TI 2005-071/4
Tinbergen Institute Discussion Paper

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

Version July 2005

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.*

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

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

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*

July 1, 2005

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 between investment grade, subinvestment grade, and default ratings for U.S. corporates. The model strongly suggests the presence of a common dynamic component that can be interpreted as the credit cycle. We also show that the impact of this credit cycle is asymmetric with respect to downgrade and upgrade probabilities.

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

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

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

1 Introduction

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), and Couderc and Renault (2004) use a duration approach conditional on observed macro-variables and they 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.

Suggestions for dynamic models with latent components are Gagliardini and Gourieroux (2004), McNeil and Wendin(2004), 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. This 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 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 CreditPro6.2 data set from Standard & Poor's, containing all issuer ratings over the period 1981 – 2002. We classify firms as Investment grade or Subinvestment grade and specify a dynamic model for upgrades, downgrades, and defaults using all

available data. This yields a data set including almost 7000 firms and more than 4000 informative rating events. The estimation results lead to some interesting empirical findings. First, there is significant evidence of a persistent common component in rating transitions. We are able to recover this component from default and rating data at daily frequencies. 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.

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 our empirical illustration. 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

[1]We assume there are no simultaneous rating transitions. In practice the S&P's database is recorded at a

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

The model specification for intensities is given by

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

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 $N_{sk}(t)$.

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).

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 superscript 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 depends on the transition type $s \in \mathbb{S}$. This implies α_s can depend on both the origin and the destination state. In the empirical literature it is common practice to have α_s parameters that 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 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

[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).

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 1.8 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

$$\psi_i = \rho\psi_{i-1} + \varepsilon_i, \tag{2}$$

where ε_i is a set of i.i.d. $\mathrm{N}(0, \sigma^2)$ innovations and the AR parameter $\rho \in [-1, 1]$. More general dynamic specifications for ψ_i can be easily incorporated in the state space framework of the next subsection.

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

$$H_{sk}(t) = H_s(t - t_{0k}\ ,\ t - t_{1k}\ ,\ \ldots\ ,\ t - t_{N_k(t),k},) \tag{3}$$

where $t - t_{ik}$ denotes the backward-recurrence time of unit k with respect to its past ith transition moment. The function $H_s(\cdot)$ 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

$$H_s(x_0, \ldots, x_N) = \sum_{i=0}^{N} a_{si} x_i^{b_s - 1}, \tag{4}$$

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 ψ_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 downgrades and $\alpha_s > 0$ for upgrades. This specification is motivated by the empirical application of Section 6. Moreover, as the parameters α_s in (1) and σ in (2) are not simultaneously identified, we normalize σ to unity. Equivalently, one can restrict one α_s parameter to unity and estimate $\sigma > 0$.

For a vector of unknown parameters θ, the likelihood function conditional on initial conditions[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

$$L\left(\theta \mid \mathcal{F}_T, \Psi_{N(T)}\right) = \prod_{i=1}^{N(T)} \prod_{k=1}^{K} \prod_{s=1}^{S} \exp\left(Y_{sk}(t_i) \ln\{\lambda_{sk}(t_i)\} - R_{sk}(t_i) \int_{t_{i-1}}^{t_i} \lambda_{sk}(t) \mathrm{d}t\right), \tag{5}$$

where dummy variable $Y_{sk}(t)$ is one if unit k at time t experiences a rating event of type s, and zero otherwise. We note that $\mathcal{F}_T$ denotes the relevant observable filtration. The likelihood function (5) 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 event for that unit. 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.

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

$$\max_{\theta} L(\theta \mid \mathcal{F}_T), \tag{6}$$

where

$$L(\theta \mid \mathcal{F}_T) = \int L\left(\theta \mid \mathcal{F}_T, \Psi_{N(T)}\right) p(\Psi_{N(T)}) \mathrm{d}\Psi_{N(T)}, \tag{7}$$

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

[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).

3 Monte Carlo Maximum Likelihood Estimation

The main difficulty with maximum likelihood estimation in (7) is the computation of the high-dimensional integral. In a typical application such as the one in section 6, this integral is more than 4000 dimensional. McNeil and Wendin (2004) 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.

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 equation (7) we use a combination of importance sampling and the Kalman filter 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 (5). In particular, for the ith event time of the pooled process, we have the conditional log-density

$$\ln p(z_i|\psi_i, \mathcal{F}_{t_i^-}) = \sum_{s=1}^{S}\sum_{k=1}^{K} Y_{sk}(t_i) \ln\{\lambda_{sk}(t_i)\} - R_{sk}(t_i) \int_{t_{i-1}}^{t_i} \lambda_{sk}(t)\mathrm{d}t, \tag{8}$$

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

$$\lambda_{sk}(t) = R_{sk}(t_i) \cdot \exp\left(Z_{ski}\nu_i\right) \cdot H_{sk}(t), \qquad \text{for } t_{i-1} < t \leq t_i, \tag{9}$$

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 (9) can be equivalent, we take

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

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

$$\nu_i = T_i \nu_{i-1} + R_i \eta_i, \qquad \eta_i \sim \text{NIID}(0, Q_i), \qquad i = 1, \ldots, N, \tag{10}$$

with initial condition $\nu_0 \sim \text{N}(a, P)$. The vector a and the matrices T_i, 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 $a = 0$, $T_i = R_i = I$, $Q_i = 0$ and $P_i = \kappa I$, 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 $a = 0$, $T_i = \rho$, $R_i = 1$, $Q_i = \sigma^2$ and $P_i = \sigma^2(1 - \rho^2)^{-1}$. A combination of unknown coefficients and latent time series processes can be incorporated in (10) 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 $a = 0$,

$$T_i = \begin{bmatrix} I_S & 0 \\ 0 & \rho \end{bmatrix}, \quad R_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q_i = \sigma^2, \quad P = \begin{bmatrix} \kappa I_S & 0 \\ 0 & \sigma^2(1-\rho^2)^{-1} \end{bmatrix}.$$

A general framework for the MLFI model can be summarized by the observation log-density for z_i conditional on the vector ν_i. This is given by (8) where $\lambda_{sk}(t)$ is given by (9) for $t_{i-1} < t \leq t_i$, and where ν_i is modeled by (10) 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). A further complexity of the model is the highly multivariate nature of the variables in z_i. However the state space framework can easily accommodate this aspect of the model. A significant advantage of this framework is the incorporation of a selection of fixed and unknown coefficients in ν_i. The size of the parameter vector θ is therefore reduced. Since θ needs to be estimated via the numerical optimization of the likelihood, computation time is also reduced as a result. The estimation procedures are developed in the next subsection.

3.2 Monte Carlo likelihood evaluation

Given the statistical model specification of the previous subsection, the likelihood function (7) can be reformulated by

$$L(\theta \mid \mathcal{F}_T) = \int \left\{ \prod_{i=1}^{N} p\left(z_i \,|\nu_i, \mathcal{F}_{i-1}\right) \right\} p(\nu \mid \mathcal{F}_T) \mathrm{d}\nu, \tag{11}$$

where $p\left(z_i \,|\nu_i, \mathcal{F}_{i-1}\right)$ is given by (8) and the model for $\nu = (\nu_1', \ldots, \nu_N')'$ is implied by (10). Both $p\left(z_i \,|\nu_i, \mathcal{F}_{i-1}\right)$ and $p(\nu|\mathcal{F}_T)$ depend on the parameter vector θ for $i = 1, \ldots, N$. An analytical expression for (11) does not exist and therefore we rely on numerical techniques for the evaluation of (11). 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 $p(\nu|\mathcal{F}_T)$ denoted by $\nu^1, \ldots, \nu^M$ where M is a large number. Second, we compute the Monte Carlo estimator of (11) given by

$$\widehat{L}(\theta \mid \mathcal{F}_T) = M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \,|\nu_i^m, \mathcal{F}_{i-1}\right) \right\} \tag{12}$$

where ν_i^m is the ith element from ν^m. The estimator (12) 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 $p(\nu|z, \mathcal{F}_T)$, but this is not feasible since no analytical expression exists for this density. The idea of importance sampling is to replace $p(\nu|z, \mathcal{F}_T)$ by the more convenient Gaussian density $p_G(\nu|z, \mathcal{F}_T)$ for simulating ν's. The basic algorithm is then adjusted as follows. First, simulate M paths of ν from $p_G(\nu|z, \mathcal{F}_T)$ denoted by $\nu^1, \ldots, \nu^M$ where M is a large number. Second, compute the Monte Carlo estimator of (11) as given by

$$\begin{aligned} \widehat{L}(\theta \mid \mathcal{F}_T) &= M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \,|\nu_i^m, \mathcal{F}_{i-1}\right) \right\} \frac{p(\nu^m|\mathcal{F}_T)}{p_G(\nu^m|z, \mathcal{F}_T)} \\ &= p_G(z|\mathcal{F}_T) M^{-1} \sum_{m=1}^{M} \left\{ \prod_{i=1}^{N} p\left(z_i \,|\nu_i^m, \mathcal{F}_{i-1}\right) \right\} \frac{1}{p_G(z|\nu^m, \mathcal{F}_T)}, \end{aligned} \tag{13}$$

since $p_G(\nu|\mathcal{F}_T) = p(\nu|\mathcal{F}_T)$ and $p_G(\nu|z, \mathcal{F}_T) = p_G(z|\nu, \mathcal{F}_T)\ p_G(\nu|\mathcal{F}_T)/p_G(z|\mathcal{F}_T)$. We refer to this estimator as the Monte Carlo likelihood. The construction of $p_G(\nu|z, \mathcal{F}_T)$ 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

$$Y_{ski} = c_{ski} + Z_{ski}\nu_i + \xi_{ski}, \qquad \xi_{ski} \sim \text{NIID}\,(0, C_{ski}), \tag{14}$$

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 (10). Therefore, observation equation (14) and the dynamic latent process (10) 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 (14) 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 $\widehat{\nu}^{(0)}$. The linear Gaussian model (14) is constructed for $j = 0$ by

$$\begin{aligned} c_{ski} &= Y_{ski} - Z_{ski}\widehat{\nu}_i^{(j)} - C_{ski} Z_{ski} \nabla \ln p(z|\nu, \mathcal{F}_T)_i, \\ C_{ski} &= -\left[Z_{ski} \nabla^2 \ln p(z|\nu, \mathcal{F}_T)_i Z_{ski}'\right]^{-1}, \end{aligned} \tag{15}$$

where

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

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

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 (14) and (10). The conditional mean of ν can be computed by the Kalman filter and smoothing (KFS) algorithm. More formally, the KFS method computes $\mathrm{E}_G(\nu|Y)$ where $\mathrm{E}_G(\cdot)$ 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 (15) 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 now consists of (10) and (14), with joint density $p_G(\nu, z|\mathcal{F}_T)$ and where (15) 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 (13)

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 (13) 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 (14). Further, the observation density of interest $p(z_i|\nu_i, \mathcal{F}_{i-1})$ is given by equation (8) 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 $p_G(\nu|z, \mathcal{F}_T)$.

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

$$\hat{\nu}_{i|N} = \left(\sum_{m=1}^{M} w_m \times \nu_i^m \right) \Bigg/ \left(\sum_{m=1}^{M} w_m \right), \tag{16}$$

where

$$w_m = \left\{ \prod_{i=1}^{N} p\left(z_i \,|\nu_i^m, \mathcal{F}_{i-1}\right) \right\} \Bigg/ p_G(z|\nu^m, \mathcal{F}_T). \tag{17}$$

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

$$\left[\left\{ \sum_{m=1}^{M} w_m \times (\nu_i^m)^2 \right\} \Bigg/ \left(\sum_{m=1}^{M} w_m \right) \right] \; - \; (\hat{\nu}_{i|N})^2. \tag{18}$$

4 Implied Short-term Transition Matrices

We now turn to the issue of estimating the short-term 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.[6] 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 the product integral[7]

$$P_k\left(T, T+\Delta\right) = \prod_{T}^{T+\Delta} \left(I_S + G_k(t)\mathrm{d}t\right). \tag{19}$$

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. Note that the methodology for computing $P_k(T, T+\Delta)$ based

[6] 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.

[7] 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 $T = t_0 < t_1 < \ldots < t_n = T+\Delta$.

on parametric specifications of $G_k(t)$ is also valid if duration and self-excitation effects are present. This is particularly interesting in the light of the empirical evidence in for example Kavvathas (2001) and Lando and Skødeberg (2002). In that case the individual entries of the matrix $G_k(t)$ are stochastic processes. We assume 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 $[T, T+\Delta]$,

$$\bar{P}_k\left(T, T+\Delta\right) = \mathrm{E}\left[P_k\left(T, T+\Delta\right) \mid \mathcal{F}_T\right] = \mathrm{E}\left[\prod_T^{T+\Delta} \left(I + G_k(t)\mathrm{d}t\right) \middle| \mathcal{F}_T\right]. \tag{20}$$

We propose a parametric bootstrap procedure for evaluating the conditional expectation in (20). 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.[8] A consistent estimator for $\bar{P}_k(T, T+\Delta)$ is given by

$$\hat{\bar{P}}_k\left(T, T+\Delta\right) = \frac{1}{M}\sum_{m=1}^{M} \prod_T^{T+\Delta} \left(I + G_k^m(t)\mathrm{d}t\right), \tag{21}$$

where $G_k^m(t)$ denotes unit k's realized matrix of intensities for replication m.

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. Each of these factors can then be evaluated separately by an appropriate truncation of the corresponding Péano series expansion,

$$\prod_{t_{i-1}^m}^{t_i^m} \left(I + G_k^m(t)\mathrm{d}t\right) = I + \sum_{p=1}^{\infty} \int \ldots \int_{t_{i-1}^m \leqslant s_1 < \ldots s_p \leqslant t_i^m} G_k^m(s_1)\mathrm{d}s_1 \ldots G_k^m(s_p)\mathrm{d}s_p,$$

see Andersen et al. (1993, p. 91). Due to the parametric assumption for $G_k(t)$, the multiple integral appearing as the general term of this series may in some cases be evaluated analytically. 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,

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

[8] 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.

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.

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 the model specification of the empirical study in Section 6. The details of the simulation design and results are presented in this section.

We consider three states. The states can be interpreted as investment grade, subinvestment grade, and default. Default is an absorbing state. The intensities are specified as

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

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. Another source of heterogeneity is the latent process $\psi(t)$ that can be interpreted as the macroeconomic effect. The benchmark model in this simulation exercise abstracts from duration dependence by setting $H_{sk}(t) \equiv 1$. Further parsimony is introduced by setting $\alpha_s = -\alpha < 0$ for downgrades, and $\alpha_s = \alpha > 0$ for upgrades. The parameter values used for the simulation can be found in the top panel of Table 1. Experiments with different parameter values yielded qualitatively similar results.

<INSERT TABLE 1 ABOUT HERE>

We simulate panels of different sizes, from 100 units up to 500 units. Each panel is generated as follows. At time $t_0 = 0$, the sample contains an equal number of firms in state 1 (investment grade) and state 2 (subinvestment grade). The unobserved process $\psi(t)$ is initialized by drawing $\psi_1 = \psi(t_1)$ from its unconditional distribution (the standard normal). 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

$$\lambda^*(t_i) = \sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_{sk}(t_i), \tag{22}$$

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

$$\pi_k(t_i) = [\lambda^*(t_i)]^{-1} \sum_{s=1}^{S} \lambda_{sk}(t_i), \qquad k = 1, \ldots, K. \tag{23}$$

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

$$\pi_{sk}(t_i) = \left[\sum_{s=1}^{S} \lambda_{sk}(t_i) \right]^{-1} \lambda_{sk}(t_i), \tag{24}$$

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 $\rho = 0.9$ 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. 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 Figure 1 and in the lower panels of Table 1. We observe that the parameters are recovered with a high degree of accuracy. This already holds for a moderate panel size K. Note that the size of K in our empirical application in Section 6 is much larger, that is $K \approx 7000$. The Monte Carlo standard errors decrease if we increase the size of the panel. This holds for all parameters. Note that if we increase the number of units in the panel, we simultaneously increase the time series dimension of the panel, as it takes longer (in expectation) for all firms to enter the absorbing default state.

<INSERT FIGURE 1 ABOUT HERE>

Table 1 and Figure 1 illustrate that the Monte Carlo likelihood procedure provides accurate parameter estimates for the model at hand. It is also interesting to see how the methodology performs in retracing the common factor $\psi(t)$ from the data. As an illustration, we take a single 'representative' simulation. The method for computing the smoothed estimate of the state vector described in Section 3 (Step 3) is used for this purpose. The smoothed estimate of $\psi(t_i)$ is presented in Figure 2. The algorithm clearly performs satisfactory in recovering the characteristics of the true, unobserved $\psi(t)$ process from the (simulated) data.

<INSERT FIGURE 2 ABOUT HERE>

6 Empirical Results

A baseline MLFI model without explanatory covariates, $w_{ks}(t) \equiv 0$, and dependence on the backward recurrence times, $H_{sk}(t) \equiv 1$, is considered for a database of ratings that covers a period of more than 23 years. The simplification of the modeling framework implies that the durations in the pooled process follow (non i.i.d.) exponential distributions. Due to the presence of a common component, the duration until a default or rating migration for an *individual* firm then follows a convolution of a (random) number of exponential distributions, see Andersen et al. (1993).

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 6.2 database. The sample period is from the end of 1980 (the left-censoring time point) until the start of 2003, covering a total of 8035 days. 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 (5). The rating histories in the data set distinguish between more than 18 different rating classes. To illustrate our methodology, we consider only two broad classes, namely investment grade (AAA down to BBB-) and subinvestment grade firms (BB+ and lower).

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.

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 of 1.85 days with a standard deviation of 0.006. There is a large number of downgrades and upgrades. The number of transitions from investment to subinvestment grades (downgrades) is 773. Vise versa (upgrades) it is 579. The number of transitions from subinvestment grade to a default state is 835 while the number of defaulting investment grade firms is only 7. This small number of 7 defaults limits the precision of estimates of the default intensity for investment grade firms.

<INSERT FIGURE 3 ABOUT HERE>

The first two plots in Figure 3 show the number of downgrades and defaults on a daily basis since December 31, 1980, respectively. We can see that downgrades and defaults tend to cluster

in time. This can be easily seen by noticing the concentration of vertical lines, which originate dark and bright areas along the horizontal (time) axis. The lower plot in Figure 3 contains the number of upgrades on any given day. Interestingly, these three plots complement each other. When downgrades and defaults cluster, upgrades are scarce, and vise versa. This suggests that the model specification used with a single common risk factor $\psi(t)$ and the sign restrictions on the factor loadings α_s already captures the most salient features of the data.

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 specification has $\lambda_{sk}(t) = \exp(\eta_s)$ and the MLFI model reduces simply to a 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

$$\hat{\eta_s} = \ln\left(\sum_{i=1}^{N}\sum_{k=1}^{K} Y_{ski}\right) - \ln\left(\sum_{i=1}^{N}\sum_{k=1}^{K} t_i R_{ski}\right). \tag{25}$$

Table 2 presents the parameter estimates based on (25) as well as the estimates obtained by using the Monte Carlo methods of Section 3. As expected, the parameter estimates are almost identical, reflecting the accuracy of the MC likelihood. The largest differences are for the estimate of the default intensity of investment grade firms and for its standard error. However, even these differences are marginal. If anything, the confidence intervals of the Monte Carlo ML estimator appear slightly larger, making the inference procedure conservative in the sense that it becomes more difficult to establish statistical significance. The estimated log-likelihood using the Monte Carlo procedure is also accurate and almost exactly coincides with the ML value.

<INSERT TABLE 2 ABOUT HERE>

6.3 Estimation results for the MLFI model

In this empirical illustration, we introduce the common component ψ_i as a random walk, i.e., with the AR parameter ρ set to unity. The common factor ψ_i is meant to capture changes in general business conditions that are typically caused by the credit or business cycle. Such changes evolve gradually over the years. The observations, on the other hand, are made at a high-frequency. The average duration of a spell of the pooled process is less than 2 days, see Subsection 6.1. To align the high-frequency nature of the data with the low-frequency

characteristics of ψ_i, a priori imposing the unit root on the AR specification appears natural. Otherwise, the ψ_i component might also pick up fluctuations at a higher frequency, which are not of prime interest in our current application. Due to the restriction $\rho = 1$, $\psi(t_1)$ and η_s for $s = 1, \ldots, S$ 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), we set $\alpha_s < 0$ for downgrades and $\alpha_s > 0$ 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.

Four different specifications for the MLFI model with respect to α_s are considered. The basic MLFI specification is with a single coefficient α_s per rating class. In the empirical illustration this means that α_s differs across the two states, Inv and Sub. The estimation results for this specification are presented in the upper-right panel of Table 3.

<INSERT TABLE 3 ABOUT HERE>

The inclusion of a common risk component leads to a huge increase of the likelihood value. The difference between the log-likelihood values that are presented in Table 2 and in the upper-right panel of Table 3, is around 174 points. This is statistically significant at any conventional level. The estimated loadings α_s of the common factor for investment and subinvestment grade firms show that the differences are small. We therefore also consider a model where the common factor loading does not depend on the initial rating class. The estimation results are presented in the upper-left part of the table. By comparing the parameter estimates for the two specifications in the top-panels, it is clear that the estimated parameters are not significantly different. This is confirmed by the likelihood values, which are also approximately equal.

The above result, however, hinges on the assumption of a single sensitivity parameter per rating class. 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). To test for this phenomenon, the α_s parameters are allowed to depend both on the input rating and the output rating. The estimation results are in the lower-left panel of Table 3. Note that we still impose the sign restrictions on α_s for upgrades and downgrades. The log-likelihood value increases by 16.6 points as a result at the expense of two additional parameters. This leads to an overall statistically significant improvement at the 1% level. For investment grade firms, the common factor sensitivity of downgrades appears much higher than for defaults. The sensitivity

for default is even estimated as zero. This is due to the sign restriction imposed on the loadings. The investment grade common factor sensitivity for a downgrade to subinvestment is -0.039 while it is -0.047 for a downgrade from subinvestment grade to default. These estimates are not similar to the upper-panel results where α_s is not different for different destination states. In the case of upgrades from subinvestment to investment grades, the estimated common factor loading is 0.017 and is less than 40% of the loading for downgrades -0.047, in absolute value. Upgrades appear to be less driven by common risk than downgrades. This result is in line with the results of Das et al. (2002) and Kavvathas (2001) based on observed macro-economic variables as proxies for the common risk factor.

Finally, we investigate whether imposing the sign restrictions on the factor loadings α_s causes significant distortions in the estimation procedure. The lower-right panel of Table 3 presents the results for unrestricted α_ss, i.e., where we only impose $\alpha_{Sub\to Dflt} < 0$ to establish identification. Comparing the log-likelihood values of the models in the lower panels clearly reveals that there is no significant difference. The estimated sensitivity of the default intensity for investment grade firms is estimated at the positive value 0.008. However, its confidence interval clearly indicates that the estimated value is not significantly different from zero.

The smoothed estimates of ψ_i can be obtained using the Monte Carlo methods of step 3 in Section 3. These estimates are based on the parameter estimates of the lower-right panel of Figure 4. The resulting estimates of ψ_i, denoted by $\widehat{\psi}_i$ for $i = 1, \ldots, N$, show clear troughs in the early and middle 80's, early and late 90's and early 2000's. This is consistent with the empirical dynamic patterns found in the empirical finance literature. These smoothed estimates confirm the salient features in the data as presented in Figure 3. The peaks in $\widehat{\psi}_i$ correspond to periods with clusters of upgrades while defaults and downgrades are relatively scarce in such periods. Conversely, in periods with many defaults and downgrades, the $\widehat{\psi}_i$ is relatively small. This feature corresponds with the interpretation of ψ_i as a common risk component in rating transitions.

<INSERT FIGURE 4 ABOUT HERE>

6.4 Forecasting rating transition probabilities

To assess the impact of fluctuations in the estimated latent risk component for the implied 1-year transition/default probabilities, a recursive out-of-sample forecasting exercise is carried out. The empirical results of this study are reported below. The details of the forecasting study are as follows. First, a data window from end 1980 to end 1990 is considered for the estimation of the parameters η_s and α_s together with the latent factor ψ_i. Next, a forecast for

the one-year transition probability matrix over the year 1991 is obtained from the MLFI model using the parametric bootstrap method described in Section 4. We can also use the estimates of the (constant) intensities from the HCTMC model over the estimation window to obtain an implied one-year transition matrix. This historical transition probability matrix can then be used as a naive estimate of the future transition rates. In order to assess the performance of these two alternative forecasting methods we compute the nonparametric Aalen Johansen (AJ) estimator of the transition probabilities over the year 1991. This empirical transition matrix can be seen in the current context as a proxy for the 'true' or realized[9] transition probabilities, see Lando and Skødeberg (2002).

This exercise is repeated by consecutively enlarging the sample by one year until end 2001. For each sample, two sets of transition probabilities for the next year are computed: the parametric HCTMC estimates over the (increased) estimation window and the bootstrapped MLFI forecasts. The resulting forecasted probabilities for the years 1991 to 2002 are presented in Figure 5 against the actual 1-year AJ estimates over each of these years.

<INSERT FIGURE 5 ABOUT HERE>

The *ex-ante* forecasts of the MLFI model resemble the nonparametric *ex-post* AJ estimates closely. Although the HCTMC model does not include a dynamic component, the forecasts still adapt over the years due to the recursive nature of the procedure. The changes are nevertheless insufficient. The forecast results in Figure 5 show that the increased flexibility of the dynamic latent factor in the MLFI model leads to more realistic forecasts. This can be seen by comparing both forecasts with the 'realized' transition probabilities as given by the nonparametric AJ historical estimates.

For investment grade firms, the MLFI model appears somewhat more conservative than the HCTMC model in that predicted default probabilities are higher. In particular this is the case when observed default frequencies are high during the previous year. The MLFI forecast and the AJ estimates are also quite similar for default probabilities of subinvestment grade firms. In the case of subinvestment firms, the MLFI model misses the large number of upgrades in the early nineties. On the other hand, it overpredicts the upgrade probabilities in the late 1990s and the early years of the 2000s. In both cases, however, the forecasts can be regarded as more prudent while still responding faster to the recently observed events as compared to those of the HCTMC model. The prudent character of the MLFI forecasts needs to be studied from a risk

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

management perspective in more detail. For example, the dynamic specification of the model can be extended, duration dependence can be included and heterogeneity can be accounted for by incorporating observed covariates in the model. This is left for future research.

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 is 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).

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, 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. Although the current specification with a single latent factor is motivated by economic intuition, generalizations toward multiple risk factors are straightforward. Furthermore it is possible to estimate and test formally the number of latent factors driving the default and rating migration intensities. To set up and use 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 in Section 6. As a final outlook, the MLFI model may also provide a useful benchmark in modeling prices of defaultable securities and credit risk premia.

The authors thank Jaap Abbring, Luc Bauwens, Arnaud De Servigny, Pieter Klaassen, Michel Mouchart, Olivier Renault, and seminar participants at the Vrije Universiteit Amsterdam and Tinbergen Institute for helpful comments. Financial support from Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), Vrije Universiteit Amsterdam and the Certified Financial Analyst executive program at the Vrije Universiteit Amsterdam is gratefully acknowledged. The rating transition data for this research was generously supplied by Standard and Poor's.

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.

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.

de Jong, P. and N. Shephard, 1995, The simulation smoother for time series models. *Biometrika 82*, p 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, 2005, A non-Gaussian panel time series model for estimating and decomposing default risk. *Working paper - Vrije Universiteit Amsterdam.*

Koopman, S.J., Lucas, A. and P. Klaassen, 2005, Empirical credit cycles and capital buffer formation. *Journal of Banking & Finance*, forthcoming.

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, 2004, Dependent credit migrations. *Working paper - NCCR FINRISK*

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

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

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

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 baseline 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, $s = 1, \ldots, 6$. The η_s and α_s estimates are presented as matrices, the row of the matrix representing the origin, and the column representing the destination state. The common factor $\psi(t)$ follows an AR(1) process with AR parameter ρ that only jumps at the event times t_i. The true parameter values are given in the top panel of the table. The bottom three panels contain simulation averages and standard errors of parameter estimates based on 500 replications for different numbers of cross-sectional units K.

K		η			α			ρ
	Origin	Destination state			Destination state			
	state	1	2	3	1	2	3	
True	1	____	-4.00	-5.00	____	-1.00	-1.00	0.90
Values	2	-4.50	____	-3.50	1.00	____	-1.00	
100	1	____	-3.94	-4.96	____	-1.07	-1.07	0.87
			(0.34)	(0.38)		(0.24)	(0.24)	(0.07)
	2	-4.62	____	-3.47	1.07	____	-1.07	
		(0.38)		(0.36)	(0.24)		(0.24)	
200	1	____	-3.94	-4.95	____	-1.02	-1.02	0.89
			(0.23)	(0.26)		(0.14)	(0.14)	(0.04)
	2	-4.55	____	-3.46	1.02	____	-1.02	
		(0.26)		(0.24)	(0.14)		(0.14)	
500	1	____	-3.97	-4.97	____	-1.00	-1.00	0.90
			(0.16)	(0.18)		(0.09)	(0.09)	(0.02)
	2	-4.48	____	-3.47	1.00	____	-1.00	
		(0.17)		(0.16)	(0.09)		(0.09)	

Table 2: Parameter estimates of the HCTMC model

This table presents estimates of a homogeneous continuous-time Markov chain (HCTMC) model with intensities $\lambda_{sk}(t) = R_{sk}(t) \exp(\eta_s)$. The left-hand panel presents the results for the Monte Carlo maximum likelihood procedure as discussed in Section 3. The right-hand panel presents the closed-form maximum likelihood (ML) estimates from (25). The rows denote the input rating: Inv for investment grade and Sub for subinvestment grade. The columns contain the output ratings: Inv, Sub or Dflt, the latter for default. Optimization is performed using 100 importance samples. Computation of standard errors is based on 1,000 importance samples.

		Monte Carlo ML			ML		
		Inv	Sub	Dflt	Inv	Sub	Dflt
η	Inv	____	-3.77	-8.54	____	-3.77	-8.48
			[-3.84,-3.70]	[-9.32,-7.76]		[-3.84,-3.70]	[-9.23,-7.72]
	Sub	-3.38	____	-3.02	-3.38	____	-3.01
		[-3.46,-3.30]		[-3.09,-2.96]	[-3.46,-3.30]		[-3.08,-2.95]
Log-lik			-9644.0			-9643.9	

Table 3: Parameter estimates of the MLFI model

This table contains the parameter estimates of the baseline MLFI model,

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

for $k = 1, \ldots, K$ with K the number of firms, $s \in \mathbb{S} = \{Inv \to Sub, Inv \to Dflt, Sub \to Inv, Sub \to Dflt\}$, with Inv, Sub, and Dflt for investment grade, subinvestment grade and default, respectively. The common factor $\psi_i = \psi(t_i)$ follows a random walk process with $\psi_0 = 0$ for $i = 1, \ldots, N$. The numbers in square brackets denote the 95% confidence intervals. Four model specifications are estimated with four different specifications for α_s. In the upper-left panel, $\alpha_s = \alpha > 0$ for upgrades and $\alpha_s = -\alpha$ otherwise. In the upper-right panel, the magnitude of α_s only depends on the origin state (the initial rating). In the lower-left panel, $\alpha_s \geq 0$ for upgrades, and $\alpha_s \leq 0$ for downgrades. In the lower-right panel, we only restrict $\alpha_{Sub \to Dflt} < 0$. Estimation is carried out by the Monte Carlo methods of Section 3 and based on 100 importance samples for each likelihood evaluation. Standard errors are computed using 1,000 samples.

		Single α			Row-wise restricted αs		
		Inv	Sub	Dflt	Inv	Sub	Dflt
η	Inv	____	-4.12	-8.91	____	-4.13	-8.91
			[-4.62,-3.63]	[-9.91,-7.91]		[-4.65,-3.61]	[-9.92,-7.90]
	Sub	-3.19	____	-3.42	-3.20	____	-3.39
		[-3.68,-2.69]		[-3.92,-2.92]	[-3.69,-2.72]		[-3.88,-2.90]
α	Inv	____	-0.034	-0.034	____	-0.036	-0.036
			[-0.054,-0.021]	[-0.054,-0.021]		[-0.060,-0.021]	[-0.060,-0.021]
	Sub	0.034	____	-0.034	0.033	____	-0.033
		[0.021,0.054]		[-0.054,0-.021]	[0.021,0.054]		[-0.054,-0.021]
Log-lik			-9470.2			-9470.0	

		Sign restricted αs			Unrestricted αs		
		Inv	Sub	Dflt	Inv	Sub	Dflt
η	Inv	____	-4.15	-8.55	____	-4.15	-8.50
			[-4.71,-3.58]	[-9.33,-7.76]		[-4.71,-3.58]	[-9.27,-7.73]
	Sub	-3.28	____	-3.56	-3.28	____	-3.56
		[3.54,-3.02]		[-4.24,-2.87]	[-3.54,-3.02]		[-4.24,-2.87]
α	Inv	____	-0.039	0.000	____	-0.038	0.008
			[-0.065,-0.023]	[-1000,0.000]		[-0.058,-0.019]	[-0.068,0.052]
	Sub	0.017	____	-0.047	0.017	____	-0.047
		[0.010,0.031]		[-0.075,-0.029]	[0.007,0.027]		[-0.075,-0.029]
Log-lik			-9453.4			-9453.4	

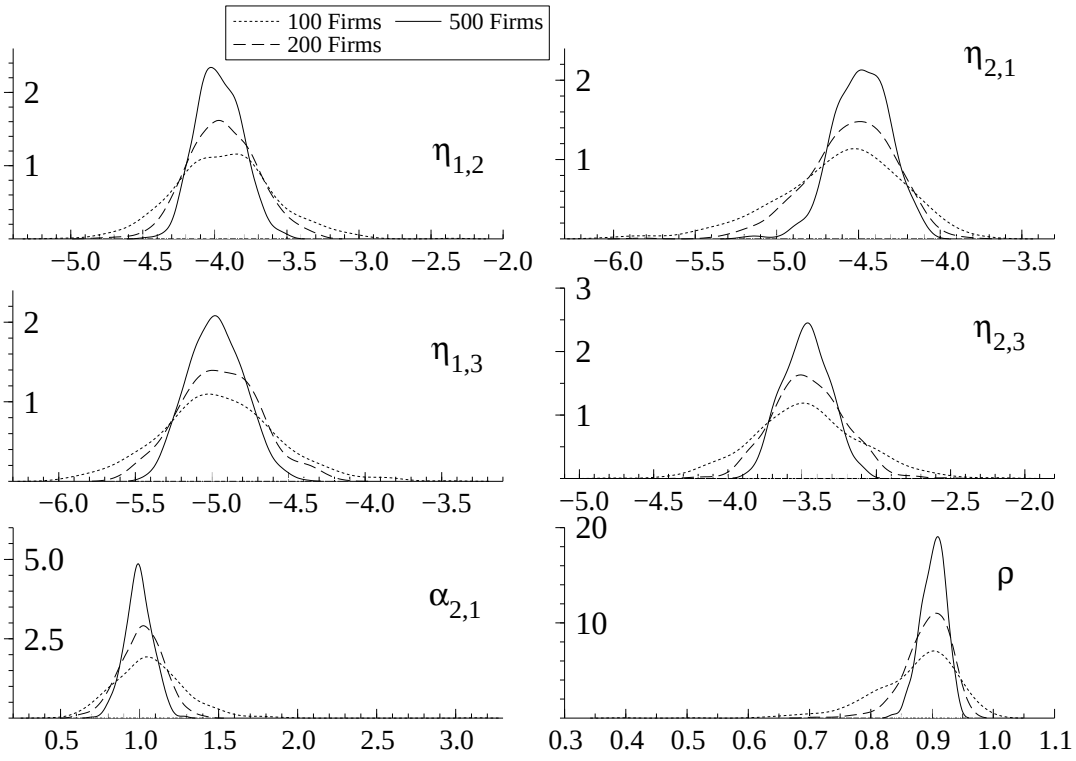

Figure 1: Empirical distributions of SML estimators for the baseline MLFI model

The baseline model and the simulation set-up are the same as explained in the note to Table 1.

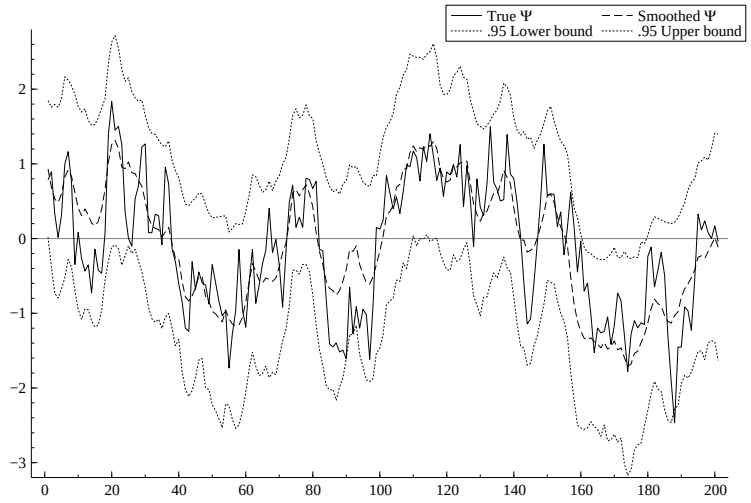

Figure 2: True versus smoothed estimate of $\psi(t)$

The baseline model and the simulation set-up are the same as explained in the note to Table 1.

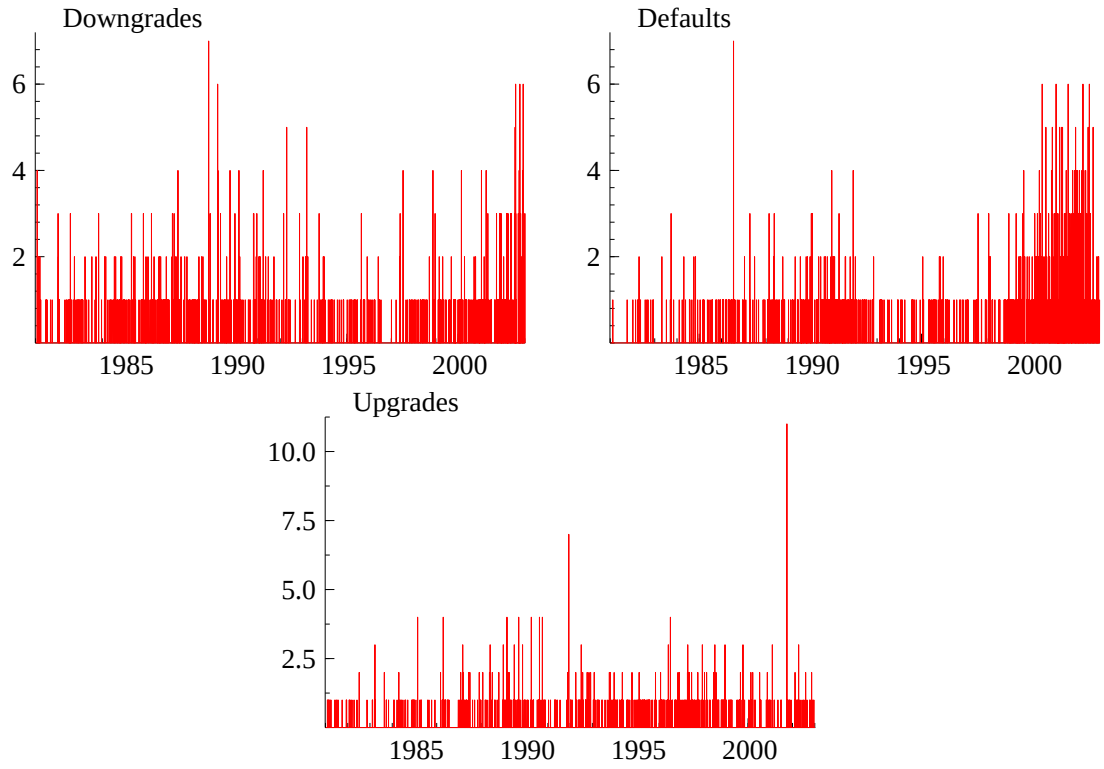

Figure 3: Daily number of rating actions and recorded defaults

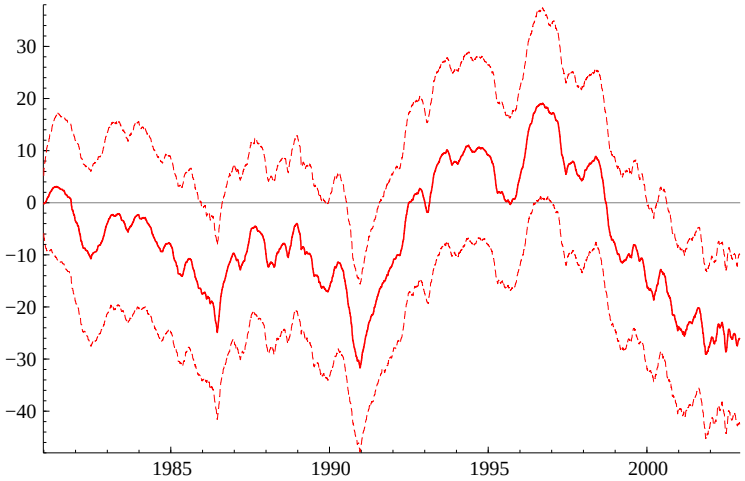

Figure 4: Smoothed credit cycle ψ_i with 95% confidence band

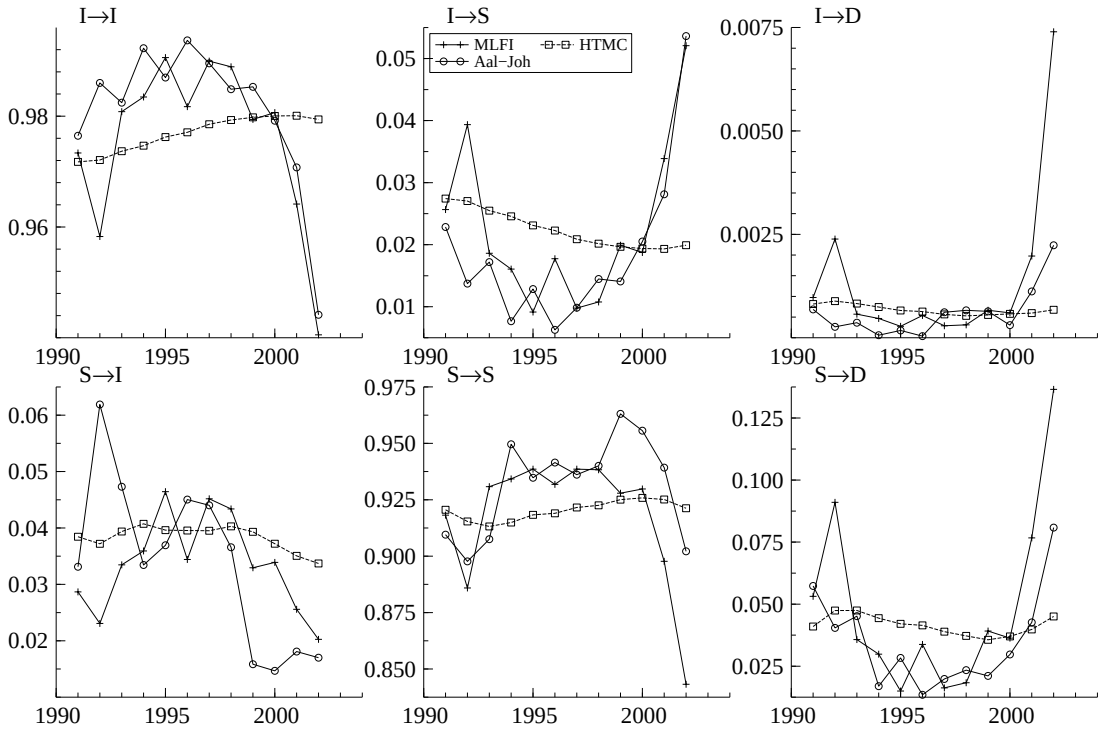

Figure 5: Forecasted transition probabilities

Using a recursive model and state estimation procedure with extending data window from Dec 1980–Dec 1990 to Dec 1980–Dec 2001. Transition probabilities are estimated using the methodology of Section 4 for the MLFI model and the homogeneous continuous-time Markov chain (HCTMC) model. The Aalen-Johansen (AJ) estimates for each year are also plotted as a proxy for the observed transition rates. The upper row of plots give the probabilities from the origin state Investment grade (I) to Investment grade (upper-left), Subinvestment grade (upper-middle), and Default (upper-right). The lower row of plots is similar, but for the origin state of Subinvestment grade (S).