



TI 2006-024/2

Tinbergen Institute Discussion Paper

Non-parametric Estimation for Non-homogeneous Semi-Markov Processes: An Application to Credit Risk

André Monteiro^{1}*

Georgi V. Smirnov²

André Lucas^{3}*

¹ Vrije Universiteit Amsterdam,

² University of Porto,

³ Vrije Universiteit Amsterdam.

* Tinbergen Institute

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Nonparametric Estimation for Non-homogeneous Semi-Markov Processes: an Application to Credit Risk

André Lucas,^(a,b) André Monteiro^{(a,b)*} and Georgi V. Smirnov^(c)

^(a) *Department of Finance, Vrije Universiteit Amsterdam,*

^(b) *Tinbergen Institute Amsterdam,*

^(c) *Department of Applied Mathematics, University of Porto*

March 13, 2006

Abstract

We propose procedures for estimating the time-dependent transition matrices for the general class of finite nonhomogeneous continuous-time semi-Markov processes. We prove the existence, and Fréchet differentiability, of a unique solution for the system of Volterra integral equations which relates the transition matrix with the subdensity functions, therefore showing that the realized transition probabilities can be consistently estimated from window censored event-history data. An implementation of the method is presented, based on nonparametric estimators of the conditional hazard rate functions in the general and separable (multiplicative) cases. We use the resulting estimators for dealing with a central issue in credit risk. We consider the problem of obtaining estimates of the historical corporate default and rating migration probabilities using a dataset on credit ratings from Standard & Poor's.

Keywords: Nonhomogeneous semi-Markov processes, transition matrix, Volterra integral equations, separability, credit risk.

JEL classification codes: C13, C14, C33, C41, G11

* Corresponding author is André Monteiro, Vrije Universiteit Amsterdam, FEWEB, De Boelelaan 1105, 1081HV Amsterdam, Netherlands, phone: +31205988626, fax: +31205986020, email: amonteiro@feweb.vu.nl.

1 Introduction

The Semi-Markov Process (SMP) introduced independently by Lévy (1954) and Smith (1955) is a generalization of the Markov process. The SMP with a finite state space can be thought of as a Markov chain (MC) with a random transformation of the time scale (Pyke, 1961a). In particular this means that the sojourn times in each state can have arbitrary distributions that may also depend on the next state to be visited. Therefore SMP's are more flexible than the homogeneous MC as a tool for the statistical analysis of transition data.

Stochastic processes evolving in continuous time with jumps over a (finite) set of states (also known in the Finance literature as transition or 'migration' models) constitute the most natural class of econometric models to consider for modeling corporate credit rating and default data, see for example Kavvathas (2000), Lando and Skødeberg (2002), D'Amico et al. (2004) and Koopman et al. (2005).

The literature on semi-Markov processes has focused mainly on the time-homogeneous case, examples are Pyke (1961a and 1961b), Limnios and Ouhbi (1999 and 2005) and Alvarez (2005). However, in the field of credit risk, several empirical studies have explicitly shown that the time-homogeneity assumption is strongly rejected (Kavvathas, 2000, Lando and Skødeberg 2002, Koopman et al. 2005). Additionally, Kavvathas (2000) and Lando and Skødeberg (2002) also present some evidence that the empirical distribution function (d.f.) of the sojourn times (or duration) of debt issuers in each rating class seems to display negative duration dependence. If the resulting 'maturity' effect is severe, then a given rating class will, in fact, be heterogeneous with respect to the associated default probabilities according to the time spent by the issuer in that rating. In this paper, in particular, we answer the following question: what is the impact over the rating migration probabilities (including the default probabilities) of this known non-Markovian feature of credit rating data? We are able to address this issue through the main methodological contribution of this paper. We develop general computational procedures for estimating the transition matrices accounting both for the presence of duration dependence and time-inhomogeneity effects. These procedures are applicable to any statistical model inside the class of finite non-homogeneous semi-Markov processes (NHSMP) in continuous time, and do not depend on the particular estimation methodology used.

The NHSMP was defined for the first time in Iosifescu Manu (1972). The discrete-time finite state case was treated in De Dominicis and Janssen (1984). However, to the best of our knowledge, we provide for the first time a formal proof of the existence and uniqueness of the corresponding (continuous-time) transition probabilities in any finite time window. We derive a computationally efficient numerical procedure for obtaining the transition matrices from knowledge of the matrix of subdensity functions, and formally prove its convergence. Additionally we show

that this procedure preserves the consistency of any such estimator of the subdensity functions that may be used as input in an empirical application.

Such a consistent estimator for the subdensity functions can be constructed (for example) using nonparametric estimators of the conditional hazard rate functions. Two leading references on this topic are Nielsen and Linton (1995) and Linton et al. (2003). Using a baseline class of hazard rate estimators we implement this new methodology for studying the CreditPro7.0 dataset from Standard & Poor's containing the detailed rating histories of more than 7300 U.S. firms recorded at a daily frequency and over a period of almost 25 years. We estimate the historical (realized) issuer rating migration probabilities considering 7 distinct rating classes. This leads to some interesting empirical findings. First, we report large differences in estimated default (and transition) probabilities with respect to what is obtained using the Aalen-Johansen estimator (Aalen and Johansen, 1978). Second, when estimating separable hazard functions, and in contrast with Lando and Skødeberg (2002), we do not find the baseline hazard to be monotonically decreasing with the elapsed duration. Instead, for the great majority of rating classes, we find strong evidence of non-monotonic behavior. After an initial period of strong increase in the downgrade and default rates, which leads to a peak located around two to three years after the date when the current rating was issued, the intensity of this type of rating event decreases significantly. In a similar way for upward rating movements the peak in the hazard rates is located, roughly, thirty months after the rating was issued.

Transition matrices are a capital input for many credit risk management tools like, for example, J.P. Morgan's Credit Metrics or McKinsey's Credit Portfolio View. In order to 'backtest' a particular credit risk model and rating system, for assessing compliance with the Basle II norms, or for comparing two or more competing models in the development stage, it is of paramount importance that one is able to consistently estimate, ex-post, the (realized) transition probabilities, accounting for the main statistical features of that rating system, but using as few assumptions as possible. The nonparametric Aalen-Johansen estimator frequently used for this purpose in the literature (see for example the studies by Lando and Skødeberg 2002, Jafry and Schuermann 2004, or Koopman et al. 2005) allows for time-inhomogeneity but is built upon the unrealistic Markov property. This assumption, as mentioned, has been strongly rejected empirically (Kavvathas, 2000, Lando and Skødeberg 2002).¹

¹Note that the results reported in Jafry and Schuermann (2004) concern only the comparison of estimated transition matrices either through the use of standard matrix norms or their proposed metric, the "average of the singular values of the mobility matrix." Comparisons, across different estimation methods or chronological periods, of estimated transition matrices, as a whole, may not capture relevant differences for particular transition probabilities (for example the default probabilities) for a given rating class. Therefore, for risk management purposes,

This paper is organized as follows. In section 2 we recall informally the main definitions and some important results from the literature on non-homogeneous semi-Markov Processes. In section 3 we review the simple class of piecewise constant nonparametric estimators of the hazard rate functions both in the general and separable (multiplicative) cases. The explicit formulas of the nonparametric estimators of the semi-Markov kernel implied by the estimators of the hazard rate functions presented in the previous section are introduced in Section 4. We then formulate rigorously and solve the problem of obtaining the empirical semi-Markov transition probabilities using window-censored event-history data. A simulation study is conducted in section 5 to assess the small-sample behavior of the hazard rate estimators employed. We report our empirical results in section 6. Section 7 concludes. In the appendix we detail the proofs of the theorems in section 4.

2 Non-homogeneous Semi-Markov Processes

In this section we briefly recall the main definitions and results from the theory of non-homogeneous Markov Renewal Processes (NHMRP) which are directly relevant for our purposes.

Consider an arbitrary bivariate stochastic process $(S, T) = (S_n, T_n)_{n \in \mathbb{N}_0}$ defined on a complete, filtered, probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, describing the evolution in time of a given discrete-event system. Directly linked to this process we explicitly consider four other processes. The process $(S_n)_{n \in \mathbb{N}_0}$, gives the sequence of states visited by the system, and takes values in the set $\mathbb{S} = \{1, 2, \dots, s\}$, ($s < \infty$), termed the state space. The nondecreasing random sequence $(T_n)_{n \geq 0}$ is built from the consecutive transition times in the set \mathbb{R}_0^+ . The bivariate process $(S_n, X_n)_{n \in \mathbb{N}_0}$ where $X_n = T_n - T_{n-1}$ is the *sojourn time* in state S_{n-1} .²

Finally we consider the continuous-time *cadlag* process $S(t)$, $t \in \mathbb{R}_0^+$ that records the current state of the observational unit (or system).

Let $Q = (Q_{ij})$ be a matrix-valued function of two arguments, Q is called a matrix of *subdistribution functions* (s.d.f.) if each entry

$$Q_{ij} : \mathbb{R}_0^+ \times \bar{\mathbb{R}}_0^+ \rightarrow [0, 1]$$

is a nondecreasing measurable function of the second argument and satisfies

$$\sum_{j=1}^s Q_{ij}(t, \infty) = \lim_{\tau \rightarrow \infty} \sum_{j=1}^s Q_{ij}(t, \tau) = 1, \quad (1)$$

the extra-computational cost implied by an estimation procedure which accounts for the time-inhomogeneity is not material.

²By convention $X_0 = T_0 = 0$ and S_{-1} is left unspecified.

for every $t \in \mathbb{R}_0^+$. The s -valued function $\mathbf{g} = (g_i)$, with components

$$g_i : \mathbb{R}_0^+ \rightarrow [0, 1]$$

is termed the (time-varying) *vector of initial probabilities* if it satisfies

$$\sum_{i=1}^s g_i(t) = 1, \forall t \in \mathbb{R}_0^+. \quad (2)$$

We say that the bivariate process $(S_n, X_n)_{n \in \mathbb{N}_0}$ is a NHMRP, and the associated process $S(t)$, $t \in \mathbb{R}_0^+$ is a NHSMP, if the following assumptions hold

$$P[S_{n+1} = j, X_{n+1} \leq \tau \mid (S_0, T_0), \dots, (S_n, T_n) = (i, t)] = Q_{ij}(t, \tau), \quad (3)$$

and

$$P[S_0 = i \mid T_0 = t_0] = g_i(t_0). \quad (4)$$

The matrix of s.d.f. (Q_{ij}) is called in this context the *semi-Markov kernel*, and together with the vector of initial probabilities completely determines the stochastic behavior of the NHSMP.³ Unlike in De Dominicis and Janssen (1984) we define the non-homogeneous semi-Markov kernel using as arguments not the last and next transition times, $T_n = t$ and $T_{n+1} = t + \tau$, but instead the last transition time and the duration τ until the next transition. In this way we make explicit the two underlying time-scales: the chronological time-scale t and the ‘age’ (or duration) scale τ . Note that we assume the non-homogeneous semi-Markov kernel to be independent of the past number of recorded transitions.

Define for every $t \in \mathbb{R}_0^+$

$$p_{ij}(t) = \lim_{\tau \rightarrow \infty} Q_{ij}(t, \tau), \quad (5)$$

and

$$H_i(t, \tau) = \sum_{j=1}^s Q_{ij}(t, \tau). \quad (6)$$

Due to condition (1), $H_i(t, \tau)$ is a d.f. on $\bar{\mathbb{R}}_0^+$ with respect to the second argument τ . This can be interpreted as the d.f. of the sojourn times in state i that start at time t . Under conditions (3) and (4) the process $(S_n)_{n \in \mathbb{N}_0}$ is a non-homogeneous (discrete time) Markov chain with transition probabilities $p_{ij}(T_n)$, this is called the *embedded* Markov chain. The d.f. of the sojourn times in state i starting at time t that finish with a transition to state j is given by

$$F_{ij}(t, \tau) = \begin{cases} p_{ij}^{-1}(t) \cdot Q_{ij}(t, \tau) & , p_{ij}(t) > 0 \\ 0 & , \text{otherwise.} \end{cases} \quad (7)$$

³We always assume that $Q_{ii}(t, \tau) = 0$, $\forall i \in \mathbb{S}$

We consider throughout this paper that $Q_{ij}(t, \tau)$, as a function of the second argument τ , is absolutely continuous with respect to Lebesgue's measure. In this case the partial derivative of $Q_{ij}(t, \tau)$ with respect to τ is well defined and we denote it by $q_{ij}(t, \tau)$,

$$q_{ij}(t, \tau) = \frac{\partial Q_{ij}(t, \tau)}{\partial \tau}.$$

In the current credit risk context it is clearly reasonable to consider that the semi-Markov kernel satisfies the additional condition

$$Q_{ij}(t, 0) = 0, \forall t \in \mathbb{R}_0^+,$$

that is, that there is no probability mass at zero. This assumption is used in section 4 for deriving our numerical procedure for computing the transition probabilities.

The hazard rate function associated with a transition to state j at time $t + \tau$ in the case where the unit has entered state i at time t is defined by

$$\rho_{ij}(t, \tau) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}[S_{n+1} = j, \tau < X_{n+1} \leq \tau + h \mid (S_0, T_0), \dots, (S_n, T_n) = (i, t), X_{n+1} > \tau] \quad (8)$$

Using assumption (3), we can obtain the deterministic function ρ_{ij} as

$$\rho_{ij}(t, \tau) = \begin{cases} \frac{q_{ij}(t, \tau)}{1 - H_i(t, \tau)} & , p_{ij}(t) > 0 \text{ and } H_i(t, \tau) < 1 \\ 0 & , \text{ otherwise.} \end{cases} \quad (9)$$

By inverting this last equation we see that we can also obtain the semi-Markov kernel (3) from knowledge of the hazard rate functions,

$$Q_{ij}(t, \tau) = \int_0^\tau \exp[-\Lambda_i(t, u)] \rho_{ij}(t, u) du, \quad (10)$$

where

$$\Lambda_i(t, \tau) = \sum_{j=1}^s \int_0^\tau \rho_{ij}(t, u) du, \quad (11)$$

is the total integrated hazard 'out' of state i , that is the total cumulative 'force' of transition for leaving state i , this relates directly to the d.f. of sojourn times in state i .⁴ This one-to-one mapping between the semi-Markov kernel and the corresponding hazard rate functions means that if we can estimate the latter ones we are also able to obtain estimates of the former. Therefore we can follow the (large) literature on conditional hazard rate estimation in order to apply empirically models with a semi-Markov structure.

⁴In fact the *survival* probability in state i , that is, $S_i(t, \tau) = 1 - H_i(t, \tau)$ is directly given by $S_i(t, \tau) = \exp[-\Lambda_i(t, \tau)]$.

2.1 Associated counting processes and their intensities

We now define the counting processes (c.p.) associated with the process (S, T) . Let

$$N(u) = \sup_n \{ n : T_n \leq u \}, u > t_0, \quad (12)$$

denote the right-continuous c.p. recording the total number of transitions experienced by the statistical unit in the interval $(t_0, u]$. We denote by $\bar{N}(u)$ the corresponding left-continuous c.p. For each pair of states i and j we similarly define two c.p., the right-continuous process

$$N_{ij}(u) = \text{nr. of direct transitions from state } i \text{ to state } j \text{ in } (t_0, u], \quad (13)$$

and the corresponding left-continuous c.p., which we denote by \bar{N}_{ij} . The set of right-continuous c.p. are related by

$$N(u) = \sum_{i \neq j} N_{ij}(u),$$

and a similar equality holds for their left-continuous counterparts.

The history of the (S, T) process over a given interval $[t_0, u]$ can be completely determined by the sample path of the multivariate c.p. $\mathbf{N} = (N_{ij})_{i,j=1}^s$ over this interval, together with information on the initial state S_0 . That is the internal filtration associated with the process (S, T) equals the internal filtration associated with the process \mathbf{N} enlarged with the σ -algebra generated by the random variable S_0 .

We consider that the compensator of the (right-continuous) c.p. N_{ij} is absolutely continuous with respect to Lebesgue's measure. Therefore there is a corresponding *caglad* intensity process λ_{ij} given by

$$\begin{aligned} \lambda_{ij}(u) &= \lim_{h \downarrow 0} \frac{1}{h} P [\bar{N}_{ij}(u+h) - \bar{N}_{ij}(u) > 0 \mid \mathcal{F}_{u-}] \\ &= Y_i(u) \rho_{ij}(T_{\bar{N}(u)}, \bar{U}(u)), \end{aligned} \quad (14)$$

where

$$Y_i(u) = \begin{cases} 1 & , S_{\bar{N}(u)} = i \\ 0 & , \text{otherwise,} \end{cases} \quad (15)$$

is an *exposure* indicator and

$$\bar{U}(u) = u - T_{\bar{N}(u)}$$

is a left-continuous version of the *backward-recurrence time*, that is, the elapsed duration in the current state.

2.2 Transition probabilities

For continuous-time Markov chains it is well known that the matrix of transition probabilities over a given time window $(t, t + \tau]$ denoted by

$$P_{ij}(t, t + \tau) = P[S(t + \tau) = j \mid S(t) = i], \tau > 0 \quad (16)$$

is the unique solution to the Kolmogorov system of ‘backward’ and ‘forward’ PDEs (Goodman and Johansen, 1973). In the current context, and due to the presence of duration dependence, we’re interested in the following ‘age-specific’ or ‘left-truncated’ transition probabilities,

$$p_{ij}^*(t, \tau_0, \tau) = P[S(t + \tau) = j \mid U(t + \tau_0) = \tau_0, S(t + \tau_0) = i]; \tau - \tau_0 > 0. \quad (17)$$

That is, from a risk management perspective we are interested in estimating the probability of finding each unit⁵ in state j by time $t + \tau$ if we know that at time $t + \tau_0$ the unit is in state i , and the elapsed duration in that state is precisely τ_0 units of time.

For obtaining these ‘left-truncated’ transition probabilities we must start by considering the related ‘renewal’ transition probabilities

$$p_{ij}(t, \tau) = P[S(t + \tau) = j \mid N(t) - \bar{N}(t) = 1, S_{N(t)} = i], \tau > 0, \quad (18)$$

these are the probabilities of finding the unit in state j , τ units of time after it entered state i , given that this transition happened at (chronological) time t .

These probabilities satisfy the following system of non-homogeneous Volterra integral equations of the second kind on two independent variables (t and τ)

$$p_{ij}(t, \tau) = \delta_{ij}(1 - H_i(t, \tau)) + \sum_{k=1}^s \int_0^\tau p_{kj}(t + u, \tau - u) q_{ik}(t, u) du, \quad i, j = 1, \dots, s, \quad (19)$$

where δ_{ij} denotes Kronecker’s symbol. These equations are the direct counterpart to the Kolmogorov ‘evolution’ equations for the Markov setting. For writing down an expression for the ‘age-specific’ transition probabilities (17) denote by $q_{ij}^*(t, \tau_0, \tau)$ the density corresponding to the ‘left-truncated’ s.d.f. of the durations in state i starting at time t which are larger than τ_0 units of time and which end with a transition to state j , that is

$$Q_{ij}^*(t, \tau_0, \tau) = \frac{Q_{ij}(t, \tau) - Q_{ij}(t, \tau_0)}{1 - Q_{ij}(t, \tau_0)}.$$

⁵In the empirical application of section 6, the statistical units are the bond issuers.

The ‘age-specific’ transition probabilities in equation (17) are given by

$$p_{ij}^*(t, \tau_0, \tau) = \delta_{ij} (1 - H_i^*(t, \tau_0, \tau)) + \sum_{k=1}^s \int_{\tau_0}^{\tau} p_{kj}(t+u, \tau-u) q_{ik}^*(t, \tau_0, u) du, \quad (20)$$

where $H_i^*(t, \tau_0, \tau)$ denotes the d.f. function of left-truncated sojourn times in state i and is obtained by adding the s.d.f. $Q_{ik}^*(t, \tau_0, \tau)$ with $k = 1, \dots, s$ and $p_{kj}(\cdot, \cdot)$ denotes the transition probabilities defined in equation (18).

3 Nonparametric estimation

In this section, we consider the problem of estimating nonparametrically the main quantities of interest associated with any NHSMP. In this first version of the current paper we base our procedures on the class of kernel hazard estimators known as ‘bin smoothers,’ this corresponds to the use of indicator functions as the underlying smoothing kernels. Hastie and Tibshirani (1990) provide a discussion of this class of smoothers in the context of nonparametric regression. However, it is conceptually straightforward to base our procedures on hazard rate estimators employing other classes of smoothing kernels. Nielsen and Linton (1995) and Linton et al. (2003) give an exposition on the general issue of conditional hazard rate estimation using kernel methods, both for the general and separable cases. Besides the general (nonseparable) case of a bivariate hazard rate function we also implement the separable case where the hazard rate functions defined in (8) have the multiplicative form⁶

$$\rho_{ij}(t, \tau) = \alpha_{ij}(\tau) \psi_{ij}(t + \tau). \quad (21)$$

This multiplicative specification for the hazard rate functions encompasses the celebrated Proportional Hazards (PH) model of Cox (1972) for the case of a homogeneous population. In this particular NHSMP we assume the existence of a time-invariant transition-specific baseline hazard $\alpha_{ij}(\tau)$ which is multiplied by a function of chronological time $\psi_{ij}(t + \tau)$. In the original PH model, without unit-specific covariates, ψ_{ij} can be seen as the combined effect of the (common across units) time-varying regression covariates.

Specification (21) encompasses both the homogeneous SMP as considered for example by Limnios and Ouhbi (1999 and 2005), and the non-homogeneous continuous-time Markov Chain (CTMC) process. The former is obtained by setting $\psi_{ij}(t + \tau) \equiv 1$, the later by setting $\alpha_{ij}(\tau) \equiv 1$.

To identify ψ_{ij} and α_{ij} simultaneously we impose the arbitrary normalization

⁶The coordinates (t, τ) in the Lexis diagram represent a straight line segment departing from the point $(t, 0)$ to the point $(t + \tau, \tau)$ in the Cartesian (t, τ) plane.

$$\psi_{ij}(t_0) = 1. \quad (22)$$

Now consider a panel of K units following independently a NHSMP characterized by a common SM kernel as in (3) and arbitrary vectors of initial probabilities. We consider estimation in the presence of left-truncation at time $t = 0$ and independent right-censoring of type I at time $t = T_1$ for the entire panel. In order to proceed we need to introduce some additional notation. Consider for unit k (right-continuous) c.p. $N^{(k)}(t)$ and $N_{ij}^{(k)}(t)$ defined as in (12) and (13). Also, let

$$N(t) = \sum_{k=1}^K N^{(k)}(t),$$

denote the *global pooled c.p.*, i.e., the process counting the total number of transitions of any type recorded across all units in the panel. Let $t_0 < t_1 < \dots < t_{N(T_1)+1} = T_1$ denote the corresponding sequence of ‘jump’ times of the process $N(t)$, together with the censoring time T_1 .

Consider for every possible transition from state i to state j , and unit k an indicator variable defined as

$$\Delta N_{ij}^{(k)}(t) = N_{ij}^{(k)}(t) - \bar{N}_{ij}^{(k)}(t),$$

where $\bar{N}_{ij}^{(k)}$ is the left-continuous c.p. associated with $N_{ij}^{(k)}$. Let $Y_i^{(k)}(t)$, defined as in (15), denote an indicator variable taking the value 1 if unit k is in state i at time t^- , where $t - t^-$ is arbitrarily small. The Likelihood function for such a panel conditional on the observed initial state of each unit can be written as,

$$L(\rho \mid \mathcal{F}_T) = \prod_{m=1}^{N(T)+1} \prod_{k=1}^K \prod_{i \neq j} \left\{ [\rho_{ij}(t_m - \bar{U}^{(k)}(t_m), \bar{U}^{(k)}(t_m))]^{\Delta N_{ij}^{(k)}(t_m)} \times \right. \quad (23)$$

$$\left. \exp \left(-Y_i^{(k)}(t_m) \int_{t_{m-1}}^{t_m} \rho_{ij}(t - \bar{U}^{(k)}(t), \bar{U}^{(k)}(t)) dt \right) \right\}.$$

We start by considering the general case where the hazard rate functions depend simultaneously on $\bar{U}^{(k)}(t + \tau)$ and $t + \tau$ without imposing any parametric assumptions over the structure of this function. We approximate this bivariate hazard rate using a piecewise constant function,

$$\rho_{ij}^*(t, \tau) = \sum_{p=0}^{M_{ij}-1} \sum_{m=0}^{L_{ij}} \rho_{ijpm} \mathbf{1}_{(x_{ij,p}, x_{ij,p+1}]}(\tau) \cdot \mathbf{1}_{(t_{ij,m}, t_{ij,m+1}]}(t + \tau). \quad (24)$$

The sequence $x_{ij,0} = 0 < x_{ij,1} < \dots < x_{ij,M_{ij}}$ consists of an arbitrary partition of the interval $[0, X_{ij}^{max}]$, where X_{ij}^{max} denotes the maximum between the largest

observed duration in state i that ended with a transition to state j and the largest right-censored duration in state i . In a similar way, the sequence $t_{ij,0} = 0 < t_{ij,1} < \dots < t_{ij,L_{ij}+1} = T_1$ denotes an arbitrary partition of the time-window $[0, T_1]$. Substituting (24) in (23) taking the natural logarithm and collecting terms yields the loglikelihood function,

$$l(\rho^* | \mathcal{F}_{T_1}) = \sum_{i \neq j} \sum_{m=0}^{L_{ij}} \sum_{p=0}^{M_{ij}-1} (o_{ijpm} \ln \rho_{ijpm} - e_{ijpm} \rho_{ijpm}), \quad (25)$$

where

$$o_{ijpm} = \sum_{k=1}^K \int_0^{T_1} \mathbf{1}_{(x_{ij,p}, x_{ij,p+1}]} (\bar{U}^{(k)}(t)) \cdot \mathbf{1}_{(t_{ij,m}, t_{ij,m+1}]}(t) dN_{ij}^{(k)}(t),$$

denotes the total number of transitions from state i to state j and

$$e_{ijpm} = \sum_{k=1}^K \int_0^{T_1} \mathbf{1}_{(x_{ij,p}, x_{ij,p+1}]} (\bar{U}^{(k)}(t)) \cdot \mathbf{1}_{(t_{ij,m}, t_{ij,m+1}]}(t) Y_i^{(k)}(t) dt,$$

is the total exposure (i.e. the cumulative length of time) observed in cell (p, m) of the bivariate grid generated by the Cartesian product of both partitions $x_{ij,0} = 0 < x_{ij,1} < \dots < x_{ij,M_{ij}}$ and $t_{ij,0} = 0 < t_{ij,1} < \dots < t_{ij,L_{ij}+1} = T_1$. Equating to zero the partial derivatives of the loglikelihood function with respect to the unknown parameters ρ_{ijpm} and solving for these unknowns, yields the following estimators,

$$\hat{\rho}_{ijpm} = \frac{o_{ijpm}}{e_{ijpm}}. \quad (26)$$

The corresponding asymptotic variance can be estimated by

$$\widehat{V}[\hat{\rho}_{ijpm}] = \frac{o_{ijpm}}{e_{ijpm}^2}. \quad (27)$$

Similarly for the separable (multiplicative) case (21) we approximate both (univariate) components of the hazard rate function ρ_{ij} by two piecewise constant functions α_{ij}^* and ψ_{ij}^* defined by,

$$\alpha_{ij}^*(\tau) = \sum_{p=0}^{M_{ij}-1} \alpha_{ijp} \mathbf{1}_{(x_{ij,p}, x_{ij,p+1}]}(\tau), \quad (28)$$

and

$$\psi_{ij}^*(t + \tau) = \sum_{m=0}^{L_{ij}} \psi_{ijm} \mathbf{1}_{(t_{ij,m}, t_{ij,m+1}]}(t + \tau). \quad (29)$$

Due to the scaling condition (22), we have $\psi_{ij0} = 1$. The sequences $x_{ij,0} = 0 < x_{ij,1} < \dots < x_{ij,M_{ij}}$ and $t_{ij,0} = t_0 < t_{ij,1} < \dots < t_{ij,L_{ij}+1} = T_1$ are similar to the ones defined for the nonseparable case.

Substituting (28) and (29) in (23), taking the natural logarithm and collecting terms yields the loglikelihood function

$$\begin{aligned} l(\alpha^*, \psi^* \mid \mathcal{F}_{T_1}) &= \sum_{i \neq j} \sum_{p=0}^{M_{ij}-1} d_{ijp} \ln \alpha_{ijp} + \sum_{i \neq j} \sum_{m=1}^{L_{ij}} b_{ijm} \ln \psi_{ijm} \\ &\quad - \sum_{i \neq j} \sum_{m=0}^{L_{ij}} \sum_{p=0}^{M_{ij}-1} \psi_{ijm} \alpha_{ijp} e_{ijpm}, \end{aligned} \quad (30)$$

where

$$d_{ijp} = \sum_{k=1}^K \int_0^{T_1} \mathbf{1}_{(x_{ij,p}, x_{ij,p+1}]} (U^{(k)}(t)) dN_{ij}^{(k)}(t), \quad (31)$$

$$b_{ijm} = \sum_{k=1}^K \int_0^{T_1} \mathbf{1}_{(t_{ij,m}, t_{ij,m+1}]} (t) dN_{ij}^{(k)}(t), \quad (32)$$

and⁷ e_{ijpm} denotes again the total exposure in cell (p, l) of the bivariate grid. Now, the loglikelihood l is strictly concave as a function of the unknown values of the step functions α_{ij}^* and ψ_{ij}^* (denote these by the vectors $\tilde{\psi}_{ij}$ and $\tilde{\alpha}_{ij}$ of dimensions L_{ij} and M_{ij} respectively). Additionally when we let the vectors $\tilde{\psi}_{ij}$ and $\tilde{\alpha}_{ij}$ increase to ∞ we see that l decreases to $-\infty$, therefore there is a unique pair $(\hat{\psi}_{ij}, \hat{\alpha}_{ij})$ which maximizes l . Numerical optimization of (30) is, under this setting, a problem for which there are available several fast and robust algorithms. A simple computational procedure is as follows. Taking partial derivatives of the loglikelihood function given in (30) with respect to its arguments and equating these to zero, yields the two following sets of estimating equations,

$$\begin{aligned} \mathbf{e}_{ij} \tilde{\psi}_{ij} &= (d_{ijp} / \alpha_{ijp}) - (e_{ijp0}) \\ \mathbf{e}'_{ij} \tilde{\alpha}_{ij} &= (b_{ijm} / \psi_{ijm}). \end{aligned} \quad (33)$$

The matrix $\mathbf{e}_{ij} = (e_{ijpm})$ contains all the e_{ijpm} elements with $p = 0, \dots, M_{ij} - 1$ and $m = 1, \dots, L_{ij}$. Iterating these two sets of nonlinear equations starting from an arbitrary (strictly positive) value for one of the unknowns, yields a sequence of vectors which converges to the unique solution $(\hat{\psi}_{ij}, \hat{\alpha}_{ij})$. This is so because this

⁷These statistics satisfy $\sum_{m=0}^{\tilde{N}_{ij}(T_1)} b_{ijm} = \sum_{p=0}^{N_{ij}(T_1)-1} d_{ijp}$.

procedure is interpretable as a gradient ascent method applied to the maximization of a concave function.

The asymptotic variance of the resulting estimators can be estimated, for example, by evaluating the symmetrical of the inverse of the Hessian matrix,

$$Hess = \begin{bmatrix} A & -\mathbf{e}_{ij} \\ -\mathbf{e}'_{ij} & C \end{bmatrix},$$

at the estimates resulting from the system of estimating equations (33), where $A = \text{diag}\left(-\frac{d_{ijp}}{\alpha_{ijp}^2}\right)$, $C = \text{diag}\left(-\frac{b_{ijm}}{\psi_{ijm}^2}\right)$ are two diagonal matrices, and again $\mathbf{e}_{ij} = (e_{ijpm})$.

4 Empirical Transition Probabilities

As we have seen in section 2 we can obtain the semi-Markov kernel (10) from the estimated hazard rate functions. The first required step consists in obtaining the integrated hazard. The estimator of the integrated hazard associated with state i over a given path (t, τ) as defined in equation (11) is obtained as,

$$\hat{\Lambda}_i(t, \tau) = \sum_{j=1}^s \sum_{n=1}^{D_{ij}} \hat{\rho}_{ij,n} (\tau \wedge x_{ij,n}^* - x_{ij,n-1}^*), \quad (34)$$

where⁸ n indexes chronologically the cells (p, m) of the ij -grid which have a non-empty intersection with the path (t, τ) and D_{ij} is the total number of these cells. The estimator of the (constant) value of the hazard rate ρ_{ij} over cell n is denoted by $\hat{\rho}_{ij,n}$. Equation (10) suggests an estimator for the subdensity functions q_{ij}

$$\hat{q}_{ij}(t, \tau) = \exp\left(-\hat{\Lambda}_i(t, \tau)\right) \hat{\rho}_{ij, D_{ij}}, \quad (35)$$

and explicitly for the semi-Markov kernel,

$$\hat{Q}_{ij}(t, \tau) = \sum_{n=1}^{D_{ij}} \left[e^{-\hat{\rho}_{i,n} x_{i,n-1}^*} - e^{-\hat{\rho}_{i,n} (\tau \wedge x_{i,n}^*)} \right] \frac{\hat{\rho}_{ij,n} \exp\left[\sum_{k=0}^{n-1} x_{i,k}^* (\hat{\rho}_{i,k+1} - \hat{\rho}_{i,k})\right]}{\hat{\rho}_{i,n}}, \quad (36)$$

where $\hat{\rho}_{i,n} = \sum_{j=1}^s \hat{\rho}_{ij,n}$. The points $x_{ij,n}^*$ result from the reunion of the set of points $x_{ij,p}$ of the partition of the duration-scale with the points $t_{ij,m} - t$ whenever these are positive and smaller than τ . In a similar way the points $x_{i,n}^*$ are obtained from

⁸ $a \wedge b$ is short-hand notation for $\min\{a, b\}$

the reunion of the sets of points $x_{ij,p}^*$ over $j = 1, \dots, s$.

We now consider formally the problem of estimating the transition probabilities over a specific time-window and for a given backward recurrence-time spent in the current state. That is we want to estimate consistently the quantity in equation (17). For doing this we need to obtain the empirical ‘renewal’ transition probabilities p_{ij} by solving the Volterra integral equations (19) using as input a consistent estimator of the subdensity functions. We start by stating an existence and uniqueness theorem regarding these transition probabilities over the time-window corresponding to the available data (the proofs are contained in the appendix).

Let $T_2 = \max_{i,j=1,\dots,s} \{X_{ij}^{max}\}$. Set $T(t) = \min\{T_2, T_1 - t\}$, $\mathcal{T} = \{(t, \tau) \mid 0 \leq t \leq T_1, 0 \leq \tau \leq T(t)\}$ and $\mathcal{T}^\circ = \{(t, \tau) \mid 0 \leq t \leq T_1, 0 \leq \tau < T(t)\}$. Consider a space $Car(\mathcal{T})$ of bounded functions $f : \mathcal{T} \rightarrow \mathbb{R}$ satisfying the following conditions:

1. $f(\cdot, \tau) : [0, T_1 - \tau] \rightarrow \mathbb{R}$ is measurable for all $\tau \in [0, T_2]$,
2. $f(t, \cdot) : [0, T(t)] \rightarrow \mathbb{R}$ is continuous for almost all $t \in [0, T_1]$,

The space $Car(\mathcal{T})$ equipped with the norm

$$\|f\|_{Car} = \sup_{\mathcal{T}} |f(t, \tau)|$$

is a Banach space. Consider also a Banach space $Car_s(\mathcal{T})$ of matrix-valued functions $p = (p)_{ij}$, $i, j = 1, \dots, s$. Every component p_{ij} , $i, j = 1, \dots, s$, belongs to the space $Car(\mathcal{T})$. The norm in $Car_s(\mathcal{T})$ is defined by

$$\|p\|_{Car_s} = \max_{1 \leq i, j \leq s} \sup_{(t, \tau) \in \mathcal{T}} |p_{ij}(t, \tau)|.$$

Assume that the functions $Q_{ij} : \mathcal{T} \rightarrow \mathbb{R}$, $i, j = 1, \dots, s$, satisfy the following conditions (C):

1. there exist non-negative functions $q_{ij} \in L_\infty(\mathcal{T})$, $i, j = 1, \dots, s$ such that

$$Q_{ij}(t, \tau) = \int_0^\tau q_{ij}(t, u) du, \quad (t, \tau) \in \mathcal{T}^\circ,$$

2. $Q_{ij}(t, T(t)) = \lim_{\tau \uparrow T(t)} Q_{ij}(t, \tau) + h_{ij}(t)$, where $h_{ij} : [0, T_1] \rightarrow \mathbb{R}$, $i, j = 1, \dots, s$, are non-negative measurable functions,
3. $\sum_{k=1}^s Q_{ik}(t, T(t)) = 1$ for all $t \in [0, T_1]$.

Now consider functions $P_{ij} : \mathcal{T} \rightarrow \mathbb{R}$, defined by

$$P_{ij}(t, \tau) = p_{ij}(t, \tau) + \sum_{k=1}^s \theta(\tau - T(t)) h_{ik}(t), \quad i, j = 1, \dots, s,$$

where $\theta(\cdot)$ is the Heaviside step function, and the functions $p_{ij} : \mathcal{T} \rightarrow \mathbb{R}$, $i, j = 1, \dots, s$, as defined in (18) satisfy the system of integral equations (19)

Theorem 1 *Equation (19) has a unique solution $\hat{p} \in Car_s(\mathcal{T})$ satisfying $\hat{p}_{ij}(t, \tau) \in [0, 1]$, $(t, \tau) \in \mathcal{T}$, $i, j = 1, \dots, s$, $\sum_{j=1}^s \hat{p}_{ij}(t, \tau) = 1$ and $\hat{p}_{ij}(t, 0) = \delta_{ij}$, $t \in [0, T_1]$, $i, j = 1, \dots, s$. This solution is Fréchet differentiable as function of the matrix of subdensity functions (q_{ij}) .*

This theorem implies that if we could directly compute \hat{p} using as input estimates of (q_{ij}) obtained from a consistent estimator, the resulting estimator of the ‘renewal’ transition matrix would not only be also consistent, but additionally, we could use the delta method for inferring its asymptotic distribution from the corresponding distribution of the estimator (\hat{q}_{ij}) employed (see for example vd Vaart, 1998, Chapters 18 and 20). In practice we need to approximate numerically the solution $\hat{p}_{ij}(t, \tau)$ over \mathcal{T} by means of a finite system of algebraic equations. We now turn to this issue.

Consider $q_{ij} \in L_\infty(\mathcal{T})$ with the norm less than or equal to $M > 0$. Assume that there exist a partition of the interval $[0, T_1]$, $0 = t_0 < t_1 < \dots < t_{\tilde{N}_1} = T_1$, a partition of the interval $[0, T_2]$, $0 = \tau_0 < \tau_1 < \dots < \tau_{\tilde{N}_2} = T_2$, a constant $L > 0$, and functions $\check{q}_{ij}^{n_1 n_2}(t, \cdot)$, $n_1 = 1, \dots, \tilde{N}_1$, $n_2 = 1, \dots, \tilde{N}_2$, satisfying the Lipschitz condition

$$|\check{q}_{ij}^{n_1 n_2}(t', \tau') - \check{q}_{ij}^{n_1 n_2}(t'', \tau'')| \leq L(|t' - t''| + |\tau' - \tau''|) \quad (37)$$

for all $i, j = 1, \dots, s$, $(t', \tau'), (t'', \tau'') \in \mathcal{T} \cap [t_{n_1-1}, t_{n_1}] \times [\tau_{n_2-1}, \tau_{n_2}]$, $n_1 = 1, \dots, \tilde{N}_1$, $n_2 = 1, \dots, \tilde{N}_2$, and such that

$$q_{ij}(t, \tau) = \check{q}_{ij}^{n_1 n_2}(t, \tau), \quad (t, \tau) \in \mathcal{T} \cap [t_{n_1-1}, t_{n_1}] \times [\tau_{n_2-1}, \tau_{n_2}],$$

with

$$n_1 = 1, \dots, \tilde{N}_1, \quad n_2 = 1, \dots, \tilde{N}_2.$$

Without loss of generality there exist positive integers $N_1 \geq \tilde{N}_1$ and $N_2 \geq \tilde{N}_2$ such that $T_1/N_1 = T_2/N_2 = \tilde{\Delta}$, $\{t_n \mid n = 1, \dots, \tilde{N}_1 - 1\} \subset \{l\tilde{\Delta} \mid l = 1, \dots, N_1 - 1\}$, and $\{\tau_n \mid n = 1, \dots, \tilde{N}_2 - 1\} \subset \{l\tilde{\Delta} \mid l = 1, \dots, N_2 - 1\}$. Put

$$\alpha_k(\xi) = s(T_2 L + 2M)k(sT_2 M)^{k-1} \xi, \quad k = 1, 2, \dots$$

and

$$\beta_n(\xi) = \frac{\alpha_n(\xi) + \sum_{k=1}^{n-1} \alpha_k(\xi)}{1 - \alpha_n(\xi) - (sT_2 M)^n / n!}, \quad n = 1, 2, \dots$$

Let N be a positive integer. Set $\Delta = \tilde{\Delta}/2^N$. Let $m_1 = 0, \dots, N_1$ and $m_2 = 0, \dots, N_2$. Consider functions \hat{p}_{ij}^Δ , $i, j = 1, \dots, s$, defined at the points $((m_1 + \bar{l} + \bar{r}/2^N)\tilde{\Delta}, (m_2 - \bar{l} - \bar{r}/2^N)\tilde{\Delta})$, $\bar{l} = 0, \dots, m_2$, $\bar{r} = 0, \dots, 2^N - 1$, by the following system of linear equations

$$\begin{aligned} & \hat{p}_{ij}^\Delta((m_1 + \bar{l} + \bar{r}/2^N)\tilde{\Delta}, (m_2 - \bar{l} - \bar{r}/2^N)\tilde{\Delta}) \\ &= \delta_{ij}(1 - H_i((m_1 + \bar{l} + \bar{r}/2^N)\tilde{\Delta}, (m_2 - \bar{l} - \bar{r}/2^N)\tilde{\Delta})) \\ &+ \Delta \sum_{k=1}^s \sum_{l=2}^{m_2 - \bar{l}} \sum_{r=0}^{\bar{r}-1} \hat{p}_{kj}^\Delta((m_1 + \bar{l} + l + r/2^N)\tilde{\Delta}, (m_2 - \bar{l} - l - r/2^N)\tilde{\Delta}) q_{ik}((m_1 + \bar{l})\tilde{\Delta}, (l-1)\tilde{\Delta}). \end{aligned} \tag{38}$$

From a computational perspective, the remarkable feature of this numerical approximation method consists in the fact that the resulting linear system of (matrix) equations can be solved simply by backward substitution. We are therefore able to avoid the need to invert the corresponding square matrix of coefficients, which is typically very large.⁹ We now provide a theoretical upper bound for the error implied by this numerical approximation. This inequality insures not only the convergence of the method to the true solution, but also the consistency of the estimator of the transition matrix which results from its application using estimates of the matrix (q_{ij}) obtained from a consistent estimator.

Theorem 2 *Let n be a positive integer such that $(sT_2M)^n/n! < 1$. If $\tilde{\Delta}$ satisfies*

$$\alpha_n(\tilde{\Delta}) < 1 - \frac{(sT_2M)^n}{n!},$$

then for all $m_1 = 1, \dots, N_1$ and $m_2 = 1, \dots, N_2$ the following inequality holds

$$|\hat{p}_{ij}^\Delta(m_1\tilde{\Delta}, m_2\tilde{\Delta}) - \hat{p}_{ij}(m_1\tilde{\Delta}, m_2\tilde{\Delta})| \leq (s + \hat{L})^2 M^2 T_2 \Delta \exp((s + \hat{L})MT_2) + \beta_n(\tilde{\Delta}),$$

where

$$\hat{L} = \max\{2L\tilde{\Delta} + M, LT_2 + M + (LT_2 + 16N_2M)(1 + \beta_n(\tilde{\Delta}))\} \frac{(1 + 4sMT_2)^{N_2-1} - 1}{4sMT_2}.$$

Obtaining a consistent estimator of the ‘left-truncated’ transition probabilities (17) can be easily done by evaluating numerically the expression (20) using the ‘grid’ of values \hat{p}_{ij}^Δ .

⁹In section 6 this matrix is of order 3500

5 Simulation Study

In this section we present the results of the Monte Carlo study conducted with the aim of assessing the small sample behavior of the estimators of the hazard rate functions presented in section 3. The simulation setup is as follows, there are 3 states $\mathbb{S} = \{1, 2, 3\}$, state 3 is absorbing and denotes default. At time $t = 0$ there are $\frac{K}{2}$ units in each one of the two non-absorbing states. For each one of the K units we consider a set of four intensity processes as defined in (14) with corresponding hazard rate functions as in (21). We start with the simplest case where the baseline hazard is a constant and the time-index is identically equal to one. This corresponds to the simple case of a homogeneous continuous time MC. We can therefore test if our estimators of the hazard rate functions are capable of recovering this particular case of a SMP. We perform 500 replications of the simulations. The results are shown in the next pictures.

<INSERT FIGURE 1 ABOUT HERE>

<INSERT FIGURE 2 ABOUT HERE>

<INSERT FIGURE 3 ABOUT HERE>

<INSERT FIGURE 4 ABOUT HERE>

It is possible to see as the number of units increases (and therefore also the number of recorded transitions of each type) the estimated components of the hazard rate functions in equation (21) seem to approach the true values with decreasing variance (as it is to be expected).

6 Empirical Application

In this section we report the empirical results obtained with a baseline SMP model for the credit rating process. We start by considering only two rating classes, investment grade (from BBB^- up to AAA) and subinvestment or speculative grade (bellow BBB^-), Default is considered an absorbing state. We do not model explicitly the non-rated (N.R.) class as a state, instead all transitions into N.R. are taken as censored observations. The ‘cells’ or ‘bins’ used for the estimation of the hazard rates were constructed using an adaptive method. The number of observations is constant across ‘bins’ while the dimensions of each ‘bin’ are random, therefore the

‘smoothing’ parameter (equivalent to the ‘bandwidth’) is in this case the number of observations used to define each bin.

<INSERT FIGURE 5 ABOUT HERE>

By looking at the number of quarterly recorded defaults by issuers rated as subinvestment grade level, we can see, by conditioning on the age of the rating (depicted at a yearly resolution), that most defaults take place within the first four to five years after receiving this rating. This is to be expected as speculative grade rated issuers represent, by definition, risky investments.

<INSERT FIGURE 6 ABOUT HERE>

The total exposure (measured in firm-years) in each ‘cell’ (covering one quarter in calendar time and one year in ‘age’) indicates the concentration of firms in that interval of time with a similar rating ‘maturity.’ In this plot it is possible to see the existence of 4 historical periods where there was a large increase in the number of firms entering the speculative grade level (both due to the downgrade from the investment level and to the request of a first rating) the initial cohort at the start of the time-window (1 January 1981, this cohort is somewhat special due to the fact that most of its firms would most probably already had their rating for an - unknown to us - period of time¹⁰), a second period from late 1986 to early 1990 (this large cohort was mostly due to firms being downgraded to the subinvestment grade in this period, this conclusion results from the analysis of the grid count of this type of event - see next graphic) a third period in the middle nineties (mostly due to initial ratings) and finally a period from 1999 to early 2001 also due to initial ratings.

<INSERT FIGURE 7 ABOUT HERE>

This plot documents the recorded number of downgrade rating actions. With exception of a high-intensity period between 1985 and 1987 the picture is here much more homogeneous throughout time than the recorded defaults for subinvestment grade issuers.

<INSERT FIGURE 8 ABOUT HERE>

The total exposure of firms in the investment level category shows that the largest such cohort corresponds to firms which entered the DB at the left-limit of the time-window.

¹⁰We can treat these durations as left-truncated, under a time homogeneity assumption - over the unknown duration to the left of the time-window - we can make the adequate correction in the likelihood function to account for this fact.

<INSERT FIGURE 9 ABOUT HERE>

<INSERT FIGURE 10 ABOUT HERE>

<INSERT FIGURE 11 ABOUT HERE>

Assuming a multiplicative structure for the hazard rates, we can extract the marginal impact of the rating ‘age’ effect. We can clearly see that there is a peak in the default intensity for subinvestment grade rated issuers roughly three years after they received that rating. This non-monotonic pattern is in contrast with the results reported in Lando and Skødeberg (2002), this is due to the fact that we have dropped the parametric assumption used in that study. The Gompertz baseline hazard used in that paper is not sufficiently flexible to accommodate non-monotonic behavior, and therefore seems to have captured only the right tail of the baseline hazard.

The time-index component expresses the relative degree of ‘risk’ of a given historical period with respect to a reference period. Due to our identifying restriction (22) this is the first chronological period (for the case of defaults by subinvestment grade issuers this was the period covering from January 1981 to late 1984). We can clearly see, for example, a very high intensity of default in late 1986. Next we present the empirical transition matrices for the 7 grades rating system (the CCC class includes all issuers with a credit rating lower than B^-),

<INSERT TABLE 1 ABOUT HERE>

<INSERT TABLE 2 ABOUT HERE>

<INSERT TABLE 3 ABOUT HERE>

<INSERT TABLE 4 ABOUT HERE>

Comparing the estimated SM ‘renewal’ transition matrix with its Markov counterpart (obtained using the Aalen-Johansen (A-J) estimator) it is possible to see the impact of the (estimated) non-constant baseline hazards on the transition probabilities. Several features are apparent. First for investment grade rated issuers it is possible to see that there is much more ‘mass’ on the main diagonal (if an issuer has just received a given rating in the investment grade level then this issuer is much less ‘mobile’ than another issuer - with the same rating - but with a longer

‘history’) than given by the A-J estimator (which effectively ‘averages’ the issuers in a given rating class over the ‘age’ of that rating). This effect is still visible for BB and B rated issuers. On the contrary for CCC rated issuers the opposite effect is present, these issuers seem to be even more ‘mobile’ than what the A-J estimator indicates. Simultaneously the A-J estimator seems to underestimate the probabilities of more extreme rating movements - this is particularly true for the default probabilities - these are systematically higher than the A-J estimator indicates, for some cases the difference corresponds to a factor larger than two. Additionally many transition probabilities away from the main diagonal are estimated as zero (at this number of digits) in the A-J estimator but not for the SM case. On the contrary the A-J estimator overestimates the probability of a rating movement of a single class. This shows clearly the large impact of the ‘rating’ age over the transition probabilities.

<INSERT TABLE 5 ABOUT HERE>

In table 5 we check the stability of the estimated SM ‘renewal’ transition matrices (corresponding to the year 2000) with respect to the number of points of the grid where the integral appearing in (19) is to be evaluated numerically, which leads to the linear system (38). There is almost no noticeable difference between the use of 500 or 2000 grid points (we therefore used 500 points for estimating the SM transition matrices).

<INSERT TABLE 6 ABOUT HERE>

In these two tables we can see two SM transition matrices from the year 2004. The first is a ‘renewal’ one (like the SM transition matrices in the previous tables). At the left point of the time window (31/12/2003) we assume there is an occurrence of a rating action which leads to the rating displayed in the rows of the matrix. Then across the columns we display the probability that the debt issuer will be in any one of the possible 8 ratings. The second table is different. Here we assume that at the left-limit (31/12/2003) each debt issuer had already been one year in its current rating (displayed in the rows). That is, the debt issuer had received its rating on 31/12/2002. The most striking feature is the dramatic reduction in the realized default probability from CCC rated issuers from 39 to only 9 percent.

7 Conclusion

In this paper we proved the existence and uniqueness of the empirical non-homogeneous semi-Markov transition probabilities for any realistic setting. An efficient numerical method for consistently estimating these time-dependent transition matrices

from consistent estimates of the subdensity functions was introduced and its convergence established. A simple implementation of the method was performed based on the piecewise-constant nonparametric estimators of the hazard rate functions. We applied the resulting estimators to the analysis of a dataset on corporate credit ratings from Standard & Poor's. We found large differences between the estimates of the probabilities of default obtained under the semi-Markov assumption and their Markov counterparts.

The class of econometric models considered in this paper, for the analysis of corporate credit rating (and default) data at a micro-level, is one of the most flexible available. Additionally the continuous-time nature of the associated estimators makes efficient use of all the information available in the sample. Another advantage of this framework is that all major stylized facts of credit rating data (as reported in the literature) can be accommodated. One such feature not yet implemented in the empirical section of the current version of the paper is the phenomenon of (downward) rating drift, however this extension is straightforward. From our theoretical results it is also easy to derive confidence intervals for the transition probabilities, this is left for future research.

The authors would like to thank André Ran and seminar participants at CORE-UCL (Belgium), Tinbergen Institute Amsterdam and Vrije Universiteit Amsterdam (The Netherlands), University of Porto, ISCTE Business School and ISEG, Technical University of Lisbon (Portugal) for helpful comments. Financial support from Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) and Vrije Universiteit Amsterdam is gratefully acknowledged. The rating transition data for this research was generously supplied by Standard and Poor's. All calculations were performed using the Ox matrix programming language of Doornik (2002)

References

- Aalen, O.O. and S. Johansen (1978): "An Empirical Transition Matrix for Non-homogeneous Markov Chains based on Censored Observations," *Scandinavian Journal of Statistics* Vol 5, pp 141-50.
- Alvarez, E.E.E. (2005): "Smoothed Nonparametric Estimation in Window Censored semi-Markov Processes," *Journal of Statistical Planning and Inference* Vol 131, pp 209-229.
- Andersen, P.K., Borgan, Ø., Gill, R.D. and N. Keiding (1993): *Statistical Models Based on Counting Processes*, Springer-Verlag.
- Cox, D.R. (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society, B* Vol 34, pp 187-200.

- Cox, D.R. (1962): *Renewal Theory*, Methuen and Co. Ltd.
- De Dominicis, R. and J. Janssen (1984): "Finite non-homogeneous semi-Markov processes: Theoretical and computational aspects," *Insurance: Mathematics and Economics* Vol 3, pp 157-165.
- D'Amico G., Janssen, J. and R. Manca (2004): "Downward Credit Risk Problem and Non-homogeneous semi-Markov Reliability Transition Models," *Working Paper - University of Rome*
- Doornik, J.A. (2002): *Object-oriented matrix programming using Ox. 3rd ed.* Timberlake Consultants Press, London and Oxford: www.nuff.ox.ac.uk/Users/Doornik.
- Ekeland, I. and R. Temam, (1976): *Convex Analysis and Variational Problems*, North-Holland Publ. Comp., Amsterdam.
- Goodman, G.S. and S. Johansen (1973): "Kolmogorov's Differential Equations for Non-Stationary, Countable State Markov Processes with Uniformly Continuous Transition Probabilities," *Proceedings of the Cambridge Philosophical Society* 73, 119-138.
- Hastie, T.J. and R.J. Tibshirani (1990): *Generalized Additive Models*, Chapman and Hall, London.
- Iosifescu Manu, A. (1972): "Non-homogeneous Semi-Markov Processes," *Studii si Cercetuari Matematice* Vol 24, pp 529-533.
- Jafry Y. and T. Schuermann (2004): "Measurement, Estimation and Comparison of Credit Migration Matrices," *Journal of Banking and Finance*, Vol 28, pp 2603-2639.
- Kavvathas, D. (2000): "Estimating Credit Rating Transition Probabilities for Corporate Bonds," *Working Paper - University of Chicago*
- Koopman, S.J., Lucas, A. and A. Monteiro (2005): "The Multi-State Latent Factor Intensity Model for Credit Rating Transitions," *Tinbergen Institute Discussion Paper* nr. 05-071/4 downloadable from ssrn.com
- Lando, D. and T.M. Skødeberg (2002): "Analyzing rating transitions and rating drift with continuous observations," *Journal of Banking and Finance*, Vol 6, pp 423-444.
- Lévy, P. (1954): "Processus Semi-Markoviens," *Proceedings from the International Congress of Mathematics*, Vol. 3, pp. 416-426.
- Limnios, N. and B. Ouhbi (1999): "Nonparametric Estimation for Semi-Markov Processes Based on its Hazard Rate Functions," *Statistical Inference for Stochastic Processes*, Vol 2, pp 151-173.

- Limnios, N. and B. Ouhbi (2005): “Nonparametric Estimation for Semi-Markov Processes Based on K-Sample Paths with Application to Reliability,” *Applied Stochastic Models and Data Analysis - Conference Proceedings*, pp 1061-1068.
- Linton, O., Nielsen J.P. and S. vd Geer (2003): “Estimating Multiplicative and Additive Hazard Functions by Kernel Methods,” *Annals of Statistics*, Vol 31, pp 464-492
- Nielsen J.P. and O. Linton (1995): “Kernel Estimation in a Nonparametric Marker Dependent Hazard Model,” *Annals of Statistics*, Vol 23, pp 1735-1748
- Pyke, R. (1961a): “Markov Renewal Processes: Definitions and Preliminary Properties,” *The Annals of Mathematical Statistics*, Vol 32, No 4 pp. 1231-1242.
- Pyke, R. (1961b): “Markov Renewal Processes with Finitely Many States,” *The Annals of Mathematical Statistics*, Vol 32, No 4 pp. 1243-1259.
- Smith, W.L. (1955): “Regenerative Stochastic Processes,” *Proceedings of the Royal Society*, Ser.A, Vol 232, pp. 6-31.
- vd Vaart, A.W. (1998): *Asymptotic Statistics*, Cambridge University Press.
- Zabreyko, P.P., A.I. Koshelev, M.A. Krasnosel'skii, S.G. Mikhlin, L.S. Rakovshchik and V.Y. Stet'senko (1975): *Integral Equations - a reference text*, Noordhoff International Publishing.

8 Appendix

Here we use mainly the notation of section 4. Let $0 < T_2 \leq T_1$. Consider a space $Car([0, T_1] \times [0, T_2])$ of bounded functions $f : [0, T_1] \times [0, T_2] \rightarrow \mathbb{R}$ satisfying the following conditions:

1. $f(\cdot, \tau) : [0, T_1] \rightarrow \mathbb{R}$ is measurable for all $\tau \in [0, T_2]$,
2. $f(t, \cdot) : [0, T_2] \rightarrow \mathbb{R}$ is continuous for almost all $t \in [0, T_1]$,

In the theory of ordinary differential equations these functions are known as *Caratheodory functions*. For a brief revision of the main properties of the functions $f \in Car([0, T_1] \times [0, T_2])$ see Ekeland and Temam (1976), Chapter 8.

Theorem 3 *Let $f \in Car([0, T_1] \times [0, T_2])$ and let $\phi : [0, T_1] \rightarrow [0, T_2]$ be measurable. Then the function $t \rightarrow f(t, \phi(t))$ is measurable.*

Theorem 4 (Scorza-Dragoni) *A bounded function f belongs to the space $Car([0, T_1] \times [0, T_2])$ if and only if for any $\epsilon > 0$ there exists a compact set $K_\epsilon \subset [0, T_1]$ such that $\text{meas}([0, T_1] \setminus K_\epsilon) < \epsilon$ and the function $f : K_\epsilon \times [0, T_2] \rightarrow \mathbb{R}$ is continuous.*

8.1 Proof of Theorem 1

Theorem 1 is an immediate consequence of Lemma 7 below. To prove Lemma 7 we need some auxiliary results.

Lemma 5 *If $q \in L_\infty(\mathcal{T})$ and $p \in Car(\mathcal{T})$, then the function*

$$I(t, \tau) = \int_0^\tau p(t+u, \tau-u)q(t, u)du, \quad (t, \tau) \in \mathcal{T},$$

belongs to $Car(\mathcal{T})$.

Proof. Set

$$\bar{p}(t, \tau) = \begin{cases} p(t, \tau), & (t, \tau) \in \mathcal{T}, \\ 0, & (t, \tau) \in \mathbb{R}^2 \setminus \mathcal{T}, \end{cases} \quad \bar{q}(t, \tau) = \begin{cases} q(t, \tau), & (t, \tau) \in \mathcal{T}, \\ 0, & (t, \tau) \in \mathbb{R}^2 \setminus \mathcal{T}, \end{cases}$$

and

$$\bar{I}(t, \tau) = \int_0^\tau \bar{p}(t+u, \tau-u)\bar{q}(t, u)du, \quad (t, \tau) \in \mathcal{T}.$$

By Theorem 3 the integral exists. We have $I(t, \tau) = \bar{I}(t, \tau)$, whenever $(t, \tau) \in \mathcal{T}$. There exist $M > 0$ such that $|p|, |q| \leq M$, and sequences of continuous functions $p_n, q_n : \mathbb{R}^2 \rightarrow [-M, M]$ converging to \bar{p} and \bar{q} , respectively, in the space $L_1(\mathbb{R}^2)$. Put

$$I_n(t, \tau) = \int_0^\tau p_n(t+u, \tau-u)q_n(t, u)du, \quad (t, \tau) \in \mathcal{T}.$$

Let $\tau \in [0, T_2]$. Combining the Scorza-Dragoni and the Lusin theorems we see that the function $(w, t) \rightarrow \bar{p}(w, \tau-w+t)$ is measurable. We have

$$\begin{aligned} |I_n(t, \tau) - \bar{I}(t, \tau)| &\leq \int_0^\tau |\bar{p}(t+u, \tau-u) - p_n(t+u, \tau-u)| |\bar{q}(t, u)| du \\ &\quad + \int_0^\tau |\bar{q}(t, u) - q_n(t, u)| |p_n(t+u, \tau-u)| du \\ &\leq M \int_0^\tau (|\bar{p}(t+u, \tau-u) - p_n(t+u, \tau-u)| + |\bar{q}(t, u) - q_n(t, u)|) du \\ &\leq M \int_0^{T_1+\tau} |\bar{p}(w, \tau-w+t) - p_n(w, \tau-w+t)| dw + M \int_0^\tau |\bar{q}(t, u) - q_n(t, u)| du = J_n(t). \end{aligned}$$

From the Fubini theorem we obtain $J_n(\cdot) \in L_1([0, T_1])$ and

$$\int_0^{T_1} J_n(t) dt \leq M \int_{\mathbb{R}^2} (|\bar{p}(\alpha, \beta) - p_n(\alpha, \beta)| + |\bar{q}(\alpha, \beta) - q_n(\alpha, \beta)|) d\alpha d\beta.$$

This implies that the sequence $J_n(\cdot)$ tends to zero in the space $L_1([0, T_1])$. Therefore there exists a subsequence $J_{n_k}(\cdot)$ converging to zero almost everywhere in $[0, T_1]$. Since the functions $I_n(\cdot, \tau)$ are measurable, we see that the function $\bar{I}(\cdot, \tau)$ and, as a consequence, the function $I(\cdot, \tau)$ are measurable in $[0, T_1]$.

Put

$$\tilde{p}(t, \tau) = \begin{cases} p(t, \tau), & (t, \tau) \in \mathcal{T}, \\ p(t, T(t)), & (t, \tau) \in ([0, T_1] \times [0, T_2]) \setminus \mathcal{T}. \end{cases}$$

Obviously $\tilde{p} \in Car([0, T_1] \times [0, T_2])$. Let $\epsilon > 0$. By the Scorza-Dragoni theorem there exists a compact set $K_\epsilon \subset [0, T_1]$ such that $\text{meas}([0, T_1] \setminus K_\epsilon) < \epsilon/M^2$ and the function $\tilde{p} : K_\epsilon \times [0, T_2] \rightarrow \mathbb{R}$ is continuous. Let $t \in [0, T_1]$. Consider $\tau, \tau + \theta \in [0, T(t)]$. We have

$$\begin{aligned} & |I(t, \tau + \theta) - I(t, \tau)| \\ & \leq \int_0^\tau |p(t+u, \tau + \theta - u) - p(t+u, \tau - u)| |q(t, u)| du + |\theta| M^2 \\ & \leq \int_{[0, \tau] \cap (K_\epsilon - t)} |p(t+u, \tau + \theta - u) - p(t+u, \tau - u)| |q(t, u)| du + \epsilon + |\theta| M^2 \quad (39) \end{aligned}$$

There exists $\eta \in]0, \epsilon/M^2[$ such that $|p(\alpha', \beta') - p(\alpha, \beta)| < \epsilon/(T_2 M)$ whenever $(\alpha', \beta'), (\alpha, \beta) \in \mathcal{T} \cap (\mathbb{R} \times K_\epsilon)$. Thus the right-hand side of inequality (39) does not exceed 3ϵ whenever $\theta \in]-\eta, \eta[$. This ends the proof. \square

For each matrix $q = (q_{ij})$, with $q_{ij} \in L_\infty(\mathcal{T})$ consider a linear operator $\mathcal{A}(q)$ defined by

$$(\mathcal{A}(q)p)_{ij}(t, \tau) = \sum_{k=1}^s \int_0^\tau p_{kj}(t+u, \tau-u) q_{ik}(t, u) du, \quad i, j = 1, \dots, s.$$

By Lemma 5 $\mathcal{A}(q)$ transforms $Car_s(\mathcal{T})$ into itself.

Lemma 6 *Assume that $|q|_{L_\infty} \leq M$. Then for any positive integer n the inequality*

$$|\mathcal{A}^n(q)| \leq \frac{(sMT_2)^n}{n!}$$

holds.

Proof. Indeed, we have

$$|\mathcal{A}^n(q)p|_{Car_s} \leq (sM)^n |p|_{Car_s} \int_0^\tau \int_0^{\tau-u_{n-1}} \cdots \int_0^{\tau-u_1-\dots-u_{n-1}} du_0 \dots du_{n-1}. \quad (40)$$

By induction one can easily prove the equality

$$\int_0^\tau \int_0^{\tau-u_{n-1}} \cdots \int_0^{\tau-u_1-\dots-u_{n-1}} du_0 \dots du_{n-1} = \frac{\tau^n}{n!}.$$

Combining this with (40) we obtain the result. \square

Consider, for each matrix q as above, $b(q) \in Car_s(\mathcal{T})$ given by

$$(b(q))_{ij}(t, \tau) = \delta_{ij}(1 - H_i(t, \tau)), \quad (t, \tau) \in \mathcal{T}, \quad i, j = 1, \dots, s,$$

and the map $\mathcal{B}(q) : Car_s(\mathcal{T}) \rightarrow Car_s(\mathcal{T})$ defined as

$$\mathcal{B}(q)p = b(q) + \mathcal{A}(q)p. \quad (41)$$

Let \mathcal{P} denote the set

$$\{p \in Car_s(\mathcal{T}) : 0 \leq p_{ij}(t, \tau), \sum_{j=1}^s p_{ij}(t, \tau) = 1, p_{ij}(t, 0) = \delta_{ij}, t \in [0, T_1], i, j = 1, \dots, s\},$$

that is, the subset of $Car_s(\mathcal{T})$ defined by the transition matrices. The set \mathcal{P} is closed.

Lemma 7 *The equation $p = \mathcal{B}(q)p$ has a unique solution $\hat{p} \in \mathcal{P}$.*

Proof. Let $p \in \mathcal{P}$. Since

$$0 \leq (\mathcal{B}(q)p)_{ij}(t, \tau) \leq \delta_{ij}(1 - H_i(t, \tau)) + \sum_{k=1}^s Q_{ik}(t, \tau) = \delta_{ij}(1 - H_i(t, \tau)) + H_i(t, \tau) \leq 1$$

(see condition (C3)), and

$$\sum_{j=1}^s \left\{ \delta_{ij}(1 - H_i(t, \tau)) + \sum_{k=1}^s \int_0^\tau p_{kj}(t+u, \tau-u) q_{ik}(t, u) du \right\} =$$

$$1 - H_i(t, \tau) + \sum_{k=1}^s \int_0^\tau \left\{ \sum_{j=1}^s p_{kj}(t+u, \tau-u) \right\} q_{ik}(t, u) du = 1,$$

$\mathcal{B}(q)$ transforms \mathcal{P} into itself. By Lemma 6 there exists n such that $\mathcal{B}^n(q)$ is contractive. This implies the existence of a unique solution to the equation

$$p = \mathcal{B}(q)p.$$

Now set $(r(q))_{ij} = -\delta_{ij}H_i(t, \tau)(q)$. Consider a map $p = G(q)$ associating to q the solution to the equation $p = \mathcal{A}(q)p + b(q)$. Note that $|p| = |G(q)| \leq 1$. Let $|q| \leq M$. Put

$$p_h = \mathcal{A}(q+h)p_h + b(q+h)$$

and $\Delta p = p_h - p$. Then we have

$$\begin{aligned} \Delta p &= \mathcal{A}(q+h)p_h + b(q+h) - \mathcal{A}(q)p - b(q) \\ &= \mathcal{A}(q)\Delta p + \mathcal{A}(h)p_h + r(h) = \Gamma_{(q,p_h,h)}(\Delta p). \end{aligned}$$

Therefore

$$\Delta p = \Gamma_{(q,p_h,h)}^n(\Delta p) = \mathcal{A}^n(q)\Delta p + \sum_{k=0}^{n-1} \mathcal{A}^k(q)(\mathcal{A}(h)p_h + r(h)).$$

Since

$$|\mathcal{A}^n(q)| \leq (sMT_2)^n/n!,$$

we see that there exists the map $(I - \mathcal{A}^n(q))^{-1}$ whenever $(sMT_2)^n/n! < 1$. Fix n satisfying this condition. Then we obtain

$$\Delta p = (I - \mathcal{A}^n(q))^{-1} \sum_{k=0}^{n-1} \mathcal{A}^k(q)(\mathcal{A}(h)p + r(h) + \mathcal{A}(h)\Delta p). \quad (42)$$

Observe that $|\Delta p| \leq |p_h| + |p| \leq 2$. Since

$$|(I - \mathcal{A}^n(q))^{-1}| = \left| \sum_{j=1}^{\infty} (\mathcal{A}^n(q))^j \right| \leq \sum_{j=1}^{\infty} \left(\frac{(sMT_2)^n}{n!} \right)^j = (1 - (sMT_2)^n/n!)^{-1},$$

from (42) we get

$$|\Delta p| \leq \left(4sT_2(1 - (sMT_2)^n/n!)^{-1} \sum_{k=0}^{n-1} \frac{(sMT_2)^k}{k!} \right) |h|.$$

Now (42) can be written as

$$G(q+h) - G(q) = \Delta p = \Lambda h + \rho(h),$$

where

$$\Lambda h = (I - \mathcal{A}^n(q))^{-1} \sum_{k=0}^{n-1} \mathcal{A}^k(q)(\mathcal{A}(h)G(q) + r(h))$$

and

$$|\rho(h)| \leq (\text{const})|h|^2.$$

This implies the Fréchet differentiability of the G map and the end of the proof. \square

8.2 Proof of Theorem 2

Define functions

$$\tilde{q}_{ij}(t, \tau) = q_{ij}(t, \max\{0, l - 2\}\tilde{\Delta}),$$

with

$$(t, \tau) \in \mathcal{T}, \quad \tau \in [(l-1)\tilde{\Delta}, l\tilde{\Delta}[, \quad i, j = 1, \dots, s, \quad l = 1, \dots, N_2.$$

Consider the linear operator $\tilde{\mathcal{A}}_{\tilde{\Delta}} : \text{Car}_s(\mathcal{T}) \rightarrow \text{Car}_s(\mathcal{T})$ defined by

$$\left(\tilde{\mathcal{A}}_{\tilde{\Delta}}(P)\right)_{ij}(t, \tau) = \sum_{k=1}^s \int_0^\tau p_{kj}(t+u, \tau-u) \tilde{q}_{ik}(t, u) du, \quad i, j = 1, \dots, s.$$

Obviously we have

$$\begin{aligned} |\mathcal{A}(p) - \tilde{\mathcal{A}}_{\tilde{\Delta}}(p)|_{\text{Car}_s} &\leq \sum_{k=1}^s |p|_{\text{Car}_s} \sum_{\{l \geq 1 | l\tilde{\Delta} \leq T(t)\}} \int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} |q_{ij}(t, u) - \tilde{q}_{ij}(t, u)| du \\ &+ \int_{l\tilde{\Delta}}^{T(t)} |q_{ij}(t, u) - \tilde{q}_{ij}(t, u)| du \leq sT_2L\tilde{\Delta}|p|_{\text{Car}_s}. \end{aligned}$$

Consider also the linear operator $\mathcal{A}_{\tilde{\Delta}} : \text{Car}_s(\mathcal{T}) \rightarrow \text{Car}_s(\mathcal{T})$ defined by

$$(\mathcal{A}_{\tilde{\Delta}}(P))_{ij}(t, \tau) = \sum_{k=1}^s \sum_{l=2}^m \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}(t+u, \tau-u) du \right) q_{ik}(t, (l-1)\tilde{\Delta}),$$

$$(t, \tau) \in \mathcal{T}, \quad \tau \in [m\tilde{\Delta}, (m+1)\tilde{\Delta}[, \quad m = 1, \dots, N_2 - 1, \quad i, j = 1, \dots, s.$$

Since

$$\left(\tilde{\mathcal{A}}_{\tilde{\Delta}}(P)\right)_{ij}(t, \tau) = \sum_{k=1}^s \sum_{l=2}^m \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}(t+u, \tau-u) du \right) q_{ik}(t, (l-1)\tilde{\Delta})$$

$$+ \sum_{k=1}^s \left(\int_0^{\tilde{\Delta}} p_{kj}(t+u, \tau-u) du \right) \tilde{q}_{ik}(t, 0) + \sum_{k=1}^s \left(\int_{m\tilde{\Delta}}^{\tau} p_{kj}(t+u, \tau-u) du \right) q_{ik}(t, (m-1)\tilde{\Delta}),$$

$$(t, \tau) \in \mathcal{T}, \quad \tau \in [m\tilde{\Delta}, (m+1)\tilde{\Delta}[, \quad m = 1, \dots, N_2 - 1, \quad i, j = 1, \dots, s,$$

we get

$$|\mathcal{A}_{\tilde{\Delta}}(p) - \tilde{\mathcal{A}}_{\tilde{\Delta}}(p)|_{Car_s} \leq 2sM\tilde{\Delta}|p|_{Car_s}.$$

Therefore we have

$$|\mathcal{A}(p) - \mathcal{A}_{\tilde{\Delta}}(p)|_{Car_s} \leq s(T_2L + 2M)\tilde{\Delta}|p|_{Car_s}. \quad (43)$$

Lemma 8 *Let n be a positive integer such that $(sT_2M)^n/n! < 1$. If $\tilde{\Delta} = T_2/N_2$ satisfies*

$$\alpha_n(\tilde{\Delta}) < 1 - \frac{(sT_2M)^n}{n!}, \quad (44)$$

then the equation

$$p = \mathcal{B}_{\tilde{\Delta}}(p) = b + \mathcal{A}_{\tilde{\Delta}}(p) \quad (45)$$

has a unique solution $p^{\tilde{\Delta}} \in Car_s(\mathcal{T})$ satisfying

$$|p^{\tilde{\Delta}} - \hat{p}|_{Car_s} \leq \beta_n(\tilde{\Delta}). \quad (46)$$

Proof. Since $|\mathcal{A}|, |\mathcal{A}_{\tilde{\Delta}}| \leq sT_2M$, from (43) we obtain

$$|\mathcal{A}_{\tilde{\Delta}}^k - \mathcal{A}^k| \leq |\mathcal{A}_{\tilde{\Delta}} - \mathcal{A}| |\mathcal{A}_{\tilde{\Delta}}^{k-1} + \mathcal{A}_{\tilde{\Delta}}^{k-2}\mathcal{A} + \dots + \mathcal{A}^{k-1}| \leq \alpha_k(\tilde{\Delta}).$$

Therefore (see Lemma 6)

$$|\mathcal{A}_{\tilde{\Delta}}^n| \leq |\mathcal{A}_{\tilde{\Delta}}^n - \mathcal{A}^n| + |\mathcal{A}^n| \leq \alpha_n(\tilde{\Delta}) + \frac{(sT_2M)^n}{n!}.$$

From this we see that condition (44) implies the existence of a unique solution $p^{\tilde{\Delta}} \in Car_s(\mathcal{T})$ to equation (45). Observe that

$$\begin{aligned} |p^{\tilde{\Delta}} - \hat{p}|_{Car_s} &= |\mathcal{B}_{\tilde{\Delta}}^n(p^{\tilde{\Delta}}) - \mathcal{B}^n(\hat{p})|_{Car_s} \leq |\mathcal{A}_{\tilde{\Delta}}^n(p^{\tilde{\Delta}}) - \mathcal{A}^n(\hat{p})|_{Car_s} + \left| \sum_{k=1}^{n-1} (\mathcal{A}_{\tilde{\Delta}}^k - \mathcal{A}^k)b \right|_{Car_s} \\ &\leq |\mathcal{A}_{\tilde{\Delta}}^n(p^{\tilde{\Delta}}) - \mathcal{A}_{\tilde{\Delta}}^n(\hat{p})|_{Car_s} + |\mathcal{A}_{\tilde{\Delta}}^n(\hat{p}) - \mathcal{A}^n(\hat{p})|_{Car_s} + \left(\sum_{k=1}^{n-1} \alpha_k(\tilde{\Delta}) \right) |b|_{Car_s} \\ &\leq |\mathcal{A}_{\tilde{\Delta}}^n| |p^{\tilde{\Delta}} - \hat{p}|_{Car_s} + |\mathcal{A}_{\tilde{\Delta}}^n - \mathcal{A}^n| |\hat{p}|_{Car_s} + \left(\sum_{k=1}^{n-1} \alpha_k(\tilde{\Delta}) \right) |b|_{Car_s}. \end{aligned}$$

$$\leq \left(\alpha_n(\tilde{\Delta}) + \frac{(sT_2M)^n}{n!} \right) |p^{\tilde{\Delta}} - \hat{p}|_{Car_s} + \alpha_n(\tilde{\Delta})|\hat{p}|_{Car_s} + \left(\sum_{k=1}^{n-1} \alpha_k(\tilde{\Delta}) \right) |b|_{Car_s}.$$

Since $|\hat{p}|_{Car_s} \leq 1$ and $|b|_{Car_s} \leq 1$, we get (46). \square

Set $\mathcal{T}_m = \{(t, \tau) \in \mathcal{T} \mid \tau \leq m\tilde{\Delta}\}$. From (45) we see that the functions $p_{ij}^{\tilde{\Delta}} : \mathcal{T}_{m+1} \rightarrow [0, 1]$ satisfy

$$p_{ij}^{\tilde{\Delta}}(t, \tau) = \delta_{ij}(1 - H_i(t, \tau)) + \sum_{k=1}^s \sum_{l=2}^m \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t+u, \tau-u) du \right) q_{ik}(t, (l-1)\tilde{\Delta}), \quad (47)$$

$$(t, \tau) \in \mathcal{T}, \quad \tau \in [m\tilde{\Delta}, (m+1)\tilde{\Delta}[, \quad i, j = 1, \dots, s,$$

and are completely determined by their values on the set \mathcal{T}_m . This implies that (47) allows to successively calculate the functions $p_{ij}^{\tilde{\Delta}} : \mathcal{T} \rightarrow [0, 1]$, $i, j = 1, \dots, s$, on the sets $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \dots \subset \mathcal{T}_{N_2}$.

Lemma 9 *The solution $p_{ij}^{\tilde{\Delta}}$ to equation (47) is Lipschitz continuous in $[m_1\tilde{\Delta}, (m_1+1)\tilde{\Delta}[\times [m_2\tilde{\Delta}, (m_2+1)\tilde{\Delta}[\cap \mathcal{T}$, $m_1 = 0, \dots, N_1 - 1$, $m_2 = 0, \dots, N_2 - 1$, with the constant*

$$\tilde{L} = \max\{2sL\tilde{\Delta} + M, sLT_2 + sM + (sLT_2 + 16sN_2M)|p^{\tilde{\Delta}}|_{Car_s}\} \frac{(1 + 4sMT_2)^{N_2-1} - 1}{4sMT_2}.$$

Proof. Consider $(t', \tau'), (t'', \tau'') \in [m_1\tilde{\Delta}, (m_1+1)\tilde{\Delta}[\times [m_2\tilde{\Delta}, (m_2+1)\tilde{\Delta}[\cap \mathcal{T}$, where $m_2 = 0$ or $m_2 = 1$. Then we have

$$\begin{aligned} |p_{ij}^{\tilde{\Delta}}(t', \tau') - p_{ij}^{\tilde{\Delta}}(t'', \tau'')| &= \delta_{ij} |H_i(t', \tau') - H_i(t'', \tau'')| \\ &= \sum_{k=1}^s \left(\int_0^{\tau'} q_{ik}(t', u) du - \int_0^{\tau''} q_{ik}(t'', u) du \right) \\ &\leq \sum_{k=1}^s \int_0^{2\tilde{\Delta}} |q_{ik}(t', u) - q_{ik}(t'', u)| du + sM|\tau' - \tau''| \leq L_1(|t' - t''| + |\tau' - \tau''|), \end{aligned}$$

where $L_1 = 2sL\tilde{\Delta} + sM$.

Suppose that $p_{ij}^{\tilde{\Delta}}$ is Lipschitzian in $[m_1\tilde{\Delta}, (m_1+1)\tilde{\Delta}[\times [m_2\tilde{\Delta}, (m_2+1)\tilde{\Delta}[\cap \mathcal{T}$, $m_1 = 0, \dots, N_1 - 1$, $m_2 = 0, \dots, m - 1$, with the constant L_{m-1} . If $(t', \tau'), (t'', \tau'') \in [m_1\tilde{\Delta}, (m_1+1)\tilde{\Delta}[\times [m\tilde{\Delta}, (m+1)\tilde{\Delta}[\cap \mathcal{T}$, we have

$$p_{ij}^{\tilde{\Delta}}(t', \tau') - p_{ij}^{\tilde{\Delta}}(t'', \tau'') = \delta_{ij}(H_i(t', \tau') - H_i(t'', \tau''))$$

$$\begin{aligned}
& + \sum_{k=1}^s \sum_{l=2}^m \left(\left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) du \right) q_{ik}(t', (l-1)\tilde{\Delta}) \right. \\
& \quad \left. - \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u) du \right) q_{ik}(t'', (l-1)\tilde{\Delta}) \right) \\
& \quad = \delta_{ij}(H_i(t', \tau') - H_i(t'', \tau'')) \\
& + \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u) du \right) (q_{ik}(t', (l-1)\tilde{\Delta}) - q_{ik}(t'', (l-1)\tilde{\Delta})) \\
& + \sum_{k=1}^s \sum_{l=2}^m \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} (p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) - p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u)) du \right) q_{ik}(t', (l-1)\tilde{\Delta}).
\end{aligned}$$

From this we obtain

$$\begin{aligned}
& |p_{ij}^{\tilde{\Delta}}(t', \tau') - p_{ij}^{\tilde{\Delta}}(t'', \tau'')| \\
& \leq s((m+1)L\tilde{\Delta} + M)(|t' - t''| + |\tau' - \tau''|) + smL|p^{\tilde{\Delta}}|_{Car_s \tilde{\Delta}}|t' - t''| \\
& \quad + M \sum_{k=1}^s \sum_{l=2}^m \int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} |p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) - p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u)| du. \quad (48)
\end{aligned}$$

Set

$$\begin{aligned}
U(t', \tau', t'', \tau'') & = \{u \mid \exists l_1, l_2, w, v : t' + u + w = t'' + v, \tau' - u + w = \tau'' - v, \\
& (t' + u, \tau' - u), (t'' + v, \tau'' - v) \in [(l_1 - 1)\tilde{\Delta}, l_1\tilde{\Delta}] \times [(l_2 - 1)\tilde{\Delta}, l_2\tilde{\Delta}]\}.
\end{aligned}$$

If $|t' - t''| + |\tau' - \tau''| < \tilde{\Delta}/4$, then from elementary geometric considerations we get

$$\text{meas}(U(t', \tau', t'', \tau'')) \geq T_2 - 2(|t' - t''| + |\tau' - \tau''|)N_2.$$

Therefore for $(t', \tau'), (t'', \tau'') \in [m_1\tilde{\Delta}, (m_1 + 1)\tilde{\Delta}] \times [m\tilde{\Delta}, (m + 1)\tilde{\Delta}] \cap \mathcal{T}$ satisfying $|t' - t''| + |\tau' - \tau''| < \tilde{\Delta}/4$ we have

$$\begin{aligned}
& \sum_{k=1}^s \sum_{l=2}^m \int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} |p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) - p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u)| du \\
& \leq \sum_{k=1}^s \left(\int_{U(t', \tau', t'', \tau'')} |p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) - p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u)| du \right. \\
& \quad \left. + \sum_{l=2}^m \int_{[(l-1)\tilde{\Delta}, l\tilde{\Delta}] \setminus U(t', \tau', t'', \tau'')} |p_{kj}^{\tilde{\Delta}}(t' + u, \tau' - u) - p_{kj}^{\tilde{\Delta}}(t'' + u, \tau'' - u)| du \right)
\end{aligned}$$

$$\leq (sL_{m-1}T_2 + 4sN_2|p^{\tilde{\Delta}}|_{Car_s})(|t' - t''| + |\tau' - \tau''|).$$

Combining this with (48) we see that the function $p_{kj}^{\tilde{\Delta}}$ is Lipschitzian in $[m_1\tilde{\Delta}, (m_1+1)\tilde{\Delta}] \times [m\tilde{\Delta}, (m+1)\tilde{\Delta}] \cap \mathcal{T}$, with the constant

$$L_m = s((m+1)L\tilde{\Delta} + M) + smL|p^{\tilde{\Delta}}|_{Car_s}\tilde{\Delta} + 4(sL_{m-1}T_2 + 4sN_2|p^{\tilde{\Delta}}|_{Car_s}).$$

By induction we obtain

$$L_{N_2-1} \leq \tilde{L}$$

$$= \max\{2sL\tilde{\Delta} + M, sLT_2 + sM + (sLT_2 + 16sN_2M)|p^{\tilde{\Delta}}|_{Car_s}\} \frac{(1 + 4sMT_2)^{N_2-1} - 1}{4sMT_2}.$$

This ends the proof. \square

Lemma 10 *The sequence*

$$\gamma_0 = 0, \quad \gamma_{\nu+1} = \alpha \left(\sum_{\mu=0}^{\nu} \gamma_{\mu} + \nu\alpha \right), \quad \nu = 0, 1, \dots,$$

satisfies the inequality

$$\gamma_{\nu+1} \leq \alpha^2\nu(1 + \alpha)^\nu.$$

Proof. Set $S_\nu = \sum_{\mu=0}^{\nu} \gamma_{\mu}$ and $\sigma_\nu = S_\nu/\nu$. We have $S_{\nu+1} = (1 + \alpha)S_\nu + \nu\alpha^2$ and

$$\sigma_{\nu+1} \leq \frac{S_{\nu+1}}{\nu} = (1 + \alpha)\sigma_\nu + \alpha^2.$$

By induction we obtain

$$\sigma_\nu \leq \alpha^2(1 + (1 + \alpha) + \dots + (1 + \alpha)^{\nu-1}) = \alpha((1 + \alpha)^\nu - 1).$$

Therefore

$$\gamma_{\nu+1} = \alpha\nu(\sigma_\nu + \alpha) \leq \alpha^2\nu(1 + \alpha)^\nu. \quad \square$$

Let $t = (m_1 + \bar{l})\tilde{\Delta}$ and $\tau = (m_2 - \bar{l})\tilde{\Delta}$. Observe that (see (47))

$$\begin{aligned} p_{ij}^{\tilde{\Delta}}(t, \tau) &= \delta_{ij}(1 - H_i(t, \tau)) + \sum_{k=1}^s \sum_{l=2}^{m_2-\bar{l}} \left(\int_{(l-1)\tilde{\Delta}}^{l\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t+u, \tau-u) du \right) q_{ik}(t, (l-1)\tilde{\Delta}) \\ &= \delta_{ij}(1 - H_i(t, \tau)) \end{aligned}$$

$$+ \Delta \sum_{k=1}^s \sum_{l=2}^{m_2-\bar{l}} \sum_{r=0}^{2^N-1} p_{kj}^{\tilde{\Delta}}(t + ((l-1) + r/2^N)\tilde{\Delta}, \tau - ((l-1) + r/2^N)\tilde{\Delta}) q_{ik}(t, (l-1)\tilde{\Delta})$$

$$+ \sum_{k=1}^s \sum_{l=2}^{m_2 - \bar{l}} \sum_{r=0}^{2^N - 1} \zeta_{iklr}, \quad (49)$$

where

$$\begin{aligned} \zeta_{iklr} &= \left(\int_{(l-1+r/2^N)\tilde{\Delta}}^{(l-1+(r+1)/2^N)\tilde{\Delta}} p_{kj}^{\tilde{\Delta}}(t+u, \tau-u) du \right) q_{ik}(t, (l-1)\tilde{\Delta}) \\ &\quad - \Delta p_{kj}^{\tilde{\Delta}}(t + ((l-1) + r/2^N)\tilde{\Delta}, \tau - ((l-1) + r/2^N)\tilde{\Delta}) q_{ik}(t, (l-1)\tilde{\Delta}). \end{aligned}$$

Put

$$\begin{aligned} &\gamma_{ij}(m_1 + \bar{l} + \bar{r}/2^N, m_2 - \bar{l} - \bar{r}/2^N) \\ &= |p_{ij}^{\tilde{\Delta}}((m_1 + \bar{l} + \bar{r}/2^N)\tilde{\Delta}, (m_2 - \bar{l} - \bar{r}/2^N)\tilde{\Delta}) \\ &\quad - \hat{p}_{ij}^{\tilde{\Delta}}((m_1 + \bar{l} + \bar{r}/2^N)\tilde{\Delta}, (m_2 - \bar{l} - \bar{r}/2^N)\tilde{\Delta})| \end{aligned}$$

(see (38)). By Lemma 9 we have $|\zeta_{ijlr}| \leq \tilde{L}M\Delta^2$. Combining (38) and (49) we obtain

$$\begin{aligned} &\gamma_{ij}(m_1 + \bar{l} + \bar{r}/2^N, m_2 - \bar{l} - \bar{r}/2^N) \\ &\leq \Delta M \sum_{k=1}^s \sum_{l=2}^{m_2 - \bar{l}} \sum_{r=0}^{\bar{r}-1} (\gamma_{kj}(m_1 + \bar{l} + \bar{r}/2^N, m_2 - \bar{l} - \bar{r}/2^N) + \tilde{L}\Delta). \end{aligned}$$

Applying Lemma 10 with $\alpha = (s + \tilde{L})M\Delta$, we get

$$\begin{aligned} \gamma_{ij}(m_1, m_2) &\leq (s + \tilde{L})^2 M^2 \Delta^2 N_2 2^N (1 + (s + \tilde{L})\Delta)^{N_2 2^N} \\ &\leq (s + \tilde{L})^2 M^2 T_2 \Delta \exp((s + \tilde{L})MT_2). \end{aligned}$$

(Here we used the well-known inequality $(1 + 1/a)^a \leq \exp(a)$.) This and the inequality

$$|p^{\tilde{\Delta}}|_{Car_s} \leq |\hat{p}|_{Car_s} + |\hat{p} - p^{\tilde{\Delta}}|_{Car_s} \leq 1 + |\hat{p} - p^{\tilde{\Delta}}|_{Car_s}$$

together with Lemma 8 imply Theorem 2.

Table 1: Empirical SM transition matrix for 1992

This table contains the nonparametric estimates of the one-year semi-Markov transition matrix for the year 1992. This matrix gives the empirical probabilities (i.e. the observed frequencies) of a firm being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/1992 given that it received a specific credit rating (represented by the rows of the table) at 31/12/1991.

	AAA	AA	A	BBB	BB	B	CCC	Dflt
AAA	0.98417	0.01297	0.00285	0.00005	0.00001	0.00000	0.00000	0.00000
AA	0.00536	0.95795	0.03336	0.00139	0.00118	0.00070	0.00004	0.00003
A	0.00108	0.00401	0.95287	0.03726	0.00413	0.00049	0.00008	0.00008
BBB	0.00001	0.00121	0.02380	0.92771	0.04122	0.00288	0.00116	0.00237
BB	0.00000	0.00100	0.00290	0.02353	0.89034	0.06412	0.00854	0.00943
B	0.00000	0.00025	0.00127	0.00439	0.02231	0.81334	0.07894	0.07832
CCC	0.00000	0.00003	0.00014	0.00446	0.02238	0.08001	0.34470	0.53435
Dflt	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 2: Aalen-Johansen estimator for 1992

This table contains the nonparametric Aalen-Johansen estimates of the one-year Markov transition matrix for the year 1992. This matrix gives the empirical probabilities (i.e. the observed frequencies) of a firm being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/1992 given that it was in a specific credit rating (represented by the rows of the table) at 31/12/1991.

	AAA	AA	A	BBB	BB	B	CCC	Dflt
AAA	0.92403	0.07458	0.00137	0.00005	0.00000	0.00000	0.00000	0.00000
AA	0.00928	0.93742	0.05191	0.00133	0.00006	0.00000	0.00000	0.00000
A	0.00013	0.01540	0.93625	0.04572	0.00242	0.00009	0.00000	0.00000
BBB	0.00000	0.00061	0.06099	0.90047	0.03644	0.00140	0.00007	0.00003
BB	0.00000	0.00006	0.00584	0.11514	0.81013	0.05227	0.01102	0.00554
B	0.00000	0.00003	0.00476	0.02748	0.11240	0.73119	0.07736	0.04678
CCC	0.00000	0.00003	0.00041	0.00274	0.03017	0.09973	0.49593	0.37102
Dflt	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 3: Empirical SM transition matrix for 1993

This table contains the nonparametric estimates of the one-year semi-Markov transition matrix for the year 1993. This matrix gives the empirical probabilities (i.e. the observed frequencies) of a firm being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/1993 given that it received a specific credit rating (represented by the rows of the table) at 31/12/1992.

	AAA	AA	A	BBB	BB	B	CCC	Dflt
AAA	0.97629	0.02063	0.00300	0.00005	0.00002	0.00001	0.00000	0.00000
AA	0.00511	0.95637	0.03528	0.00134	0.00118	0.00072	0.00002	0.00002
A	0.00074	0.00370	0.95881	0.03336	0.00291	0.00041	0.00005	0.00006
BBB	0.00001	0.00120	0.02375	0.92131	0.04757	0.00312	0.00117	0.00203
BB	0.00000	0.00101	0.00339	0.02088	0.90965	0.05787	0.00354	0.00363
B	0.00000	0.00026	0.00136	0.00515	0.02291	0.90651	0.03219	0.03143
CCC	0.00000	0.00004	0.00020	0.00509	0.02302	0.18516	0.38943	0.39272
Dflt	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 4: Aalen-Johansen estimator for 1993

This table contains the nonparametric Aalen-Johansen estimate of the one-year Markov transition matrix for the year 1993. This matrix gives the empirical probabilities (i.e. the observed frequencies) of a firm being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/1993 given that it was in a specific credit rating (represented by the rows of the table) at 31/12/1992.

	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.95487	0.04361	0.00149	0.00003	0.00000	0.00000	0.00000	0.00000
AA	0.00011	0.94244	0.05639	0.00104	0.00002	0.00000	0.00000	0.00000
A	0.00333	0.00663	0.95886	0.03031	0.00084	0.00002	0.00000	0.00000
BBB	0.00009	0.00033	0.04699	0.89641	0.05163	0.00424	0.00024	0.00007
BB	0.00001	0.00620	0.00845	0.08211	0.82690	0.06865	0.00652	0.00116
B	0.00001	0.00044	0.00505	0.01420	0.12686	0.82013	0.01681	0.01651
CCC	0.00000	0.00004	0.00015	0.00179	0.03572	0.31275	0.49087	0.15866
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5: Empirical SM transition matrices for 2000

This table contains the nonparametric estimates of the one-year semi-Markov transition matrix for the year 2000. Here we have increased the resolution of the discrete grid used to apply the numerical integration algorithm discussed in section 4. We are thus able to see the impact of the number of points of the grid over the final estimates. This matrix gives the empirical probabilities (i.e. the observed frequencies) of a firm being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/2000 given that it received a specific credit rating (represented by the rows of the table) at 31/12/1999.

500 points								
	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.98926	0.00781	0.00286	0.00005	0.00001	0.00000	0.00000	0.00000
AA	0.00643	0.94603	0.04396	0.00189	0.00116	0.00073	0.00003	0.00003
A	0.00068	0.00532	0.95399	0.03659	0.00284	0.00048	0.00006	0.00011
BBB	0.00001	0.00092	0.01359	0.94701	0.02973	0.00440	0.00109	0.00347
BB	0.00000	0.00071	0.00150	0.00993	0.89104	0.08014	0.00721	0.00932
B	0.00000	0.00024	0.00062	0.00140	0.00630	0.86747	0.05977	0.06343
CCC	0.00000	0.00001	0.00004	0.00366	0.00230	0.04137	0.33102	0.60614
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

1000 points								
	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.98926	0.00781	0.00286	0.00005	0.00001	0.00000	0.00000	0.00000
AA	0.00643	0.94603	0.04401	0.00189	0.00116	0.00073	0.00003	0.00003
A	0.00068	0.00533	0.95399	0.03656	0.00284	0.00048	0.00006	0.00011
BBB	0.00001	0.00092	0.01358	0.94701	0.02973	0.00440	0.00108	0.00343
BB	0.00000	0.00071	0.00150	0.00993	0.89104	0.08018	0.00721	0.00932
B	0.00000	0.00024	0.00062	0.00140	0.00630	0.86747	0.05977	0.06337
CCC	0.00000	0.00001	0.00004	0.00366	0.00230	0.04139	0.33102	0.60554
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

2000 points								
	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.98926	0.00781	0.00286	0.00005	0.00001	0.00000	0.00000	0.00000
AA	0.00643	0.94603	0.04398	0.00189	0.00116	0.00073	0.00003	0.00003
A	0.00068	0.00533	0.95399	0.03657	0.00284	0.00048	0.00006	0.00011
BBB	0.00001	0.00092	0.01359	0.94701	0.02973	0.00440	0.00109	0.00343
BB	0.00000	0.00071	0.00150	0.00993	0.89104	0.08017	0.00721	0.00931
B	0.00000	0.00024	0.00062	0.00140	0.00630	0.86747	0.05977	0.06335
CCC	0.00000	0.00001	0.00004	0.00366	0.00230	0.04138	0.33102	0.60564
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 6: Empirical SM transition matrices for 2004

The first table below contains the nonparametric estimates of the one-year semi-Markov transition matrix for the year 2004. This matrix gives the empirical probabilities of a debt issuer being found in each one of the 8 rating classes (ordered through the columns of the table) at 31/12/2004 given that it received a specific credit rating (represented by the rows of the table) at 31/12/2003. The second table contains the equivalent ‘left-truncated’ semi-Markov transition matrix. Here the initial rating was received at 31/12/2002 and by 31/12/2003 the issuer was still in that same rating.

‘renewal’ transition matrix								
	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.97868	0.01821	0.00305	0.00007	0.00002	0.00001	0.00000	0.00000
AA	0.00722	0.93740	0.05120	0.00219	0.00129	0.00074	0.00003	0.00001
A	0.00067	0.00530	0.94608	0.04024	0.00715	0.00065	0.00008	0.00010
BBB	0.00006	0.00091	0.01020	0.94776	0.03308	0.00400	0.00130	0.00292
BB	0.00000	0.00071	0.00158	0.01076	0.89995	0.07767	0.00503	0.00413
B	0.00000	0.00025	0.00064	0.00170	0.00981	0.91680	0.04880	0.02203
CCC	0.00000	0.00002	0.00006	0.00448	0.00317	0.09810	0.48754	0.39976
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

‘left-truncated’ transition matrix								
	AAA	AA	A	BBB	BB	B	CCC	Dft
AAA	0.96332	0.03397	0.00325	0.00007	0.00003	0.00001	0.00000	0.00000
AA	0.00615	0.90949	0.07935	0.00273	0.00140	0.00069	0.00003	0.00002
A	0.00093	0.01766	0.91809	0.05869	0.00477	0.00057	0.00009	0.00014
BBB	0.00001	0.00090	0.01748	0.92622	0.04986	0.00424	0.00065	0.00085
BB	0.00000	0.00064	0.00328	0.03001	0.88430	0.07521	0.00396	0.00286
B	0.00000	0.00023	0.00050	0.00244	0.04147	0.89297	0.04299	0.01885
CCC	0.00000	0.00001	0.00005	0.00356	0.00370	0.06070	0.83702	0.09212
Dft	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

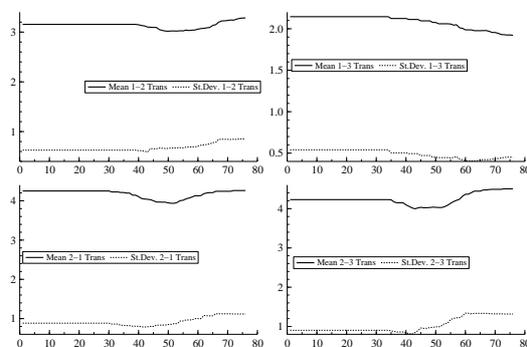


Figure 1: Empirical distributions of the NPML estimators of the baseline hazard 200 units - the SM model and the simulation set-up are explained in the text. Depicted on the top left picture is the estimated baseline hazard for a transition from state 1 to state 2 (the true baseline hazard was constant and equal to 3). The top right plot corresponds to transitions from state 1 to state 3 (true value was equal to 2). On the lower left picture we have the estimated baseline hazard for transitions from state 2 to state 1 (true value is 4). Finally the lower right graph depicts the estimated baseline hazard for transitions from state 2 to state 3 (underlying value is 4).

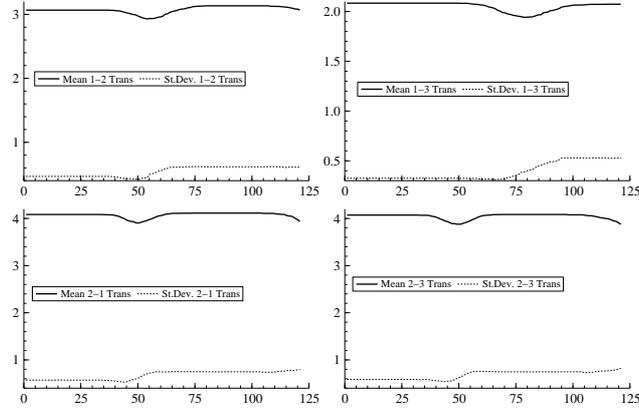


Figure 2: Empirical distributions of the NPML estimators of the baseline hazard (400 units - the baseline model and the simulation set-up are explained in the text. The structure of the plot is as in figure 1)

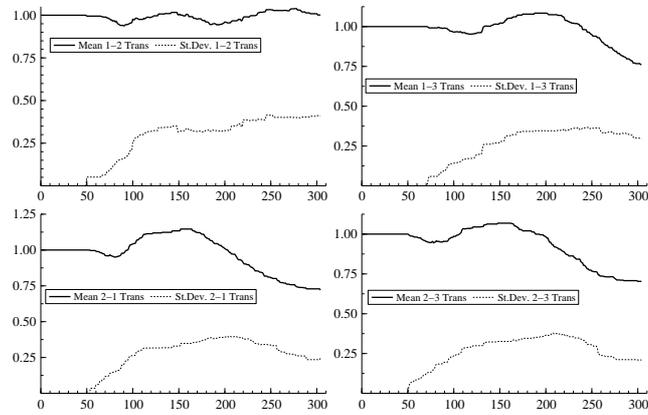


Figure 3: Empirical distributions of the NPML estimators of the time-index (200 units - the baseline model and the simulation set-up are explained in the text. The structure of the plot is as in figure 1 but the underlying value of the index is 1 throughout)

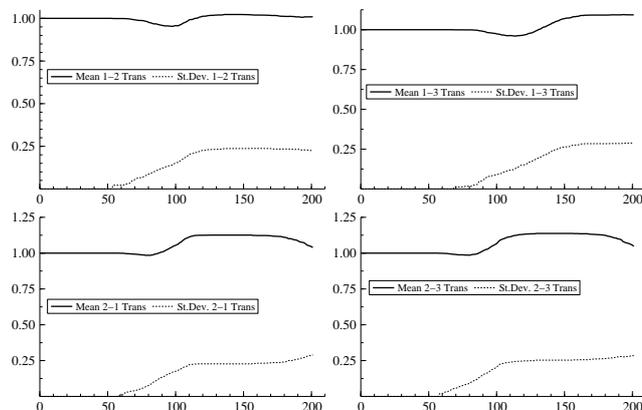


Figure 4: Empirical distributions of the NPML estimators of the time-index (400 units - the baseline model and the simulation set-up are explained in the text. The structure of the plot is as in figure 1 but the underlying value of the index is 1 throughout)

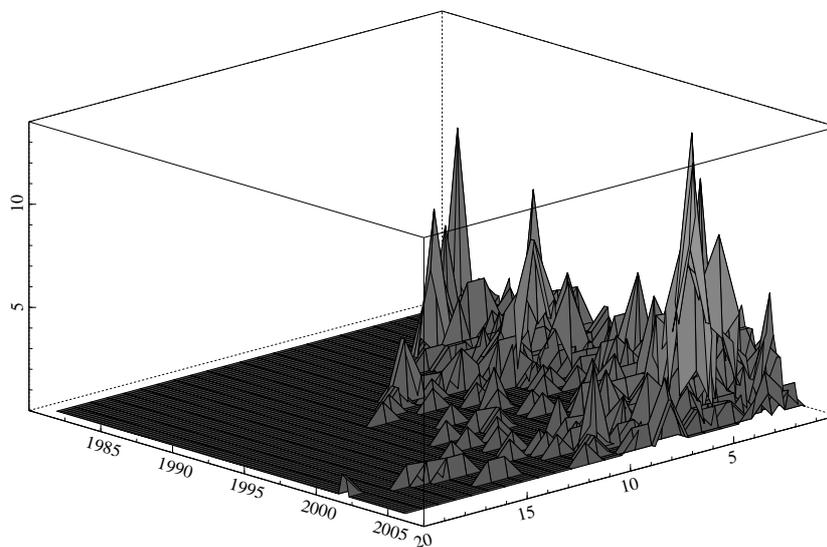


Figure 5: Recorded Number of Defaults: subinvestment grade rated issuers (each grid cell covers a quarter in calendar time and one year in age-time.)

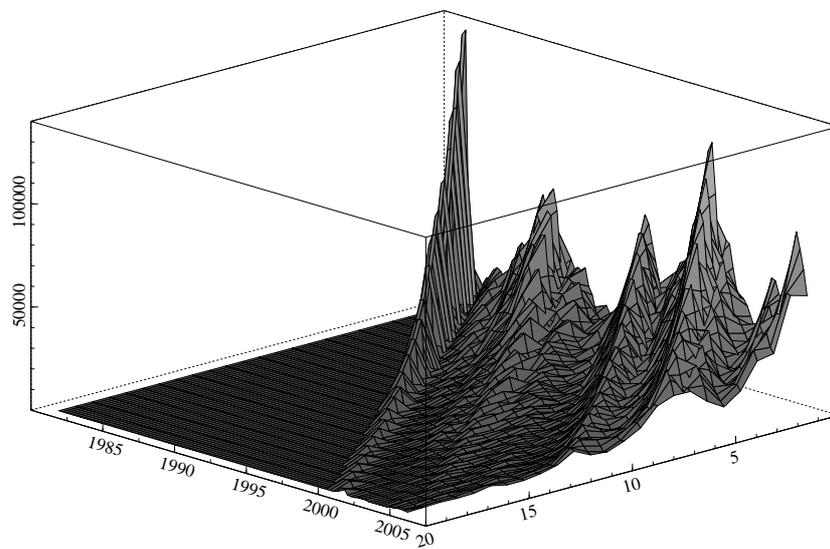


Figure 6: Total Exposure: subinvestment grade rated issuers
 (each grid cell covers a quarter in calendar time and one year in age-time)

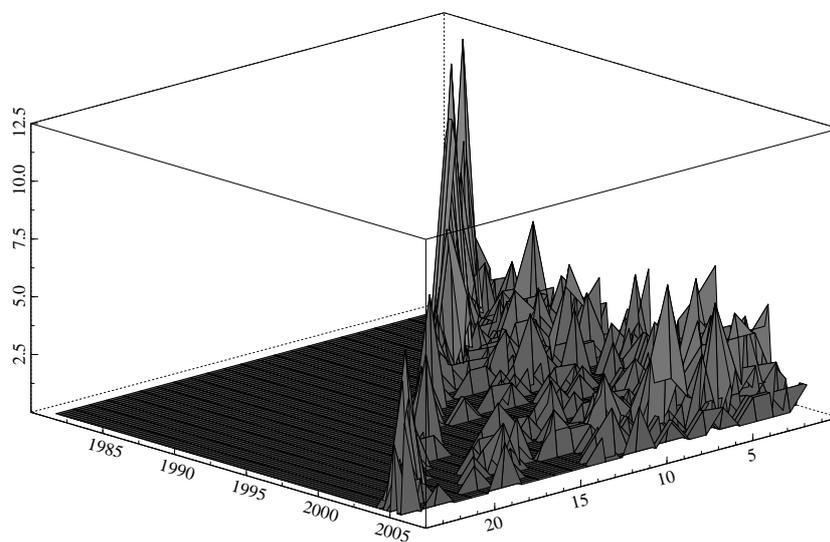


Figure 7: Downgrade from Investment to speculative grade: number of events
 (each grid cell covers a quarter in calendar time and one year in age-time)

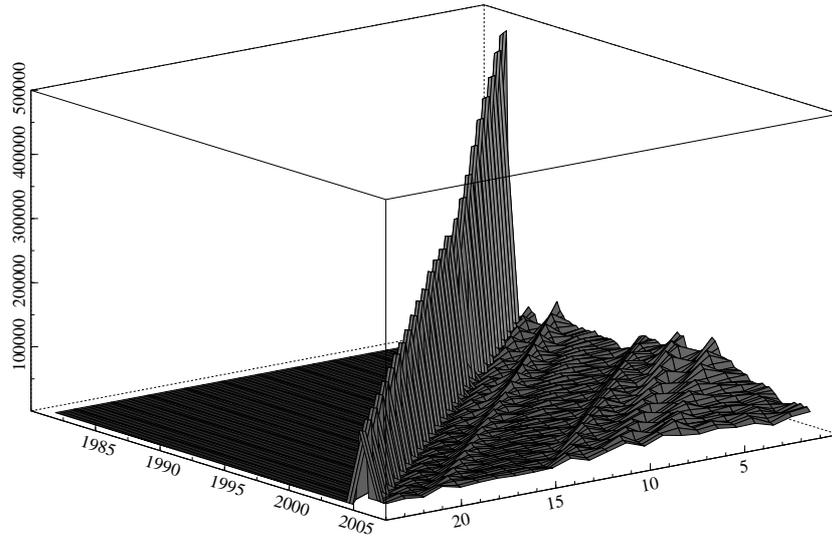


Figure 8: Total Exposures: investment grade rated issuers
(each grid cell covers a quarter in calendar time and one year in age-time)

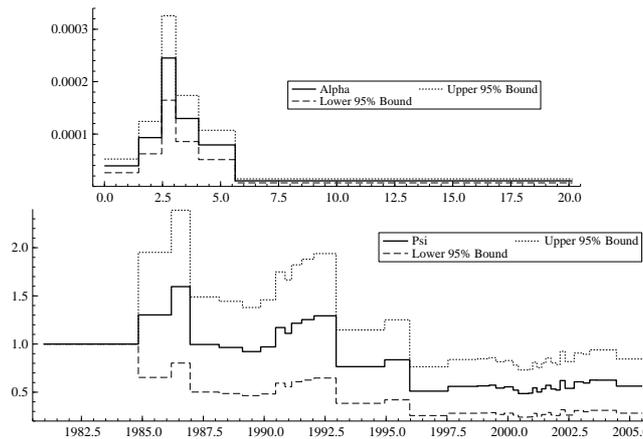


Figure 9: Recorded defaults by subinvestment grade rated issuers
(The topmost plot depicts the baseline hazard estimated with 6 ‘bins’, each cell or ‘bin’ encompasses 139 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that the strong non-monotonic behavior of the baseline hazard is statistically significant at the .95 level. On the lower pane we plot the estimated time index ψ . The first time period is taken as the basis for comparison, the time index was estimated using 36 ‘time bins’ each one of these ‘bins’ encompasses 23 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that, although the average amplitude of these confidence bands is rather large, the time-homogeneity is strongly rejected at this level, in particular in the period from early 96 to late 2001 the overall intensity of this type of event was significantly lower - when compared to the reference period - from early 1981 to the end of 1983.)

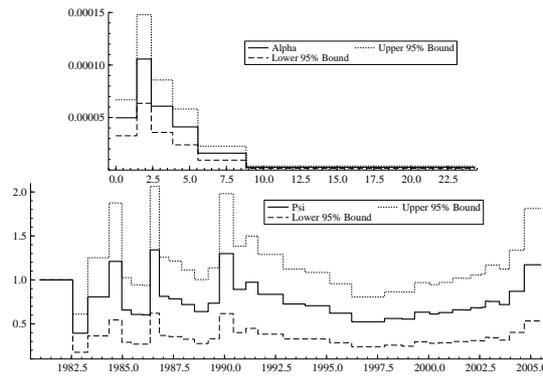


Figure 10: Downgrade from investment to subinvestment grade

(The topmost plot depicts the baseline hazard estimated with 6 ‘bins’, each cell or ‘bin’ encompasses 128 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that the strong non-monotonic behavior of the baseline hazard is statistically significant at the .95 level. On the lower pane we plot the estimated time index ψ . The first time period is taken as the basis for comparison, the time index was estimated using 36 ‘time bins’ each one of these ‘bins’ encompasses 21 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that, although the average amplitude of these confidence bands is rather large, the time-homogeneity is strongly rejected at this level, this takes place for several periods in 1983, 1986, 1989 and from 1996 to 1999 the overall intensity of this type of event was significantly lower - when compared to the reference period - from early 1981 to middle 1982.)

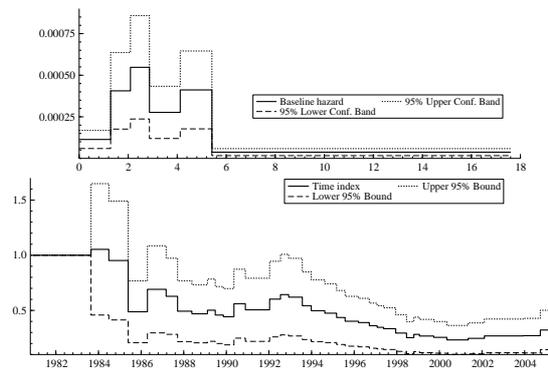


Figure 11: Upgrade from subinvestment to investment grade

(The topmost plot depicts the baseline hazard estimated with 6 ‘bins’, each cell or ‘bin’ encompasses 96 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that the strong non-monotonic behavior of the baseline hazard is statistically significant at the .95 level. On the lower pane we plot the estimated time index ψ . The first time period is taken as the basis for comparison, the time index was estimated using 36 ‘time bins’ each one of these ‘bins’ encompasses 16 observations. Also plotted are the upper and lower .95 confidence bands. It is possible to see that the time-homogeneity is strongly rejected at this level, this takes place for several periods like 1986, from 1988 to 1992 and from 1994 onwards the overall intensity of this type of event was significantly lower - when compared to the reference period - from early 1981 to late 1983.)