The multi-state latent factor intensity model for credit rating transitions

Siem Jan Koopman, André Lucas, André Monteiro

Author's Accepted Manuscript

The multi-state latent factor intensity model for credit rating transitions

Siem Jan Koopman, André Lucas, André Monteiro

The Multi-State Latent Factor Intensity Model

for Credit Rating Transitions

Siem Jan Koopman, André Lucas, André Monteiro

Corresponding author:

Andre Lucas

FEWEB/FIN

VU University Amsterdam

De Boelelaan 1105

NL-1081HV Amsterdam

THE NETHERLANDS

voice: +31 20 598 6039

fax: +31 20 598 6020

mail: alucas@feweb.vu.nl

The Multi-State Latent Factor Intensity Model
for Credit Rating Transitions

Siem Jan Koopman[a,c], André Lucas[b,c], André Monteiro[b,c]

[a] *Department of Econometrics, Vrije Universiteit Amsterdam,*
[b] *Department of Finance, Vrije Universiteit Amsterdam,*
[c] *Tinbergen Institute Amsterdam*

May 23, 2007

Abstract

A new empirical reduced-form model for credit rating transitions is introduced. It is a parametric intensity-based duration model with multiple states and driven by exogenous covariates and latent dynamic factors. The model has a generalized semi-Markov structure designed to accommodate many of the stylized facts of credit rating migrations. Parameter estimation is based on Monte Carlo maximum likelihood methods for which the details are discussed in this paper. A simulation experiment is carried out to show the effectiveness of the estimation procedure. An empirical application is presented for transitions in a 7 grade rating system. The model includes a common dynamic component that can be interpreted as the credit cycle. Asymmetric effects of this cycle across rating grades and additional semi-Markov dynamics are found to be statistically significant. Finally, we investigate whether the common factor model suffices to capture systematic risk in rating transition data by introducing multiple factors in the model.

Keywords: unobserved components, credit cycles, duration model, generator matrix, Monte Carlo likelihood.

JEL classification codes: C15, C33, C41, C43, G11, G21

1   Introduction

Ratings play a prominent role in the credit industry. Their key purpose is to provide a simple qualitative classification of the solidity, solvency and prospects of a debt issuer. The importance of credit ratings has increased significantly with the introduction of the new regulatory framework known as Basel II (BCBS, 2004). In this framework, ratings can be used directly to determine the size of a bank's capital buffer. As capital constitutes a relatively costly source of funding for a bank, ratings and rating changes directly affect the banks' willingness to grant credit to individual firms. Moreover, if ratings and thus capital requirements co-vary with the business cycle, economic fluctuations may be exacerbated by capital becoming increasingly scarce in adverse economic conditions, precisely when it is needed most. It is clear that a good understanding of the dynamic behavior of ratings and rating changes is therefore important from both a regulatory and financial industry perspective.

In this paper we introduce a new model for rating transitions. The main novelty of our model is that rating transitions are modeled continuously in event time rather than calendar time and are subject to common dynamic latent factors. Although the model is relatively complex, we show that it can be estimated efficiently using modern importance sampling techniques for non-Gaussian models in state space form.

The literature on modeling credit events such as defaults and rating changes has grown rapidly over the past 10 years. Wilson (1997a,b) modeled default rates using logistic regressions with macroeconomic explanatory variables. Nickell, Perraudin and Varotto (2000) and Bangia et al. (2002) show that upgrade, downgrade, and default probabilities differ over different economic regimes, whether characterized by NBER business cycle classifications or by GDP growth rates. Default and downgrade intensities are higher during recessions. In the same spirit, Kavvathas (2001), Carling, Jacobson, Lindé and Roszbach (2002), Couderc and Renault (2004), and Duffie, Saita, and Wang (2006a) use a duration approach conditional on observed macroeconomic and firm characteristics and show that average times-to-default decrease if economic activity decreases. Koopman and Lucas (2005) and Koopman, Lucas and Klaassen (2005) have adopted a direct time series approach and identified the time-varying cyclical nature of default rates over a long historical period. Also Fledelius, Lando and Nielsen (2004) corroborate the existence of time-fluctuations for credit rating migration rates.

Whereas some of the contributions in the literature introduce observed macro-variables to capture co-variation in default intensities between firms and industries, an alternative approach is to estimate the common components of default risk directly from the data. An advantage of such an approach is that one is less prone to misspecification caused by the use of an incorrect

macroeconomic proxy for the credit cycle. Couderc and Renault (2004) tested a large number of macroeconomic variables for their predictive ability and found five significant factors. Still, a large part of the fluctuations in systematic default probabilities could not be accounted for. Second, by estimating the default dynamics directly from the data, one obtains an integrated framework for capital determination and risk management, see Koopman, Lucas and Klaassen (2005). By contrast, if observed macroeconomic variables are used, one needs an auxiliary forecasting model for such variables, see for example Duffie et al. (2006a,b).

Suggestions for dynamic models with latent components are Gagliardini and Gourieroux (2004), McNeil and Wendin (2006), and Koopman, Lucas and Daniels (2005). These models, however, are all set in a calendar time framework: rating transitions are observed empirically over discrete time slots, e.g., years or quarters. The observed frequencies are subsequently modeled by non-Gaussian time series processes. By contrast, in this paper we use a duration model with unobserved components. The duration (continuous time) approach is the more natural approach in the current context, where durations to transitions are endogenous rather than exogenous. In this way, we are able to use all the information in the data-set Lando and Skødeberg (2002) provide a further detailed discussion of the advantages of the continuous-time approach. Our model can be regarded as a multi-state extension of the Latent Factor Intensity (LFI) model of Bauwens and Hautsch (2003). The LFI model is a point process model for stock transactions in tick-time. Durations in the LFI model are the time to the next trade. By contrast, in our model it is not only the time to the next rating event that is unknown, but also the type of event that is going to occur, e.g., upgrade, downgrade, or default. In that sense, our model is set in the so-called competing risks framework. Given a firm's initial rating, there are multiple states for the firm's next rating. Each of these states has its own duration process and we observe only the minimum of those. This leads to a more complicated likelihood structure than considered by Bauwens and Hautsch (2003).

The likelihood function of our model contains a high dimensional integral involving the latent common risk factor. In this way, our parameter driven model differs from well-known observation driven counterparts like the Autoregressive Conditional Duration model (ACD) of Engle and Russell (1998), or the Autoregressive Conditional Intensity model (ACI) of Russell (1999). We evaluate the likelihood using a multivariate extension of the Monte Carlo techniques that are developed by Durbin and Koopman (1997, 2001). We demonstrate the effectiveness of the method by means of a simulation experiment.

The model is estimated for the CreditPro7.0 data set from Standard & Poor's, containing all issuer ratings over the period 1981 – 2005. We classify firms into 7 standard rating categories and specify a dynamic model for upgrades, downgrades, and defaults using all available

data. This yields a data set including almost 7000 firms and almost 14000 informative rating events (more than 25000 if we also count sample extension and attrition). We obtain some interesting empirical findings. First, there is significant evidence of a persistent common component in rating transitions. We further show that the impact of this common component with respect to downgrade and upgrade probabilities is asymmetric. Upgrades are idiosyncratic to a large extent, whereas downgrades and defaults tend to cluster together in time. Further, experiments with multiple (latent) factor models cast doubt on the suitability of (widespread) single factor models for risk management and capital buffer determination. Finally, we find statistically significant evidence of semi-Markov effects in transitions and defaults. Investment grade transition intensities tend to increase over the first few years and become more stable thereafter. Sub-investment grade companies, by contrast, show increased transition activity at short durations, probably due to momentum effects, and at longer durations, possibly due to debt roll-over.

The paper is organized as follows. In Section 2, the model is presented. In Section 3 we develop the estimation methodology for this model. Section 4 discusses how to obtain default probabilities over finite time periods from the event time specification. Section 5 contains the results of a Monte Carlo study. Section 6 presents the results of our empirical study. We conclude in Section 7.

2   The Multi-State Latent Factor Intensity model

The multi-state latent factor intensity (MLFI) model is a multi-state generalization for multivariate point processes of the latent factor intensity (LFI) model of Bauwens and Hautsch (2003). Consider a set of K units (or firms) whose event-histories can be adequately described by the history of transitions between a finite set of states. The states in our empirical application will be the set of credit ratings for issuers as assigned by Standard and Poor's (S&P). The data set has a clear panel structure and consists of the exact dates and the corresponding type of the rating changes recorded for each firm in the sample. In order to account for unobserved dependence between the transition histories in a parsimonious way, we introduce a common factor $\psi(t)$. We assume that conditional on $\psi(t)$, rating events are independent across firms (i.e., along the cross section dimension). This assumption is standard in the credit risk literature and is used to prevent the model's corresponding joint state-space becoming quickly unmanageable due to its size. Gagliardini and Gourieroux (2004) provide a short discussion of this curse of dimensionality problem.

The multi-state feature of the model is represented as a set $\mathbb{S}$ of transition types, $\mathbb{S} =$

$\{1, 2, \ldots, S\}$. For example, in the case of three rating classes (AAA,AA,A), $s = 1$ denotes a downgrade from AAA to AA, $s = 2$ from AAA to A, $s = 3$ an upgrade from AA to AAA, ..., up to $s = S = 6$ an upgrade from A to AA. Next, define the right-continuous counting processes $N_k(t)$ and $N(t)$. The process $N(t)$ makes a jump of unit size at each time there is a rating event for one of the K units.[1] Similarly, $N_k(t)$ jumps at the times there is a credit event for unit k such that

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

These point processes are marked because at each event time we also observe the transition type of the unit, i.e., the specific type of upgrade or downgrade. In fact, the counting process $N_k(t)$ can be expressed as the sum of S counting processes $N_{sk}(t)$ that keep track of the total number of transitions of type s for firm k. It follows that

$$N_k(t) = \sum_{s=1}^{S} N_{sk}(t), \qquad N(t) = \sum_{k=1}^{K} N_k(t) = \sum_{s=1}^{S} \sum_{k=1}^{K} N_{sk}(t).$$

Corresponding to each of these point processes $N_{sk}(t)$ we assume there is a finite stochastic intensity $\lambda_{sk}(t)$. In practical terms this intensity describes the instantaneous probability of unit k experiencing a type s rating transition at time t conditional on the information available just before time t. Naturally, such transition intensities are only defined at time t if the unit actually is 'at risk' for transition type s at $t^- < t$, where $t - t^-$ is arbitrarily small. For example, the downgrade intensity from AAA to AA for firm k at time t is only defined if firm k actually has an AAA rating just prior to t. The intensity for each point process[2] $\lambda_{sk}(t)$ can be (informally) defined by

$$\lambda_{sk}(t) = \lim_{\Delta \downarrow 0} \frac{P\left[N_{sk}((t + \Delta)^-) - N_{sk}(t^-) > 0 \mid \mathcal{F}_{t^-}\right]}{\Delta},$$

see for example Andersen et al. (1993, p. 51). The conditional information up to (but not including) time t is represented by $\mathcal{F}_{t^-} = \cup_{\tau < t} \mathcal{F}_\tau$ for an appropriate filtration $\mathcal{F}_\tau$.

Define $R_{sk}(t)$ as a dummy variable that takes the value one if unit k is 'at risk' for transition type $s \in \mathbb{S}$ at time t^-, and zero otherwise. Note that unit k can be at risk for multiple transition types at the same time. For example, both the AAA to AA and the AAA to A transitions may be at risk simultaneously. Obvious reasons for a transition type not to be at risk for firm k at

---

[1]We assume there are no simultaneous rating transitions. In practice the S&P's database is recorded at a daily frequency. This means multiple rating actions can be observed on a single day (for distinct firms). Our likelihood specification in Section 3 incorporates this phenomenon.

[2]We assume $N_{sk}(t)$ to be a conditionally orderly process, i.e., it satisfies $P\left[N_{sk}((t + \Delta)^-) - N_{sk}(t^-) > 1 \mid \mathcal{F}_{t^-}\right] = o(\Delta) P\left[N_{sk}((t + \Delta)^-) - N_{sk}(t^-) = 1 \mid \mathcal{F}_{t^-}\right]$, such that we can discard the probability of a jump larger than 1 in $N_{sk}(t)$.

time t are that unit k has the incorrect current initial rating, has defaulted, or dropped out of the sample earlier for other reasons.

We adopt a proportional hazards specification. The model specification for the conditional hazards is given by

$$\lambda_{sk}(t) = \exp\left[\eta_s + \gamma_s' w_k(t) + \alpha_s \psi(t)\right] \cdot H_{sk}(t), \tag{1}$$

with $s = 1, \ldots, S$ and $k = 1, \ldots, K$, where (i) scalar η_s, $m \times 1$ vector γ_s, and scalar α_s are fixed unknown coefficients, (ii) $m \times 1$ vector $w_k(t)$ contains explanatory variables (covariates), (iii) scalar $\psi(t)$ represents a latent dynamic factor, and (iv) scalar function $H_{sk}(t)$ represents the generalized baseline hazard function, which can be used to model duration dependence of the multivariate type. This specification encompasses for example the homogeneous continuous-time Markov chain model that is frequently used in the empirical credit risk literature, see, for example, Kavvathas (2001) and Lando and Skødeberg (2002). Note we assume $\lambda_{sk}(t)$ to be only defined if the corresponding $R_{sk}(t) = 1$.

A more detailed discussion of the intensity specification (1) follows below. The parameter η_s represents the reference-level log-intensity of transition type s. It is independent of time and common across all units $k = 1, \ldots, K$. The parameter vector γ_s and scalar α_s measure the sensitivity of unit k's log-intensity for transition type s with respect to observed covariates $w_k(t)$ and the unobserved process $\psi(t)$, respectively. The m-dimensional vector of covariates $w_k(t)$ can contain unit-specific information such as leverage and profitability ratios, industry dummies, stock volatilities or statistics depending on the rating process.[3] Further, $w_k(t)$ can include macroeconomic information such as economic growth rates, interest rate levels and term structure variables. In this case subscript k can be dropped from the notation. Note that phenomena like rating momentum can also be included in $w_k(t)$ such that past downgrades and upgrades make subsequent downgrades and upgrades more likely, respectively.

The coefficients α_s depend on the transition type $s \in \mathbb{S}$. This implies that α_s can depend on both the origin and the destination state. In the empirical literature it is common practice to let α_s parameters depend on the origin state, i.e., the initial rating, only. Here, however, the impact of the common risk factor $\psi(t)$ depends on the type of transition, and therefore on the destination state as well. For example, upgrades might be less subject to common risk factors than downgrades, see Gagliardini and Gourieroux (2005). Restrictions on α_s can be tested explicitly using the maximum likelihood based procedure of Section 3.

Following the empirical work in the credit risk literature, we assume all intensities are subject

---

[3]The possible endogenous nature of a selection of (time-varying) covariates leads to an inference procedure that can no longer be interpreted as full (conditional) maximum likelihood. Instead, we then have a partial likelihood inference framework, see Lancaster (1990).

to the same unobserved dynamic common factor $\psi(t)$. Relaxing this assumption by making $\psi(t)$, for example, rating or industry specific is conceptually straightforward in our modeling framework. The latent process might even be unit specific as in Bauwens and Hautsch (2003). In the case of rating transition data, however, specifying unit-specific processes is not really feasible. The number of rating events for an individual firm is usually too small, even over a prolonged period of time. This is a direct consequence of the rating agencies' policy to provide stable ratings to the investment community.

Assume that $\psi(t)$ only changes at observed event times t_i for $i = 1, \ldots, N(T)$ where T denotes the time index of the last observation (right-censoring of type I). The specification of $\psi(t)$ as a stochastic process with piecewise constant (left-continuous) sample paths is intuitive since the intensity of the pooled process (pooled over firms and transition types) is not identified between two consecutive events. Moreover, in the context of credit rating transitions, $\psi(t)$ is intended to capture low-frequency co-movements in the vector of migration intensities. In the empirical illustration of Section 6, the average duration of the pooled process is 1.2 days. Therefore, no serious bias will arise from disregarding possible changes in the macroeconomic variables over the almost bi-daily spells of the pooled process.

Let $\psi_i = \psi(t_i)$ denote the value of the common risk factor $\psi(t)$ over the interval $t \in (t_{i-1}, t_i]$. In order to capture serial correlation in the intensity of the pooled process, the dynamic process for ψ_i can be specified, for example, by a first order autoregressive (AR) equation

$$\psi_i = \rho_s^{t_i - t_{i-1}} \psi_{i-1} + \sigma_i \varepsilon_i, \tag{2}$$

where ε_i is a set of i.i.d. $N(0,1)$ innovations, the AR parameter $\rho \in [-1, 1]$, and where σ_i may depend on the length of the spell $(t_i - t_{i-1})$. The specification in (2) has small changes in $\psi(t)$ over short spells. This is in line with our interpretation of $\psi(t)$ as an economy wide risk factor, which one would not expect to vary wildly at high frequencies. As not all α_s parameters from (1) and σ_i can be identified simultaneously, we normalize σ_i to

$$\sigma_i^2 = \begin{cases} (1 - \rho^{2(t_i - t_{i-1})/D})/(1 - \rho^{2/D}) & \text{for } -1 < \rho < 1, \\ t_i - t_{i-1} & \text{for } \rho = 1, \end{cases} \tag{3}$$

in our empirical work, with t_i measured in days of a $D = 260$ day business year. This brings $\psi(t)$ close to an Ornstein-Uhlenbeck (for $|\rho| < 1$) or Brownian Motion (for $\rho = 1$) process observed at the event days t_i. More general dynamic specifications for ψ_i can be easily incorporated in the state space framework of the next subsection. For example, different AR processes can be considered for specific rating transitions. See also the empirical application in Section 6.

The baseline hazard $H_{sk}(t)$ is specified by the deterministic function

$$H_{sk}(t) = H_s(t - t_{0k} \,,\ t - t_{1k} \,,\ \ldots \,,\ t - t_{N_k(t),k}), \tag{4}$$

8

where $t - t_{ik}$ denotes the backward-recurrence time of unit k with respect to its past ith transition moment. The function $H_{sk}(\cdot)$ can be any non-negative function of its arguments. The inclusion of $H_{sk}(t)$ introduces duration dependence into the model and, therefore, relaxes the Markov assumption. More precisely, if $H_{sk}(t)$ is allowed to depend only on $t - t_{N_k(t),k}$, then each unit follows a semi-Markov process. In the general case a generalized semi-Markov process is obtained, see Glynn (1988). Possible choices for $H_{sk}(t)$ include the hazard function of a multivariate Weibull distribution, given by

$$H_s(x_0, \ldots, x_N) = \sum_{i=0}^{N} a_{si} x_i^{b_s - 1}, \tag{5}$$

with $x_i \geq 0$ and fixed coefficients $a_{si} > 0$ and $b_s > 0$ for $i = 0, 1, \ldots, N$. Another valid alternative is the self-excitation mechanism introduced by Hawkes (1971) and also considered for the LFI model by Bauwens and Hautsch (2003).

We note that k's observed duration or spell $t_{N_k(t),k} - t_{N_k(t)-1,k}$ is the minimum of $\sum_s R_{sk}(t)$ latent durations corresponding to the set of feasible transitions 'at risk' for unit k at time t. We adopt the standard practice of assuming that the latent duration processes are mutually independent conditional on the common factor $\psi(t)$.[4] See van den Berg (2001) for a detailed discussion on identification problems in this setting.

To complete the model specification, an additional set of identifying assumptions for the parameters is required. The global identification of intensity specification (1) requires a sign restriction for α_s. Changing the sign simultaneously for all α_s's and for the complete path of $\psi(t)$ clearly yields the same path for intensity $\lambda_{sk}(t)$. We therefore set $\alpha_s < 0$ for $s = S$.

For a vector of unknown parameters θ, the likelihood function conditional on the initial ratings, pre-sample event histories,[5] and on the complete path of the unobserved process, as defined by $\Psi_{N(T)} = \{\psi_i\}_{i=0}^{N(T)}$, can be written as

$$L\left(\theta \mid \mathcal{F}_T, \Psi_{N(T)}\right) = \prod_{i=1}^{N(T)} \prod_{k=1}^{K} \prod_{s=1}^{S} \exp\left(Y_{sk}(t_i) \ln\{\lambda_{sk}(t_i)\} - R_{sk}(t_i) \int_{t_{i-1}}^{t_i} \lambda_{sk}(t) \mathrm{d}t\right), \tag{6}$$

where dummy variable $Y_{sk}(t)$ is one if unit k at time t experiences a rating event of type s, and zero otherwise, and $\mathcal{F}_T$ denotes the relevant observable filtration. The likelihood function (6) has an intuitive interpretation. Unit k only contributes to the (conditional) likelihood if it is at risk, that is if $R_{sk}(t_i) = 1$. In this case, the likelihood contains the probability of survival of unit k in its current state over each spell of the pooled point process if there was no rating

---

[4]If no exogenous covariates are included, as in the empirical illustration of Section 6, this is an innocuous assumption, see Tsiatis (1975, Theorem 2).

[5]A discussion of the initial conditions problem in event-history models is provided by van den Berg (2001).

event for this unit at risk. When rating event i takes place at the end of the spell of the pooled process for firm k, that is if $Y_{sk}(t_i) = 1$, the survival probability is multiplied by the hazard rate to yield the probability density of the rating event.

The likelihood in (6) can be decomposed in a likelihood of a spell length in the pooled point process, and a likelihood of the associated mark. The spell length follows from the cumulated (over all s and k) intensities at risk, $R_{sk}(t)\lambda_{sk}(t)$. The mark then follows from the multinomial distribution with probabilities $R_{sk}(t)\lambda_{sk}(t)/\sum_{s,k} R_{sk}(t)\lambda_{sk}(t)$ for firm k experiencing rating event type s. See also Section 5. This is the most general specification.[6] More restrictive specifications can of course be accommodated as well. For example, one might take the ordered nature of ratings into account by restricting the α_s parameters to depend on the initial/input rating only, and not on the output rating. In this paper, however, we stick to the more general specification.

In order to estimate the parameter vector θ, the conditional likelihood function must be integrated with respect to the complete path $\Psi_{N(T)}$ of the unobserved process $\psi(t)$. The maximum likelihood problem becomes

$$\max_{\theta} L(\theta \mid \mathcal{F}_T), \tag{7}$$

where

$$L(\theta \mid \mathcal{F}_T) = \int L\left(\theta \mid \mathcal{F}_T, \Psi_{N(T)}\right) p(\Psi_{N(T)}) \mathrm{d}\Psi_{N(T)}, \tag{8}$$

and $p(\Psi_{N(T)})$ denotes the density function of $\Psi_{N(T)}$.

3 Monte Carlo Maximum Likelihood Estimation

The main difficulty with maximum likelihood estimation in (8) is the computation of the high-dimensional integral. In a typical application such as the one in the next section, this integral is much more than 4000 dimensional. McNeil and Wendin (2006) address a similar problem by adopting a Bayesian perspective, albeit in a lower dimensional space (around 50). Bauwens and Hautsch (2003) adopt the simulated maximum likelihood method of Liesenfeld and Richard (2003) for the estimation of a single-state LFI model.

---

[6]Given the close resemblence of the process for the marks with qualitative response models, well-known problems for those classes of models like (in)dependence of irrelevant alternatives (IIR), will be an issue here as well, at least for the general multinomial specification. Though not the issue of this paper, the current methodology is also applicable in more general classes of point process models that are less subject to such problems. Moreover, in our specific application to credit risk modeling within a 7 grade ratings system, IIR is not a key issue.

10

By contrast, in this paper, we adopt the general method of Monte Carlo maximum likelihood for a multi-state LFI (MLFI) model. To overcome the inefficiency problem of direct Monte Carlo estimation of the high-dimensional integral in (8) we use a combination of importance sampling and the Kalman filter and smoother as described in Durbin and Koopman (Part II, 2001). It is shown that the methodology can be made applicable for high-dimensional problems. In this section, the model is formulated in state space form in Subsection 3.1. The Monte Carlo simulation method for likelihood evaluation is discussed in Subsection 3.2.

3.1 Statistical model specification

The MLFI model considers the following three sources of stochastic variation: (i) the duration between events in the pooled process, denoted by $\tau_i = t_i - t_{i-1}$; (ii) the transition types s being at risk at t_i^- for unit k, denoted by $R_{sk}(t_i)$; (iii) the specific transition type s at time t_i for unit k, denoted by $Y_{sk}(t_i)$. These stochastic variables are collected in the vector z_i for $i = 1, \ldots, N$ with $N = N(T)$, where z_i is defined as

$$z_i = \{\tau_i \ , \ R_{11}(t_i) \ , \ \ldots \ , \ R_{SK}(t_i) \ , \ Y_{11}(t_i) \ , \ \ldots \ , \ Y_{SK}(t_i)\}'.$$

The vector z_i can be constructed (or observed) at each event $i = 1, \ldots, N$. The analogue of the observation equation for z_i is implied by the non-Gaussian conditional likelihood in (6). In particular, for the ith event time of the pooled process, we have the conditional log-density

$$\ln p(z_i | \psi_i, \mathcal{F}_{t_i^-}) = \sum_{s=1}^{S} \sum_{k=1}^{K} Y_{sk}(t_i) \ln\{\lambda_{sk}(t_i)\} - R_{sk}(t_i) \int_{t_{i-1}}^{t_i} \lambda_{sk}(t)\mathrm{d}t, \tag{9}$$

for $i = 1, \ldots, N$.

The intensity specification (1) can be formulated more generally via vector ν_i that contains latent processes and fixed effects. We have

$$\lambda_{sk}(t) = \exp\left(Z_{ski}\nu_i\right) \cdot H_{sk}(t), \qquad \text{for } t_{i-1} < t \leq t_i, \tag{10}$$

where Z_{ski} is a fixed and known 'selection' vector, for $s = 1, \ldots, S$, $k = 1, \ldots, K$, and $i = 1, \ldots, N$. In case $H_{sk}(t) = 1$, intensity $\lambda_{sk}(t)$ is constant for $t_{i-1} < t \leq t_i$. To show that specifications (1) and (10) can be equivalent, we take

$$\nu_i \ = \ \{\eta_1 \ , \ \ldots \ , \eta_S \ , \ \gamma_1' \ , \ \ldots \ , \ \gamma_S' \ , \ \psi(t_i)\}',$$
$$Z_{ski} \ = \ \{e_s' \ , \ e_s' \otimes w_k(t_i)' \ , \ \alpha_s\},$$

where e_s is the s-th column of I_S. It follows that $Z_{ski}\nu_i = \eta_s + \gamma_s' w_k(t_i) + \alpha_s \psi_i$. If another specification for $\lambda_{sk}(t)$ is considered, the specifications for Z_{ski} and ν_i need to be adjusted accordingly.

The vector ν_i can contain both fixed unknown coefficients and dynamic latent processes. We therefore model ν_i by the general Markovian process

$$\nu_i = F_i \nu_{i-1} + \tilde{R}_i \eta_i, \qquad \eta_i \sim \text{NIID}(0, Q_i), \qquad i = 1, \ldots, N, \tag{11}$$

with initial condition $\nu_0 \sim \text{N}(a, P)$. The vector a and the matrices F_i, $\tilde{R}_i$, Q_i and P are fixed matrices that may depend on the parameter vector θ. If the vector ν_i only consists of fixed unknown coefficients, we set $a = 0$, $F_i = \tilde{R}_i = I$, $Q_i = 0$ and $P = \kappa I$, where κ is the so-called diffuse prior constant. Usually, κ is set to some large value in numerical software, see Harvey (1989, pp. 367-8). Exact solutions for $\kappa \to \infty$ are available as well, see Durbin and Koopman (2001, Ch. 4). If the vector ν_i only contains the latent autoregressive process (2), that is $\nu_i = \psi_i$, we set $a = 0$, $F_i = \rho^{t_i - t_{i-1}}$, $\tilde{R}_i = 1$, $Q_i = \sigma_i^2$ and $P = (1 - \rho^2)^{-1}$. A combination of unknown coefficients and latent time series processes can be incorporated in (11) in a straightforward way. For example, in the case of (1) with $w_k(t) = 0$, we have $\nu_i = (\eta_1, \ldots, \eta_S, \psi_i)'$ with $a = 0$,

$$F_i = \begin{bmatrix} I_S & 0 \\ 0 & \rho^{t_i - t_{i-1}} \end{bmatrix}, \quad \tilde{R}_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q_i = \sigma_i^2, \quad P = \begin{bmatrix} \kappa I_S & 0 \\ 0 & (1 - \rho^2)^{-1} \end{bmatrix}.$$

In case multiple latent factors or higher order dynamics of the latent factors are part of the model, the state vector ν_i can be extended in a natural way. The state space matrices need to be adjusted accordingly. A general framework for the MLFI model can be summarized by the observation log-density for z_i conditional on the state ν_i. This is given by (9) where $\lambda_{sk}(t)$ is given by (10) for $t_{i-1} < t \leq t_i$, and where ν_i is modeled by (11) with $i = 1, \ldots, N$. This set of equations makes up a nonlinear non-Gaussian state space model as considered by Shephard and Pitt (1997) and Durbin and Koopman (1997, 2001). Importance sampling methods enable the evaluation of the likelihood function but also the estimation of ν_i and the computation of the corresponding standard errors for $i = 1, \ldots, N$. The details of these methods are given in Subsection 3.2.

We notice the high dimension of the state vector ν_i since it includes the scalars $\eta_1, \ldots, \eta_S$, where S can be as large as $49 = 7^2$ in a 7 ratings system. However, our state space analysis can accommodate this aspect of the model in a feasible way since it relies on computationally efficient methods such as the Kalman filter and associated algorithms. Furthermore, it leads to a significant reduction of the size of the parameter vector θ. Since θ needs to be estimated via the numerical optimization of the likelihood, computation time is also reduced as a result. The remaining parameters in θ are ρ and α_ss. These coefficients can be placed in the state vector ν_i as well although in this case the model becomes nonlinear in the state equation. The linear Markovian process (11) for the state vector ν_i is not applicable and we cannot rely

on the computationally efficient Kalman filter methods as we do below. The treatment of nonlinear state processes is more involved and computationally more demanding. The details of the estimation procedures for ν_i and θ in the current framework are presented in the next subsection.

3.2 Monte Carlo likelihood evaluation

Given the statistical model specification of the previous subsection, the likelihood function (8) can be reformulated by

$$L(\theta \mid \mathcal{F}_T) = \int \left\{ \prod_{i=1}^{N} p\left(z_i \mid \nu_i, \mathcal{F}_{i-1}\right) \right\} p(\nu \mid \mathcal{F}_T) \mathrm{d}\nu, \tag{12}$$

where $p\left(z_i \mid \nu_i, \mathcal{F}_{i-1}\right)$ is given by (9) and the model for $\nu = (\nu_1', \ldots, \nu_N')'$ is implied by (11). Both $p\left(z_i \mid \nu_i, \mathcal{F}_{i-1}\right)$ and $p(\nu \mid \mathcal{F}_T)$ depend on the parameter vector θ for $i = 1, \ldots, N$. An analytical expression for (12) does not exist and therefore we rely on numerical techniques for the evaluation of (12). For this purpose we explore the technique of Monte Carlo integration using the method of importance sampling. The basic idea is simple. First, we simulate M paths of ν from $p(\nu \mid \mathcal{F}_T)$ denoted by $\nu^1, \ldots, \nu^M$ where M is a large number. Second, we compute the Monte Carlo estimator of (12) given by

$$\widehat{L}(\theta \mid \mathcal{F}_T) = M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \mid \nu_i^m, \mathcal{F}_{i-1}\right) \right\} \tag{13}$$

where ν_i^m is the ith element from ν^m. The estimator (13) is poor since ν^m is simulated 'unconditionally' and is therefore likely to make little contribution to the likelihood. A more efficient approach is to simulate from $p(\nu \mid z, \mathcal{F}_T)$, but this is not feasible since no analytical expression exists for this density. The idea of importance sampling is to replace $p(\nu \mid z, \mathcal{F}_T)$ by the more convenient Gaussian density $p_G(\nu \mid z, \mathcal{F}_T)$ for simulating ν's. The basic algorithm is then adjusted as follows. First, simulate M paths of ν from $p_G(\nu \mid z, \mathcal{F}_T)$ denoted by $\nu^1, \ldots, \nu^M$ where M is a large number. Second, compute the Monte Carlo estimator of (12) as given by

$$\begin{aligned} \widehat{L}(\theta \mid \mathcal{F}_T) &= M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \mid \nu_i^m, \mathcal{F}_{i-1}\right) \right\} \frac{p(\nu^m \mid \mathcal{F}_T)}{p_G(\nu^m \mid z, \mathcal{F}_T)} \\ &= p_G(z \mid \mathcal{F}_T) M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \mid \nu_i^m, \mathcal{F}_{i-1}\right) \right\} \frac{1}{p_G(z \mid \nu^m, \mathcal{F}_T)}, \end{aligned} \tag{14}$$

since $p_G(\nu \mid \mathcal{F}_T) = p(\nu \mid \mathcal{F}_T)$ and $p_G(\nu \mid z, \mathcal{F}_T) = p_G(z \mid \nu, \mathcal{F}_T) \, p_G(\nu \mid \mathcal{F}_T)/p_G(z \mid \mathcal{F}_T)$. We refer to this estimator as the Monte Carlo likelihood. The construction of $p_G(\nu \mid z, \mathcal{F}_T)$ and the evaluation of the different densities is described in detail below.

Step 1: Simulate paths of ν from $p_G(\nu|z, \mathcal{F}_T)$

To build a device for simulating from the conditional Gaussian density $p_G(\nu|z, \mathcal{F}_T)$, an approximating linear Gaussian model needs to be formulated that represents the joint density $p_G(\nu, z|\mathcal{F}_T)$. This density for the linear Gaussian model ideally resembles the true density $p(\nu, z|\mathcal{F}_T)$ as close as possible because samples generated from the conditional Gaussian density $p_G(\nu|z, \mathcal{F}_T)$ may then be similar to samples from the conditional density $p(\nu|z, \mathcal{F}_T)$. An appropriate linear Gaussian model can be obtained using the method described in Durbin and Koopman (2001, Part II) and is based on the linearization of the observational log-density using a second-order Taylor expansion.

In the context of the model described in Subsection 3.1, the basic idea is to construct a linear Gaussian state space model for the series of rating event indicators at event i as given by

$$\{Y_{11i}, \ldots, Y_{S1i}, Y_{12i}, \ldots, Y_{SKi}\},$$

where $Y_{ski} \equiv Y_{sk}(t_i)$ and $Y_{sk}(t_i)$ is one or zero to indicate whether a rating event of type s has taken place for unit k at time t_i. Such a rating event is triggered by the signal $Z_{ski}\nu_i = \eta_s + \gamma_s' w_k(t_i) + \alpha_s \psi_i$ which determines the intensity $\lambda_{sk}(t)$ for $t_{i-1} < t \leq t_i$, see Subsection 3.1. To establish an approximating Gaussian model that relates the signal $Z_{ski}\nu_i$ to Y_{ski}, we consider the linear Gaussian observation equation

$$Y_{ski} = c_{ski} + Z_{ski}\nu_i + \xi_{ski}, \qquad \xi_{ski} \sim \text{NIID}\left(0, C_{ski}\right), \tag{15}$$

for $s = 1, \ldots, S$, $k = 1, \ldots, K$ and $i = 1, \ldots, N$, where scalar constant c_{ski} and scalar variance C_{ski} are considered as auxiliary and unknown variables that need to be constructed in a consistent fashion as is shown below. The observation Y_{ski} is linear in vector ν_i and modeled by the linear Gaussian process (11). Therefore, observation equation (15) and the dynamic latent process (11) make up a standard linear Gaussian state space model, see Durbin and Koopman (2001, part I) for a detailed discussion on this class of models.

The constant c_{ski} and variance C_{ski} of the observation equation (15) are constructed in such a way that the conditional density of the model of interest $p(z|\nu, \mathcal{F}_T)$ and the conditional density of the approximating model $p_G(Y|\nu, \mathcal{F}_T)$ have the same mode for ν, where $Y = (Y_{111}, \ldots, Y_{SKN})'$. The joint solution for c_{ski} and C_{ski} to obtain the mode denoted by $\bar{\nu}$ can be obtained recursively, see the treatment in Durbin and Koopman (2001, Chapter 11).

The implementation of this procedure is relatively simple. An initial guess for the mode $\bar{\nu}$ needs to be found that is denoted by $\hat{\nu}^{(0)}$. The linear Gaussian model (15) is constructed for

$j = 0$ by

$$c_{ski} = Y_{ski} - Z_{ski}\widehat{\nu}_i^{(j)} - C_{ski}Z_{ski}\nabla \ln p(z|\nu, \mathcal{F}_T)_i,$$

$$\tag{16}$$

$$C_{ski} = -\left[Z_{ski}\nabla^2 \ln p(z|\nu, \mathcal{F}_T)_i Z'_{ski}\right]^{-1},$$

where

$$\nabla \ln p(z|\nu, \mathcal{F}_T)_i = \left.\frac{\partial \ln p(z|\nu, \mathcal{F}_T)}{\partial \nu_i}\right|_{\nu=\widehat{\nu}^{(j)}},$$

$$\nabla^2 \ln p(z|\nu, \mathcal{F}_T)_i = \left.\frac{\partial^2 \ln p(z|\nu, \mathcal{F}_T)}{\partial \nu_i \partial \nu'_i}\right|_{\nu=\widehat{\nu}^{(j)}}.$$

A new guess of the mode for ν is obtained by estimating the conditional mean of ν conditional on Y for the approximating linear Gaussian state space model (15) and (11). The conditional mean of ν can be computed by the Kalman filter and smoothing (KFS) algorithm. More formally, the KFS method computes $E_G(\nu|Y)$ where $E_G(\cdot)$ is with respect to the approximating linear Gaussian model. It is well-known that the mode and the mean are equivalent in a Gaussian model. The new estimate of ν is denoted by $\widehat{\nu}^{(j+1)}$. New guesses for the mode are obtained by the KFS based on (16) for $j = 1, 2, \ldots$ until convergence is reached according to some metric. Usually convergence takes place after 5 to 10 iterations.

The approximating linear Gaussian model consists of (11) and (15), with joint density $p_G(\nu, z|\mathcal{F}_T)$ and where (16) is evaluated at $\nu = \widehat{\nu}$ with $\widehat{\nu}$ as the estimated mode. We adopt this model to generate conditional samples for ν from $p_G(\nu|z, \mathcal{F}_T)$. Direct sampling from such a high-dimensional Gaussian density requires many high-dimensional matrix operations. These numerical problems can be overcome because the model is formulated as a linear Gaussian state space model. Therefore, the simulation smoothing algorithms of de Jong and Shephard (1995) or Durbin and Koopman (2002) can be used to generate conditional samples for ν, denoted as ν^m for $m = 1, \ldots, M$.

Step 2: Compute the Monte Carlo likelihood (14)

Given a set of simulated samples from $p_G(\nu|z, \mathcal{F}_T) \equiv p_G(\nu|Y, \mathcal{F}_T)$ and denoted by ν^m, the computation of the Monte Carlo likelihood (14) is relatively simple. The Gaussian density $p_G(z|\nu, \mathcal{F}_T) \equiv p_G(Y|\nu, \mathcal{F}_T)$ is conditional on ν and its expression is well-known for the linear model (15). Further, the observation density of interest $p(z_i|\nu_i, \mathcal{F}_{i-1})$ is given by equation (9) and can also be computed straightforwardly.

The Monte Carlo likelihood is then maximized with respect to θ for a particular choice of M. The maximization can be carried out by a numerical optimization procedure. For example, a quasi-Newton method can be used for this purpose. To ensure a likelihood surface that is continuous (or smooth) in θ, the same random numbers are used for the sampling in Step 1

of the M signals from $p_G(\nu|z,\mathcal{F}_T)$. The optimization procedure requires an initial estimate of θ that is chosen ideally in the neighborhood of its final estimate. However, even when the initialisation is poor, several modifications in numerical optimization procedures exist that make global convergence attainable.

Step 3: Smoothed estimates of the state vector

The state vector ν_i contains fixed unknown coefficients and dynamic latent processes. Estimating the state vector for each i leads to estimates of regression parameters and latent processes such as ψ_i. A straightforward estimate of the state vector, given the data, is obtained by weighting each simulated state vector ν_i^m by its contribution to the likelihood function, that is

$$\hat{\nu}_{i|N} = \left(\sum_{m=1}^{M} w_m \times \nu_i^m\right) \bigg/ \left(\sum_{m=1}^{M} w_m\right), \tag{17}$$

where

$$w_m = \left\{\prod_{i=1}^{N} p\left(z_i \,|\nu_i^m, \mathcal{F}_{i-1}\right)\right\} \bigg/ p_G(z|\nu^m, \mathcal{F}_T). \tag{18}$$

Standard errors for $\hat{\nu}_{i|N}$ are obtained by taking the square root of

$$\left[\left\{\sum_{m=1}^{M} w_m \times (\nu_i^m)^2\right\} \bigg/ \left(\sum_{m=1}^{M} w_m\right)\right] - (\hat{\nu}_{i|N})^2. \tag{19}$$

4 Implied Transition Matrices

We now turn to the issue of estimating the transition matrix given the Monte Carlo maximum likelihood estimates of the parameters. Typical examples include 1-year transition matrices as the ones published by Standard & Poor's and Moody's. We start by recalling the connection for unit k between the *infinitesimal generator matrix* $G_k(t)$ and the implied matrix P_k of transition probabilities for a continuous-time finite-state Markov process. The case of semi-Markov processes is dealt with later on.[7] The matrix $G_k(t)$ contains the hazard rates for each origin and destination state combination. In particular, the (i,j)th element of $G_k(t)$ equals $\lambda_{sk}(t)$ for s corresponding to a transition from origin state i to destination state j. The diagonal elements of $G_k(t)$ are such that the rows of $G_k(t)$ sum to zero. Consider an interval $[T, T+\Delta]$. Then the matrix of transition probabilities over the interval $[T, T+\Delta]$ is given by

---

[7]For a Markov chain, the entries of the generator matrix are either constants or (deterministic) functions of time. However, for generalized semi-Markov processes the entries of the generator matrix are, in general, stochastic processes.

the product integral[8]

$$P_k (T, T + \Delta) = \prod_{T}^{T+\Delta} (I_S + G_k(t)\mathrm{d}t). \tag{20}$$

For the MLFI model, a parametric form for $G_k(t)$ conditional on observed regressors and an unobserved factor is assumed. In Aalen and Johansen (1978), by contrast, $G_k(t)$ is left completely unspecified under the assumption that duration and self-excitation effects are absent. We therefore use the Aalen Johansen estimator in our empirical section as a benchmark for evaluating model adequacy. We assume that the elements of $G_k(t)$ are adapted to the observable filtration $\mathcal{F}_t$. In this situation $P_k(T, T+\Delta)$ becomes a random variable, and we want to compute its expectation conditional on $\mathcal{F}_T$. This expectation can be interpreted as the transition matrix over the interval $[T, T + \Delta]$,

$$\bar{P}_k (T, T + \Delta) = \mathrm{E}\left[P_k (T, T + \Delta) \mid \mathcal{F}_T\right] = \mathrm{E}\left[\left.\prod_{T}^{T+\Delta} (I + G_k(t)\mathrm{d}t)\right| \mathcal{F}_T\right]. \tag{21}$$

To estimate the conditional expectation in (21), we suggest two approximating schemes. In the first scheme, we generate a large number of high frequency (say daily) paths from the process $\psi(t)$, compare (2) and (3) in Section 2. Using these paths, a consistent estimator for $\bar{P}_k(T, T + \Delta)$ is given by

$$\hat{\bar{P}}_k (T, T + \Delta) = \frac{1}{M} \sum_{m=1}^{M} \prod_{T}^{T+\Delta} (I + G_k^m(t)\mathrm{d}t), \tag{22}$$

where $G_k^m(t)$ denotes unit k's realized matrix of intensities for replication m. This approximation skips the fact that in our empirical model specification, $\psi(t)$ only jumps at event times. This last assumption, however, is not material in the model's specification, and jumps at higher frequencies (such as every day) are easily allowed for. The second scheme to obtain estimates of annual transition probabilities uses the bootstrap. Here, we build on the empirical model's assumption that the common factor $\psi(t)$ only jumps at event times. We start with the estimates of the unknown model parameters and the smoothed estimates of the latent process, $\mathrm{E}[\psi_i|\mathcal{F}_T]$ for $i = 1, \ldots, N(T)$. Next, we simulate a large number M of possible future sample paths over the $[T, T + \Delta]$ interval for the *full panel* of K firms as well as for the unobserved risk factor ψ_i.[9]

---

[8]See Gill (2001) for an exposition on product integration. The product integral is the continuous counterpart of the standard, discrete product operator, just as the integral is the continuous counterpart of the summation operator. Informally, the product integral of a function $f(t)$ over the interval $[T, T + \Delta]$ is $\prod_{T}^{T+\Delta}(1 + \mathrm{d}f(t)) = \lim_{n\to\infty} \prod_{i=1}^{n}(1 + f(t_i) - f(t_{i-1}))$ for a partition $T = t_0 < t_1 < \ldots < t_n = T + \Delta$.

[9]If weakly exogenous covariates were included in equation (1), then an auxiliary model is needed to forecast the future path of these covariates (as mentioned in the introduction). One resulting possibility is the estimation of scenario forecasts.

The pooled process over $[T, T + \Delta]$ for replication m provides a partition $T = t_0^m < t_1^m < \ldots < t_n^m = T + \Delta$, over which the product integral can be factored, see Andersen et al. (1993, p. 91). For the empirical model in Section 6 these calculations become particularly manageable.

The estimates of $\bar{P}_k[T, T + \Delta]$ can be used to compute several interesting risk measures. For example, one can compute the average transition probabilities over a specific time interval for a portfolio of firms,

$$\bar{P}[T, T + \Delta] = K^{-1} \sum_{k=1}^{K} \bar{P}_k[T, T + \Delta].$$

One can also compute (non-linear) functions of the default probabilities in $\bar{P}_k[T, T + \Delta]$ to obtain direct estimates of capital requirements according to the official Basel II regulations. This is especially interesting if one does not average over simulations of $\psi(t)$, but considers quantiles instead. See the forecasting exercise in Section 6. This also allows one to consider stress scenarios in terms of extreme $\psi(t)$ realizations.

So far, we have discussed how to obtain estimates of transition matrices for Markov processes. If we move on to semi-Markov processes, the equations become more involved. In the current case of semi-Markov behavior and a common risk factor $\psi(t)$, the relevant equations are worked out in Monteiro, Smirnov, and Lucas (2006). If we follow the first approximation scheme above, we then have to solve a system of Volterra integral equations for a fixed path of $\psi(t)$ to obtain the transition matrix. This matrix subsequently has to be averaged over different simulated paths of ψ. Alternatively, we could follow the second approach and simulate the complete panel of firms to obtain a realized path of ψ that only jumps at event times. This could then be used in the Volterra equations to obtain a transition matrix. The resulting matrix again would have to be averaged over many replications. Working out the finite sample properties of these different approaches is beyond the scope of the current paper. Here we first focus on estimating the empirical dynamics of systematic credit risk. We leave the implication of these dynamics for one-year default probabilities and risk measures for future research.

5   Simulation Results

To assess the performance of the Monte Carlo maximum likelihood method in a controlled environment, a simulation experiment is carried out. The modeling framework resembles closely one of the model specifications in the empirical study of Section 6. We consider 7 rating classes. The states can be interpreted as the familiar grades AAA, AA, A, BBB, . . . , CCC, and default. Default is modeled as an absorbing state. In the simulation section, we restrict the intensities

to

$$\lambda_{sk}(t) = R_{sk}(t) \cdot \exp\left[\eta_s + \alpha_s \psi(t)\right],$$

where $\psi(t)$ is a step function that jumps at the endogenous event times t_i as in (2). The firm heterogeneity in this specification enters through the different parameters η_s for the different transition types $s = 1, \ldots, S$. Another source of heterogeneity is the latent process $\psi(t)$ that can be interpreted as the (unobserved) macroeconomic effect. The benchmark model in this simulation exercise abstracts from duration dependence by setting $H_{sk}(t) \equiv 1$. This assumption is relaxed in the empirical section. Further parsimony is introduced by setting $\alpha_s = \alpha^{down} < 0$ for downgrades, and $\alpha_s = \alpha^{up} > 0$ for upgrades. The parameter values used for the simulation can be found in the first column of Table 1.

<INSERT TABLE 1 ABOUT HERE>

Note that the number of parameters, even in this simple model specification, is large. For 7 rating classes, we have 49 possible rating transitions. Given the underlying data generating process (dgp) in this simulation study, some of the transitions are extremely unlikely, e.g., from AAA to default. As a result, in a particular simulation run there may be no transitions of this type. In such cases, the corresponding η_s parameter is not estimated. This means that not all η_s parameters can be estimated for every simulation. Table 1 only reports the simulation results for those η_s parameters for which a reasonable (50 out of 500) number of simulations exists for at least one of the four data generating processes (DGPs) presented in Table 1. The model settings we consider are $K = 70$ and $K = 700$ combined with autoregressive parameters $\rho = 1$ and $\rho = 0.9$.

For both $K = 70$ and $K = 700$, a panel of firms and rating transitions is generated as follows. At time $t_0 = 0$, the sample contains an equal number of firms in each rating category. The unobserved process $\psi(t)$ is initialized at zero. Given the parameters, this completely specifies the intensities up to the event date t_1. For the time interval $(t_{i-1}, t_i]$, the intensity of the *pooled* process is defined by

$$\lambda^*(t_i) = \sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_{sk}(t_i), \tag{23}$$

with $\lambda^*(t_1)$ applicable over the first spell $(t_0, t_1]$. The length of any spell in the pooled process can therefore be drawn from the exponential distribution with intensity parameter $\lambda^*(t_i)$. Given the durations of the spells $(t_{i-1}, t_i]$ for $i = 1, \ldots, N(T)$, the firm experiencing a rating event is drawn from the univariate Multinomial$\{\pi_1(t_i), \ldots, \pi_K(t_i)\}$ distribution where the probability

of drawing unit k is given by

$$\pi_k(t_i) = [\lambda^*(t_i)]^{-1} \sum_{s=1}^{S} \lambda_{sk}(t_i), \qquad k = 1, \ldots, K. \tag{24}$$

Next, the type of rating event for unit k is drawn from the multinomial distribution with the probability of state s being drawn for unit k given by

$$\pi_{sk}(t_i) = \left[\sum_{s=1}^{S} \lambda_{sk}(t_i) \right]^{-1} \lambda_{sk}(t_i), \tag{25}$$

for $s = 1, \ldots, S$ and $k = 1, \ldots, K$. If the event is a default, the dummy variable $R_{ik}(t)$ jumps to zero. Finally, the unobserved common risk factor $\psi_i = \psi(t_i)$ is updated using (2) with $\rho = 1$ or $\rho = 0.9$ and where the disturbances ε_i, $i = 1, \ldots, N(T)$, are drawn from a standard normal distribution. This process is repeated until all units have entered the absorbing default state, or until the event time t_i exceeds the maximum period of 25 years. For each panel size, we performed 500 replications of the simulations. All calculations in this paper were performed using the `Ox` matrix programming language of Doornik (2002) and the estimation and smoothing routines in the package `SsfPack` of Koopman, Shephard and Doornik (1999).

The simulation results for the Monte Carlo maximum likelihood procedure discussed in Section 3 are shown in Table 1. We first concentrate on the model with a random walk factor ($\rho = 1$). We see that many of the parameters are estimated accurately for a panel with 70 firms. The parameters that are estimated less accurately, correspond to larger rating transitions (e.g, AAA to BB or AA to CCC). As the larger rating transitions are much less likely, the Monte Carlo averages of the corresponding η_s parameters are based on less replications and, therefore, less accurate themselves. By contrast, the presented averages for the smaller rating transitions are all very close to the true parameter values. The accuracy of the η_s parameters increases further if the number of units is increased to $K = 700$. This is due to the fact that for a larger panel, we will observe more types of transitions in the same period of 25 years. The increase in precision also holds for the α_s parameters. By considering the average estimate, it appears that the estimator for α_s is somewhat biased towards zero. This implies the magnitude of the common risk factor is underestimated. Consequently, it is more difficult to find significant evidence of such a factor in the empirical section later on. The strength of the common factor actually found in the empirical section might thus be a lower bound on its true value.

If we consider a stationary specification for ψ with $\rho = 0.9$, the results are similar. Parameters for small rating transitions are estimated accurately. Parameters for the larger rating transitions are more difficult to estimate for the smaller panel due to the limited number of observations. For the large panel, the bias in the estimates of α_s appears smaller than for a

non-stationary $\psi(t)$. For the stationary model, we estimate the additional AR(1) parameter ρ. For the small panel of $K = 70$, the average estimate is near its true value of 0.9. The slight negative finite-sample bias is well-known from the linear model context. If the panel size increases to $K = 700$, the estimate of ρ remains stable, while its Monte Carlo standard deviation decreases somewhat. The mild improvement is largely due to the parameterization chosen. We increase the cross-sectional dimension of the panel while keeping T at 25 years. This increases the number of event times, without increasing the calendar time available. Given the scaling of the parameters of the autoregressive process with the length of the spell intervals, there is limited additional information on the long-term dynamics of the AR process. There is only a moderate precision gain caused by a more precise estimate of the signal given the increased number of events.

As explained in Section 3 (Step 3), we can use smoothing techniques to obtain an estimate of the unobserved $\psi(t)$ factor. Figure 1 illustrates the result for a single 'representative' simulation. For this simulation, we plotted the true value of $\psi(t)$ against its estimated (smoothed) counterpart using our importance sampling scheme. The 95% confidence bounds are also provided. The algorithm clearly performs adequately in recovering the characteristics of the true, unobserved $\psi(t)$ process from the observed data. As expected, the true $\psi(t)$ is much more volatile at high frequencies than its smoothed counterpart. Local and global peaks and troughs of the series, however, appear correctly positioned in calendar time. The true process also falls inside the 95% confidence interval most of the time.

<center><INSERT FIGURE 1 ABOUT HERE></center>

6 Empirical Results

6.1 Data

The data consist of rating transitions obtained from Standard & Poor's. The rating histories of all issuers are recorded in the CreditPro 7.0 database. The sample period is from the end of 1980 (the left-censoring time point) until June 2005. We express the durations of the pooled process as a fraction of the business year. Note that there may be multiple rating events on a single day. This is captured by the variables $Y_{sk}(t_i)$ in (6). The rating histories in the data set distinguish between more than 18 different rating classes. To illustrate our methodology, we consider only seven broad classes, namely AAA, AA, A, BBB, BB, B, CCC (and lower), and default. This produces 49 possible rating transitions. Out of these 49, 42 are observed in

the current sample. In terms of model (9), we therefore have $S = 42$, K is almost 7000, and $N > 25000$ (or 14000 if the $\psi(t)$ process does not jump at transitions involving the non-rated class). Even with these sample sizes, the model can be implemented on a standard desktop PC.

New firms enter the sample when they receive a rating for the first time. Firms leave the sample when they enter the default state or when their rating is withdrawn. However, S&P continues to track firms whose ratings are withdrawn. It is notified in the database when such firms default at a later stage. This should substantially mitigate any biases caused by strategic behavior of firms in maintaining a rating at S&P. If a firm first enters the non-rated class and later defaults, we skip the transition to the non-rating class. In total, we observe 7000 firms, though many of these firms are only observed over part of the sample. The number of rated firms increases over time. Accounting for attrition (defaults and transitions to non-rated) and sample extension, the time series average of the number of firms available at any moment is around 2200.

Given our interest in the clustering effect of defaults, we first clean the database for alternative forms of clustering. We use two filters. First, we check the data for firms that have long histories of coincident rating increases and decreases. Typical examples of these are firms that have merged during the time of the database. In such cases, we leave out one of the two firms from the time of the merger onwards. Second, there appeared to be some policy of the rating agencies in clustering re-ratings, e.g., centered around meeting times of the committee. To account for this, we Winsorized the 'number of events per day' to 3 by replacing $Y_{sk}(t_i)$ by $Y_{sk}(t_i) \cdot \min(3, Y_s(t_i))/Y_s(t_i)$, with $Y_s(t_i) = \sum_k Y_{sk}(t_i)$. Experiments with Winsorizing values between 1 and 5 yielded similar results. The Winsorizing procedure has the largest effect on the estimation of the dynamics of the latent component, summarized by the AR parameter ρ. Erratic clustering due to rating agencies' policies, e.g., around committee meetings, corrupts the dynamics of ψ and causes a downward bias in the estimate of ρ. This effect is mitigated considerably by the Winsorizing procedure. In future research, these anomalous clustering effects can be dealt with differently, e.g., by trying to model them explicitly.

Some descriptive features of the data are as follows. The pooled process has a high intensity of migrations, resulting in an average duration between transitions, births, or withdrawals of 1.2 days. There is a large number of downgrades and upgrades. The counts of transition events is given in Table 2. It is clear that most transitions take place to adjacent rating categories. Some of the transitions are very rare, e.g., large down-grades or up-grades. In order to check the sensitivity of our results to these rare events, we also performed a robustness check where we only incorporated transition types with more than 20 events. The results are not sensitive

to this.

The data are visualized in Figure 2. In order to keep the number of graphs manageable, we temporarily clustered the ratings further into investment grade (AAA–BBB) and sub-investment grade (BB–CCC). The upper two plots in Figure 2 show the number of investment grade downgrades and defaults on a daily basis since December 31, 1980, respectively. We see that downgrades and defaults tend to cluster in time. This follows from the concentration of vertical lines, which originate dark and bright areas along the horizontal (time) axis. The lower two plots in Figure 2 contain the number of sub-investment grade upgrades and downgrades on any given day, respectively. Interestingly, the plots complement each other. When downgrades and defaults cluster, upgrades are more sparse, and vise versa. This suggests that the model specification used with a single common risk factor $\psi(t)$ might already captures the most salient features of the data. We also experiment, however, with multiple factor models at a later stage in the analysis.

<center><INSERT FIGURE 2 ABOUT HERE></center>

6.2 Homogeneous continuous-time Markov chain model

To get a first impression of the adequacy of the empirical model specification, we consider the MLFI model without any latent dynamics. In this case, the model has $\lambda_{sk}(t) = \exp(\eta_s)$ and the MLFI model reduces to a simple homogeneous continuous-time Markov chain (HCTMC) model. The maximum likelihood (ML) estimator of η_s for the HCTMC model has a closed-form expression and is given by

$$\hat{\eta}_s = \ln\left(\sum_{i=1}^{N}\sum_{k=1}^{K} Y_{ski}\right) - \ln\left(\sum_{i=1}^{N}\sum_{k=1}^{K} t_i R_{ski}\right). \tag{26}$$

Table 3 presents the parameter estimates based on (26) as well as the estimates obtained by using the Monte Carlo methods of Section 3. As expected, the parameter estimates are almost identical for all transition types for which the number of observations is sufficiently large. But even for the transition types with only one observation, the importance sampling estimates (SML) never differ more than 10% from their closed form counterparts. This is well within a bound of two standard errors. As stated before, we also performed a robustness check only including the transition types with 20 observations or more. Here the differences between the sampling and analytic approach appear negligible. Also note that the estimate of the likelihood differs by less than 0.01% from its analytic counterpart. Again, this difference shrinks to zero if only transition types with more than 20 observations are used.

<center>23</center>

<INSERT TABLE 3 ABOUT HERE>

6.3 Estimation results for the MLFI model

We start our empirical exercise with the introduction of a single random walk component ψ_i capturing systematic credit risk. This implies the AR parameter ρ is set to unity. The estimation results are in Table 4, model B. Due to the restriction $\rho = 1$, $\psi(t_1)$ and η_s for $s = 1, \ldots, S(= 42)$ are not jointly identified. Therefore we start the latent process $\psi(t)$ at $\psi_1 = \psi(t_1) = 0$. This means that $\psi(t)$ can be interpreted as a relative credit index compared to its starting level in December 31, 1980.

For the intensity specification (1) in model B, we set $\alpha_s = -\bar{\alpha} < 0$ for downgrades and $\alpha_s = \bar{\alpha} > 0$ for upgrades. Since ψ_i is interpreted as the (unobserved) credit cycle, these sign restrictions on α_s imply an increase in the probability of downgrades and defaults if ψ_i is negative, and a simultaneous decrease in the probability of rating upgrades. Conversely, if ψ_i is positive, it leads to an increase in the probability of firms being upgraded. The sign restrictions are relaxed later on.

It is worth mentioning here that all models reported in Table 4 have an additional 42 (unreported) η_s parameters that need to be estimated. This is the number of non-zeros in Table 2. Given space constraints, we do not report them here. They are however a part of the estimation problem as sketched in Section 3.

<INSERT TABLE 4 ABOUT HERE>

The likelihood increase from model A (no common factor) to model B (a single common random walk $\psi(t)$) is large: almost 350 points upon adding one parameter. We conclude that there is ample evidence of a systematic credit risk component in the data at hand. This also appears from the standard error of the estimate of α_s.

The above result, however, hinges on the assumption of a single sensitivity parameter for all transition types. This may not hold empirically. For example, upgrades can be less sensitive to common risk factors than downgrades and defaults, see the discussion in Kavvathas (2001). As a first interesting refinement of model B, we allow in model C the single systematic credit risk factor $\psi(t)$ to have a different impact on downgrade intensities compared to upgrade intensities. We again restrict the α_s coefficient of downgrades to be negative. For the upgrade α_s, no sign restriction is imposed. We find that the improvement of model C compared to B is again statistically significant: an almost 18 point increase with one additional parameter. The sign of the upgrade α_s is intuitive. The $\psi(t)$ factor has the opposite effect on downgrade versus upgrade

24

intensities. Moreover, the systematic risk in downgrade intensities appears to be significantly higher than in upgrades. The α_s for upgrades (1.61) is about half of that for downgrades (3.13). This result is in line with the results of Das et al. (2002) and Kavvathas (2001) based on observed macroeconomic variables as proxies for the common risk factor.

Given the significant difference between systematic risk in upgrades and downgrades, we make a further distinction. We allow for a different systematic risk sensitivity for investment grade (AAA–BBB) companies that (i) upgrade, (ii) downgrade to investment grade, and (iii) downgrade to sub-investment grade. Similarly, for sub-investment grade companies we allow for a different sensitivity for downgrades, upgrades within the sub-investment grade class, and upgrades to the investment grade class. The results are presented as model D in Table 4. The likelihood increase from model C to D is again statistically significant: a 7.9 points increase upon adding 4 parameters. This gives a likelihood ratio test statistic of 15.8 with p-value 0.003. Again, we see the largest systematic credit risk sensitivity for sub-investment grade downgrades and defaults, followed by small and large investment grade downgrades. The smallest systematic risk exposure appears to be for investment grade upgrades. Though still significant, it is less than 50% of any of the other sensitivities.

The left-hand panel in Figure 2 presents the smoothed estimate of the factor $\psi(t)$ for model D. The factor is obviously low (so credit risk is high) in the mid 80s, early 90s, and early 2000s. Given the 95% confidence bands, at each of these three episodes the credit risk factor is significantly negative. Given the signs of the α_s coefficients, this implies that downgrade and default intensities were significantly higher in those periods. The converse holds for the upgrade intensities.

<INSERT FIGURE 3 ABOUT HERE>

In order to investigate whether a single risk factor suffices to model systematic credit risk, we allow for a different factor $\psi_s(t)$ for each of the 6 blocks used in model D. The results are presented as model E. The likelihood increases by about 110 points from model D to E. This cannot be a formal test, however, as the models are non-nested. The results are, however, highly suggestive of different systematic risk factors being relevant for different transition types. This also appears from the smoothed estimates of the common risk factors to be presented later. By looking at the sensitivity parameters α_s, we can get a first idea of which transition types might show a deviating systematic risk exposure. For investment grade companies, we see large increases in the (absolute) magnitude of the α_s coefficients compared to model D for rating migrations within the investment grade class. This signals that some clustering may take place, but that the clustering need not coincide with the systematic default risk component.

The same holds for limited sub-investment grade upgrades, i.e., within the sub-investment grade class. The coefficient almost doubles compared to model D. In short, the results are suggestive of a difference in clustering effects between defaults and large rating movements on the one hand, and limited re-ratings on the other hand. The difference might be attributable to the through-the-cycle policy of rating agencies. Here large rating movements and especially defaults might be attributable to economic up and down-turns, whereas limited re-ratings might be more company, industry, or rating policy specific, thus causing a different type of clustering.

So far, we imposed the restriction of a unit root on the latent component(s) $\psi(t)$. Given the short average spell length of 1.2 days, this should give a good in-sample description of the systematic risk factor. Out-of-sample, however, the restriction of a unit root may be less realistic. It implies that even when the credit risk factor is very low, it can still move up or down with equal probability. This may result in over-estimates of credit Value-at-Risks (VaRs) for risk management purposes. To check the sensitivity of the results for the unit root assumption, we re-estimate all models using stationary AR(1) models. We use the specification in (2) and (3) in Section 2. The autoregressive parameter and the innovation variance are now tuned automatically to the spell length. Short durations have a high persistence parameter and a small innovation variance. By using this specification and the data at hand, the ρ parameter can readily be interpreted as an autoregressive parameter at an 'annual' (260 business days) frequency.

The results for the single factor model with stationary latent component and common sensitivity parameter α_s are presented as model F. The likelihood increases by 4.5 points compared to model B upon adding one parameter. We cannot say, however, whether this is statistically significant, because this is a unit root test in a class of non-linear models. Working out the statistical properties of this test is well beyond the scope of the current paper. A similar increase of 4.1 points is seen for the case with unrestricted α_ss, compare model G to D.

The estimated value of the AR parameter in models F (0.38) and G (0.37) is very similar. The estimates are much smaller than those in Koopman et al. (2005a). The latter, however, are not based on an intensity framework and use annual default counts. The estimates in McNeil and Wendin (2006) are based on quarterly counts and are also somewhat higher, around 0.6 to 0.7. Comparing models G and D, we see that the reduced persistence of ψ by the lower autoregressive parameter causes the sensitivity coefficients α_s to increase. This is due to the fact that the model also tries to match the unconditional in-sample variance in default intensities.

A graph of the smoothed AR(1) factor for model G is presented in the right-hand panel of Figure 3. Comparing this to the smoothed estimate for the random walk model in the left-hand panel (model D), we notice that the characteristics of the two paths in terms of peaks

and troughs are roughly similar. There are three main differences: the location, scale, and standard error bands. It must be kept in mind, however, that $\psi(t)$ enters the intensity process as $\eta_s + \alpha_s \psi(t)$. The location and scale effect is thus partly off-set by the differences in estimates of α_s and η_s between the models. On average, the η_s in model D is about 0.4 lower than in model G for downgrades and defaults. For upgrades, the average difference is -0.2. Similarly, the increase in the (absolute) magnitude of the α_ss from model D to G off-sets the smaller scale for the stationary $\psi(t)$ factor in model G. As a result, given the sign differences in α_s between upgrades and downgrades, the difference in impact of the systematic risk factor between models D and G appears negligible.

In model H we again look at a multiple factor model. The six ψ factors now all follow separate (independent) stationary autoregressive processes with parameters ρ_s. Again we see the familiar effect that lower persistence in ψ is matched by a higher α_s. More interestingly, however, are the differences in the estimated autoregressive parameters between transition types. The persistence of the sub-investment grade downgrade factor (0.40) is similar as in model G. The persistence in within investment grade downgrades and within sub-investment grade upgrades, however, is virtually absent. The common factor here does not appear to be sticky at all. Though not apparent immediately, this implies that there is not really a systematic credit risk factor for these types of transitions. If ρ_s is close to zero, the ψ factor essentially collapses to another source of firm (and time) specific error given its frequency of updating (i.e., at every event). Consequently, the durations in the pooled process are completely idiosyncratic, though no longer (conditionally) exponential. By contrast, we see that the persistence of the common factor for large up or downgrades is high (0.93 and 0.92). This again underlines our earlier finding that the common risk factors may be different across the different transition types. This is also visualized in Figure 4. The default and downgrade factors $\psi(t)$ for IS^{down} and SS^{down} clearly reveal a different pattern from the upgrades, II^{up} and SI^{up}.

To conclude our set of model specifications, we also consider a semi-Markov model. In particular, we allow the intensities to depend on the time spent in the current rating. We capture this effect through the baseline hazard functions H_{sk} in (1). We allow for a different baseline hazard for investment grade versus sub-investment grade companies. The baseline hazards are flexibly parameterized by cubic splines with knots at 0, 1, 3, and 6 years. The spline passes through (0,0) in order to identify both the spline parameters and the η_s coefficients. The results are presented as model I in Table 4. The increase in likelihood from model H to I is again statistically significant: 76.9 points upon adding 6 parameters. The persistence parameters ρ and sensitivity parameters α_s are very similar between the two models.

Figure 5 presents the log baseline hazard functions (cubic splines) capturing the semi-Markov

behavior in model I. There is a clear difference between the investment grade and sub-investment grade splines. During the initial two years, the default intensity of investment grade firms sharply increases. After that, the intensity again slowly decreases. For sub-investment grade companies, the pattern is more volatile. Two peaks can be distinguished. First, firms with very short durations in the sub-investment grade ratings have high re-ratings and default intensities. This is probably due to the rating momentum effect: firms that have been downgraded recently are more likely to downgrade in the near future because of the smooth rating revision policy by rating agencies. A second peak emerges after 3 to 5 years. Here, default typically occurs because initial debt taken by these companies has to be rolled over. This causes a peak in defaults because of some sub-investment quality firms getting into financial distress at that time.

6.4 Forecasting transition probabilities

To illustrate the impact of fluctuations in the estimated latent risk component for the implied 1-year transition/default probabilities, a recursive (partially) out-of-sample forecasting exercise is carried out. The details of the forecasting study are as follows. We consider the data windows Dec 1980–June 1992, Dec 1980–June 1993, ..., Dec 1980–June 2005. For each data window, we estimate the parameters η_s, the smoothed risk factor estimate $\hat{\psi}$ and its covariance matrix. The estimates of the loadings α_s and ρ_s are kept fixed at their estimates from Table 4 in this illustration. This reduces the computations needed. Note that in the forecasting exercise, we do not Winsorize the realizations in the forecasting period as we did during the estimation stage.

Given the estimates in a specific year, we simulate daily paths of the common risk factor(s) $\psi_s(t)$ over the subsequent year, starting with a draw from the smoothed distribution. For each path ψ, the product integral in (20) is computed. This gives us a conditional (on ψ) annual transition matrix. Finally, we compute the 99% quantile of the default probabilities from these simulated transition matrices for each initial rating category. The 99% quantile is taken over the different simulated paths of ψ. We do so for models A, D, E, G, and H from Table 4. As mentioned in Section 4, if a semi-Markov specification were used like model I, computing the annual transition matrix would be much more involved and dependent on the current portfolio composition. See Monteiro et al. (2006) for more details on the semi-Markov case.

To benchmark all model forecasts, we also compute the nonparametric Aalen Johansen (AJ) estimator of the transition probabilities over each year. This empirical transition matrix can be seen in the current context as a proxy for the 'true' or realized[10] transition probabilities, see

---

[10]The term realized is chosen deliberately here. Its well-known analogue is the nonparametric computation

Lando and Skødeberg (2002). Figures 6 and 7 presents the results.

<INSERT FIGURES 6 AND 7 ABOUT HERE>

The forecasts from the MLFI model move much more in accordance with the nonparametric *ex-post* AJ estimates than the HCTMC model (HMC). Note that though the HMC does not include a dynamic component,the forecasts still adapt over the years due to the recursive nature of the estimation procedure for η_s. The changes are nevertheless insufficient to keep up with the changes in the true default rates (see AJ curve).

The 99% quantiles for the MLFI models with 1 or 6 AR risk factors appear reasonable upper bounds for all initial rating categories except single A. By contrast, the forecasted default probabilities for the non-stationary models are much larger. This is due to the random walk nature of the factors. Even if the common risk factor is already very low (so default probabilities are high), it is still equally likely that the factor goes further down or up. Excessively large forecasts for default probabilities are the result and these subsequently lead to inefficient (though highly prudent) capital buffers at most times. For the stationary risk factors, low factors tend to mean revert, resulting in less extreme estimates of future default probabilities. This comes at the cost of an increased probability of violating the 99% quantile bounds around early 2000s for some initial rating categories.

An interesting feature of the MLFI specifications is its increase in predicted default probabilities before the actual crash. This holds in many cases, especially around the record default years in the early 2000s. The HMC model with recursively estimated parameters, however, obviously lags the dynamic pattern in defaults. This suggests that for risk management purposes, models that account for the dynamic pattern in systematic risk factors are very important.

Our forecasting procedure produces much more than only the default rates. Also all rerating probabilities are available. To get a (parsimonious) impression of how well the different models perform with respect to all the remaining transition probabilities, a summary statistic of the transition matrix is needed. Jafry and Schuermann (2004) discuss a number of metrics to compare the forecasts of transition matrices. In particular, they propose the average of singular values of the transition matrix minus the unit matrix. We computed this metric over time for the different model specifications used. The results are in the lower-right panel of Figure 7. Again, we see a much larger resemblance between the dynamics of the AJ realizations with the MLFI specifications compared to the HMC model. It appears that the random walk models are the only ones able to explain the large increase in singular values in the early 2000s. As

of realized volatility in the empirical finance literature, see Andersen, Bollerslev, Diebold, Labys (2003). The AJ estimator is its counterpart in the context of transition models.

seen earlier, however, this comes at the cost of very prudent capital standards in other years.

It is clear that the current forecasting exercise can be extended further. Different specifications for the MLFI could be tried. Also, the semi-Markov property could be included at the cost of a significant increase in the required computations in line with Monteiro et al. (2006). Moreover, dollar based metrics could be easily developed. For example, given the conditional forecasts of transition matrices from the product integral expression, one can construct (conditional) forecasts of capital requirements in line with the New Basle Capital Accord, see also Jafry and Schuermann (2004). We leave such extensions to future applications.

7 Conclusion

In this paper we have motivated and introduced a multi-state latent factor intensity (MLFI) model for credit rating transitions. The model can be regarded as a generalization of the latent factor intensity point process introduced in Bauwens and Hautsch (2003) to a situation with multiple origin and destination states. However, the econometric issues related to this generalization are intricate and the computational consequences are severe. We have discussed the details for the estimation of the MLFI model using a Monte Carlo maximum likelihood procedure that consists of a combination of importance sampling techniques and state space methods, as outlined in Durbin and Koopman (1997, 2001). It is shown that this procedure can be extended successfully to a multivariate class of non-Gaussian models. A simulation study was carried out to show that the estimation procedure works well in recovering the parameters of the MLFI model. Finally, we have applied the model to a real-world dataset of credit rating migrations. A significant common risk factor in credit rating migrations is found. The impact of this risk factor is higher for downgrades than for upgrades. This empirical result suggests that upgrades are more subject to idiosyncratic shocks than downgrades. This finding is consistent with the conclusions in the earlier studies of Kavvathas (2001) and Das et al. (2002). Furthermore, our results pointed towards important potential differences in the systematic risk of different transition types. This can have important implications for single factor models for risk management, as they are currently widely employed in the industry.

Generalizations with respect to the current empirical specification are easily incorporated in the structure of the MLFI model. For example, the general specification allows for the inclusion of observed firm-specific and economic variables as in Duffie et al. (2006b) and Koopman et al. (2006), self-exciting processes, and additional dynamic components. A further interesting feature of the model is that it produces a high-frequency credit cycle index estimated directly from default and rating migration data. Finally, within the model specification proposed in

this paper it is possible to estimate and test formally for the number of latent factors driving the default and rating migration intensities. The implementation of the model for credit risk simulations is straightforward given the integrated structure of the model. The dynamics of the common risk factor are estimated simultaneously with the development of default events conditional on this common risk factor. Therefore, they can also be easily integrated in a forecasting exercise as the one presented in Section 6.

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 5*, 141-50.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003, Modelling and forecasting volatility, *Econometrica 71*, 579-625.

Andersen, P.K., Borgan, Ø., Gill, R.D., and N. Keiding, 1993, *Statistical models based on counting processes*. Springer-Verlag, New York.

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

Bauwens, L., and N. Hautsch, 2003, Dynamic latent factor models for intensity processes, *Working paper - CORE*.

Carling, K., Jacobson, T., Lindé, J., and K. Roszbach, 2002, Capital charges under Basel II: Corporate credit risk modeling and the macro economy, *Working paper - Sveriges Riksbank*.

Couderc, F., and O. Renault, 2004, Time-to-default: Life cycle, global and industry cycle impacts, *Working paper - University of Geneva*.

Cox, D.R. 1962, *Renewal theory*. Methuen and Co. Ltd. London.

Das, S.R., Freed, L., Geng, G. and N. Kapadia, 2002, Correlated default risk, *Working paper - Santa Clara University*.

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.

Duffie, D., L. Saita, and K. Wang, 2006a, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics*, forthcoming.

Duffie, D., A. Eckner, G. Horel, and L. Saita, 2006b, Frailty correlated default, *Working paper, Graduate School of Business, Stanford University.*

Durbin, J. and S.J. Koopman, 1997, Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*, pp 669-684.

Durbin, J. and S.J. Koopman, 2001, *Time series analysis by state space methods.* Oxford University Press, Oxford.

Durbin, J. and S.J. Koopman, 2002, A simple and efficient simulation smoother for state space time series models. *Biometrika 89*, 603-616

Engle, R.F. and J.R. Russell, 1998, Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica Vol. 66*, No. 5, pp 1127-1162.

Fledelius, P., Lando, D. and J.P. Nielsen, 2004, Non-parametric analysis of rating transition and default data. *Journal of Investment Management Vol. 2*, No.2.

Gagliardini, P. and C. Gourieroux, 2004, Stochastic migration models with application to corporate risk. *Working paper - CREST.*

Gagliardini, P. and C. Gourieroux, 2005, Migration correlation: Definition and efficient estimation. *Journal of Banking & Finance 29*, 865-894.

Gill, R.D. 2001, Product integration. *Working paper - University of Utrecht*

Glynn, P.W. 1988, A GSMP formalism for discrete-event systems. *Proceedings of the IEEE 77*, 14-23.

Harvey, A.C. 1989, *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press.

Hawkes, A.G. 1971, Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*, 83-90.

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

de Jong, P. and N. Shephard, 1995, The simulation smoother for time series models. *Biometrika 82*, 339-350.

Kavvathas, D. 2001, Estimating credit rating transition probabilities for corporate bonds, *Working paper - University of Chicago.*

Koopman, S.J. and A. Lucas, 2005, Business and default cycles for credit risk. *Journal of Applied Econometrics 20*, 311-323.

Koopman S.J., Lucas, A. and R.J.O. Daniels, 2005a, A non-Gaussian panel time series model for estimating and decomposing default risk. Tinbergen Institute Discussion paper, TI 05-060/4.

Koopman, S.J., A. Lucas, and P. Klaassen, 2005b, Empirical credit cycles and capital buffer formation. *Journal of Banking & Finance*, 29, 3159-3179.

Koopman S.J., R. Kraeussl, A. Lucas, and A.B. Monteiro, 2006, Credit Cycles and Macro Fundamentals, Tinbergen Institute Discussion paper TI06-023/2.

Koopman, S.J., Shephard, N. and J. Doornik, 1999, Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal*, 113-166.

Lancaster, T. 1990, *The econometric analysis of transition data.* Cambridge University Press.

Lando, D. and T.M. Skødeberg, 2002, Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance 26*, 423-444.

Liesenfeld, R. and J.F. Richard, 2003, Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance 10*, 505-531.

McNeil, A.J. and J. Wendin, 2006, Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk, *Journal of Empirical Finance*, forthcoming.

Monteiro, A., G.V. Smirnov, and A. Lucas, 2006, Nonparametric Estimation for Non-Homogeneous Semi-Markov Processes: An Application to Credit Risk. Tinbergen Institute Discussion paper TI06-024/2.

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

Russell, J.R. 1999, Econometric Modelling of Multivariate Irregularly spaced High-frequency data. *Working Paper - University of Chicago.*

Shephard, N. and M.K. Pitt, 1997, Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653-667.

Tsiatis, A. 1975, A nonidentifiability aspect of the problem of competing risks. *Proceedings of the Nacional Academy of Sciences of the United States of America vol. 72*, pp. 20-22.

van den Berg, G.J. 2001, Duration Models: Specification, identification, and multiple durations, in Heckman, J.J. and E. Leamer, (Eds.), *Handbook of Econometrics, Vol.*

V, North Holland, Amsterdam

Wilson, T. 1997a, Portfolio credit risk part I. *Risk*, 111-117.

Wilson, T. 1997b, Portfolio credit risk, part II. *Risk*, 56-61.

Table 1: Monte Carlo Results

This table contains parameter estimates for the MLFI model, $\lambda_{sk}(t) = R_{sk}(t) \cdot \exp[\eta_s + \alpha_s \psi(t)]$, for $k = 1, \ldots, K$ with K the number of units, $s = 1, \ldots, 49$. The true parameters are taken in accordance with the empirical results in Section 6, with different loading α_s for up and down-grades, model C in Table 4. There are 7 rating classes. Initial ratings are distributed evenly over these classes. We estimate the model for $K = 70$ and $K = 700$, both with a random walk common factor ψ and with a stationary AR, parameter $\rho = 0.9$. The maximum time T is set to 25 years, unless the complete sample has defaulted at an earlier stage. We performed 500 replications for each parameter combination. Monte-Carlo averages and standard errors (in parentheses) are presented for those parameters that have a sufficient number of occurrences over all simulations.

	True	K=70 $\rho^{true}=1$		K=700 $\rho^{true}=1$		K=70 $\rho^{true}=0.9$		K=700 $\rho^{true}=0.9$	
$\eta_{AAA \to AA}$	-3.47	-3.48	(0.79)	-3.51	(0.27)	-3.55	(0.88)	-3.50	(0.16)
$\eta_{AAA \to A}$	-5.88	-5.20	(1.36)	-5.99	(0.64)	-5.35	(1.32)	-6.04	(0.51)
$\eta_{AAA \to BBB}$	-8.38	-5.39	(1.10)	-7.67	(1.02)	-5.52	(1.37)	-7.72	(0.93)
$\eta_{AAA \to BB}$	-7.55	-5.73	(1.10)	-7.36	(1.08)	-5.64	(0.89)	-7.44	(0.75)
$\eta_{AA \to AAA}$	-5.04	-4.81	(0.93)	-5.05	(0.23)	-4.84	(1.00)	-5.07	(0.15)
$\eta_{AA \to A}$	-3.04	-3.10	(0.80)	-3.07	(0.26)	-3.12	(0.63)	-3.08	(0.14)
$\eta_{AA \to BBB}$	-5.84	-5.38	(1.40)	-5.92	(0.55)	-5.49	(1.38)	-6.00	(0.41)
$\eta_{AA \to BB}$	-8.47	-5.65	(1.23)	-8.01	(1.00)	-5.74	(1.30)	-8.06	(0.72)
$\eta_{AA \to B}$	-7.59	-5.78	(1.44)	-7.49	(1.08)	-5.54	(1.32)	-7.65	(0.83)
$\eta_{AA \to CCC}$	-9.63	-6.03	(1.57)	-8.03	(0.90)	-5.50	(1.05)	-8.20	(0.69)
$\eta_{A \to AAA}$	-7.06	-6.27	(1.65)	-7.03	(0.50)	-5.93	(0.82)	-7.22	(0.59)
$\eta_{A \to AA}$	-3.96	-3.99	(0.63)	-3.93	(0.10)	-3.96	(0.56)	-3.96	(0.08)
$\eta_{A \to BBB}$	-3.38	-3.49	(0.78)	-3.41	(0.25)	-3.49	(0.72)	-3.40	(0.14)
$\eta_{A \to BB}$	-6.18	-5.91	(1.41)	-6.30	(0.52)	-5.80	(1.32)	-6.27	(0.33)
$\eta_{A \to B}$	-6.89	-6.34	(1.65)	-6.95	(0.73)	-6.16	(1.23)	-7.06	(0.56)
$\eta_{A \to D}$	-7.75	-6.52	(1.69)	-7.77	(0.98)	-6.34	(1.28)	-7.86	(0.78)
$\eta_{BBB \to AAA}$	-8.78	-6.17	(1.18)	-8.09	(0.92)	-5.91	(0.74)	-8.02	(0.64)
$\eta_{BBB \to AA}$	-5.88	-5.66	(1.15)	-5.94	(0.36)	-5.39	(0.87)	-6.00	(0.26)
$\eta_{BBB \to A}$	-3.08	-3.12	(0.42)	-3.06	(0.08)	-3.06	(0.29)	-3.07	(0.05)
$\eta_{BBB \to BB}$	-3.41	-3.50	(0.84)	-3.44	(0.26)	-3.49	(0.64)	-3.43	(0.15)
$\eta_{BBB \to B}$	-5.81	-5.61	(1.31)	-5.87	(0.54)	-5.68	(1.26)	-5.90	(0.35)
$\eta_{BBB \to CCC}$	-7.94	-6.51	(1.26)	-7.89	(0.92)	-6.23	(1.02)	-7.92	(0.79)
$\eta_{BBB \to D}$	-6.51	-6.19	(1.25)	-6.61	(0.56)	-5.97	(1.39)	-6.65	(0.52)
$\eta_{BB \to AAA}$	-7.62	-5.60	(1.15)	-7.36	(0.81)	-5.41	(0.96)	-7.45	(0.68)
$\eta_{BB \to AA}$	-6.75	-5.46	(0.92)	-6.74	(0.79)	-5.27	(1.12)	-6.91	(0.71)
$\eta_{BB \to A}$	-5.20	-4.86	(1.11)	-5.26	(0.28)	-4.93	(0.94)	-5.24	(0.20)
$\eta_{BB \to BBB}$	-2.61	-2.60	(0.35)	-2.61	(0.08)	-2.59	(0.28)	-2.59	(0.05)
$\eta_{BB \to B}$	-3.02	-3.13	(0.85)	-3.06	(0.27)	-3.12	(0.75)	-3.05	(0.14)
$\eta_{BB \to CCC}$	-5.83	-5.53	(1.63)	-5.87	(0.51)	-5.32	(1.48)	-5.94	(0.43)
$\eta_{BB \to D}$	-5.51	-5.26	(1.17)	-5.61	(0.49)	-5.26	(1.30)	-5.66	(0.34)
$\eta_{B \to AA}$	-7.06	-5.04	(0.76)	-6.98	(0.84)	-5.12	(1.15)	-7.10	(0.70)
$\eta_{B \to A}$	-6.14	-5.03	(1.10)	-6.24	(0.65)	-4.90	(1.17)	-6.41	(0.60)
$\eta_{B \to BBB}$	-5.37	-4.79	(1.10)	-5.47	(0.46)	-4.86	(0.96)	-5.44	(0.28)
$\eta_{B \to BB}$	-2.64	-2.65	(0.48)	-2.62	(0.08)	-2.59	(0.28)	-2.63	(0.05)
$\eta_{B \to CCC}$	-3.14	-3.23	(0.93)	-3.19	(0.28)	-3.22	(0.75)	-3.17	(0.16)
$\eta_{B \to D}$	-3.97	-4.03	(0.89)	-4.02	(0.33)	-4.03	(1.03)	-4.01	(0.18)
$\eta_{CCC \to A}$	-5.24	-3.40	(1.00)	-5.21	(0.75)	-3.63	(1.11)	-5.30	(0.70)
$\eta_{CCC \to BBB}$	-4.84	-3.48	(1.12)	-4.89	(0.74)	-3.49	(0.86)	-5.05	(0.62)
$\eta_{CCC \to BB}$	-4.12	-3.42	(0.90)	-4.24	(0.45)	-3.36	(1.03)	-4.30	(0.43)
$\eta_{CCC \to B}$	-1.74	-1.76	(0.57)	-1.73	(0.10)	-1.79	(0.51)	-1.72	(0.05)
$\eta_{CCC \to D}$	-1.24	-1.37	(0.73)	-1.29	(0.22)	-1.32	(0.55)	-1.28	(0.14)
α^{up}	1.60	1.43	(0.81)	1.47	(0.36)	1.46	(1.40)	1.56	(0.36)
α^{down}	-3.20	-2.64	(1.05)	-2.90	(0.64)	-2.83	(1.41)	-3.12	(0.63)
ρ		___		___		0.85	(0.20)	0.83	(0.15)

35

Table 2: Event counts

This table contains the number of events over the sample period from (row) a specific rating to a new (column) rating. D denotes default.

From	To							
	AAA	AA	A	BBB	BB	B	CCC	D
AAA	___	95	8	1	2	0	0	0
AA	34	___	475	25	2	5	1	0
A	10	219	___	714	40	19	0	9
BBB	2	27	457	___	605	53	6	24
BB	4	8	36	467	___	769	48	65
B	0	6	16	34	524	___	855	349
CCC	1	0	5	8	13	135	___	678

Table 3: Parameter estimates of the HCTMC model

This table presents estimates of a homogeneous continuous-time Markov chain (HCTMC) model with intensities $\lambda_{sk}(t) = \exp(\eta_s)$. The transition types are sorted on their number of observations. The first two columns provide the rating transition type. The third column indicates the number of steps taken for this transition type in the 7 grade system. The fourth column contains the number of observations for this transition type. The column headed ML-$\hat{\eta}_s$ contains the closed form maximum likelihood (ML) estimates from (26). Its ML asymptotic standard error is in parentheses in column six. The Monte Carlo maximum likelihood estimates SML-$\hat{\eta}_s$ and their simulated standard errors (using 100 samples) follow in columns seven and eight. The column $\Delta\hat{\eta}_s$ and $\Delta s.e.$ give the difference between the ML and SML estimates, and between their standard errors, respectively.

From		To	#steps	#obs	ML-$\hat{\eta}_s$		SML-$\hat{\eta}_s$		$\Delta\hat{\eta}_s$	$\Delta s.e.$
B	$\rightarrow$	CCC	1	855	-2.50	(0.04)	-2.50	(0.03)	0.00	0.00
BB	$\rightarrow$	B	1	769	-2.39	(0.04)	-2.40	(0.03)	0.00	0.01
A	$\rightarrow$	BBB	1	714	-2.86	(0.04)	-2.86	(0.04)	0.00	0.00
CCC	$\rightarrow$	D	1	678	-0.53	(0.04)	-0.52	(0.04)	-0.01	0.00
BBB	$\rightarrow$	BB	1	605	-2.93	(0.04)	-2.94	(0.04)	0.01	0.00
B	$\rightarrow$	BB	1	524	-2.94	(0.04)	-2.96	(0.04)	0.02	0.00
AA	$\rightarrow$	A	1	475	-2.55	(0.05)	-2.56	(0.04)	0.00	0.00
BB	$\rightarrow$	BBB	1	467	-2.89	(0.05)	-2.88	(0.04)	-0.01	0.00
BBB	$\rightarrow$	A	1	457	-3.20	(0.05)	-3.21	(0.05)	0.01	0.00
B	$\rightarrow$	D	2	349	-3.35	(0.05)	-3.34	(0.06)	-0.01	0.00
A	$\rightarrow$	AA	1	219	-4.06	(0.07)	-4.07	(0.07)	0.01	0.00
CCC	$\rightarrow$	B	1	135	-2.12	(0.09)	-2.11	(0.08)	-0.01	0.01
AAA	$\rightarrow$	AA	1	95	-2.96	(0.10)	-2.97	(0.10)	0.00	0.01
BB	$\rightarrow$	D	3	65	-4.85	(0.12)	-4.88	(0.10)	0.03	0.03
BBB	$\rightarrow$	B	2	53	-5.34	(0.14)	-5.32	(0.14)	-0.02	0.00
BB	$\rightarrow$	CCC	2	48	-5.16	(0.14)	-5.18	(0.13)	0.02	0.02
A	$\rightarrow$	BB	2	40	-5.72	(0.16)	-5.70	(0.15)	-0.02	0.01
BB	$\rightarrow$	A	2	36	-5.44	(0.17)	-5.46	(0.17)	0.02	-0.01
AA	$\rightarrow$	AAA	1	34	-5.14	(0.17)	-5.14	(0.11)	0.00	0.06
B	$\rightarrow$	BBB	2	34	-5.67	(0.17)	-5.63	(0.16)	-0.04	0.01
BBB	$\rightarrow$	AA	2	27	-6.02	(0.19)	-6.01	(0.15)	0.00	0.04
AA	$\rightarrow$	BBB	2	25	-5.45	(0.20)	-5.38	(0.20)	-0.07	0.00
BBB	$\rightarrow$	D	4	24	-6.13	(0.20)	-6.04	(0.21)	-0.10	-0.01
A	$\rightarrow$	B	3	19	-6.46	(0.23)	-6.44	(0.22)	-0.02	0.00
B	$\rightarrow$	A	3	16	-6.43	(0.25)	-6.43	(0.24)	0.01	0.01
CCC	$\rightarrow$	BB	2	13	-4.46	(0.28)	-4.49	(0.27)	0.03	0.01
A	$\rightarrow$	AAA	2	10	-7.11	(0.32)	-7.16	(0.30)	0.06	0.02
A	$\rightarrow$	D	5	9	-7.21	(0.33)	-7.31	(0.33)	0.10	0.01
AAA	$\rightarrow$	A	2	8	-5.39	(0.35)	-5.40	(0.26)	0.01	0.09
BB	$\rightarrow$	AA	3	8	-6.95	(0.35)	-7.06	(0.31)	0.11	0.04
CCC	$\rightarrow$	BBB	3	8	-4.95	(0.35)	-5.17	(0.36)	0.22	-0.01
BBB	$\rightarrow$	CCC	3	6	-7.52	(0.41)	-7.48	(0.40)	-0.04	0.01
B	$\rightarrow$	AA	4	6	-7.41	(0.41)	-7.38	(0.36)	-0.03	0.05
AA	$\rightarrow$	B	4	5	-7.06	(0.45)	-7.20	(0.45)	0.14	0.00
CCC	$\rightarrow$	A	4	5	-5.42	(0.45)	-5.59	(0.40)	0.17	0.05
BB	$\rightarrow$	AAA	4	4	-7.64	(0.50)	-7.93	(0.61)	0.29	-0.11
AAA	$\rightarrow$	BB	4	2	-6.78	(0.71)	-7.14	(0.73)	0.36	-0.03
AA	$\rightarrow$	BB	3	2	-7.97	(0.71)	-7.97	(0.59)	-0.01	0.12
BBB	$\rightarrow$	AAA	3	2	-8.62	(0.71)	-8.93	(0.88)	0.32	-0.17
AAA	$\rightarrow$	BBB	3	1	-7.47	(1.00)	-7.88	(0.91)	0.40	0.09
AA	$\rightarrow$	CCC	5	1	-8.67	(1.00)	-9.43	(1.15)	0.76	-0.15
CCC	$\rightarrow$	AAA	6	1	-7.03	(1.00)	-7.63	(1.03)	0.61	-0.03
log-likelihood					-25582.7		-25584.4		1.7	(<0.01%)

Table 4: Parameter estimates of the MLFI model

This table contains the parameter estimates ρ_s and $\tilde{\alpha}_s = 100\alpha_s$ of the MLFI model, $\lambda_{sk}(t) = R_{sk}(t)\exp[\eta_s + \alpha_s\psi_s(t) + h_s(u_{sk}(t))]$, for $k = 1,\ldots,K$ with K the number of firms, $s \in \mathbb{S} = \{AAA \rightarrow AA, AAA \rightarrow A,\ldots,CCC \rightarrow B, CCC \rightarrow Dflt\}$, the 49 different possible rating transitions for the 7 rating classes $AAA,\ldots,CCC$. The factor $\psi_{is} = \psi_s(t_i)$ follows an AR(1) process $\psi_{si} = \rho_s^{t_i-t_{i-1}}\psi_{s,i-1} + \sigma_{si}\varepsilon_{si}$, with ε_i i.i.d. standard normal random variables, t_i for $i = 1,\ldots,N$ the event times of the pooled process measured in fractions of the 260 day business year (instead of in days as in (3)), and $\sigma_{si}^2 = (1 - \rho_s^{t_i-t_{i-1}})/(1 - \rho_s^{2/260})$. For identification, $\psi_{s0} = 0$. The log baseline hazard function h_s depends on the time spent in the current rating and follows a cubic spline function with knots at 0, 1, 3, and 10 years. Given the presence of the η_s factor, we restrict $h_s(0) = 0$. Spline values at spell lengths exceeding 10 years are set to the spline value at 10 years. The η_s are allowed to be different for all transition types (yielding almost 49 parameter estimates that are not presented here). The log baseline hazard h_s only differs between investment grade (AAA-BBB) and sub-investment (BB-CCC) grade firms (and has 3 parameters per spline). We allow for a different α_s for within investment grade upgrades (II^{up}) and downgrades (II^{down}), within sub-investment grade upgrades (SS^{up}) and downgrades (SS^{down}), and across investment-sub-investment grade upgrades (SI^{up}) and downgrades (IS^{down}). A similar block structure is imposed on the ψ_s and ρ_s if we allow for multi-factor models. For models A–E, the latent component(s) $\psi_s(t)$ are random walks, i.e., $\rho_s \equiv 1$. Standard errors (in parentheses) and log-likelihood values are based on 1,000 importance samples.

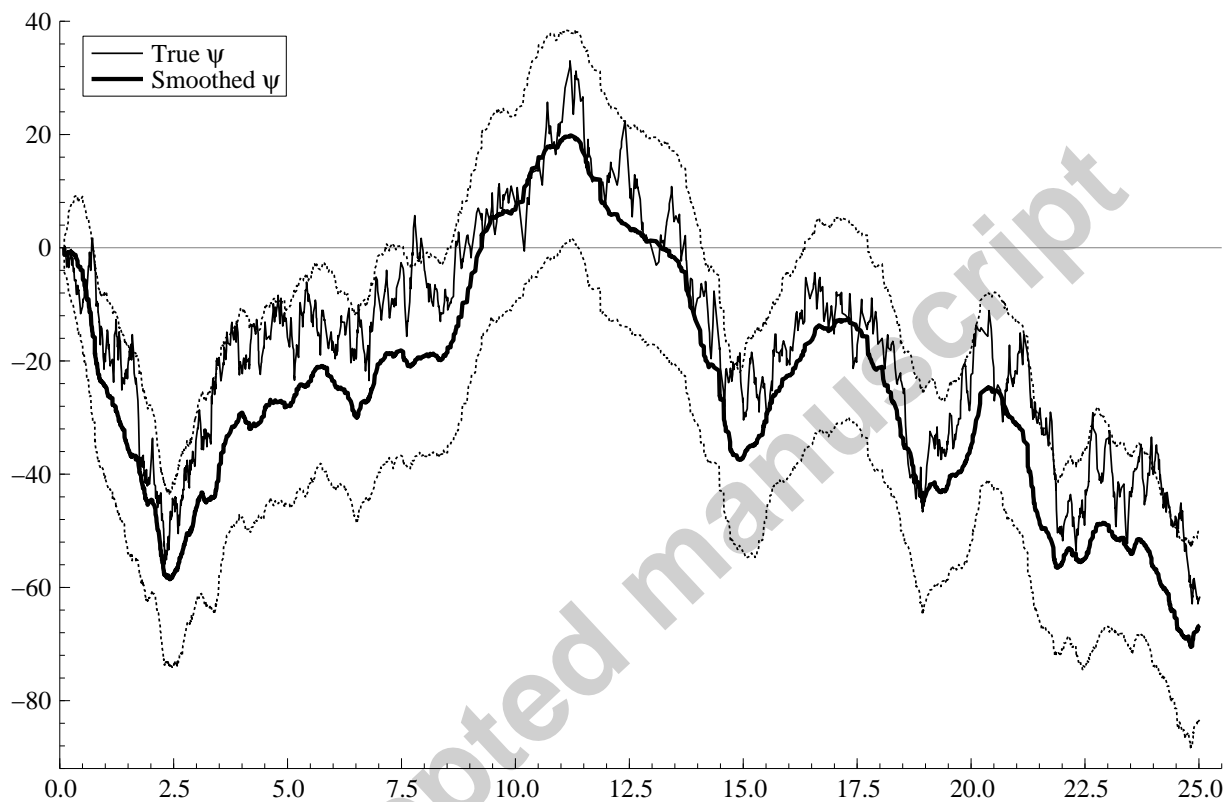| Model		II_{up}	II_{down}	IS_{down}	SI_{up}	SS_{up}	SS_{down}	log-lik	#spline pars	#α pars	#ρ pars	#ψ fact.
A	$\tilde{\alpha}$	0.00	0.00	0.00	0.00	0.00	0.00	-25584.4	0	0	0	0
B	$\tilde{\alpha}$	2.56	-2.56	-2.56	2.56	2.56	-2.56	-25236.4	0	1	0	1
		(0.36)	(0.36)	(0.36)	(0.36)	(0.36)	(0.36)					
C	$\tilde{\alpha}$	1.61	-3.13	-3.13	1.61	1.61	-3.13	-25217.7	0	2	0	1
		(0.29)	(0.46)	(0.46)	(0.29)	(0.29)	(0.46)					
D	$\tilde{\alpha}$	0.84	-2.82	-2.54	1.92	2.17	-3.27	-25209.8	0	6	0	1
		(0.33)	(0.47)	(0.47)	(0.42)	(0.44)	(0.48)					
E	$\tilde{\alpha}$	1.93	-4.37	-2.26	1.27	3.72	-3.32	-25097.2	0	6	0	6
		(0.64)	(0.83)	(0.53)	(0.40)	(0.85)	(0.56)					
F	$\tilde{\alpha}$	3.11	-3.11	-3.11	3.11	3.11	-3.11	-25231.9	0	1	1	1
		(0.56)	(0.56)	(0.56)	(0.56)	(0.56)	(0.56)					
	ρ	0.38	0.38	0.38	0.38	0.38	0.38					
		(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)					
G	$\tilde{\alpha}$	1.03	-3.41	-3.09	2.30	2.64	-3.93	-25205.7	0	6	1	1
		(0.43)	(0.75)	(0.74)	(0.59)	(0.64)	(0.80)					
	ρ	0.37	0.37	0.37	0.37	0.37	0.37					
		(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)					
H	$\tilde{\alpha}$	2.29	-8.17	-2.65	1.42	6.81	-3.93	-25072.0	0	6	6	6
		(0.79)	(1.80)	(0.77)	(0.52)	(1.77)	(0.80)					
	ρ	0.80	0.04	0.79	0.92	0.10	0.43					
		(0.19)	(0.05)	(0.20)	(0.14)	(0.10)	(0.20)					
I	$\tilde{\alpha}$	2.07	-7.80	-2.50	1.43	6.95	-3.93	-24995.1	6	6	6	6
		(0.73)	(1.51)	(0.66)	(0.53)	(2.04)	(0.80)					
	ρ	0.82	0.05	0.92	0.93	0.09	0.40					
		(0.17)	(0.05)	(0.15)	(0.14)	(0.12)	(0.19)					

Figure 1: True versus smoothed estimate of $\psi(t)$

The baseline model and the simulation set-up are the same as explained in the note to Table 1. The thick, solid curve is the smoothed estimate of $\psi(t)$. The thin solid curve is the true $\psi(t)$ path in a typical simulation. The dotted lines give the 95% confidence band of the smoothed $\psi(t)$.
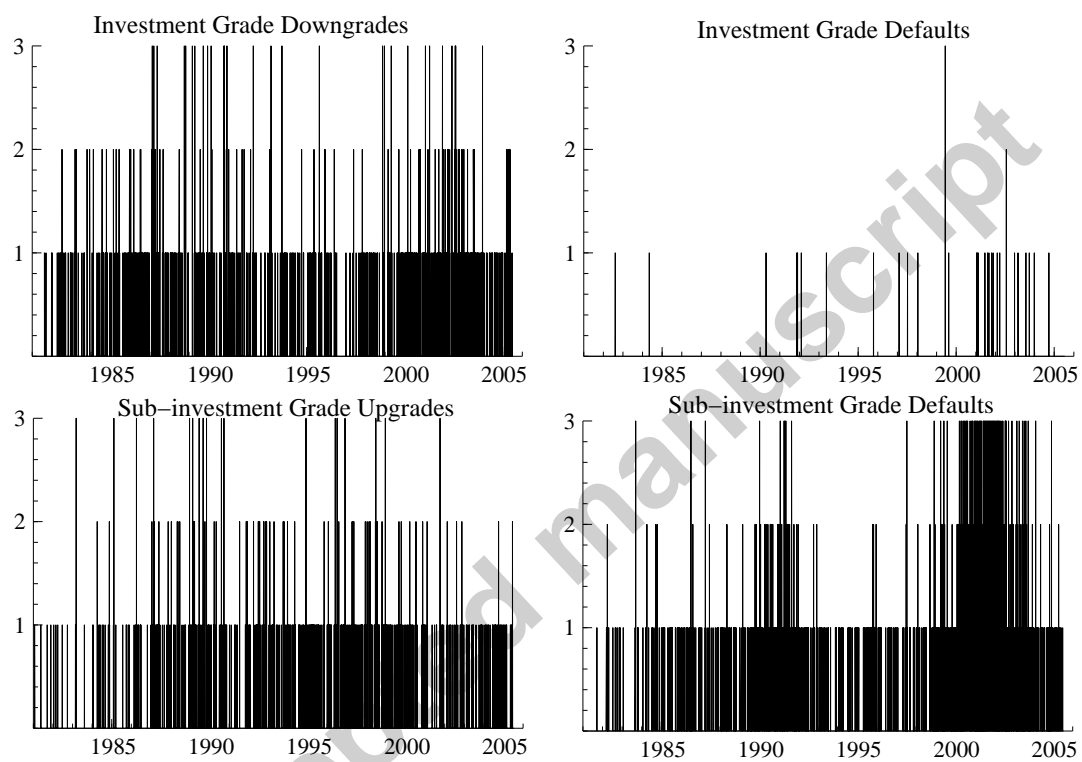
39

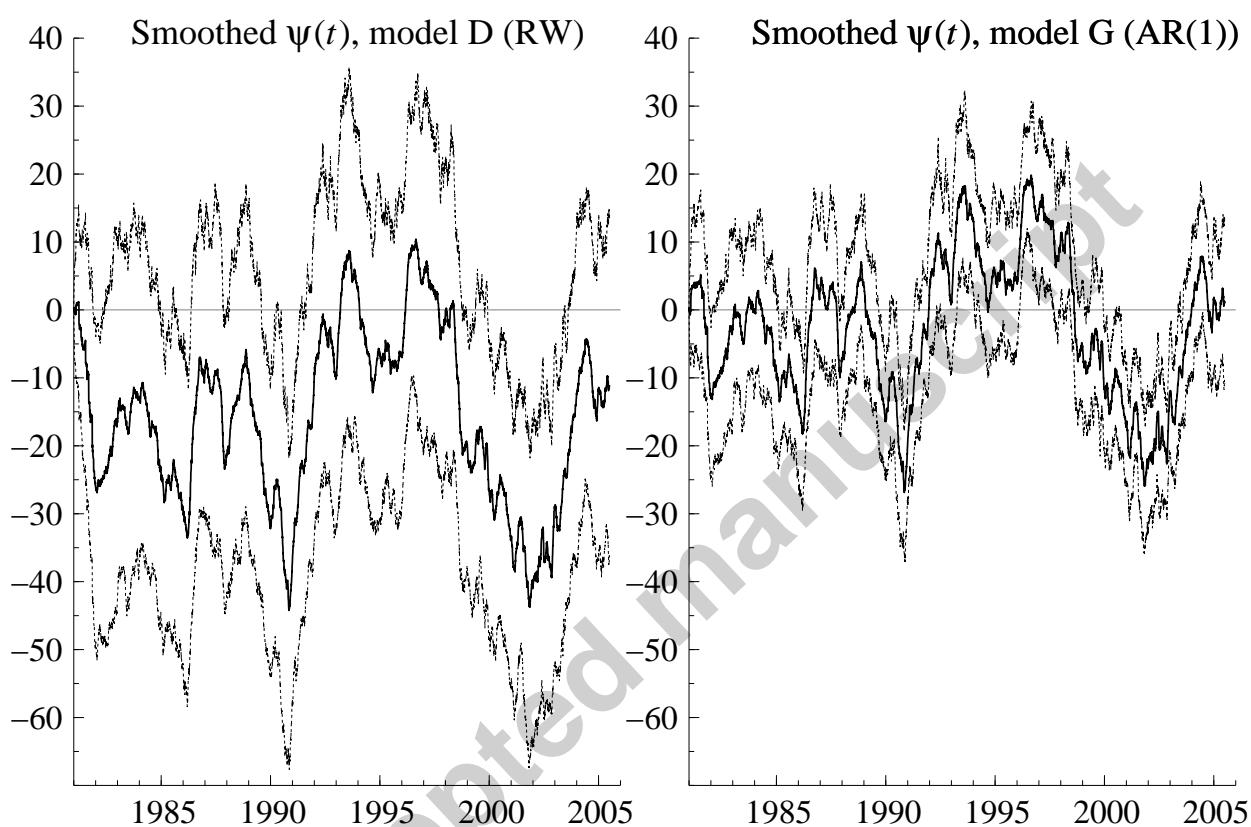Figure 2: Daily number of rating actions and recorded defaults

Figure 3: Smoothed credit cycle ψ_i with 95% confidence band for model D and G
The left-hand panel contains the smoothed estimate of ψ_i for a model with a single common random walk component (model D). The right-hand panel gives the smoothed estimate for a model with a single common AR(1) credit risk component (model G).
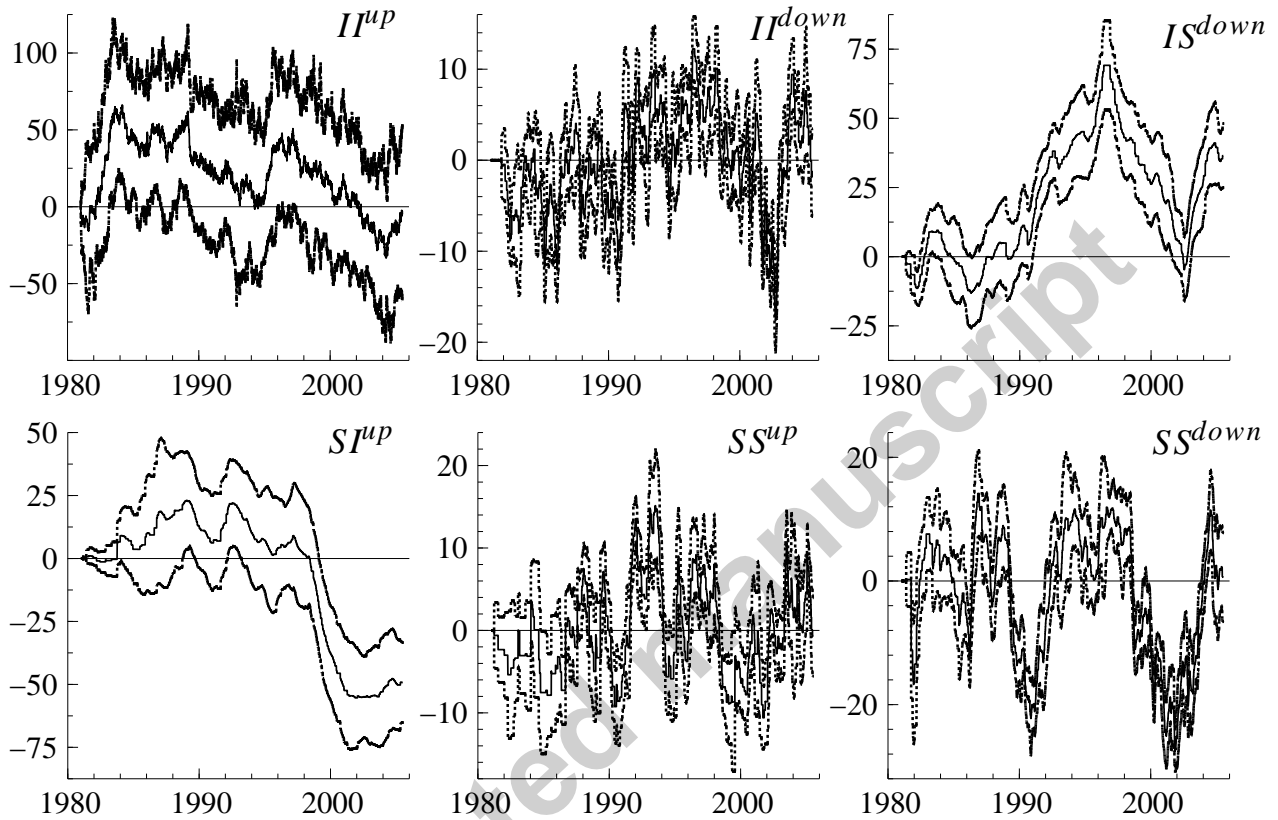
41

Figure 4: Smoothed credit cycles $\psi_s(t_i)$ with 95% confidence band for model I

This figure contains the smoothed estimates of $\psi_s(t_i)$ for a model with 6 independent AR(1) systematic credit risk factors. The 6 factors describe the systematic credit risk in upgrades from investment grade to investment grade (upper-left: II^{up}), downgrades from investment grade to investment grade (upper-middle: II^{down}), downgrades from investment grade to sub-investment grade and default (upper-right: IS^{down}), upgrades from sub-investment grade to investment grade (lower-left: SI^{up}), upgrades from sub-investment grade to sub-investment grade (lower-middle: SS^{up}), and downgrades from sub-investment grade to sub-investment grade and default (lower-right: SS^{down}).
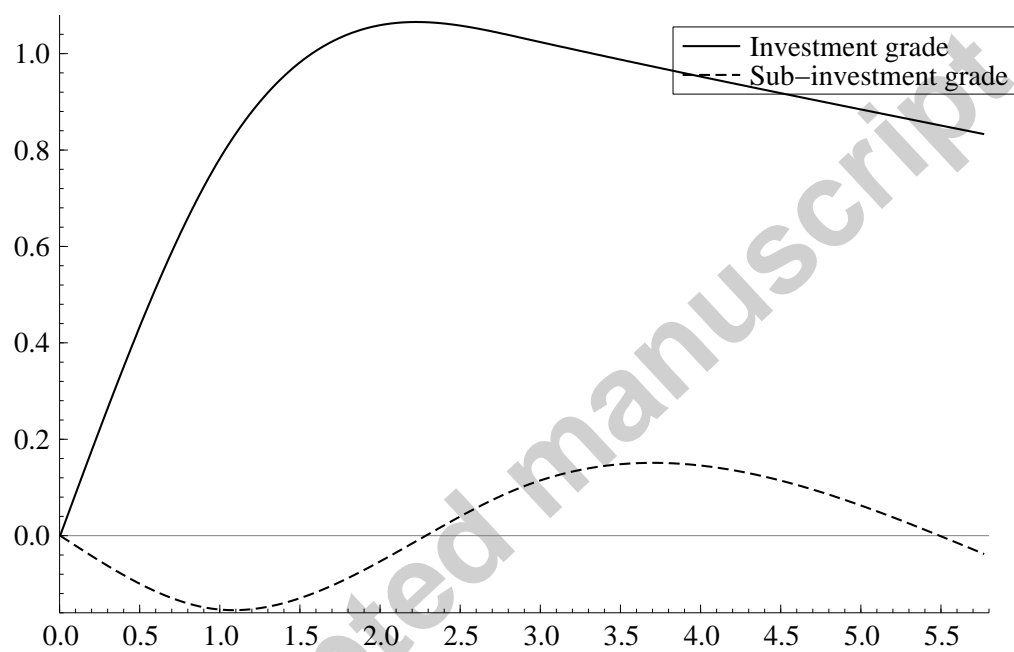
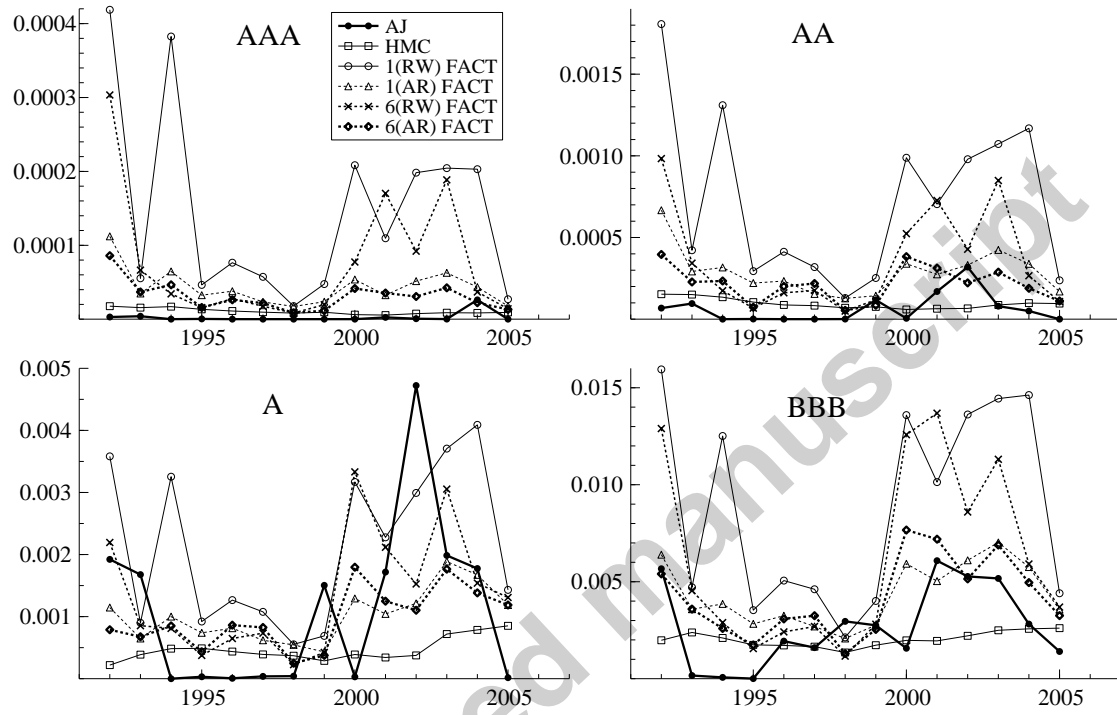Figure 5: Log baseline hazard functions

Figure 6: Forecasted Investment Grade default probabilities

Using a recursive state estimation procedure with extending data window from Dec 1980–June 1992 to Dec 1980–June 2005. Transition probabilities are estimated using the methodology of Section 4 for the MLFI model and the homogeneous continuous-time Markov chain (HMC) model. The MLFI specifications used from Table 4 are models D (1 common random walk (RW) factor), G (1 common autoregressive (AR) factor), E (6 independent RW factors), and H (6 independent AR factors). The Aalen-Johansen (AJ) estimates for each year are also plotted as a proxy for the observed transition rates. There are 4 plots for the rating categories AAA, AA, A, and BBB, respectively.
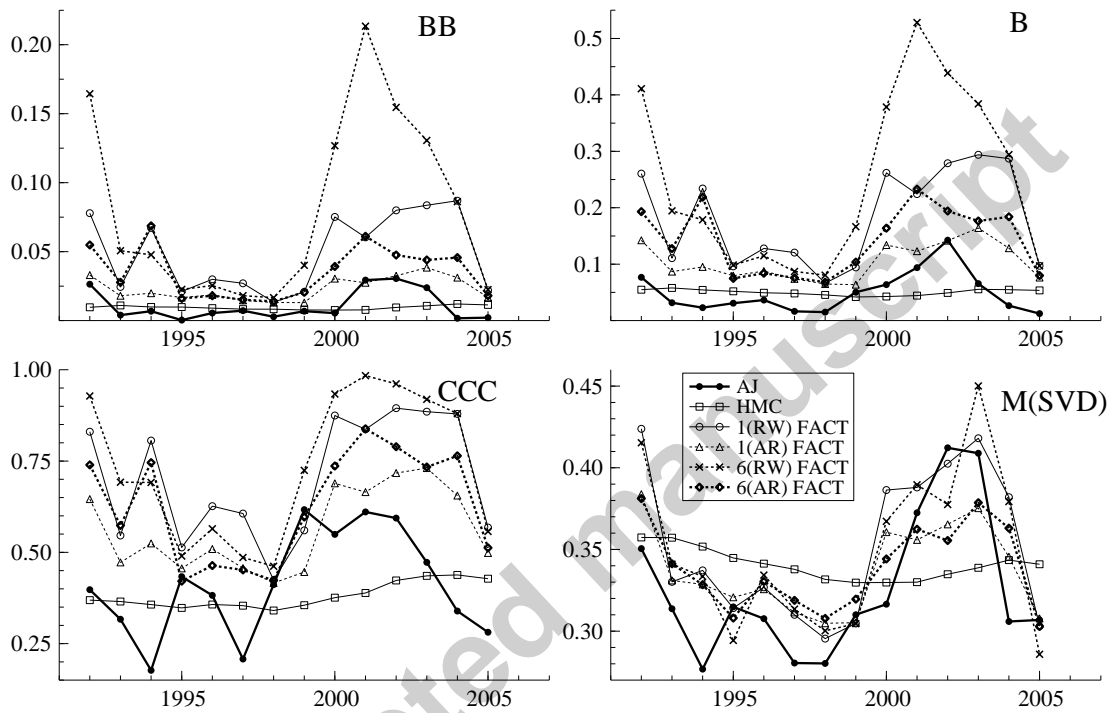
44

Figure 7: Forecasted Sub-Investment Grade default probabilities

See the note to Figure 6 for further explanation. There are 3 plots for the rating categories BB, B, and CCC, respectively. The lower right-hand plot provides the average singular value measure for transition matrices as proposed in Jafry and Schuermann (2004) as a general measure for the forecasting accuracy of transition matrices.