper bivariate intensity. Our framework is closely related to the early paper of Dabrowska (1988) and matches works of Fermanian (1997). Dabrowska proposed an estimator of the joint distribution of times-to-default considering the partial ordering $(\mathbf{s} \leq \mathbf{u}) = (s_i \leq u_i \,\cap\, s_l \leq u_l)$. This set up allows to define a bivariate survival function by $S^{i,l}(\mathbf{u}) = P(D_i > u_i; D_l > u_l)$. Dabrowska built his estimator on the ground of a multivariate version of the Kaplan-Meier (1958) estimator which becomes tricky. On the contrary, Fermanian proposed an estimator using the nice features of a smooth version of the Nelson-Aalen estimator. The estimator we propose below is similar but replaces the symmetric kernel with the Gamma kernel, exactly as we modified the standard Ramlau-Hansen estimator. As in the univariate case we consider the bivariate counting process over the entire classes $N^{I,L}(\mathbf{u}) = N^{I,L}(u_I, u_L) = \sum_{i \in I} \sum_{l \in L} N^{i,l}(u_i, u_l)$. Under the filtration $\mathcal{F}_u = \mathcal{F}_u^I \vee \mathcal{F}_u^L$, the intensity of this process can be defined as

$$\lambda^{I,L}(\mathbf{u}) = \alpha(\mathbf{u}) Y^{I,L}(\mathbf{u})$$

where $Y^{I,L}(\mathbf{u}) = \sum_{i \in I} \sum_{l \in L} Y^{i,l}(\mathbf{u})$. Then the baseline hazard can be estimated through observation of $N^{I,L}$.

Definition IV.1 *Under Assumptions I.1-I.3 and IV.1, we define the bivariate Gamma kernel estimator $\widehat{\alpha}(\mathbf{u}) = \widehat{\alpha}(u_I, u_L)$ of the bivariate hazard rate (Bivariate Gamma Ramlau-Hansen Estimator, BGRHE) by*

$$\widehat{\alpha}(\mathbf{u}) = \int_0^\infty \int_0^\infty K\left(s_I, \frac{u_I}{b} + 1, b\right) K\left(s_L, \frac{u_L}{b} + 1, b\right) \frac{J^{I,L}(\mathbf{s})}{Y^{I,L}(\mathbf{s})} N^{I,L}(ds_I, ds_J) \tag{4}$$

where $J^{I,L}(\mathbf{s}) = \mathbb{I}_{(Y^{I,L}(\mathbf{s}) > 0)}$

Convergence and asymptotic properties of the estimator follow from the study of Fermanian (1997) and techniques used in appendix to prove the properties of the GRHE. This bivariate version clearly smooths over the durations from two risk classes. Previous interpretations still hold. Remark that Fermanian investigates higher dimensional scalar-valued hazard rates. We do not extend the definition of the BGRHE to the general case because kernel estimators suffer from the curse of dimensionality as the convergence rate critically decreases with the number of dimensions, and because in practice we are not able to easily analyze (and visualize) higher dimensional dependence structures.

B. Dependence Structure

In this setting, so as to derive the bivariate survival function, we also require estimation of the hazard rate of the risk class I given survival of individuals from the class L.

Definition IV.2 *Under assumptions I.1-I.3 and IV.1, we define the Gamma kernel estimator* $\widehat{\alpha}_I(\mathbf{u}) = \widehat{\alpha}(u_I, u_L)$ *of the I-class hazard rate conditional on the L-class survival (Conditional Gamma Ramlau-Hansen Estimator, CGRHE) by*

$$\widehat{\alpha}_I(\mathbf{u}) = \int_0^\infty K\left(s_I, \frac{u_I}{b} + 1, b\right) \frac{J^{I,L}(s_I, u_L)}{Y^{I,L}(s_I, u_L)} N^{I,L}(ds_I, u_L) \tag{5}$$

There is indeed no direct relationship between the bivariate survival function and the bivariate hazard. However, Dabrowska (1988) proposes simple derivations using conditional hazard rates. The survival function can be expressed in the two dimensional continuous case by

$$S(\mathbf{u}) = S_I(u_I)\; S_L(u_L)\; \exp\left(\int_0^{u_i}\int_0^{u_l} \alpha(\mathbf{s})\; ds_l\, ds_i - \int_0^{u_i}\int_0^{u_l} \alpha_I\;(\mathbf{s})\; \alpha_L\;(\mathbf{s})\; ds_l\, ds_i\right)$$

$S_{.}(u_{.})$ being marginal survival functions, and estimated as $\widehat{S}(\mathbf{u})$ using the BGRHE, the CGRHE and the GRHE while keeping the same degree of smoothing.

From that stage, using the survival function version of the famous Sklar's Theorem, the survival copula function as well as the corresponding copula function can easily be infered. Remark that methods to estimate nonparametrically the copula function from kernel estimators of density functions have already been proposed (e.g. see Fermanian & Scaillet (2003)). However, these methods cannot handle for the censoring properties of duration samples.

Summary

We examine default intensities, default probabilities and the role of ratings from the Standard & Poor's rated universe. Even if standard cohort or homogenous hazard estimators are highly convenient, they rely on an assumption that is not required by rating systems and which is known to be empirically violated: the Markov property. In particular, recent studies show that the Markov property is not verified when conditioning on the business cycle, or

conditioning on the previous rating. As cohort and homogenous hazard approaches, or more generally the Markov chain representation, are only reporting tools, we examine whether these patterns are artefacts due to misspecified estimation techniques of default probabilities or not.

We thus propose a nonparametric estimator of instantaneous default probabilities which do not necessitate a Markovian framework so as to investigate the evolution of default probabilities through time. We build this estimator in a duration framework. It is free of bias, highly tractable and allows for straightforward estimation of default intensities across rating categories, industries and time periods, taking into account censoring properties of credit events and credit data. Unlike the homogenous hazard estimator, the default intensity can vary as the time spent in a rating class increases. Through simulations, we prove the superior behaviour of this Gamma kernel estimator (GRHE) in comparison with existing smooth hazard estimators. We show that its unbiased feature is of first importance. Indeed, we find that other smooth estimators (not relying on a Markovian setup) are not successful in estimating realistic default probabilities because the boundary bias that they suffer from, has large impacts over horizons up to 2 to 3 years. Moreover, looking at standard estimators we show that both relaxing the Markovian setup and taking into account censoring matter. The GRHE achieves a better accuracy and can provide reasonable pictures of default probabilities, on the short, mid and long run.

We investigate the behaviour of firms in the rating process focusing on the primary goal of ratings, i.e. on the basis of the cross-sectional risk ranking that they deliver. We find that ratings are successful not only in ranking firms when they enter into the rating process but also in preserving this rank afterwards. Through-the-cycle ratings correctly split the credit risk universe at a 3 to 5 year horizon. We exhibit intensities shapes and show that they are far from being constant as assumed by cohort and homogenous hazard estimation. Capturing these movements in intensities is crucial to obtain finer probability estimates and to differentiate the risk of a new A issuer with respect to an old one. Our results support an increasing exposition to the risk of default for AAA to BB issuers, but a decreasing exposition for B and overall CCC issuers. However, studying more specifically vintages, we show that time-to-default distributions change through calendar time. In other words macro-economic conditions seem to induce large variations in intensities. These results advocate for modelling default probabilities taking into account ageing effects but conditionally on time series of covariates.

Finally, the flexible setting of the GRHE allows us to investigate conditional patterns avoiding spurious inferences. Our empirical results reveal time spans of non Markovian patterns of default probabilities. In particular we show that the rating momemtum phenomenon reverses on the long run. Moreover we do not find evidence of lag in the reaction of ratings to significant changes in default likelihood. Results suggest that rating reactivity problems and rating class heterogeneity evidences, are attributable to under reactions in magnitude rather than to timing issues. These drawbacks may have been solved by the higher granularity brought by +/- or 1-3 rating variants, but we are not able to adress this question with this dataset. Nevertheless, our findings justify these "step by step" rating adjustments since the mean reversal feature of default intensities seems to match the target horizon of rating agencies.

References

AcharyaV.V., Das S.R. and R.K. Sundaram, 2002, Arbitrage-free pricing of credit derivatives with rating transitions, Financial Analyst Journal 58, n°3, 28-44

Altman E.I., 1989, Measuring Corporate Bond Mortality and Performance, Journal of Finance 44, 909-922

Altman E.I. and D.L. Kao, 1992, Rating drift in high-yield bonds, Journal of Fixed Income 2, 15–20

Altman E.I. and H.A. Rijken, 2004, How Rating Agencies Achieve Rating Stability, Journal of Banking and Finance 28, 2679-2714

Altman E.I. and A. Saunders, 2001, An analysis and critique of the BIS proposal on capital adequacy and ratings, Journal of Banking and Finance 25, 25-46

Andersen P., Borgan Ø., Gill R. and N. Keiding, 1997, Statistical Models Based on Counting Processes, Springer, 2^{nd} Edition

Bahar R. and K.M. Nagpal, March 1999, Modeling the Dynamics of Rating Transitions, Credit, 57-63

Bangia A., Diebold F.X., Kronimus A., Schagen C. and T. Schuermann, 2002, Ratings migration and the business cycle, with application to credit portfolio stress testing, Journal of Banking and Finance 26, 445-474

Bielecki T., Jeanblanc M. and M. Rutkowski, 2004, Replication of Defaultable Claims within the Reduced-Form Framework, Forthcoming Paris-Princeton Series, Springer

Bielecki T., Jeanblanc M. and M. Rutkowski, 2004, Mean-Variance Hedging of Defaultable Claims, Forthcoming Paris-Princeton Series, Springer

Bouezmarni T. and O. Scaillet, 2001, Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, Forthcoming Econometric Theory

Brémaud P., 1981, Point Processes and Queues: Martingale Dynamics, Springer Series in Statistics, Springer-Verlag New York

Brown B.M. and S.X. Chen, 1999, Beta-Bernstein smoothing for regression curves with compact supports, Scandinavian Journal of Statistics 26, 47-59

Chen S.X., 2000, Probability Density Functions Estimation Using Gamma kernels, Ann. Inst. Statis. Math. 52, 471-480

Christensen J.H.E., Hansen E. and D. Lando, 2004, Confidence sets for continuous-time rating transition probabilities, Journal of Banking and Finance 28, 2575-2602

Dabrowska D.M., 1988, Kaplan-Meier Estimate on the Plane, The Annals of Statistics 16, n°4, 1475-1489

D'Amato J. and C.H. Furfine, 2003, Are Credit Ratings Procyclical?, BIS Working Paper 129

Efron B., June 1981, Censored Data and the Bootstrap, Journal of the American Statistical Association 76, n°374, 312-319

Fermanian J.D., 1997, Multivariate Hazard Rates under Random Censorship, Journal of Multivariate Analysis 62, 273-309

Fermanian J.D. and O. Scaillet, 2003, Nonparametric estimation of copulas for time series, Journal of Risk 5, 25-54

Fernandes M. and J. Grammig, October 2002, Nonparametric Specification Tests for Conditional Duration Models, Forthcoming Journal of Econometrics

Fledelius P., Lando D. and J.P. Nielsen, 2004, Non-parametric analysis of rating transition and default data, Journal of Investment Management 2, 71-85

Gordy M. and E. Heitfield, September 2004, Default Probability and Dependence in Credit Rating Systems: Efficient Estimators for Cohort Performance Data, Federal Reserve Board, Washington DC

Gouriéroux C. and A. Monfort, 2002, Equidependence in Qualitative and Duration Models with Application to Credit Risk, CREST DP

Hamilton D.T. and R. Cantor, February 2004, Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History, Moody's Investors Service, Special Comment

Jafry Y. and T. Schuermann, 2004, Measurement, estimation and comparison of credit migration matrices, Journal of Banking & Finance 28, 2603-2639

Jarrow R.A., Lando D. and S.M. Turnbull, 1997, A Markov Model for the Term Structure of Credit Risk Spreads, Review of Financial Studies 10, n°2, 481-523

Kaplan E.L. and P. Meier, June 1958, Nonparametric Estimation from Incomplete Observations, American Statistical Association Journal 53, 457-481

Kavvathas D., May 2000, Estimating Credit Rating Transition Probabilities for Corporate Bonds, Working Paper, University of Chicago

Kijima M. and K. Komoribayashi, Fall 1998, A Markov chain model for valuing credit risk derivatives, The Journal of Derivatives, 97-108

Koopman S.J. and A. Lucas, 2004, Business and Default Cycles for Credit Risk, Forthcoming Journal of Applied Econometrics

Lando D. and T.M. Skodeberg, 2001, Analyzing Rating Transitions and Rating Drift with Continuous Observations, Journal of Banking and Finance 26, 423-444

Löffler G., 2004, An Anatomy of Rating Through the Cycle, Journal of Banking and Finance 28, 695-720

Muller H.-G. and J.-L. Wang, March 1994, Hazard Rate Estimation Under Random Censoring with Varying Kernels and Bandwidths, Biometrics 50, 61-76

Nickell P., Perraudin W. and S. Varotto, 2000, Stability of ratings transitions, Journal of Banking and Finance 24, 203-227

Patil P.N., 1993, On the least squares cross-validation bandwidth in hazard rate estimation, The Annals of Statistics 21, 1792-1810

Pedroso De Lima A.C. and K.S. Pranab, 1997, A Matrix-Valued Counting Process With First-Order Interactive Intensities, The Annals of Applied Probability 7, n°2, 494-507

Ramlau-Hansen H., 1983, Smoothing Counting Process Intensities by Means of Kernel Functions, The Annals of Statistics 11, n°2, 453-466

Strawderman R.L. and M.T. Wells, December 1997, Accurate Bootstrap Confidence Limits for the Cumulative Hazard and Survivor Functions Under Random Censoring, Journal of American Statistical Association 92, n°440, 1356-1374

Tanner M.A. and W.H. Wong, 1983, The estimation of the hazard function from randomly censored data by the kernel method, The Annals of Statistics 11, n°3, 989-993

Wand M.P. and M.C. Jones, 1995, Kernel Smoothing, Monographs on Statistics and Applied Probability 60, Chapman & Hall

Yandell B. S., 1983, Nonparametric Inference for Rates with Censored Survival Data, The Annals of Statistics 11, n°4, 1119-1135

Appendix

Asymptotic Properties of the GRHE

Properties of the Gamma hazard rate estimator are derived in the usual set up of counting processes. This methodology can be considered as extensions of the work of Ramlau & Hansen (1983) or of Andersen & al. (1997). As underlined, the GRHE offers at least two advantages in comparison to their proposition. Indeed for the same degree of smoothness, the Gamma kernel induces thiner confidence intervals when the duration increases. Furthermore, our estimator is consistent at the boundary.
Before turning on proofs, let us define other quantities for mathematical convenience. As usual we will note the Gamma kernel as

$$K\left(s,\frac{u}{b}+1,b\right)=\frac{s^{u/b}\,e^{-s/b}}{b^{u/b+1}\,\Gamma\left(\frac{u}{b}+1\right)}.$$

Then let $\widetilde{\alpha}\left(u\right)$ denote a smooth version of the true hazard rate

$$\widetilde{\alpha}\left(u\right)=\int_{0}^{\infty}K\left(s,\frac{u}{b}+1,b\right)\alpha\left(s\right)ds \tag{6}$$

and $\alpha^{*}\left(u\right)$ which guarantees estimations on the support of $J\left(s\right)=\mathbb{I}_{(Y(s)>0)}$

$$\alpha^{*}\left(u\right)=\int_{0}^{\infty}K\left(s,\frac{u}{b}+1,b\right)\alpha\left(s\right)J\left(s\right)ds. \tag{7}$$

Even if smoothed intensity estimation has been widely studied in the statistical literature, Ramlau-Hansen proposition is much more powerful because their estimator has been derived in the multiplicative intensity framework. However, proofing techniques used by Ramlau-Hansen cannot be applied here. The main advantage of their estimator consists of the fact that the support of their kernel can be as small as required when the number of observations tends to infinite. Hence, convergence in the interior of the intensity domain can easily be handled. In our case the support of the kernel is fixed and for that reason, we need different assumptions.

Definition *Integration compatibility. The Gamma kernel is said to be integration compatible with a distribution if* $\exists B$ *such that* $\forall 0<b\leq B$, $K\left(s,\frac{u}{b}+1,b\right)\frac{f(s)}{1-F(s)}=o\left(s^{-r_B^H}\right)$ *when* $s\longrightarrow\infty$ *for a* $r_B^H>1$ *and* $K\left(s,\frac{u}{b}+1,b\right)\frac{f(s)}{1-F(s)}=o\left(s^{-r_B^L}\right)$ *when* $s\longrightarrow 0$ *for a* $r_B^L<1$*, where* F *is the cumulative distribution function and* f *its density.*

Definition represents a constraint both on the tail behaviour and on the boundary between the Gamma kernel and the hazard rate of the distribution, meaning in particular that the product should be integrable on $[0,\infty[$. This is required for Equation (6) to be defined. However let us notice that this is a very weak constraint. Indeed, for instance, one can easily show that the Gamma kernel is integration compatible with any exponential, Weibull, loglogistic or lognormal distribution.

Assumption (A1)

- *α is $\mathcal{C}^2([0,\infty[)$*
- *$b = b_{(n)} \longrightarrow 0$ when $n \longrightarrow \infty$,*
- *The Gamma kernel is integration compatible with the distribution of default durations $D_1, .., D_n$ with cumulative distribution function F_D.*
- *$\forall 0 < u < \infty$, $F_C(u) < 1$ where F_C is the cumulative empirical distribution of censoring durations.*

The last assumption simply means that the whole distribution can be observed, provided that the number of observations tends toward infinity.

Proof of Proposition I.1

We need to show that $\mathbb{E}\left[\widehat{\alpha}_{(n)}(u)\right] \longrightarrow \alpha(u)$ as $n \longrightarrow \infty$. Consider a sequence of aggregate counting processes $N_{(n)}(u)$ and let us compute,

$$\begin{aligned}
\mathbb{E}\left[\widehat{\alpha}_{(n)}(u)\right] &= \mathbb{E}\left[\int_0^\infty K\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right)\frac{J(s)\, dN_{(n)}(s)}{Y_{(n)}(s)}\right] \\
&= \mathbb{E}\left[\int_0^\infty K\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right)\frac{J(s)\ \ dM_{(n)}(s)}{Y_{(n)}(s)}\right] + \mathbb{E}\left[\alpha^*_{(n)}(u)\right] \text{ from 1.}
\end{aligned}$$

The first term is a martingale and its expectation is null. The second term still involves a stochastic component through the positivity of the aggregate specific part. At this stage, notice that $Y_{(n)}(u)$ is positive for all u, that $Y_{(n)}(0) > 0$ and that 0 is an absorbing barrier for this process[15]. Hence, setting $a_n = \sup\left\{s \in \Re_+; Y_{(n)}(s) > 0\right\}$, we can write

$$\mathbb{E}\left[\widehat{\alpha}_{(n)}(u)\right] = \widetilde{\alpha}_{(n)}(u) - \int_{a_n}^\infty K\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right)\alpha(s)\, ds\ . \tag{8}$$

From usual technics of asymmetric kernels[16], $\widetilde{\alpha}_{(n)}(u)$ can be understood as an expectation:

$$\widetilde{\alpha}_{(n)}(u) = \mathbb{E}\left[\alpha\left(\xi_{u,(n)}\right)\right]$$

[15] In fact, in the full unconditional setting, the process Y only tackles censoring and as a consequence is a decreasing function of u. However, the decreasing shape is not necessary and does not hold in a semi-parametric framework.

[16] See Chen (2000).

where $\xi_{u,(n)}$ is a random variable with $\Gamma\left(\frac{u}{b_{(n)}}+1, b_{(n)}\right)$ distribution. By Taylor expansion as $n \longrightarrow \infty$ quoting $\mu_{u,(n)} = \mathbb{E}\left[\xi_{u,(n)}\right]$,

$$\begin{aligned}
\widetilde{\alpha}_{(n)}(u) &= \alpha\left(\mu_{u,(n)}\right) + \frac{1}{2}\alpha''(u) Var\left(\xi_{u,(n)}\right) + o\left(b_{(n)}\right) \\
&= \alpha(u) + b_{(n)}\left(\alpha'(u) + \frac{1}{2}u\alpha''(u)\right) + o\left(b_{(n)}\right) \ . \qquad (9)
\end{aligned}$$

As from the latest assumption $a_n \longrightarrow \infty$ as $n \longrightarrow \infty$, from the integration compatibility condition and positivity the second term on the right hand side of equation (8) is defined and converges to 0 as $n \longrightarrow \infty$ (remainder of a convergent positive generalized integral).

Last point and Equation (9) show that $\mathbb{E}\left[\widehat{\alpha}_{(n)}(u)\right] \longrightarrow \alpha(u)$ as $n \longrightarrow \infty$ for all $u \in [0, \infty[$, meaning that the GRHE is asymptotically unbiased both in the interior and at the boundary. ■

Next definition has already been proposed by Tanner & Wong (1983) in a simpler framework:

Definition *Compatibility. The Gamma kernel is said to be compatible with a distribution if for all $M > 0$, $\exists B$ such that $\forall 0 < b \leq B$, $\frac{K\left(s, \frac{u}{b}+1, b\right)}{1-F(s)}$ is uniformly bounded for all s such that $|s-u| > M$, where F is the cumulative distribution function.*

Let us remark that if the Gamma kernel is compatible with $F_X(s)$ and $F_Z(s)$, then it is compatible with the independent product $F_{XZ}(s)$.

Assumption (A2)
- *(A1)*
- $nb_{(n)} \longrightarrow \infty$ *when* $n \longrightarrow \infty$,
- $\sup P(Y_{(n)}(u) = 0) = o\left(n^{-1}\right)$,
- $\mathbb{E}\left[\frac{nJ_{(n)}(u)}{Y_{(n)}(u)}\right] \longrightarrow \frac{1}{\tau(u)}$ *as* $n \longrightarrow \infty$ *uniformly on compact sets for which* $Y_{(n)}(u) > 0$,
- *The Gamma kernel is compatible with the distribution of default (resp. censoring) durations* $D_1, .., D_n$ *(resp.* $C_1, .., C_n$*).*

$\tau(u)$ represents the survival probability at time u. Indeed, if F_D and F_C are the cumulative functions corresponding respectively to default durations and censoring, then it is obvious that $\mathbb{E}\left[\frac{nJ_{(n)}(u)}{Y_{(n)}(u)}\right] \longrightarrow \left((1-F_D(u))\left(1-F_C\left(u^-\right)\right)\right)^{-1}$.

Proof of Proposition I.2
Let us derive the asymptotic variance of the GRHE. We consider the squared bias term between

the smooth version of the hazard rate and the true hazard $\left(\widetilde{\alpha}_{(n)}(u)-\alpha(u)\right)^2$ which from (9) is equal to $\left|b_{(n)}\left(\alpha'(u)+\frac{1}{2}u\alpha''(u)\right)+o\left(b_{(n)}\right)\right|^2$ and therefore,

$$\mathbb{E}\left[\left(\widetilde{\alpha}_{(n)}(u)-\alpha(u)\right)^2\right]=O\left(b_{(n)}^2\right) \tag{10}$$

meaning that this component tends to zero as n tends towards infinity. The first variance component comes from the non capacity to observe the whole default distribution in finite samples:

$$\begin{aligned}\mathbb{E}\left[\left(\alpha^*_{(n)}(u)-\widetilde{\alpha}_{(n)}(u)\right)^2\right] &= \mathbb{E}\left|\int_0^\infty K\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\alpha(s)\left(J_{(n)}(s)-1\right)ds\right|^2 \\ &= \mathbb{E}\left|\int_0^\infty K\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\alpha(s)\,\mathbb{I}_{\left(J_{(n)}(s)=0\right)}ds\right|^2\end{aligned}$$

Using the integration compatibility condition and the fact that the whole default distribution can be observed as $n\longrightarrow\infty$, we get that for $M>0$, $\exists N$ s.t. $\forall n\geq N$,

$$\mathbb{E}\left[\left(\alpha^*_{(n)}(u)-\widetilde{\alpha}_{(n)}(u)\right)^2\right]\leq\left|\int_0^\infty K\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\alpha(s)\,P(Y_{(n)}(s)=0)ds\right|^2$$

and from assumptions made,

$$\mathbb{E}\left[\left(\alpha^*_{(n)}(u)-\widetilde{\alpha}_{(n)}(u)\right)^2\right]=o\left(\frac{1}{n}\right) \tag{11}$$

when $n\longrightarrow\infty$, and thus $\mathbb{E}\left[\left(\alpha^*_{(n)}(u)-\widetilde{\alpha}_{(n)}(u)\right)^2\right]=o\left(\frac{1}{b_{(n)}^{\gamma}n}\right)$ for $\gamma=0.5$ and 1.

Finally, we focus on the second variance component $\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u)-\alpha^*_{(n)}(u)\right)^2\right]$. For that purpose we note that

$$\int_0^\infty K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)ds=\frac{\Gamma\left(\frac{2u}{b}+1\right)}{b2^{\frac{2u}{b}+1}\Gamma^2\left(\frac{u}{b}+1\right)}=G_b(u)\ .$$

We now show that $n\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u)-\alpha^*_{(n)}(u)\right)^2\right] \sim \frac{\alpha(u)}{\tau(u)}G_b(u)$ as $n \longrightarrow \infty$. Indeed, let us fix $\varepsilon > 0$,

$$
\begin{aligned}
|A| &= \left|\frac{n\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u)-\alpha^*_{(n)}(u)\right)^2\right]}{G_b(u)} - \frac{\alpha(u)}{\tau(u)}\right| \\
&= G_b^{-1}(u)\left|\int_0^\infty K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\left\{\alpha(s)\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right]-\frac{\alpha(u)}{\tau(u)}\right\}ds\right| .
\end{aligned}
$$

As α is uniformly continuous on a neighbourhood of u, for ε_α, $\exists\delta_\alpha$ such that $|\alpha(s)-\alpha(v)| < \varepsilon_\alpha$ for $|s-v| < \delta_\alpha$. On the same way, for ε_τ, $\exists\delta_\tau$ such that $\left|\frac{1}{\tau(s)}-\frac{1}{\tau(v)}\right| < \varepsilon_\tau$ for $|s-v|<\delta_\tau$. Let us denote $\delta = \delta_\alpha \wedge \delta_\tau$. In addition, for ε_Y, $\exists N$ such that $\forall n \geq N$ $\sup\limits_{[0,a_n]}\left|\mathbb{E}\left[\frac{nJ_{(n)}(x)}{Y_{(n)}(x)}\right]-\frac{1}{\tau(x)}\right| < \varepsilon_Y$. Therefore using the same notations, for $n \geq N$,

$$
\begin{aligned}
|A| &\leq G_b^{-1}(u)\left|\int_{\left|s-\mu_{u,(n)}\right|<\delta} K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\left\{\alpha(s)\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right]-\frac{\alpha(u)}{\tau(u)}\right\}ds\right| \\
&\quad +G_b^{-1}(u)\left|\int_{\left|s-\mu_{u,(n)}\right|>\delta} K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)\left\{\alpha(s)\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right]-\frac{\alpha(u)}{\tau(u)}\right\}ds\right| \\
|A| &\leq G_b^{-1}(u)\int_{\left|s-\mu_{u,(n)}\right|<\delta}|A_1| + G_b^{-1}(u)\int_{\left|s-\mu_{u,(n)}\right|>\delta}|A_2| .
\end{aligned}
$$

Introducing required terms,

$$
\begin{aligned}
\int_{\left|s-\mu_{u,(n)}\right|<\delta}|A_1| &\leq \int_{\left|s-\mu_{u,(n)}\right|<\delta}|\alpha(s)-\alpha(u)|\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right]K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)ds \\
&\quad + \int_{\left|s-\mu_{u,(n)}\right|<\delta}\alpha(u)\left|\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right]-\frac{1}{\tau(s)}\right|K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)ds \\
&\quad + \int_{\left|s-\mu_{u,(n)}\right|<\delta}\alpha(u)\left|\frac{1}{\tau(s)}-\frac{1}{\tau(u)}\right|K^2\left(s,\frac{u}{b_{(n)}}+1,b_{(n)}\right)ds
\end{aligned}
$$

which gives for suitable choices of ε_α, ε_Y and ε_τ, and for a constant CT_Y depending both on N and the neighbourhood $\mu_{u,(n)}$,

$$G_b^{-1}(u) \int_{\left|s-\mu_{u,(n)}\right|<\delta} |A_1| \leq \frac{G_b(u)}{G_b(u)} \left\{\varepsilon_\alpha CT_Y + \alpha(u)\,\varepsilon_Y + \alpha(u)\,\varepsilon_\tau\right\} \leq \frac{\varepsilon}{2} \, . \tag{12}$$

Furthermore we have,

$$\begin{aligned}
\int_{\left|s-\mu_{u,(n)}\right|>\delta} |A_2| \;&\leq \int_{\left|s-\mu_{u,(n)}\right|>\delta} K^2\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) \alpha(s)\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right] ds \\
&\quad + \frac{\alpha(u)}{\tau(u)} \int_{\left|s-\mu_{u,(n)}\right|>\delta} K^2\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) ds \\
&\leq \int_{\left\{\left|s-\mu_{u,(n)}\right|>\delta\right\}\cap[0,a_n]} K^2\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) \alpha(s)\,\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right] ds \qquad (13) \\
&\quad + \frac{\alpha(u)}{\tau(u)} G_b(u) \int_{\left|s-\mu_{u,(n)}\right|>\delta} f_{\eta_{u,(n)}}(s)\, ds
\end{aligned}$$

where $\eta_{u,(n)}$ is a random variable with $\Gamma\left(\frac{2u}{b_{(n)}}+1, \frac{b_{(n)}}{2}\right)$ distribution and mean $\mu_{u,(n)}$ and variance $\sigma^2_{u,(n)} = \frac{bu}{2} + \frac{b^2}{4}$, hence using Tchebyshev inequality,

$$\int_{\left|s-\mu_{u,(n)}\right|>\delta} f_{\eta_{u,(n)}}(s)\, ds \leq \frac{\sigma^2_{u,(n)}}{\delta} \, .$$

The first component on the right hand side of equation (13) can be written as

$$\begin{aligned}
&\leq \int_{\left\{\left|s-\mu_{u,(n)}\right|>\delta\right\}\cap[0,a_n]} K^2\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) \alpha(s) \left|\mathbb{E}\left[\frac{nJ_{(n)}(s)}{Y_{(n)}(s)}\right] - \frac{1}{\tau(s)}\right| ds \\
&\quad + \int_{\left\{\left|s-\mu_{u,(n)}\right|>\delta\right\}\cap[0,a_n]} K^2\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) \frac{\alpha(s)}{\tau(s)} ds
\end{aligned}$$

thus using integration compatibility and compatibility assumptions, it exists CT_1 and $CT_2 > 0$, such that

$$\leq CT_1 \varepsilon_Y \int\limits_{\left\{\left|s-\mu_{u,(n)}\right|>\delta\right\}} K\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) ds$$
$$+CT_2 \int\limits_{\left\{\left|s-\mu_{u,(n)}\right|>\delta\right\}} K\left(s, \frac{u}{b_{(n)}}+1, b_{(n)}\right) ds$$

and once again using Tchebyshev inequality, we get,

$$G_b^{-1}(u) \int\limits_{\left|s-\mu_{u,(n)}\right|>\delta} |A_2| \leq \frac{\alpha(u)}{\tau(u)} \frac{\sigma^2_{u,(n)}}{\delta} + G_b^{-1}(u)\left(CT_1 \varepsilon_Y + CT_2\right) \frac{v^2_{u,(n)}}{\delta}$$

where $v^2_{u,(n)} = bu + b^2$ is the variance of $\xi_{u,(n)}$. Therefore since[17] $G_b^{-1}(u) \longrightarrow 0$, for a suitable $N' \geq N$, as $b \longrightarrow 0$ when $n \longrightarrow \infty$ we proved that,

$$G_b^{-1}(u) \int\limits_{\left|s-\mu_{u,(n)}\right|>\delta} |A_2| \leq \frac{\varepsilon}{2}. \tag{14}$$

From equations (12) and (14), we showed that for $n \geq N'$

$$|A| \leq \varepsilon$$

meaning that it tends to zero as n tends to infinity and that

$$n\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u) - \alpha^*_{(n)}(u)\right)^2\right] \sim \frac{\alpha(u)}{\tau(u)} G_b(u) \tag{15}$$

As a consequence using limiting properties of $G_b(u)$, equivalency (15), equation (11) and Cauchy-Schwarz inequality, we proved that

$$n\sqrt{b}\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u) - \widetilde{\alpha}_{(n)}(u)\right)^2\right] \sim \frac{\alpha(u)}{2\sqrt{\pi}\tau(u)\sqrt{u}} \text{ if } \frac{u}{b} \to \infty \text{ and}$$
$$nb\mathbb{E}\left[\left(\widehat{\alpha}_{(n)}(u) - \widetilde{\alpha}_{(n)}(u)\right)^2\right] \sim \frac{\alpha(u)\,\Gamma(2\kappa+1)}{\tau(u)\,2^{1+2\kappa}\,\Gamma^2(\kappa+1)} \text{ if } \frac{u}{b} \to \kappa\ .$$

Proof of asymptotic normality does not involve further complexity and can be obtain in the same way than Ramlau-Hansen using the same techniques than we have just presented. As this is not the main goal of this paper we leave the proof to interested reader. ■

We point out that asymptotic properties of the GRHE do not require non standard assumptions. Besides, the asymptotic convergence of the estimator has been derived with weaker assumptions, allowing for instance for linear-exponentially increasing asymptotics for the haz-

[17] Limits of $G_b(u)$ on the interior and at the boundary has been studied by Brown and Chen (1999).

ard rate at the infinity and unbounded hazard at zero. On the contrary standard estimators are painful on the tails of the hazard rate. Solutions to this problem have been proposed in the statistical literature but they require changing kernels and bandwidths which make them harder to implement. The GRHE is a simpler and more accurate alternative.

Adaptative Kernels

Müller & Wang (1994) provide the reference contribution concerning adaptative kernels and bandwidths. They propose such estimators to improve basic kernel estimators of intensities. They specify a new form for the kernel $K^*(u,s,b)$ by $K^*(u,s,b) = \frac{1}{b(u)} K\left(q(u), \frac{u-s}{b(u)}\right)$. It modifies the Ramlau-Hansen type of estimators using a different kernel function in the interior and at the boundary. More precisely, near the boundary the kernel support becomes asymmetric so as to reduce the boundary bias. Secondly, they proposed to use local bandwidths instead of a global bandwidth, in order to diminish the long run variance of the estimator while avoiding oversmoothing. This local bandwidths minimize pointwise mean squared errors. In the following, we recall the basic definition of adaptative kernels estimators, as well as the explicit form of the particular epanechnikov and biquadratic cases. Müller & Wang (1994) give a complete description of possible kernel choices. At last we present the estimation procedure that we used in the comparison, as proposed by Müller & Wang.

Definition *Under assumptions I.1 to I.4, the adaptative kernel $\widehat{\alpha}_K(\cdot)$ with bandwidth function $b(\cdot)$ is defined setting $K^*(u,s,b) = \frac{1}{b(u)} K\left(q(u), \frac{u-s}{b(u)}\right)$, where $q(u) = 1$ if $u \in I = \{x; x \geq b(x)\}$ and $q(u) = \frac{u}{b(u)}$ if $u \in B = \{x; 0 \leq x \leq b(x)\}$*

Let us remark that this definition differs sligthly from the definition of Müller & Wang as they do not rely on a multiplicative intensity framework. However, the definitions are equivalent under the independent censoring assumption.

Definition *The adaptive epanechnikov estimator of intensities is given by*

$$K(q,z) = \frac{12}{(1+q)^4}(z+1)\left[z(1-2q) + \left(3q^2 - 2q + 1\right)/2\right]$$

and the adaptative biquadratic estimator by

$$K(q,z) = \frac{15}{(1+q)^5}(z+1)^2(q-z)\left[2z\left(5\frac{1-q}{1+q} - 1\right) + (3q-1) + 5\frac{(1-q)^2}{1+q}\right]$$

For estimation purposes, the difficulty lies in the fact that the local bandwidth determines the local kernel expression. The process can be decomposed as follows:

1. For each duration point of a first equidistant grid G1, compute the leading intensity estimate setting $b(\cdot) \equiv b_0$.

2. Using the leading intensities estimates, search for each point of G1 the bandwidth minimizing the estimated MSE criterion over a bandwidth grid B which ranges from $2b_0/3$ and $4b_0$. It provides an optimal bandwidth function $\widehat{b}_1^*(\cdot)$ on the grid G1.

3. On a final grid G2, compute local bandwidths by smoothing $\widehat{b}_1^*(\cdot)$ so as to obtain a new bandwidth function $\widehat{b}_2^*(\cdot)$ on the grid G2. $\widehat{b}_2^*(\cdot)$ approximates the optimal bandwidth function.

4. Estimates final intensity on G2 using the bandwidth function $\widehat{b}_2^*(\cdot)$.

Notice that in some circumstances (in particular for bell-shape intensities) the adaptative Epanechnikov and the adaptative biquadratic kernels lead to negative intensities on the short run which is incompatible with hazard rate theory and reality as intensities can never be negative by definition. A second comment comes from the intricate estimation procedure of these adaptative kernels: simulation results were highly sensitive with respect to the initial bandwidth we used. In particular optimal bandwidths (b_{epa}^* and b_{biq}^*) derived from the basic correponding kernels were not always the best choice (we tried alternatively $2b_{xxx}^*$ and $b_{xxx}^*/2$ as b_0). In other words, we repeated steps 1 to 3 three times with different initial bandwidths b_0 and kept best minimizers of the estimated MSE criterion.

Bandwidth Choice for Smooth Hazard Estimators

Kernel estimators are known to be quite sensitive to the choice of the bandwidth (i.e. the degree of smoothness of the estimator). In our empirical applications, we computed the optimal bandwidth by cross-validation using a grid-search algorithm. More specifically, we used the well-tried leave-one-out method[18] to determine the bandwidth which minimises the mean of squared in-sample prediction errors. We give below the main lines of the method. If λ is the true intensity and $\widehat{\lambda}_b$ is the estimated intensity using the bandwidth b, the optimal bandwidth is obtained as:

$$\begin{aligned} b_{opt} &= \arg\min_b \mathrm{E}\left[\int \left(\widehat{\lambda}_b - \lambda\right)^2\right] \\ b_{opt} &= \arg\min_b \mathrm{E}\left[\int_0^\infty \left(\int_0^\infty K^*(u,s,b)\, d\widehat{\Lambda}(s) - \lambda(u)\right)^2 du\right] \end{aligned}$$

where Λ is the true cumulative hazard function. Expanding this expression, one can only focus on the two first terms:

$$b_{opt} = \arg\min_b \mathrm{E}\left[\int_0^\infty \widehat{\lambda}_b^2(u)\, du - 2\int_0^\infty \widehat{\lambda}_b(u)\,\lambda(u)\, du\right]$$

[18] See Wand & Jones (1995) for discussion of bandwidth selection methods.

The leave-one-out cross-validation consists of replacing the intensity at each duration by an estimation on the subsample which leaves that observation out. Calling CVL the corresponding empirical criterion, we get

$$CVL = \int_0^\infty \widehat{\lambda}_b^2 (u)\, du - 2 \sum_{i=1}^n \widehat{\lambda}_{b,-i} (U_i)\, \lambda (U_i)\, \mathbb{I}_{(U_i = D_i)}$$

Patil (1993) showed the validity of this criterion. Its empirical counterpart is given by replacing $\lambda (U_i)$ by another estimation. We can either choose the leave-one-out Nelson-Aalen estimator,

$$\widetilde{\lambda} (U_i) = \frac{d\widehat{\Lambda} (U_i)}{du} = \frac{dN_{U_i,-i}}{Y_{-i} (U_i)}$$

where $N_{\cdot,-i}$ (resp. $Y_{-i}(\cdot)$) is the class counting processes (resp. the risk set size) where the firm i has been removed, or from a smoothed estimator[19] as proposed by Ramlau-Hansen (1983) and Andersen & al. (1997):

$$\widehat{\widehat{\lambda}} (U_i) = \frac{1}{n} \frac{1}{\widehat{S}_{D,-i} (U_i)} \frac{dN_{U_i,-i}}{\widehat{S}_{C,-i} (U_i)}$$

where $S_{D,-i} (\cdot)$ (resp. $S_{C,-i} (\cdot)$) is the default (resp. censoring) survival function estimated leaving the firm i out through $\widehat{S}_{D,-i} (u) = \exp \left(- \int_0^u \widehat{\lambda}_{b,-i} (s)\, ds \right)$

(resp. $\widehat{S}_{C,-i} (u) = \prod_{U_j \leq u}^{j \neq i} \left(\frac{(n-1)-j+1}{(n-1)-j+2} \right)^{1-dN_{j,-i}}$ the corresponding product limit estimator).

In empirical applications the optimal bandwidth has always been chosen from the CVL criterion, computing a first grid of computations with 50 days spaced out bandwidths, and a second grid around the first optimal point with 1 day spaced out bandwidths.

[19] In all applications we choose to use this last estimator.

Table I

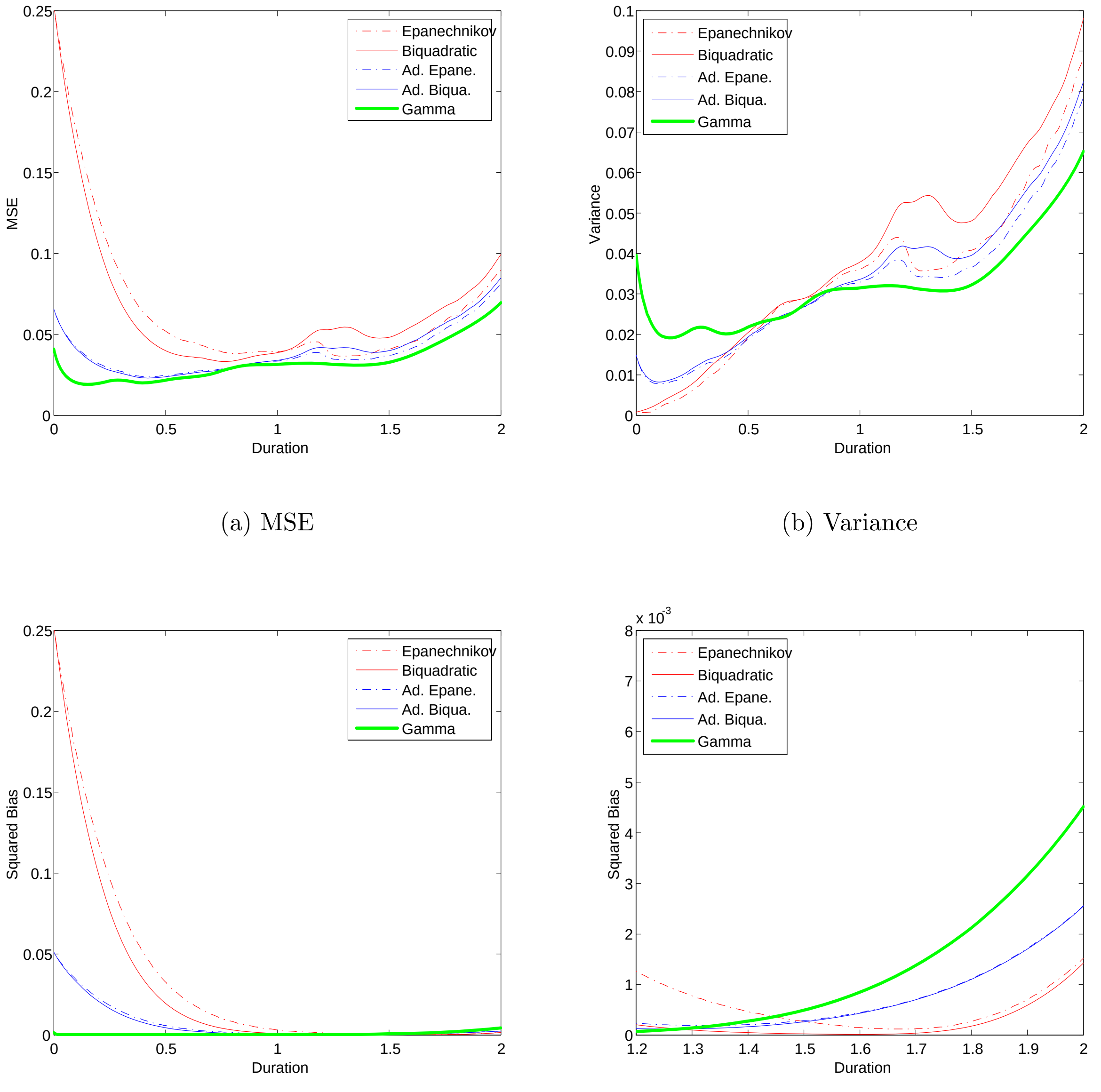
Estimators Performance on Simulated Samples

Unit Exponential	Average MSE	Average Variance	Average Squared Bias
Censoring 10%			
Epanechnikov	0.060416	0.037998	0.030679
Biquadratic	0.058383	0.041543	0.025100
Adaptative Epanechnikov	0.041137	0.034916	0.006220
Adaptative Biquadratic	0.041997	0.036334	0.005663
Gamma	0.033885	0.033157	0.000728
Censoring 20%			
Epanechnikov	0.07974	0.051142	0.028598
Biquadratic	0.079958	0.057002	0.022956
Adaptative Epanechnikov	0.055655	0.049461	0.006193
Adaptative Biquadratic	0.058435	0.052805	0.005629
Gamma	0.048292	0.046780	0.001512
Censoring 40%			
Epanechnikov	0.092349	0.066949	0.025399
Biquadratic	0.10553	0.083883	0.021646
Adaptative Epanechnikov	0.073962	0.062389	0.011573
Adaptative Biquadratic	0.079196	0.068842	0.010354
Gamma	0.067355	0.062535	0.004820

Comparison of smooth hazard rate estimators. Results are computed from 1000 simulated samples of 100 observations for different uniform censoring schemes. The true intensity is constant and equal to 1 (a unit exponential distribution). Averages are computed over an equidistant grid (250 points) between [0,2].

Figure 1

Pointwise Estimators Performance on Simulated Samples



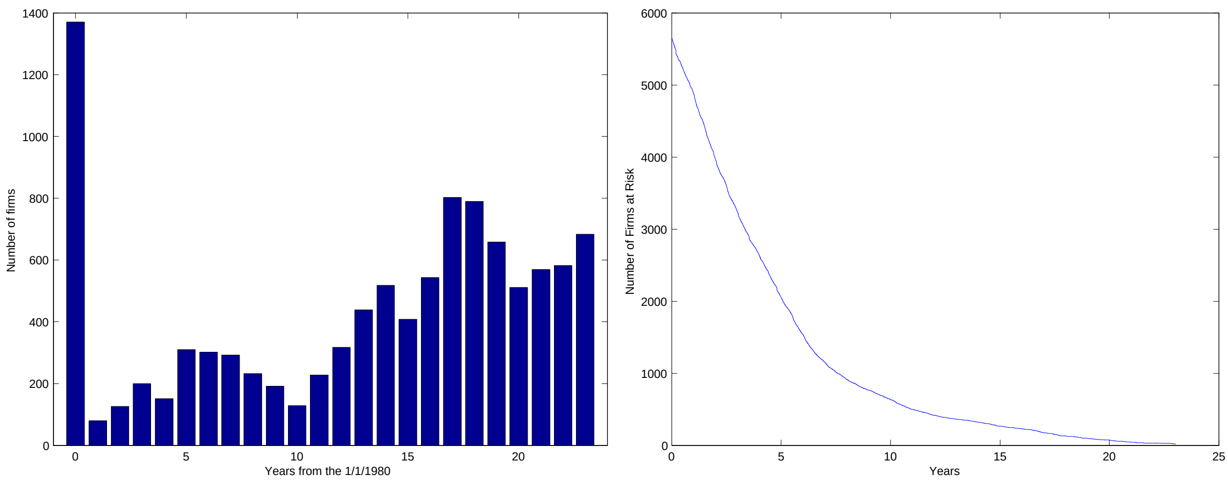
(a) MSE (b) Variance

(c) Squared Bias (d) Long Term Sq. Bias

Estimation errors for various smooth hazard rate estimators on 1000 simulated samples of 100 observations from a unit exponential distribution under a 10% uniform censoring scheme.

Figure 2

Censoring Structure of the Database

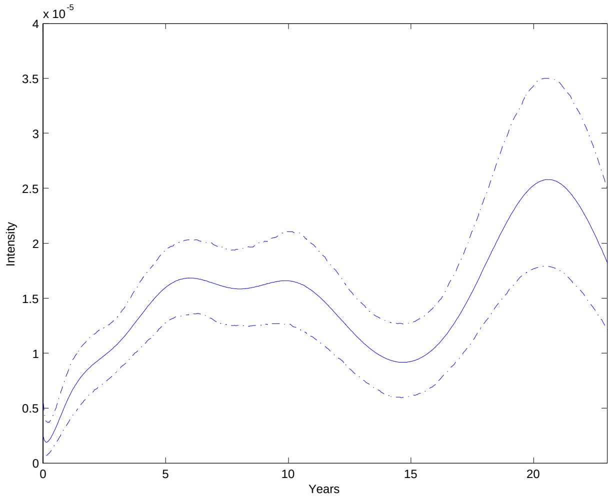


(a) Number of new firms rated by S&P each year (b) Risk set size $Y(u)$ of the NIG risk class

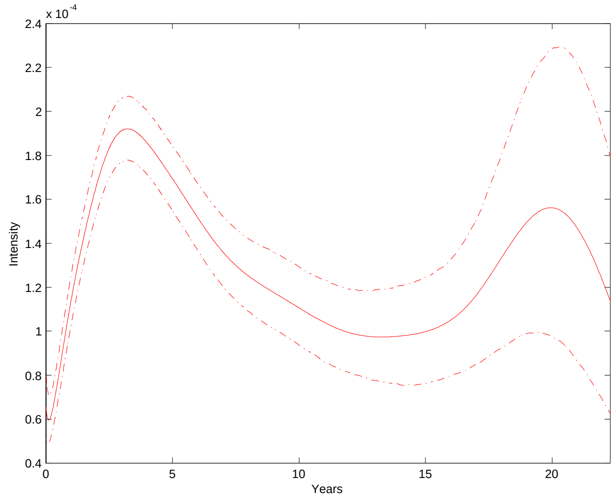
Database information: number of firms entering the rating process and lengths of observed durations.

Figure 3

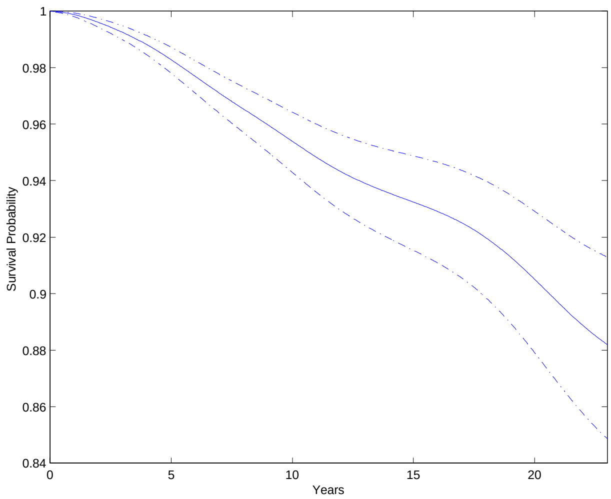
IG and NIG Unconditional Probabilities of Default



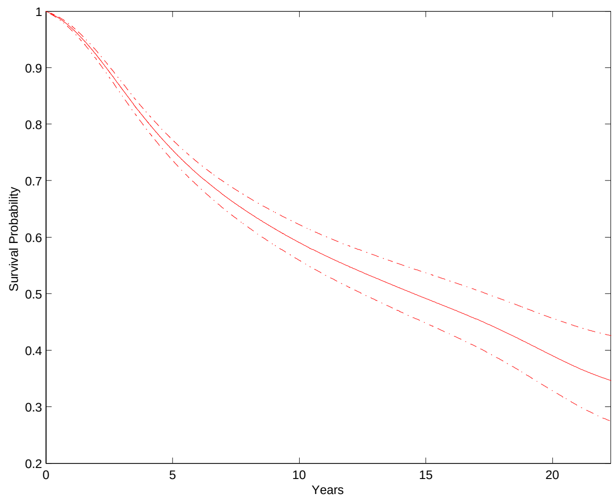
(a) IG - Intensity



(b) NIG - Intensity



(c) IG - Survival probability

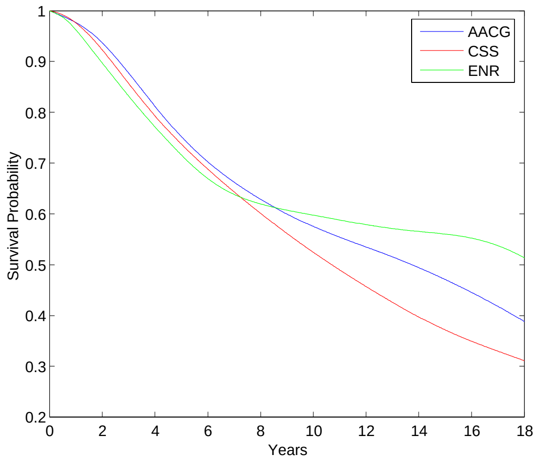


(d) NIG - Survival probability

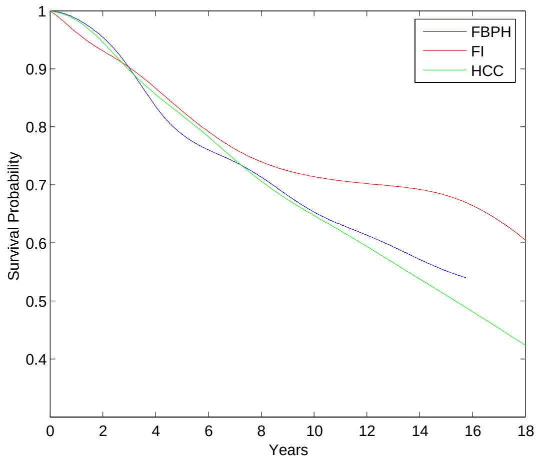
Intensities and implied survival probabilities of Investment Grades (IG, ratings from AAA to BBB) and Non Investment Grades (NIG, ratings from BB to CCC) risk classes from entry in the class to the last day of observation. Dotted lines represent 95% bootstrap confidence intervals. For IG (resp. NIG), we observed 235 (1363) defaults - average duration among defaults is about 3228 (1635) days for 3038 (1789) among exits - the optimal bandwidth counts 111 (106) days.

Figure 4

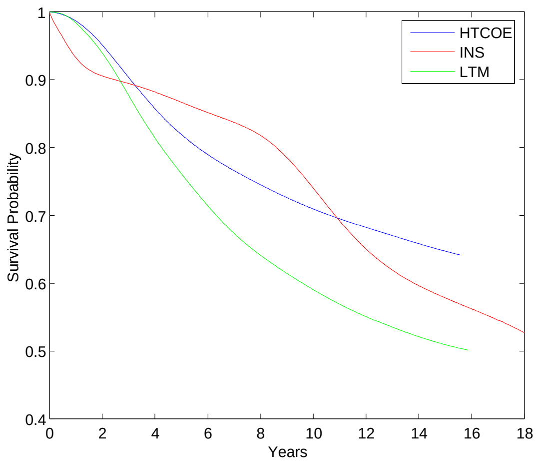
Survival Probabilities over Industrial Classes



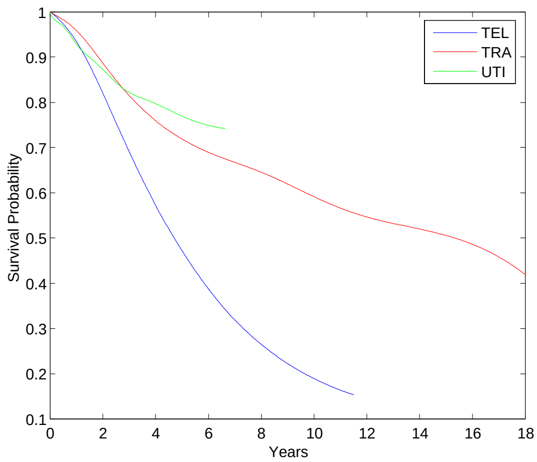
(a)



(b)



(c)

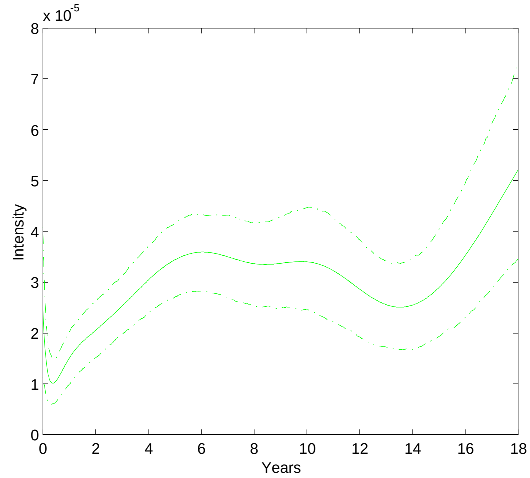


(d)

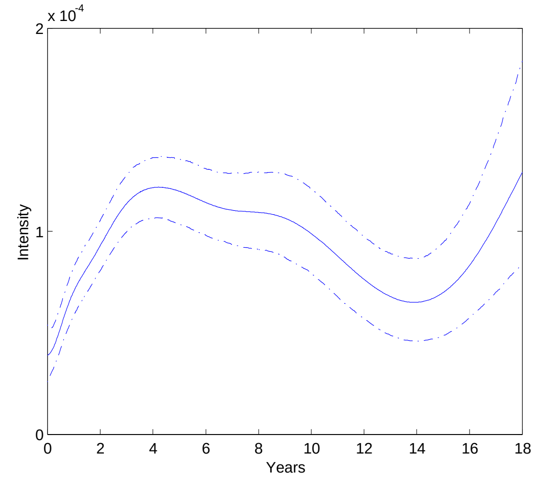
Survival probabilities oves industrial risk classes from entry in the class to the last day of observation. Sectors are broad sectors defined by Standard & Poor's : Aeromotive, Automobile, Capital Goods and Metal (AACG) - Consumer and Services (CSS) - Energy and Natural Ressources (ENR) - Forest Product and Building (FPBH) - Financial Institutions (FI) - Health Care and Chemicals (HCC) - High Tech and Office Equipments (HTCOE) - Insurance (INS) - Leisure, Time and Media (LTM) - Telecommunications (TEL) - Transports (TRA) - Utility (UTI). A last sector, namely Real Estate, has not been estimated because of minimum population requirements.

Figure 5

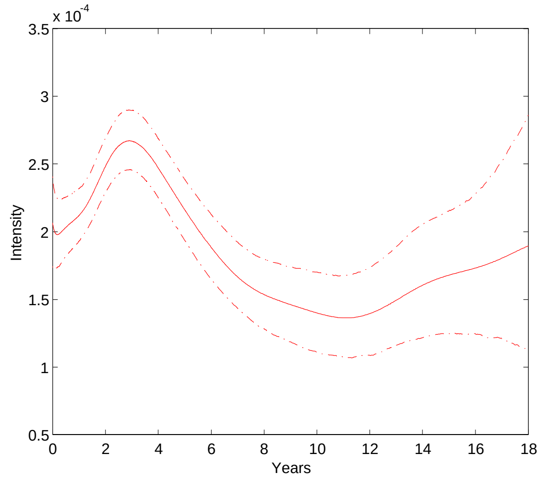
Intensities of Default over Rating Classes



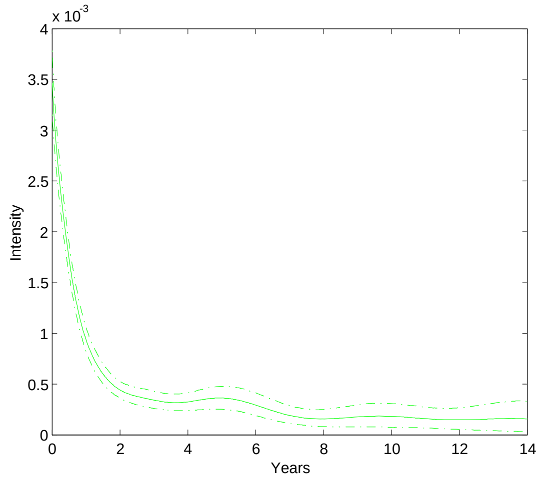
(a) BBB



(b) BB



(c) B

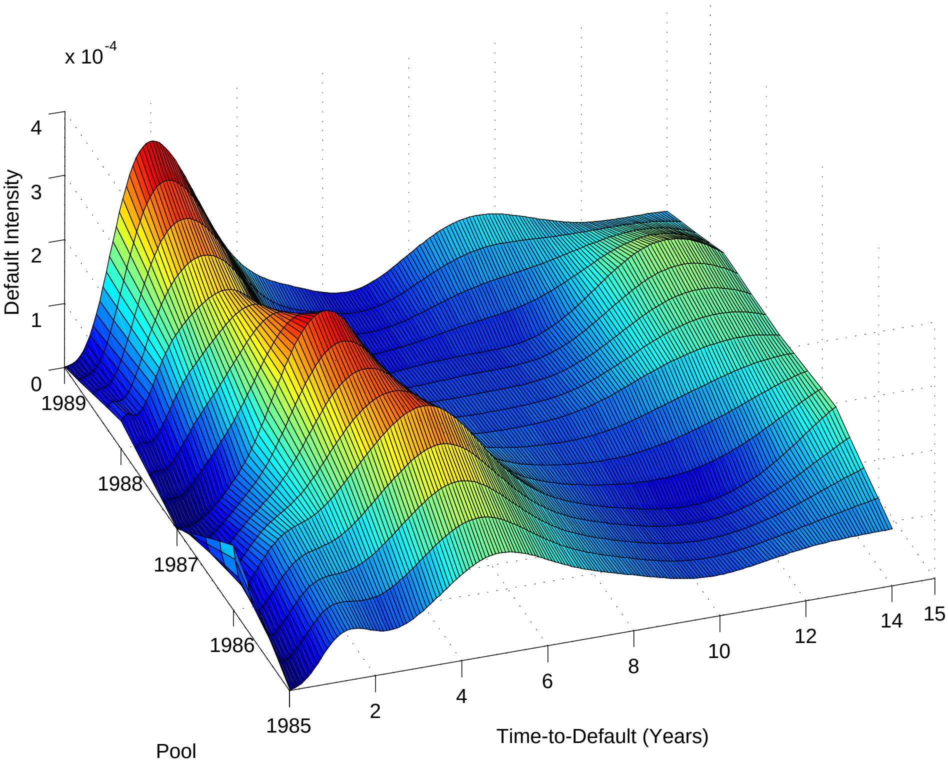


(d) CCC

Intensities of default over BBB, BB, B and CCC risk classes from entry in the class to the last day of observation. Dotted lines represent 95% bootstrap confidence intervals. We observed respectively 217, 574, 1199 and 879 defaults - average duration among defaults is respectively about 2671, 1932, 1326 and 335 days for 2293, 1878, 1605 and 1071 among exits - the optimal bandwidth counts 108, 106, 101 and 66 days.

Figure 6

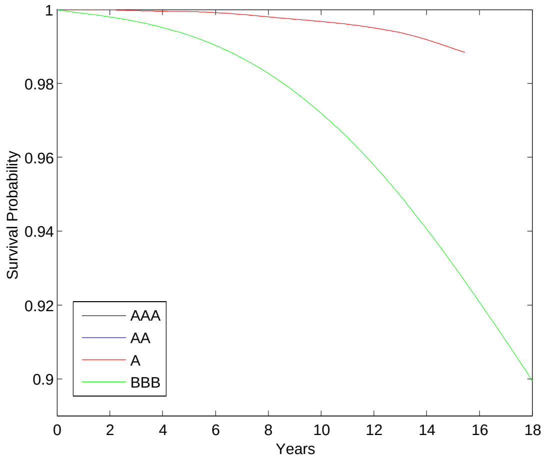
Impact of Vintages over NIG Default Intensities



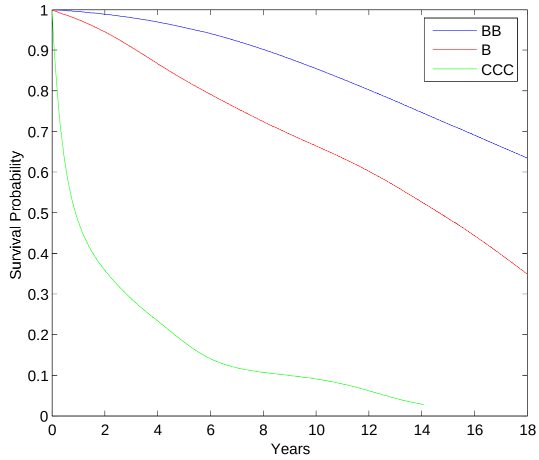
Non Investment Grades (NIG) default intensity for distinct vintages up to the last day in the sample, sliding the reference month quaterly. For instance the first pool is made up of firms entered into the rating process between the 1/1/1985 and the 12/31/1985, whereas firms entered between the 1/4/1985 and the 3/31/1986 have been collected in the second one.

Figure 7

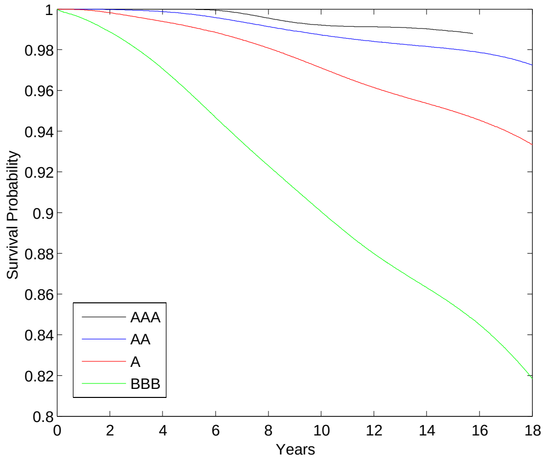
Survival Probabilities over Rating Classes



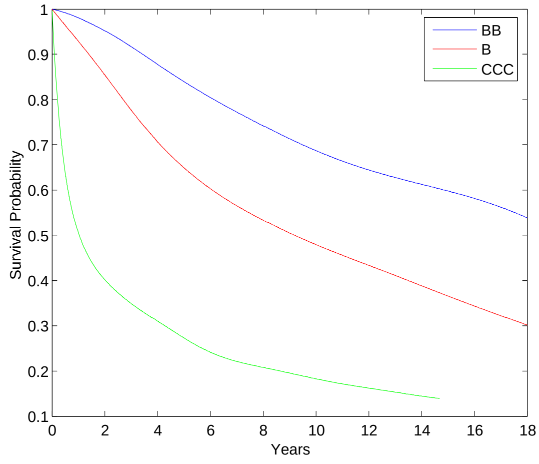
(a) Within IG - Survival Prob.



(b) Within NIG - Survival Prob.



(c) IG - Survival Prob.

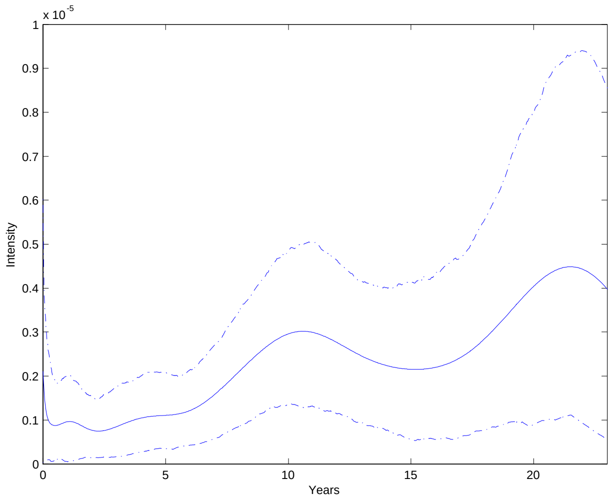


(d) NIG - Survival Prob.

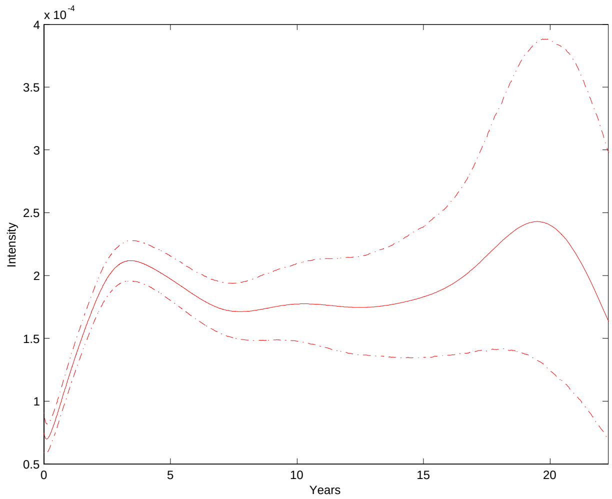
Survival probabilities over rating classes both within classes (conditional on staying in the class up to default) and from entry in the class to the last day of observation (unconditional).

Figure 8

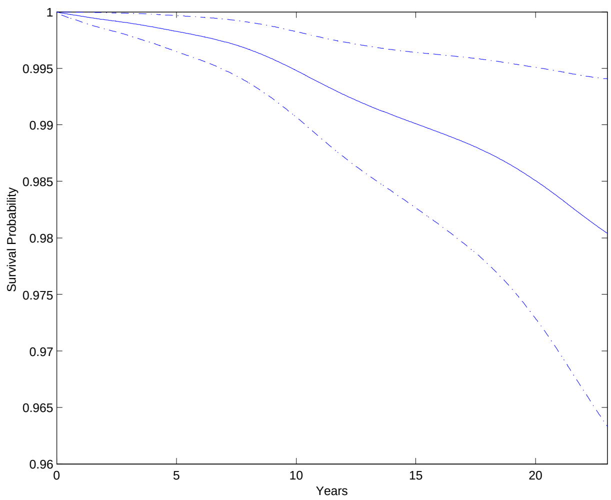
IG and NIG Probabilities of Straight Default



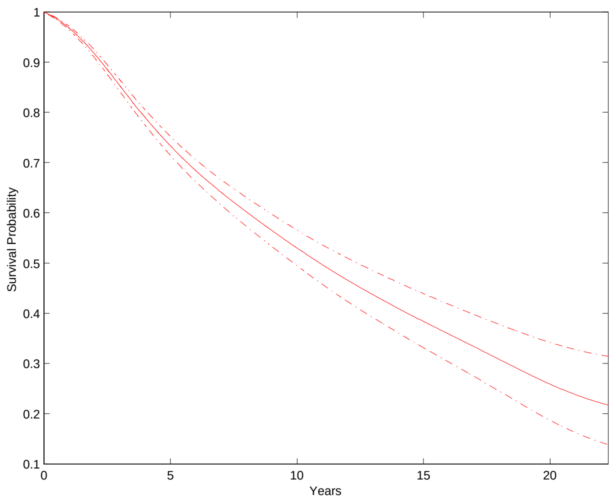
(a) IG - Intensity



(b) NIG - Intensity



(c) IG - Survival probability

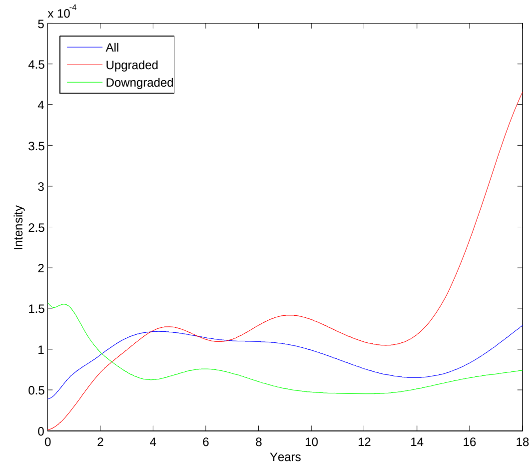


(d) NIG - Survival probability

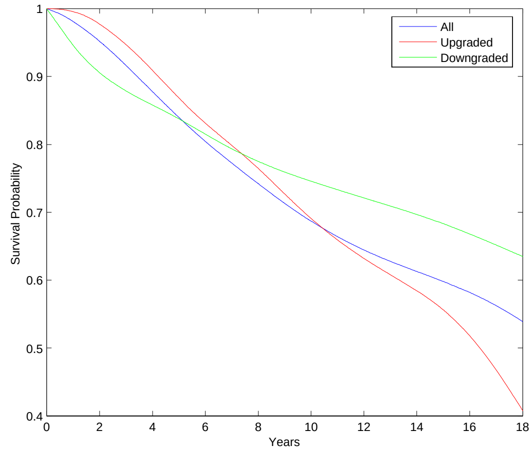
Intensities and implied survival probabilities of Investment Grades (IG, ratings from AAA to BBB) and Non Investment Grades (NIG, ratings from BB to CCC) risk classes from entry in the class to first exit. Dotted lines represent 95% bootstrap confidence intervals. For IG (resp. NIG), we observed 29 (1357) defaults - average duration to default is about 3508 (1513) days, and 2704 (1451) days for durations to exit - the optimal bandwidth counts 111 (106) days.

Figure 9

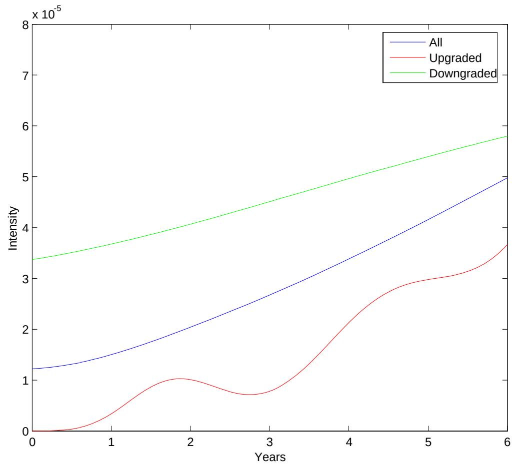
BB Default Behaviour w.r.t Previous Move



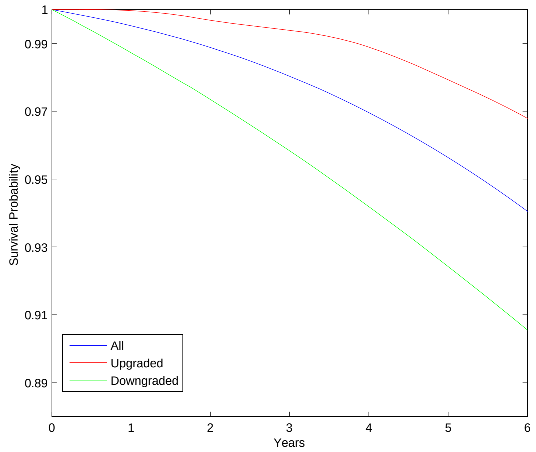
(a) Intensity (U)



(b) Survival Probabilities (U)



(c) Intensity (CS)

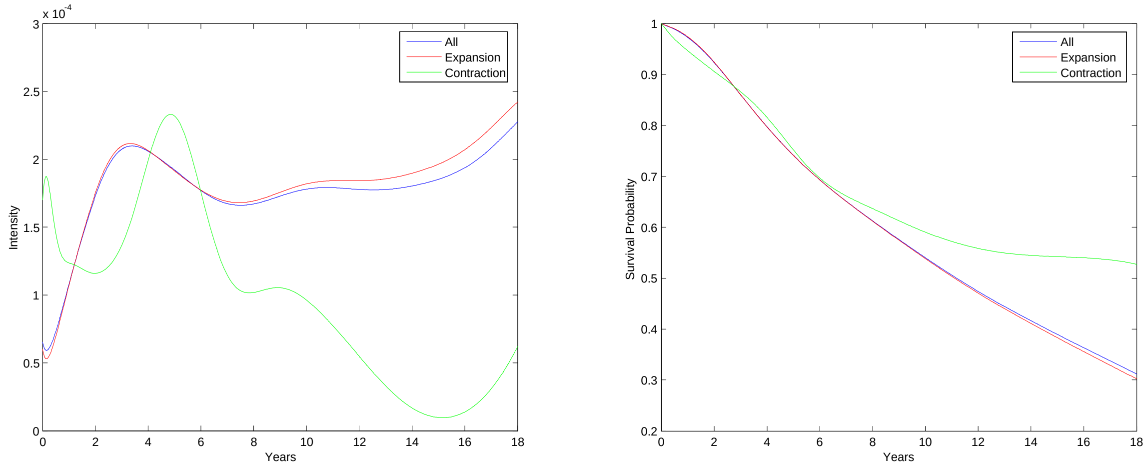


(d) Survival Probabilities (CS)

Default intensities and survival probabilities on BB firms conditioning by previous grades, up to the last day in the sample (Unconditional, U) and up to the first exit (Conditional on Staying in the class, CS). For unconditional observations, from upgrades (resp. downgrades), we observed 67 (133) defaults - average duration among defaults is about 2233 (1428) days for 1691 (2020) among exits - the optimal bandwidth counts 91 (96) days. For conditional observations, from upgrades (resp. downgrades), we observed 16 (19) defaults - average duration among defaults is about 2955 (1012) days for 1069 (824) among exits - the optimal bandwidth counts 66 (65) days. Notice that the total unconditional (resp. conditional) population counts 574 (80) defaults.

Figure 10

Non Investment Grades Default Behaviour and the Business Cycle



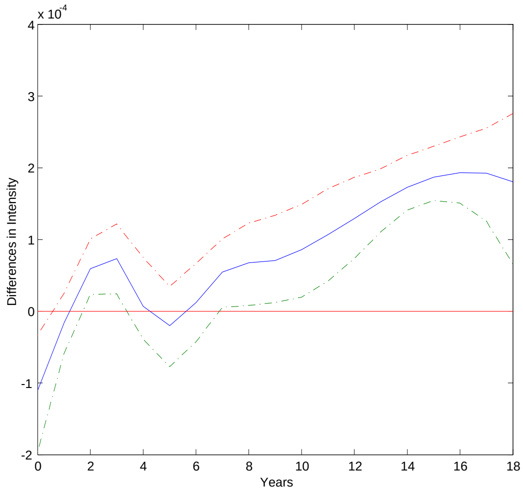
(a) Intensity

(b) Survival Probabilities

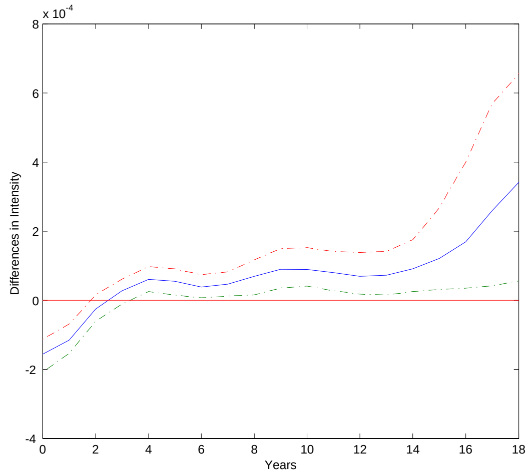
Non Investment Grades default intensities and survival probabilities conditional on economic conditions, up to the first exit. During expansions (resp. contraction), we observed 1058 (39) defaults - average duration among defaults is about 1654 (1237) days for 1656 (1421) among exits - the optimal bandwidth counts 106 (97) days.

Figure 11

Bootstrap Differences in Conditional Hazard Rates



(a) Expansion vs contraction on NIG firms



(b) Upgrades vs downgrades on BB firms

Differences between two intensity curves with 95% confidence intervals. These intervals have been computed from differences between 1000 bootstrap samples of both curves.

The FAME Research Paper Series

*The International Center for Financial Asset Management and Engineering (FAME) is a private foundation created in 1996 on the initiative of 21 leading partners of the finance and technology community, together with three Universities of the Lake Geneva Region (Switzerland). FAME is about **Research, Doctoral Training**, and **Executive Education** with "interfacing" activities such as the FAME lectures, the Research Day/Annual Meeting, and the Research Paper Series.*

*The **FAME Research Paper Series** includes three types of contributions: First, it reports on the research carried out at FAME by students and research fellows; second, it includes research work contributed by Swiss academics and practitioners interested in a wider dissemination of their ideas, in practitioners' circles in particular; finally, prominent international contributions of particular interest to our constituency are included on a regular basis. Papers with strong practical implications are preceded by an Executive Summary, explaining in non-technical terms the question asked, discussing its relevance and outlining the answer provided.*

Martin Hoesli is acting Head of the Research Paper Series. Please email any comments or queries to the following address: Martin.Hoesli@hec.unige.ch.

The following is a list of the 10 most recent FAME Research Papers. For a complete list, please visit our website at www.fame.ch under the heading 'Faculty and Research, Research Paper Series, Complete List'.

N°140 **Robust Mean-Variance Portfolio Selection**
Cédric PERRET-GENTIL, Union Bancaire Privée, Maria-Pia VICTORIA-FESER, HEC, University of Geneva

N°139 **Trading Volume in Dynamically Efficient Markets**
Tony BERRADA, HEC Montréal, CIRANO and CREF, Julien HUGONNIER, University of Lausanne, CIRANO and FAME, Marcel RINDISBACHER, Rotman School of Management, University of Toronto and CIRANO.

N°138 **Growth Options in General Equilibrium: Some Asset Pricing Implications**
Julien HUGONNIER, University of Lausanne and FAME, Erwan MORELLEC, University of Lausanne, FAME and CEPR, Suresh SUNDARESAN, Graduate School of Business, Columbia University.

N°137 **On the Demand for Budget Constrained Insurance**
Richard WATT, Universidad Autónoma de Madrid, Henri LOUBERGÉ, University of Geneva and FAME

N°136 **Direct Prference for Wealth in Aggregate Household Portfolios**
Pascal ST-AMOUR, HEC, University of Lausanne, FAME, CIRANO and CIRPEE

N°135 **Indirect Robust Estimation of the Short-Term Interest Rate Process**
Veronika CZELLAR, Dept. of Econometrics, University of Geneva, G. Andrew KAROLYI, Fisher College of Business, Ohio State University, Elvezio RONCHETTI, Dept. of Econometrics, University of Geneva.

N°134 **Do Major Financial Crises Provide Information on Sovereign Risk to the Rest of the World? A Look at Credit Default Swap Markets**
Didier COSSIN, IMD International and FAME, Gero JUNG, Graduate Institute of International Studies.

N°133 **Are European Corporate Bond and Default Swap Markets Segmented?**
Didier COSSIN, IMD International, Hongze LU, IMD International, HEC, University of Lausanne.

N°132 **Conditional Asset Allocation Under Non-Normality: How Costly is the Mean-Variance Criterion?**
Eric JONDEAU, HEC Lausanne and FAME, Michael ROCKINGER, HEC Lausanne and FAME.

N°131 **Capital Gains Taxes, Irreversible Investment, and Capital Structure**
Norman SCHURHOFF, HEC, University of Lausanne and FAME.

International Center FAME - Partner Institutions

The University of Geneva

The University of Geneva, originally known as the Academy of Geneva, was founded in 1559 by Jean Calvin and Theodore de Beze. In 1873, The Academy of Geneva became the University of Geneva with the creation of a medical school. The Faculty of Economic and Social Sciences was created in 1915. The university is now composed of seven faculties of science; medicine; arts; law; economic and social sciences; psychology; education, and theology. It also includes a school of translation and interpretation; an institute of architecture; seven interdisciplinary centers and six associated institutes.

More than 13'000 students, the majority being foreigners, are enrolled in the various programs from the licence to high-level doctorates. A staff of more than 2'500 persons (professors, lecturers and assistants) is dedicated to the transmission and advancement of scientific knowledge through teaching as well as fundamental and applied research. The University of Geneva has been able to preserve the ancient European tradition of an academic community located in the heart of the city. This favors not only interaction between students, but also their integration in the population and in their participation of the particularly rich artistic and cultural life. ***http://www.unige.ch***

The University of Lausanne

Founded as an academy in 1537, the University of Lausanne (UNIL) is a modern institution of higher education and advanced research. Together with the neighboring Federal Polytechnic Institute of Lausanne, it comprises vast facilities and extends its influence beyond the city and the canton into regional, national, and international spheres.

Lausanne is a comprehensive university composed of seven Schools and Faculties: religious studies; law; arts; social and political sciences; business; science and medicine. With its 9'000 students, it is a medium-sized institution able to foster contact between students and professors as well as to encourage interdisciplinary work. The five humanities faculties and the science faculty are situated on the shores of Lake Leman in the Dorigny plains, a magnificent area of forest and fields that may have inspired the landscape depicted in Brueghel the Elder's masterpiece, the Harvesters. The institutes and various centers of the School of Medicine are grouped around the hospitals in the center of Lausanne. The Institute of Biochemistry is located in Epalinges, in the northern hills overlooking the city. ***http://www.unil.ch***

The Graduate Institute of International Studies

The Graduate Institute of International Studies is a teaching and research institution devoted to the study of international relations at the graduate level. It was founded in 1927 by Professor William Rappard to contribute through scholarships to the experience of international co-operation which the establishment of the League of Nations in Geneva represented at that time. The Institute is a self-governing foundation closely connected with, but independent of, the University of Geneva.

The Institute attempts to be both international and pluridisciplinary. The subjects in its curriculum, the composition of its teaching staff and the diversity of origin of its student body, confer upon it its international character. Professors teaching at the Institute come from all regions of the world, and the approximately 650 students arrive from some 60 different countries. Its international character is further emphasized by the use of both English and French as working languages. Its pluralistic approach - which draws upon the methods of economics, history, law, and political science - reflects its aim to provide a broad approach and in-depth understanding of international relations in general. ***http://heiwww.unige.ch***

INTERNATIONAL CENTER FOR
FINANCIAL ASSET MANAGEMENT AND ENGINEERING

40, Bd. du Pont d'Arve
PO Box, 1211 Geneva 4
Switzerland
Tel (++4122) 312 09 61
Fax (++4122) 312 10 26
http: //www.fame.ch
E-mail: admin@fame.ch

UNIVERSITÉ DE GENÈVE

THE GRADUATE INSTITUTE OF
INTERNATIONAL STUDIES

UNIVERSITÉ DE LAUSANNE