

Estimation methods for complex models with latent factors

Importance sampling and MCMC methods

Application to credit risk modelling

Siem Jan Koopman

joint work with Borus Jungbacker, André Lucas & André Monteiro

VU University Amsterdam

Tinbergen Institute

- - -

Statistical Modelling of Complex Systems, Munich, 2006

- - -

Modelling Credit Rating Transitions

- Ratings play a prominent role in the credit industry: a simple qualitative classification of the solidity, solvency and prospects of a debt issuer. Note the importance of new regulatory framework known as Basel II where ratings determine bank's capital buffer.
- A good understanding of dynamics in ratings and changes is important from regulatory and financial industry perspectives.
- Literature on modeling credit events, defaults and rating changes: Wilson (1997a,b) uses logistic regressions with macroeconomic explanatory variables; Nickell et al (2000) and Bangia et al (2002) show that upgrade, downgrade, and default probabilities differ over different economic regimes; in the same spirit, Kavvathas (2001), Carling et al (2002) and Couderc and Renault (2004) use a duration approach conditional on observed macro-variables, average times-to-default decrease if economic activity decreases. Koopman and Lucas (2005) and Koopman, Lucas and Klaassen (2005) identify a time-varying cycle for default rates over a long historical period, see also Fledelius et al (2004).

Methodology of estimation

The motivation is to estimate non-Gaussian nonlinear time series model by classical and Bayesian methods.

In many cases, as for credit risk modelling, much data is at hand.

Therefore efficient estimation methods are needed;

Bayesian methods such as MCMC and particle filtering tend to be quite slow in practice.

On the other hand, IS **does not always “work”**. However, we have diagnostics to check the validity of the IS methods.

But when it works, it works and it works **fast**.

This presentation gives a general treatment and reports some recent developments on the subject. Results are relevant for IS and for MCMC methods.

Model for the state vector: linear Gaussian

The linear Gaussian state process is defined as

$$\alpha_{t+1} = d_t + T_t \alpha_t + \eta_t, \quad \eta_t \sim NID(0, Q_t), \quad t = 1, \dots, n,$$

where system vector d_t and system matrices T_t and Q_t are fixed and known for $t = 1, \dots, n$.

The initial state vector is normally distributed with mean a and variance matrix P , that is $\alpha_1 \sim N(a, P)$.

The disturbances η_t ($t = 1, \dots, n$) are serially independent and are independent of the initial state vector.

A vector stochastic process with these properties is defined as a *linear Gaussian state process*.

Model for the state vector: linear Gaussian

The *linear Gaussian state process* can be expressed by the multivariate normal density $\alpha \sim N(d, \Omega)$, where $\alpha = (\alpha'_1, \dots, \alpha'_n)'$ and

$$d = T (a', d'_1, \dots, d'_{n-1})', \quad \Omega = T \text{diag}(P_1, Q_1, \dots, Q_{n-1}) T',$$

and

$$T = \begin{bmatrix} I & 0 & 0 & \cdots & 0 & 0 \\ T_1 & I & 0 & \cdots & 0 & 0 \\ T_2 T_1 & T_2 & I & & 0 & 0 \\ & & & \ddots & & \vdots \\ T_{n-2} \cdots T_1 & T_{n-2} \cdots T_2 & T_{n-2} \cdots T_3 & & I & 0 \\ T_{n-1} \cdots T_1 & T_{n-1} \cdots T_2 & T_{n-1} \cdots T_3 & \cdots & T_{n-1} & I \end{bmatrix}.$$

Furthermore, $\log p(\alpha) = -\frac{qn}{2} \log 2\pi - \frac{1}{2} \log |\Omega| - \frac{1}{2} (\alpha - d)' \Omega^{-1} (\alpha - d)$.

Gaussian observation model

The linear Gaussian observation model for vector Y_t is given by

$$Y_t = c_t + \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, H_t), \quad t = 1, \dots, n,$$

where vector c_t and matrix H_t are fixed and known for $t = 1, \dots, n$.
For data vector $Y = (Y_1', \dots, Y_n')'$, we have

$$Y|\alpha \sim N(c + \alpha, H),$$

with $c = (c_1', \dots, c_n')'$ and block diagonal matrix $H = \text{diag}(H_1, \dots, H_n)$.
Since $E(Y) = c + d$, $\text{Var}(Y) = \Sigma = \Omega + H$ and $\text{Cov}(\alpha, Y) = \Omega$, it follows from a standard lemma of the multivariate normal density that the conditional mean and variance are given by

$$E(\alpha|Y) = d + \Omega\Sigma^{-1} (Y - c - d), \quad \text{Var}(\alpha|Y) = \Omega - \Omega\Sigma^{-1}\Omega.$$

Mean and mode estimation for Gaussian model

Kalman filter and smoother evaluate the mean $E(\alpha_t|Y)$ and $Var(\alpha_t|Y)$ in a recursive and computationally efficient way, see Durbin and Koopman (2001) for a review.

Since all densities are Gaussian, the **mode** of $p(\alpha|Y)$, denoted by $\hat{\alpha}$, is equivalent to the **mean** of $p(\alpha|Y)$.

Applying a standard inversion lemma, it follows that

$$\hat{\alpha} = d + \Omega \Sigma^{-1} (Y - c - d), \text{ with } \Sigma = \Omega + H,$$

\Leftrightarrow

$$\hat{\alpha} = [\Omega^{-1} + H^{-1}]^{-1} (H^{-1}\{Y - c\} + \Omega^{-1}d),$$

It should be emphasized that the Kalman filter and smoother effectively computes $\hat{\alpha}$.

The mode for nonlinear non-Gaussian observations

Consider nonlinear non-Gaussian density $p(Y|\alpha)$ with α as before and with the independent channel assumption

$$p(Y|\alpha) = \prod_{t=1}^n p(Y_t|\alpha_t).$$

The **mode** is obtained by maximising $p(\alpha|Y)$ numerically since an analytical solution is not available.

The standard Newton-Raphson method is adopted: for a given guess g of mode $\hat{\alpha}$, a new guess is

$$g^+ = g - \left[\ddot{p}(\alpha|Y)|_{\alpha=g} \right]^{-1} \dot{p}(\alpha|Y)|_{\alpha=g},$$

where the step-length is one and

$$\dot{p}(\cdot|\cdot) = \frac{\partial \log p(\cdot|\cdot)}{\partial \alpha}, \quad \ddot{p}(\cdot|\cdot) = \frac{\partial^2 \log p(\cdot|\cdot)}{\partial \alpha \partial \alpha'}.$$

The mode for nonlinear non-Gaussian observations

Since $\log p(\alpha|Y) = \log p(Y|\alpha) + \log p(\alpha) - \log p(Y)$, we have

$$\dot{p}(\alpha|Y) = \dot{p}(Y|\alpha) - \Omega^{-1}(\alpha - d), \quad \ddot{p}(\alpha|Y) = \ddot{p}(Y|\alpha) - \Omega^{-1}.$$

Independent channel assumption implies $\ddot{p}(Y|\alpha)$ block diagonal. Then,

$$\begin{aligned} g^+ &= g - \left[\ddot{p}(Y|\alpha)|_{\alpha=g} - \Omega^{-1} \right]^{-1} \left(\dot{p}(Y|\alpha)|_{\alpha=g} - \Omega^{-1} \{g - d\} \right) \\ &= \left[\Omega^{-1} - \ddot{p}(Y|\alpha)|_{\alpha=g} \right]^{-1} \left(\dot{p}(Y|\alpha)|_{\alpha=g} - \ddot{p}(Y|\alpha)|_{\alpha=g} g + \Omega^{-1} d \right) \\ &= \left[\Omega^{-1} + A^{-1} \right]^{-1} \left(A^{-1} x + \Omega^{-1} d \right), \end{aligned}$$

where $A = - \left[\ddot{p}(Y|\alpha)|_{\alpha=g} \right]^{-1}$, $x = g + A \dot{p}(Y|\alpha)|_{\alpha=g}$.

Since structures of Gaussian mode estimator and nonlinear non-Gaussian mode “steps” are similar, compute g^+ by KFS with

$$Y = x, \quad c = 0, \quad H = A.$$

The mode for nonlinear non-Gaussian observations

To summarize,

- For Gaussian model, $Y = c + \alpha + \varepsilon$ with $\varepsilon \sim N(0, H)$ and H block diagonal, the KFS computes

$$\hat{\alpha} = [\Omega^{-1} + H^{-1}]^{-1} (H^{-1}\{Y - c\} + \Omega^{-1}d),$$

- For a non-Gaussian model, the mode is obtained numerically via Newton-Raphson where at each step we compute, for a given g ,

$$g^+ = [\Omega^{-1} + A^{-1}]^{-1} (A^{-1}x + \Omega^{-1}d),$$

where $A = -[\ddot{p}(Y|\alpha)|_{\alpha=g}]^{-1}$, $x = g + A \dot{p}(Y|\alpha)|_{\alpha=g}$.

- By considering Gaussian model with

$$Y = x, \quad c = 0, \quad H = A.$$

it follows that the KFS computes g^+ .

What if $p(Y|\alpha)$ is not log-concave ???

The arguments used are only valid when the block elements in A are positive semi-definite ($H = A$ is the variance matrix for the Gaussian observation model).

Therefore, $p(Y|\alpha)$ must be log-concave and all block elements of $\ddot{p}(Y|\alpha)$ need to be negative definite.

In case elements of $\ddot{p}(Y|\alpha)$ are positive, the three propositions below claim that KFS can still be used although no appeal can be made to the linear Gaussian model.

Nevertheless, the numerical algorithm for computing the next guess of the mode is equivalent to the KFS equations.

So the KFS can work with “variance matrix” H being negative definite. This sounds **strange** initially but viewing the KFS as an operator for specially structured matrices may help in the appreciation of this claim.

Justification for KFS when $p(Y|\alpha)$ is not log-concave

Proposition 1: Consider $\alpha \sim NID(d, \Omega)$ as before, $\Sigma = A + \Omega$ for **any** block diagonal matrix A with appropriate dimensions. If the LU decomposition

$$\Sigma = LU,$$

exists, with lower block unity triangular L and upper block triangular U , the forwards recursion

$$\begin{aligned} v_t &= x_t - a_t, & F_t &= A_t + P_t, \\ K_t &= T_t P_t F_t^{-1}, & L_t &= T_t - K_t, \\ a_{t+1} &= d_t + T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + Q_t^*, \end{aligned}$$

solves the set of linear equations $Lv = x$ for v with x given. This result holds even when A_t is negative definite for any $t = 1, \dots, n$.

Although F_t is not necessarily pd, this algorithm is the same as the Kalman filter.

Justification for KFS when $p(Y|\alpha)$ is not log-concave

Proposition 2: Given the definitions, condition and result in Proposition 1, the computation $u = \Sigma^{-1}x$ is carried out by the backwards recursion

$$u_t = F_t^{-1}v_t - K_t' r_t, \quad r_{t-1} = u_t + T_t' r_t, \quad ,$$

for $t = n, n - 1, \dots, 1$ with $r_n = 0$.

The algorithm is carried out after the forwards recursion of Proposition 1 and the storage of v_t, F_t and K_t for $t = 1, \dots, n$.

Justification for KFS when $p(Y|\alpha)$ is not log-concave

Proposition 3: Given the definitions, conditions and results in Propositions 1 and 2, the computation $a^* = d + \Omega\Sigma^{-1}(x - d)$ is carried out by the forwards recursion

$$a_{t+1}^* = d_t + T_t a_t^* + Q_t r_t, \quad t = 1, \dots, n,$$

with $a_1^* = a_1 + P_1 r_0$. The algorithm is carried out after the earlier recursions and storage of r_t for $t = 0, \dots, n - 1$.

Again, it can be shown that

$$d + \Omega\Sigma^{-1}(x - d) = [\Omega^{-1} + A^{-1}]^{-1} (A^{-1}x + \Omega^{-1}d).$$

The right-hand side is the computation of the next mode.

We have therefore shown that the standard KFS can be used for finding the mode.

The derivations of the Propositions 1, 2 and 3 are in Jungbacker and Koopman (2005).

Line search for finding mode

- In current procedures for finding the mode using Newton-Raphson, a line search is not implemented.
- Introducing scalar $0 < \lambda \leq 1$ in Newton-Raphson step

$$g_{\lambda}^{+} = g - \lambda \left[\ddot{p}(\alpha|Y)|_{\alpha=g} \right]^{-1} \dot{p}(\alpha|Y)|_{\alpha=g},$$

where the line search procedure consists of finding a λ such that $p(\alpha|Y)|_{\alpha=g_{\lambda}^{+}} > p(\alpha|Y)|_{\alpha=g}$.

- Equivalently, we have

$$g_{\lambda}^{+} = g + \lambda(g^{+} - g),$$

where g^{+} is computed as before. Note that $g^{+} = g_{\lambda}^{+}$ with $\lambda = 1$.

- After calculation of $g^{+} - g$, line search procedure for λ can start.
- Note that score $\dot{p}(\theta|y)$ can also be evaluated by KFS.

Simulation for linear Gaussian observation models

- Simulation smoothing refers to drawing samples from $p_G(\alpha|Y)$ where $p_G(\alpha, Y)$ is implied by linear Gaussian state space model.
- From earlier results, it follows that $p_G(\alpha|Y) = N(\hat{\alpha}, V)$ with

$$\begin{aligned}\hat{\alpha} &= d + \Omega\Sigma^{-1}(Y - d) &= (\Omega^{-1} + H^{-1})^{-1} (\Omega^{-1}d + H^{-1}Y), \\ V &= \Omega - \Omega\Sigma^{-1}\Omega &= (\Omega^{-1} + H^{-1})^{-1},\end{aligned}$$

where $\Sigma = \Omega + H$.

- The standard procedure for simulating from a multivariate normal distribution is to carry out a Cholesky on $V = LL'$ and to compute

$$\tilde{\alpha} = \hat{\alpha} + Lu, \quad u \sim N(0, I),$$

such that $\tilde{\alpha} \sim p_G(\alpha|Y) = N(\hat{\alpha}, V)$.

- Efficient algorithms for sampling from $N(\hat{\alpha}, V)$ are developed by de Jong and Shephard (1995) and Durbin and Koopman (2002).

Sampling for nonlinear non-Gaussian models

- Direct sampling from $p(\alpha|Y)$ is not possible since in most cases no analytical expressions are available for $p(\alpha|Y)$.
- Since the mode is computed via Newton-Raphson for which a linear Gaussian model is considered at each step, we can also adopt this model at the mode of $p(\alpha|Y)$.
- Then, by adopting the simulation smoothing methods of previous slide, we sample effectively around the mode of $p(\alpha|Y)$ with a curvature implied by $-\ddot{p}^{-1}(Y|\alpha)$ for $\alpha = \hat{\alpha}$.
- It is anticipated that these samples from $p_G(\alpha|Y)$ (yes, p_G refers to the “mode” model) mimic samples from $p(\alpha|Y)$ for obvious reasons.
- These ideas and concepts are used for computing the likelihood function $p(y)$ via Monte Carlo integration.

Sampling for nonlinear non-Gaussian models

- Sampling around the mode of $p(\alpha|Y)$ with a curvature implied by $-\ddot{p}^{-1}(Y|\alpha)$ for $\alpha = \hat{\alpha}$ is done by considering $p_G(\alpha, y)$ and applying the simulation smoothing method of de Jong and Shephard (1995) or of Durbin and Koopman (2002).
- However, what to do if $p(Y|\alpha)$ is not log-concave ($\ddot{p}(Y|\alpha) > 0$) ???
 - Durbin and Koopman (2002) can not be used since it requires the unconditional sampling from $p_G(Y|\alpha)$. Although this is simple, it is not possible with negative variances !
 - Proof of de Jong and Shephard (1995) is also based on observation densities that do not exist in this case.
 - Jungbacker and Koopman (2005) have dealt with the case of $\ddot{p}(Y|\alpha)$ being positive.
 - Some of these results are presented below. An important consequence of these developments is that we can work with negative variances !

Theorem 2: simulation smoothing

Consider the linear Gaussian signal vector $\alpha \sim N(\mu, \Omega)$. Further, consider x and A as defined earlier and evaluated at $\theta = \hat{\theta}$. Both A and $\Sigma = \Omega + A$ can be nd while $V = \Omega - \Omega\Sigma^{-1}\Omega = A - A\Sigma^{-1}A$ is pd. Sampling from

$$N(\hat{\alpha}, A - A\Sigma^{-1}A),$$

is carried out by Kalman filter and simulation smoothing equations

$$\begin{aligned} C_t &= A_t^{-1} - F_t^{-1} - K_t'N_tK_t, & R_t &= C_t^{-1}(A_t^{-1} - K_t'N_tT_t), \\ w_t &\sim N(0, C_t), & u_t &= A_t(w_t + F_t^{-1}v_t - K_t'r_t), \\ r_{t-1} &= A_t^{-1}u_t - R_t'w_t + T_t'r_t, & N_{t-1} &= R_t'C_tR_t - A_t^{-1} + T_t'N_tT_t, \end{aligned}$$

for $t = n, n - 1, \dots, 1$ and with the initialisations $r_n = 0$ and $N_n = 0$. It can be shown that these equations are equivalent to the ones of de Jong and Shephard (1995). However, the equations here are computationally more efficient.

Likelihood evaluation via importance sampling

- IS is based on the simulation of α_t given the observations Y .
- Simulations can be used to evaluate the likelihood function and to estimate α_t (signal extraction).
- When the simulations are conditional on y , the samples are informative with respect to y .
- A naive Monte Carlo estimator of the likelihood function $p(y) = \int p(y|\alpha)p(\alpha)d\alpha$ is based on the unconditional density

$$\hat{p}(y) = \sum_{i=1}^M p(y|\alpha^i), \quad \text{where } \alpha^i \sim p(\alpha).$$

This simple estimator is poor since many simulations will make no contribution to $p(y|\alpha)$.

- Therefore a very large number of simulations M is needed to obtain only an inaccurate Monte Carlo estimate $\hat{p}(y)$.
- We therefore use IS !

Likelihood evaluation via importance sampling

- Given draws $\alpha^i \sim p_G(\alpha|y)$ for $i = 1, \dots, M$, an efficient Monte Carlo estimator of the likelihood is given by

$$\begin{aligned}\widehat{p}(y) &= M^{-1} \sum_{i=1}^M p(y|\alpha^i)p(\alpha^i) / p_G(\alpha^i|y) \\ &= p_G(y)M^{-1} \sum_{i=1}^M p(y|\alpha^i) / p_G(y|\alpha^i),\end{aligned}$$

where $p_G(\alpha) = p(\alpha)$ and $p_G(\alpha|Y)$ is the approximating model based on the mode and the curvature $-\ddot{p}^{-1}(Y|\alpha)$.

- In a similar way, we obtain estimates of α :

$$\bar{\alpha} = \int \alpha p(\alpha|y) d\alpha = p(y)^{-1} \int \alpha p(\alpha, y) d\alpha = p(y)^{-1} \int \alpha p(y|\alpha) p(\alpha) d\alpha.$$

The Monte Carlo estimator of $\bar{\alpha}$ is given by

$$\widehat{\bar{\alpha}} = \frac{\sum_{i=1}^M \alpha^i p(y|\alpha^i) / p_G(y|\alpha^i)}{\sum_{i=1}^M p(y|\alpha^i) / p_G(y|\alpha^i)}.$$

Multi-state Latent Factor Intensity (MLFI) Model

- Earlier relevant work is done by Gagliardini and Gouriéroux (2004), McNeil and Wendin(2004) and Koopman, Lucas and Daniels (2005).
- In this paper, we consider a duration model with unobserved components.
- Here, durations to transitions are endogenous rather than exogenous so we use all the information in the data-set.
- It is a multi-state extension of the Latent Factor Intensity (LFI) model of Bauwens and Hautsch (2003), a point process model for stock transactions in tick-time.
- In our model, the time to the next rating event and the type of event that is going to occur (upgrade, downgrade, or default) are modelled simultaneously.
- This leads to a complex likelihood structure !

Credit risk modelling

- Likelihood function consists of high dimensional integral involving latent common risk factor.
- Our parameter driven model differs from well-known observation driven counterparts: ACD of Engle and Russell (1998) or ACI of Russell (1999).
- Likelihood evaluation via a multivariate extension of Monte Carlo techniques discussed earlier.
- We demonstrate effectiveness of method by simulation experiment.
- Model is estimated for CreditPro6.2 data set from Standard & Poor's containing all issuer ratings over the period 1981 – 2002.
- Firms are classified as Investment 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 MLFI model for credit risk

- The multi-state feature of the model is represented as a set \mathbb{S} of transition types, $\mathbb{S} = \{1, 2, \dots, S\}$.
- Define the right-continuous counting processes $N_k(t)$ and $N(t)$.
- The process $N(t)$ makes a unity jump at time of rating event for any of K units (no simultaneous rating transitions).
- Count process is marked: $N_{sk}(t)$ is for transition s and firm k .
- $N_k(t) = \sum_{s=1}^S N_{sk}(t)$, $N(t) = \sum_{k=1}^K N_k(t) = \sum_{s=1}^S \sum_{k=1}^K N_{sk}(t)$.
- Finite stochastic intensity $\lambda_{sk}(t)$ describes instantaneous probability of unit k experiencing a type s rating transition at time t conditional upon the information available just before time t :

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

The MLFI model for credit risk, cont.

Dummy $R_{sk}(t)$ indicates unit k being 'at risk' for transition type $s \in \mathbb{S}$ at time t^- (can be at risk for multiple transition types).

The model specification for intensities is given by

$$\lambda_{sk}(t) = R_{sk}(t) \cdot \exp[\eta_s + \gamma'_s w_k(t) + \alpha_s \psi(t)] \cdot H_{sk}(t),$$

with $s = 1, \dots, S$ and $k = 1, \dots, K$, where

1. scalar η_s , $m \times 1$ vector γ_s , and scalar α_s are fixed unknown coefficients,
2. $m \times 1$ vector $w_k(t)$ contains explanatory variables (covariates),
3. scalar $\psi(t)$ is latent dynamic factor, and
4. scalar $H_{sk}(t)$ is generalized baseline hazard function, to model duration dependence (multivariate type).

This specification encompasses homogeneous continuous-time Markov chain model, used in empirical credit risk literature, see Kavvathas (2001) and Lando and Skødeberg (2002).

The MLFI model for credit risk, cont.

- Assume $\psi(t)$ only changes at event times $t = t_i$ for $i = 1, \dots, N(T)$ (right-censoring of type I) and follows stochastic process with piecewise constant (left-continuous) sample paths.
- Let $\psi_i = \psi(t_i)$ be common risk factor $\psi(t)$, $t \in (t_{i-1}, t_i]$,

$$\psi_i = \rho\psi_{i-1} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

- Baseline hazard is fixed non-neg function for duration dependence

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

$t - t_{ik}$ is backward-recurrence time (unit k , past i th transition).

- For example,

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

with $x_i \geq 0$ and fixed coeff $a_{si} > 0$, $b_s > 0$ for $i = 0, 1, \dots, N$.

Likelihood function

The likelihood function conditional on initial conditions and on complete path of unobserved process $\Psi_{N(T)} = \{\psi_i\}_{i=0}^{N(T)}$ is given by

$$L(\theta \mid \mathcal{F}_T, \Psi_{N(T)}) =$$

$$\prod_{i=1}^{N(T)} \prod_{k=1}^K \prod_{s=1}^S \exp \left(Y_{sk}(t_i) \ln \{ \lambda_{sk}(t_i) \} - R_{sk}(t_i) \int_{t_{i-1}}^{t_i} \lambda_{sk}(t) dt \right),$$

with dummy $Y_{sk}(t)$ (one if unit k at time t reports rating event of type s) and \mathcal{F}_T denotes the relevant observable filtration.

Likelihood needs integrating wrt $\Psi_{N(T)}$. Then $\hat{\theta} = \max_{\theta} L(\theta \mid \mathcal{F}_T)$ and

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

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

Likelihood evaluation

The stochastic variables are in z_i :

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

for $i = 1, \dots, N$. The conditional log-density for the i th event is

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

for $i = 1, \dots, N$ and with intensity

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

where Z_{ski} selects from ν_i . In case $H_{sk}(t) = 1$, $\lambda_{sk}(t)$ is constant for $t_{i-1} < t \leq t_i$. In case, $\nu_i = \{\eta_1, \dots, \eta_S, \gamma'_1, \dots, \gamma'_S, \psi(t_i)\}'$ and $Z_{ski} = \{e'_s, e'_s \otimes w_k(t_i)', \alpha_s\}$, we have $Z_{ski} \nu_i = \eta_s + \gamma'_s w_k(t_i) + \alpha_s \psi_i$.

Monte Carlo likelihood evaluation

The likelihood function can be given by

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

No analytical expression, we rely on numerical techniques:

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

is poor. Simulating from $p(\nu | z, \mathcal{F}_T)$ is not feasible.

Importance sampling

To evaluate

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

do importance sampling and simulate from Gaussian density $p_G(\nu|z, \mathcal{F}_T)$. Then compute

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

since $p_G(\nu | \mathcal{F}_T) = p(\nu | \mathcal{F}_T)$ and

$$p_G(\nu | z, \mathcal{F}_T) = p_G(z | \nu, \mathcal{F}_T) p_G(\nu | \mathcal{F}_T) / p_G(z | \mathcal{F}_T).$$

Monte Carlo study: for a three-rating system

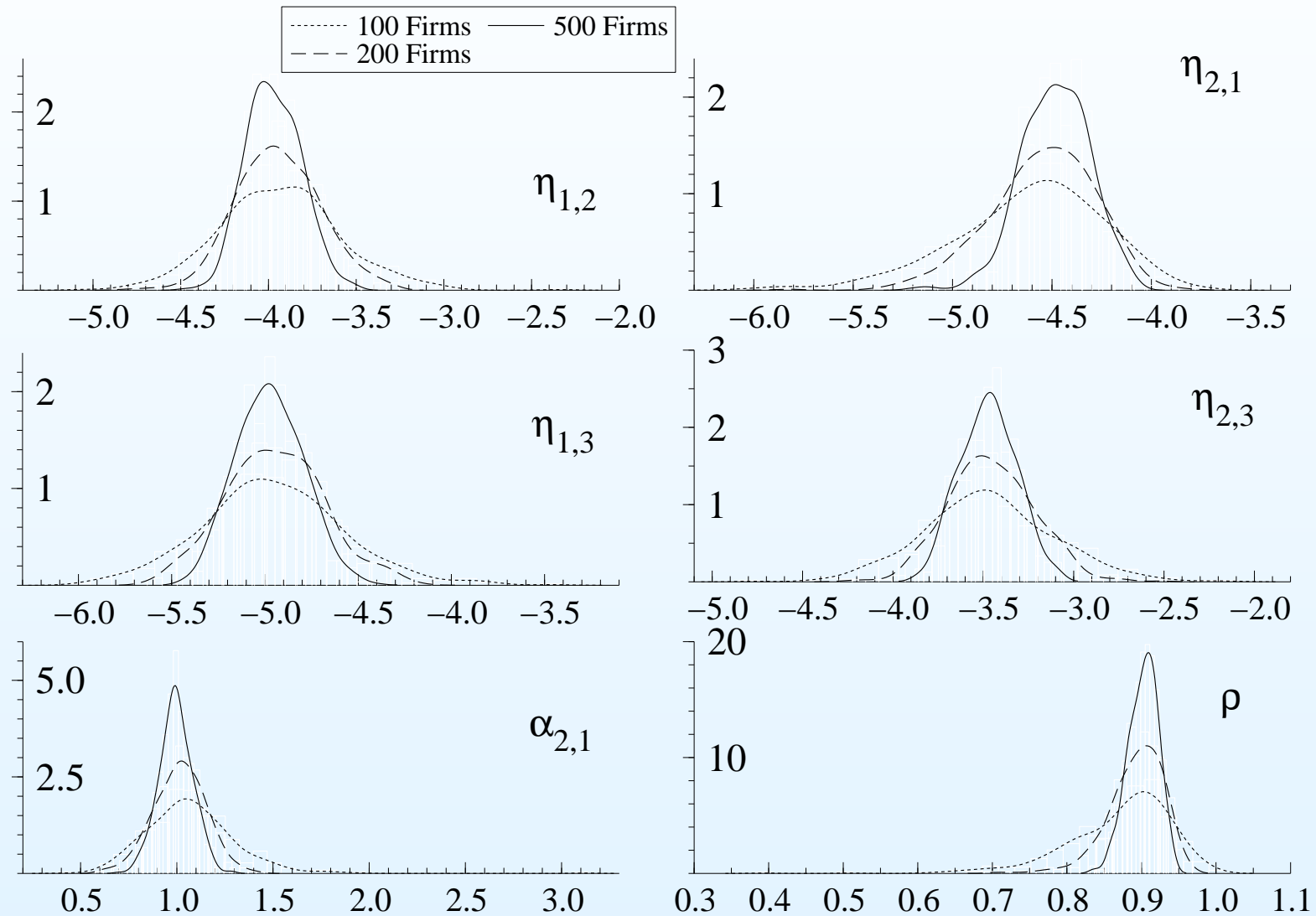
We consider three states: investment grade, subinvestment grade, and default (absorbing state) with intensities

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

where

- $\psi(t)$ is a step function, jumps at the endogenous event times t_i . The firm heterogeneity enters through η_s for the different transition types s .
- $\psi(t)$ is an AR(1)
- the benchmark model abstracts from duration dependence by setting $H_{sk}(t) \equiv 1$.
- further we set $\alpha_s = -\alpha < 0$ for downgrades, and $\alpha_s = \alpha > 0$ for upgrades.

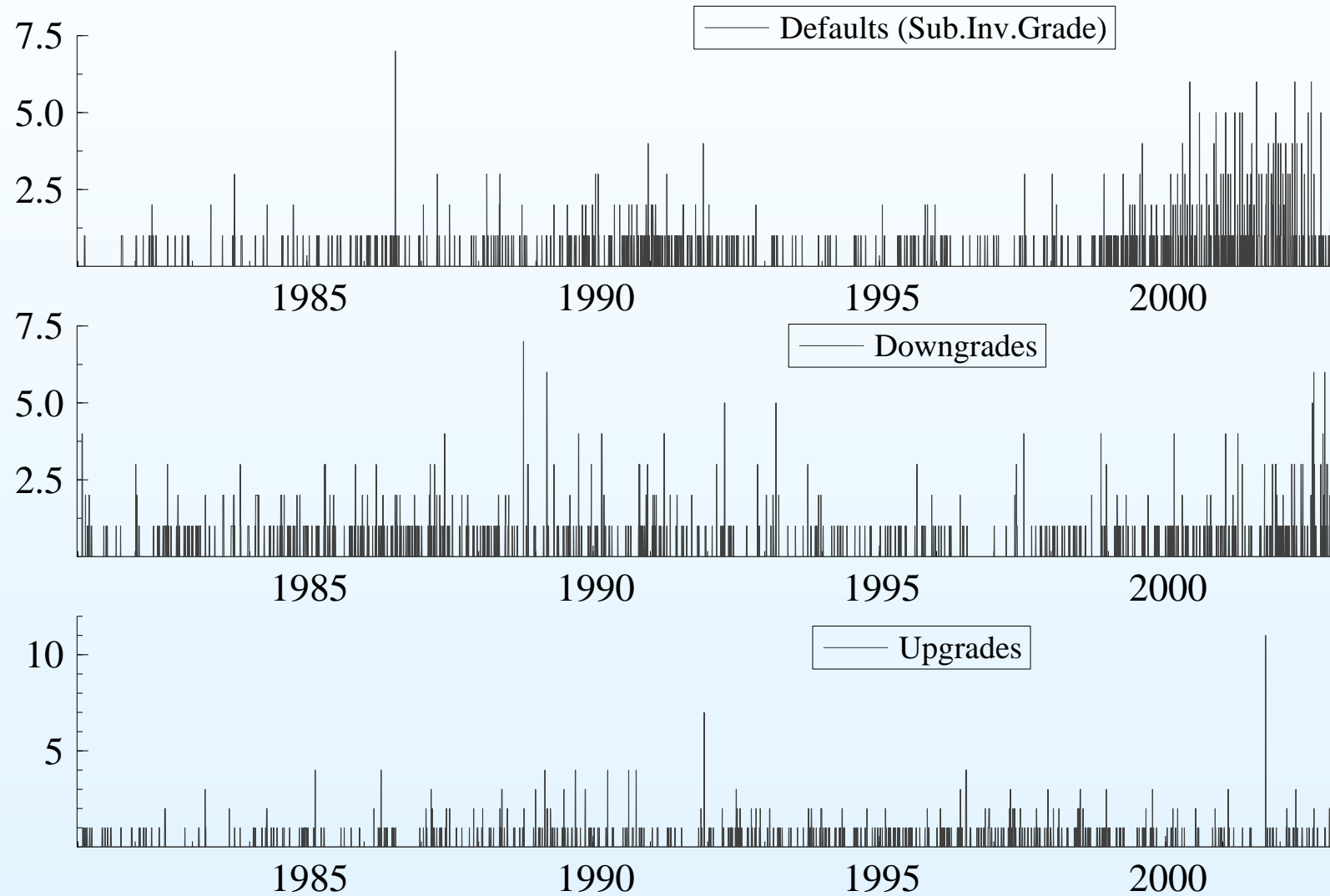
Monte Carlo study: for a three-rating system



Empirical study

- data consist of rating transitions obtained from Standard & Poor's.
- rating histories of all issuers are recorded in the CreditPro 6.2 database.
- sample period is from the end of 1980 (the left-censoring time point) until the start of 2003, covering a total of 8035 days.
- durations of the pooled process are fraction of the business year. (multiple rating events occur on a single day).
- 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.

Empirical study



Empirical study

Descriptive Statistics Rating Data:

	Investment Grade	Subinvestment Grade
Total Number of Defaults	7	835
Total Number of Downgrades	773	–
Total Number of Upgrades	–	579
Durations (pooled process)	1.85 days (Mean)	0.006 (St.Dev.)

Latent factor: η AR(1)

