ARE DEFAULT CORRELATIONS TIME DEPENDENT? A BAYESIAN APPROACH

CHRISTINA R. NIETHAMMER

ABSTRACT. In this paper we discuss if asset correlations in a Gaussian one factor model are time dependent. Starting from S & P's default study, we examine a random factor loading type model (two correlations for two stages of the economy) and a model with time dependent variance of the common factor. As classical estimation is not very flexible, we apply Markov Chain Monte Carlo methods: An approach proposed by Wendin (2006) (Binomial approach) can be adopted to analyze a two stage correlation model. However an extended version of this Binomial algorithm becomes unstable in more complex models. We therefore develop an asset value approach to handle e.g. time dependence in the variance of the factor. Surprisingly, neither a two stage correlation nor a time dependence of the variance of the factor can be detected from loss rates!

1. INTRODUCTION

Constant volatility in the standard Black Scholes Model corresponds to a constant correlation between different assets in the underlying portfolio of a CDO - tranche. Option prices of different strikes yield different volatilities, we have a volatility smile. Similar for CDOs, different tranches yield different correlations, we have a correlation skew. Moreover in extreme market situations (e.g. May 2005, so called correlation crisis) implied correlations of certain tranches do not exist. To overcome these problems, so called base correlations were introduced by Ahluwalia et al. (JPMorgan) (2004), i.e. implied correlations of all equity tranches, e.g. for itraxx (CDO)-tranches 0-3%, 0-6% and so forth are taken. [Itraxx is an European index of 125 Credit Default Swaps of firms.] By interpolating these points, non standard tranches like 4-5% can be priced. An appropriate mapping for bespoke CDOs (non standardized underlying portfolio) yields a price for such products. However, base correlations are not based on a model, many drawbacks

Date: July 7, 2006, revised: June 30, 2008.

We thank Sascha Meyer-Dautrich, Matthias Heck, Thomas Seifert, and Christoph Wagner for helpful comments and suggestions concerning this work and Mikhail Krayzler for programming parts of the algorithms. CN gratefully acknowledges financial support by UniCredit, Markets and Investment Banking. However, this paper does not reflect the opinion of UniCredit, Markets and Investment Banking, it is the personal view of the authors.

E-mail address: Christina.Niethammer@uni-konstanz.de.

are presented e.g. in Willemann (2004). Moreover, what is to do when simultaneously buying and selling two different tranches, if in addition there is a not-linear dependency between the tranches. The trade cannot be divided in two equity tranches. The base correlation concept again fails. A concept called Random Factor Loading was introduced (see e.g. Andersen and Sidenius (2004)). Essentially, one proposes two asset correlations, one for bad and one for normal/good states of the economy (symbolized by a common factor Y). A time dependence in correlations is introduced to extend the number of degrees of freedom to better fit market prices.

In general one can ask, if asset or implied default correlations are time dependent. Within the framework of Bayesian analysis, we discuss a random factor loading approach as well as a time lagged stochastic volatility model up on the common economy factor Y. Surprisingly, our empirical study on historical loss rates does not yield any evidence for either model. Clearly, this might be a consequence of rare data in credit risk in general and one can claim to use stock market data or other macroeconomic data to determine asset correlations. In fact, there are studies detecting time dependence in equity correlations, see e.g. Buraschi et al. (2006) for a survey. Moreover, Das et al. (2005) put some effort on calculating default correlations from estimated intensity time series derived from equity data and some other financial information. Again time dependent default correlations are detected. Nevertheless, we think that loss rates are the natural source to examine default correlations. So indirectly we include macroeconomic and equity data as a driving risk factor on loss rates. In this version, we again see no significant difference to our correlation study. The major difference to the mentioned studies on equity time series is the data source, the time intervals, and time horizon data. We conclude that in several model variants based on the Gaussian one factor model the information contained in loss rates does not yield a time dependence in correlations of the unknown underlying asset value process of the firms.

Finally, we give a short word on the applied Bayesian estimation techniques. To account for relatively rare data and to cover more complex models, we have to set up our estimation within a Bayesian methodology and apply Markov Chain Monte Carlo methods (MCMC) for our estimation procedure. By applying classical estimation techniques in Niethammer and Overbeck (2008), we have already seen that it is difficult to gain intuition how good our point estimator is, i.e. the extent of the corresponding model risk on risk figures depending on their parameters is quite unknown. Asymptotic considerations or strong assumptions often have to be proposed to derive confidence intervals in classical statistics. Moreover, in Niethammer and Overbeck (2008) we see that the reliability is questionable, because confidence intervals are valid for a large number of observations only. The number of observed periods has to go to infinity although we barely get 25 periods. In particular for small default probabilities as the parameter comes close to the boundary, the number of periods has to be amazingly huge.

In contrary, Bayesian analysis automatically yields these desired confidence intervals, because a whole distribution for parameters is developed providing much more information than a point estimate. A distribution of the risk figures is implied. An asymptotic analysis of the estimators is not necessary.

Moreover, MCMC methods yield efficient methods to calculate maximum likelihood estimation when classical numerical methods are almost intractable and flexible extensions are possible, e.g. a non Gaussian common factor or an incorporation of a prior belief can be included, updated by the included data set. We compare classical estimation with a Bayesian perspective and the corresponding risk figures. At first we rely on a procedure proposed by Gössl (2005); McNeil and Wendin (2006); Wendin (2006). They present a similar methodology to include MCMC in credit risk models - in the sequel called Binomial approach. We examine the binomial approach and suggest extensions appropriate for our correlation experiment. As this algorithm does not perform well in complex models, we propose a more natural asset value algorithm. Essentially, we sample an underlying world index. Autocorrelation within a Markov chain is reduced enormously and in addition the algorithm is very robust to changes in initial values of our Markov chains. In a simulation study, parameters are recovered in all considered models, whereas the Binomial approach fails to recover parameters in more complex models.

The paper is organized as follows. We start with a review on Bayesian methodology and Markov chain Monte Carlo methods (MCMC). Important links to the literature are given. Simple examples are provided to prepare an intuitive way to our new stable asset value algorithm. In Section 1.2, we then show how classical maximum likelihood estimators can be obtained by means of MCMC. A comparison to the estimators gained in Niethammer and Overbeck (2008) is given. We review a general one factor model and confront classical estimation with Bayesian methodology. In Section 1.3, extensions to the Gaussian model are presented, in particular the two-stage-correlation model. An empirical study on time dependence is performed. In Section 2, we suggest our mentioned asset value process, advantages are discussed. Results on time dependent variances are provided on the basis of S&P's default study. Afterwards a sensitivity analysis is performed and a possible extension to the multivariate case is introduced. Finally, we compare our results on risk figures with classical methods as given in Niethammer and Overbeck (2008).

1.1. **A short Review of Bayesian Analysis.**

We start with the promised introduction to Bayesian analysis and MCMC. In Bayesian Analysis, the model depends on a **random** parameter vector θ and the given data. Estimation is as follows:

(1) Specify prior beliefs $p(\theta)$ about θ (prior distribution, $p(\theta)$)

(2) Calculate posterior beliefs π by Bayes' Theorem by inserting the data:

$$\pi(\theta|Data) = \frac{\mathrm{L}(Data|\theta)p(\theta)}{\int \mathrm{L}(Data|\theta)p(\theta)d\theta} \propto \mathrm{L}(Data|\theta)p(\theta)$$

If θ is uniformly distributed, we obtain the usual ML-estimator by taking the mode.

Remark 1.1 (Notation)**.** *In our simulation studies, we say that parameters are recovered if their mode and mean is close to the initial parameter value set in the simulation study. $\theta \approx 2$ means that the mean and the mode are close to the fixed number 2 set in the simulation study.*

■

In some cases π can be calculated analytically, (if L and p_θ are conjugate). In earlier days this was the only possibility to aim information about the distribution of the parameters. Bayesian analysis was not very flexible. Nowadays, MCMC methods solve this problem. Moreover, a clever calculation of ML-estimators is included. We give a short overview by following an introduction given in Chib and Greenberg (1996).

The aim of MCMC is to construct a Markov chain $(\theta^{(i)})_i$ with transition kernel P and stationary distribution π (posterior distribution). From ergodic theory, we then know that under suitable regularity assumptions (see (Chib and Greenberg, 1996, Proposition 1)) for all initial values $\theta^{(0)}$ and measurable sets A :

$$|P^{(m)}(\theta^{(0)}, A) - \pi(A)| \to 0,\ m \to \infty \tag{1}$$

where $P^{(m)}$ is the m-th step transition kernel:

$$P^{(m)}(\theta^{(0)}, A) = \int P(\theta^{(0)}, ds)P^{(m-1)}(s, A).$$

Suppose we have found an m such that $P^{(m)}$ is already close enough to π. Let's say for all m greater than a burn in $\bar{b}$. We can then calculate the empirical distribution function from $(\theta^{(m)})_{m\geq\bar{b}}$ and other useful statistics from the sample and analyze the approximate posterior. Often not every sample after the burn in is taken because Markov chains can be quite autocorrelated (autocorrelation adjustment).

The next assertion is important for calculating functions of our parameters (e.g. risk figures): for all π-integrable real valued functions f and a Markov chain $(\theta^{(i)})_i$ with stationary distribution π we have

$$\frac{1}{m}\sum f(\theta^{(i)}) \overset{a.s.}{\to} \int f(\theta)\pi(d\theta),\ m \to \infty. \tag{2}$$

Our risk figures are functions, say f, of default probabilities and correlation. That means if the function f is sufficiently integrable and our parameters $p_k,\ k = 1, ..., K$ and ρ are consistent, then the risk figure is consistent as well.

There are two main algorithm to generate such a Markov chain. Besides the so called Metropolis-Hastings algorithm (see e.g. Chib and Greenberg (1995)), Gibbs sampling is applied. In fact, Gibbs Sampling is a special case of the MH-algorithm. Moreover, many combinations and further developments were introduced, e.g. we will come back to Adaptive Rejection Metropolis Sampling, short ARMS-algorthm, see Gilks et al. (1995) for a detailed introduction. We could now start to introduce several methods and argue which algorithm performs best. Actually this really much depends on the special situation. As we almost exclusively work with Gibbs-Sampling (except from ARMS), we refrain from such a discussion and refer to e.g. Lancaster (2004); Gamerman and Lopez (2006). Assumptions for the convergence of the MH- and Gibbs-sampling are given in Chib and Greenberg (1996), see also Jones (2004). We only state conditions for the convergence of the Gibbs-Sampler in Proposition 1.1 as we focus on Gibbs Sampling next.

We continue with Gibbs-Sampling of a Markov chain $(\theta^{(i)})_i$. The corresponding transition kernel for Gibbs Sampling possesses a density p_G :

$$p_G(\theta,\xi)=\prod_{k=1}^{d}\pi(\xi_k|\xi_1,...,\xi_{k-1},\theta_{k+1},...,\theta_d),\ \theta=\theta^{(i)},\ \xi=\theta^{(i+1)}$$

The algorithm then works as follows:

(1) Potentially block θ appropriately in $\tilde{\theta}=(\tilde{\theta}_1,...,\tilde{\theta}_a),\ a\leq d$, e.g. $(\tilde{\theta}_1,\tilde{\theta}_2)=((\theta_1,\theta_2),\theta_3)$.
(2) Specify starting values $\tilde{\theta}^{(0)}=(\tilde{\theta}_1^{(0)},..,\tilde{\theta}_a^{(0)})$, set $i=0$
(3) Simulate

$$\begin{aligned}
\tilde{\theta}_1^{(i+1)} &\quad \text{from} \quad \pi(\tilde{\theta}_1|\tilde{\theta}_2^{(i)},..,\tilde{\theta}_a^{(i)},Data)\\
\tilde{\theta}_2^{(i+1)} &\quad \text{from} \quad \pi(\tilde{\theta}_2|\tilde{\theta}_1^{(i+1)},\tilde{\theta}_3^{(i)},..,\tilde{\theta}_a^{(i)},Data)\\
&\quad\ \vdots\\
\tilde{\theta}_a^{(i+1)} &\quad \text{from} \quad \pi(\tilde{\theta}_a|\tilde{\theta}_1^{(i+1)},\tilde{\theta}_2^{(i+1)},..,\tilde{\theta}_{a-1}^{(i+1)},Data)
\end{aligned}$$

(4) Set i=i+1, if $i+1<I$, go to step 3, where I is the number of iterations.

$\theta^{(j)}$ is then a sample of the jth step transition of the Gibbs Sampler $P_G^{(j)}(\theta^{(0)},\cdot)$, $j=1,...,I$. For i large enough ($i\geq\bar{b}$ =burn in), $\theta^{(i)}$ is then by Equation 1 approximately a sample of $\pi(\theta|Data)=\pi(\tilde{\theta}|Data)$. $\pi(\tilde{\theta}_1|\tilde{\theta}_2,..,\tilde{\theta}_a,Data)$ are called full conditional distributions. These are usually easier to sample as the unconditional posterior distribution $\pi(\tilde{\theta}_1,...,\tilde{\theta}_a|Data)$. The above mentioned suitable conditions for the convergence of the Gibbs Sampler (i.e. Equation 1/2) are given in (Chib and Greenberg, 1996, Proposition 2):

Proposition 1.1. *Let a data set - Data - be given. Suppose that the following conditions are satisfied:*

(1) $\pi(\theta|Data) > 0$ *implies there exists an open neighborhood* U_θ *with* $\theta \in U_\theta$ *and an* $\epsilon > 0$ *such that for all* $\zeta \in U_\theta$, $\pi(\zeta|Data) \geq \epsilon > 0$.
(2) $\int \pi(\theta|Data)d\theta_k$ *is bounded.*
(3) *The support of* $\theta|Data$ *is arc-connected.*

Then Equation 1 and 2 hold.

We will see that all assumptions trivially hold in all our algorithms as long as prior distributions are sufficiently smooth because $|Bin(\cdot)| \leq 1$.

As Gibbs Sampling is essential for our further analysis, we give a simple regression example. Afterwards, we expand the example by assuming the dependent variable is a Bernoulli variable driven by an additional variable (the dependent variable from the first example). In the estimation procedure this additional random variable is sampled given the original regression parameters and the Bernoulli data. Estimation of the original parameters is then performed given the additional variable as in the first example. So in principle, there is a close connection to credit risk. In Section 2, we apply this idea. The asset value process X is the additional variable and can be sampled given the common factor y, the correlation ρ, and the (default) data $(\mathbf{l}, \mathbf{N})$. Gibbs Sampling assumes in every other step that the remaining parameters are known, e.g. X the asset value process is known when sampling y and ρ. Estimation of ρ and y given the asset value process X boils down to the simple regression example. Afterwards X is then again sampled given ρ,y, and $(\mathbf{l}, \mathbf{N})$ and so forth. So the Bernoulli or Probit example can be seen as a toy example for the later algorithm:

Example 1.1 (Simple Regression Example)**.** *Assume the data* $(\tilde{y}_j, \breve{x}_j)_{j\leq n}$ *follow the model:*

$$\breve{X}_j = \beta\tilde{Y}_j + \epsilon_j,\ \epsilon_j \sim \mathcal{N}(0, \sigma^2),\ \sigma^2 > 0$$

To perform Bayesian analysis we need a prior information about β *and* σ^2*. The prior of* β *is set normal with mean* β_0 *and variance* b_0^{-1}*, where* β_0 *and* b_0^{-1} *have to be known in advance. An inverted gamma distribution is a common choice for a prior on* σ^2*:* $\sigma^2 \sim IG(w_0/2, \delta_0/2)$*. We further assume independent prior information, i.e. the joint prior of* (β, σ^2) *is equal to*

$$f_{\mathcal{N}}(\beta|\beta_0, b_0^{-1}) f_{IG}(\sigma^2|w_0/2, \delta_0/2).$$

The posterior is then given by (up to a normalizing constant)

$$\pi(\beta, \sigma^2|\breve{x}, \tilde{y}) \propto f_{\mathcal{N}}(\beta|\beta_0, b_0^{-1}) f_{IG}(\sigma^2|w_0/2, \delta_0/2)\sigma^n \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(\breve{x}_j - \beta\tilde{y}_j)^2\right).$$

The full conditional distributions can be derived analytically, see e.g. (Gamerman and Lopez, 2006; Lancaster, 2004; Chib and Greenberg, 1996):

$$\beta|\breve{x}, \tilde{y}, \sigma^2 \sim \mathcal{N}(\hat{\beta}, (b_0 + \sigma^{-2}\sum \tilde{y}_j^2)^{-1}),\ \sigma^2|\breve{x}, \tilde{y}, \beta \sim IG\left(\frac{w_0 + n}{2}, \frac{\delta_0 + \delta_n}{2}\right),$$

where

$$\hat{\beta} = (b_0 + \sigma^{-2}\sum \tilde{y}_j^2)^{-1}(b_0\beta_0 + \sigma^{-2}\sum \tilde{y}_j\breve{x}_j),\ \delta_n = \sum(\breve{x}_j - \beta\tilde{y}_j)^2.$$

At first b_0, δ_0, β_0, and w_0 have to be chosen. The corresponding Gibbs sampler then works as follows

(1) *No blocking: $\theta = (\beta, \sigma^2)$ in $\tilde{\theta} = (\beta, \sigma^2)$, $a = d = 2$*
(2) *Specify starting values $\beta^{(0)}$ and $(\sigma^2)^{(0)}$ set $i = 0$*
(3) *Simulate*

$$\beta^{(i+1)} \quad \text{from} \quad f_{\mathcal{N}}\left(\beta|\hat{\beta}((\sigma^2)^{(i)}, \tilde{y}), (b_0 + \frac{1}{(\sigma^2)^{(i)}}\sum \tilde{y}_j^2)^{-1}\right)$$

$$(\sigma^2)^{(i+1)} \quad \text{from} \quad f_{IG}\left(\sigma^2|\frac{w_0 + n}{2}, \frac{\delta_0 + \delta_n(\beta^{(i+1)}, \tilde{y})}{2}\right)$$

(4) *Set i=i+1, if $i + 1 < I$, go to step 3, where I is the number of iterations, otherwise stop.*

As mentioned samples are usually quite autocorrelated, we only take every ath sample starting from a burn in $\bar{b}$. Assume there exists an m such that $I = \bar{b} + ma$. We then consider $(\beta^{(i)}, (\sigma^2)^{(i)})$ for $i = \bar{b}, \bar{b} + a, \bar{b} + 2a, ..., I$ as a sample of the posterior distribution $\pi(\beta, \sigma^2|\breve{x}, \tilde{y})$. The mode of this sample is then approximately the maximum likelihood estimator of the regression equation. Alternatively, one can sample several chains with different initial values and take the first sample after the burn in.

In this simple example, we obtain a Gaussian full conditional distribution from which we can easily sample. Later on, if full conditional distributions are non standard, this is replaced by an algorithm called ARMS (Adjusted Rejection Metropolis Sampling, see Gilks et al. (1995)).

■

Example 1.2 (Simple Probit Example)**.** *Recall, in a Probit model, we are given data $(\tilde{h}_i, \tilde{y}_i)$, where the model is reproduced by the following equation*

$$\breve{X}_j = \tilde{Y}_j\beta + \epsilon_j,\ \epsilon_j \sim \mathcal{N}(0, \sigma^2),\ \tilde{H}_j = 1_{\breve{X}_j < 0}.$$

If $\breve{X} = \breve{x}$ were known, we could apply the results from Example 1.1 without knowledge of $\tilde{h}$: $\beta|\breve{x}, \tilde{y} \overset{d}{=} \beta|\breve{x}, \tilde{y}, \tilde{h}$. We do not know $\breve{X}$, but given $\beta, \tilde{h}, \tilde{y}_j$ we can derive its distribution:

$$F_j(\cdot|\tilde{h}_j, \tilde{y}_j, \beta) = \begin{cases} T\mathcal{N}_{(-\infty,0)}(\beta\tilde{y}_j, \sigma^2) & \text{if} \quad \tilde{h}_j = 1 \\ T\mathcal{N}_{[0,\infty)}(\beta\tilde{y}_j, \sigma^2) & \text{if} \quad \tilde{h}_j = 0 \end{cases}$$

where $T\mathcal{N}_{(a,b)}$ denotes the normal distribution truncated to (a, b). f_j denotes the density of F_j. The trick is to sample $\breve{X}$ in addition. So we could also replace the normal distribution of ϵ by another distribution. We obtain the following algorithm with $\breve{X} = (\breve{X}_1, ..., \breve{X}_n)$:

(1) *Blocking: $\theta = (\beta, \sigma^2, \breve{X}_1, ..., \breve{X}_n)$ in $\tilde{\theta} = (\beta, \sigma^2, (\breve{X}_1, ..., \breve{X}_n))$, $a = 3$, $d = n + 2$*

(2) *Specify starting values* $\beta^{(0)}$, $(\sigma^2)^{(0)}$, *and* $\breve{X}_j^{(0)}$, $j = 1, .., n$ *set* $i = 0$
(3) *Simulate*

$$\begin{aligned}
\beta^{(i+1)} \quad &\text{from} \quad f_{\mathcal{N}}\left(\beta|\hat{\beta}(\breve{X}_1^{(i)}, ..., \breve{X}_n^{(i)}), (b_0 + \frac{}{(\sigma^2)^{(i)}}\sum \tilde{y}_j^2)^{-1}\right) \\
(\sigma^2)^{(i+1)} \quad &\text{from} \quad f_{IG}\left(\sigma^2|\frac{w_0+n}{2}, \frac{\delta_0 + \delta_n(\beta^{(i+1)}, \breve{X}^{(i)}, \tilde{y})}{2}\right) \\
\breve{X}_j^{(i+1)} \quad &\text{from} \quad f_j(\cdot|\tilde{h}_j, \tilde{y}_j, \beta^{(i+1)}, (\sigma^2)^{(i+1)}) \text{ for } j = 1, .., n
\end{aligned}$$

(4) *Set i=i+1, if* $i + 1 < I$, *go to step 3, where* I *is the number of iterations.*

Taking again the mode of β *and* σ^2 *gives the ML-estimator. Sampling is very easy, no complicated full conditional distribution is involved. We do not have to derive or sample from the full conditional, here of* β*:*

$$\begin{aligned}
p(\beta|\sigma^2, \tilde{h}) \quad &\propto \quad f_{IG}(\sigma^2|w_0/2, \delta_0/2) f_{\mathcal{N}}(\beta|\beta_0, b_0^{-1}) \prod_j \Phi(\beta\tilde{y}_j/\sigma)^{\tilde{h}_j}(1 - \Phi(\beta\tilde{y}_j/\sigma))^{1-\tilde{h}_j} \\
&\propto \quad f_{\mathcal{N}}(\beta|\beta_0, b_0^{-1}) \prod_j \Phi(\beta\tilde{y}_j/\sigma)^{\tilde{h}_j}(1 - \Phi(\beta\tilde{y}_j/\sigma))^{1-\tilde{h}_j}
\end{aligned}$$

In addition, one gains an estimator for the latent variable! So one knows how close a switch from 0 to 1 actually is!

■

In the next section, at first we start by directly sampling the parameters of the loss distribution without sampling the asset value process X. This was originally suggested by Gössl (2005); Wendin (2006). Later on $\breve{X}$ will symbolize the unknown asset value process X.

1.2. **Comparison in a One Factor Model.** The advantage of the Bayesian approach is that it can be easily extended to more general distributions, whereas e.g. the moment estimator (BO) in Bluhm and Overbeck (2003) breaks down if the distribution changes or even if we want to estimate one asset correlation over the whole portfolio not depending on rating classes. The latter makes sense as correlation usually does not depend on the rating class. Classical ML-estimation (ML-MCMC indexed by MCMC) with respect to the likelihood $\mathrm{L} = P_{p,\rho}$ is slightly more flexible (see (4) below). We have to integrate over Y. Consequently, numerical methods have to be used. However, if the distribution of the common factor Y becomes complex, numerical estimation is almost intractable. In a Bayesian approach both problems, numerical evaluation of the classical estimator and more complex distributions, are solved by simultaneously sampling the latent factor y. The trick here is to additionally estimate the latent variable y by including it into the parameter space. Then the obligors are independent of each other and the integral over y does not have to be evaluated. We consider $\pi(y_1, ..., y_n, p_A, .., p_{CCC}, \rho, |\mathbf{l}, \mathbf{N})$. Note, y is not the asset value process.

y can be compared with $\tilde{y}$ (but unknown!) and not with $\breve{X}$ from Example 1.2. We next recall the Gaussian framework and compare classical and Bayesian estimation.

1.2.1. *Model.* We formulate the model slightly more general: as usual on a one year horizon the asset value of firm i at t is given by

$$X_{i,t} = \sqrt{\rho}Y_t + \sqrt{1-\rho}Z_{i,t},\ i = 1,..,N_t,\ t = 1,...,n$$

where $\rho \in (0,1)$, $Y_t \overset{iid}{\sim} F$, and $Z_{i,t} \overset{iid}{\sim} G$ and F and G distribution functions. If $F = G$ and F stable under convolution (denoted by $\tilde{F}$), we have

$$D_{i,t} = \tilde{F}^{-1}(p_{i,t}) = \tilde{F}^{-1}(p_{R_j}) =: D_j, \tag{3}$$

where firm i is in rating class R_j at t, and p_{R_j} the constant default probability of rating class R_j. We have

$$\begin{aligned} P(\text{asset } i \text{ defaults at } t|Y_t) &= P(X_{i,t} < D_{i,t}|Y_t) = p_{j|Y_t} \\ \Rightarrow p_{j|Y_t} &= G\left(\frac{D_j - \sqrt{\rho}Y_t}{\sqrt{1-\rho}}\right). \end{aligned}$$

With the loss indicator $L_{j,t|Y_t} = \sum_{i:\ k_{i,t}=R_j}^{N_t} 1_{(X_{i,t}<D_j|Y_t)}$, we obtain:

$$P_{p,\rho}(L_{1,t} = l_{1,t},...,L_{K,t} = l_{K,t}|Y_t) = \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|Y_t}, l_{R_j,t})$$

$$P_{p,\rho}(L_{1,t} = l_{1,t},...,L_{K,t} = l_{K,t}) = \int \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t})dF(y_t) \tag{4}$$

where $N_t := (N_{1,t},...,N_{K,t})$ represents the number of companies in one rating class and $l_t := (l_{1,t},...,l_{K,t})$ the number of defaults in these rating classes taken from S&P's default study (the study provides the number of obligors and defaults in the rating buckets A,BBB, BB, B, CCC). As already mentioned a numerical treatment of classical ML- estimation (ML-MCMC) is difficult already for the standard normal distribution, $F = \Phi$. If Y_t assumed to have an arbitrary distribution, optimization can be become infeasible. Bayesian algorithms offer a smart alternative for every distribution:

1.2.2. *Estimation.* Estimations works as follows. Prior distributions of $\mathbf{p}$ and ρ are set to be uniform on $[0,1]$. The prior distribution of y_t depends on the assumption on F, we make on Y_t in the one factor model, so e.g. set $F = \Phi$. We then generate samples from $\pi(\mathbf{p}, \rho, \mathbf{y}|\mathbf{l}, \mathbf{N})$: $(\mathbf{p}^{(i)}, \rho^{(i)}, \mathbf{y}^{(i)})_i$, i=burn,..., Iterations=I by Gibbs Sampling. For a large enough burn in, the $I - burn \times 2$-vector $(\mathbf{p}^{(i)}, \rho^{(i)})_i$, is then approximately a sample of $\pi(\mathbf{p}, \rho|\mathbf{l}, \mathbf{N})$ and taking the mode of $(\mathbf{p}^{(i)}, \rho^{(i)})_i$, is by definition a good approximation of

the ML- estimator:

$$\max_{\mathbf{p},\rho} \prod_{t=1}^{n} P_{\mathbf{p},\rho}^{F_{Y_t}}(L_{1,t} = l_{1,t}, ..., L_{K,t} = l_{K,t})$$

$$P_{\mathbf{p},\rho}^{F_{Y_t}}(L_{1,t} = l_{1,t}, ..., L_{K,t} = l_{K,t}) = \int \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) dF_{Y_t}(y_t)$$

for an arbitrary distribution F_{Y_t} with density f_{Y_t}. As (Y_t) are iid, we have $f := f_{Y_t}$ for all t. The algorithm works because for a constant $C > 0$ we obtain:

$$\pi(\mathbf{p}, \rho, \mathbf{y}|\mathbf{N}, \mathbf{l}) = \frac{\prod_{t=1}^{n} \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) f(y_t)}{\int \prod_{t=1}^{n} \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) f(y_t) d(\mathbf{p}, \rho, \mathbf{y})}$$

$$\Rightarrow \quad \pi(\mathbf{p}, \rho|\mathbf{N}, \mathbf{l}) = C \int \prod_{t=1}^{n} \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) f(y_t) d\mathbf{y}$$

$$\overset{Y_t \sim iid}{\Rightarrow} \quad \pi(\mathbf{p}, \rho|\mathbf{N}, \mathbf{l}) = C \prod_{t=1}^{n} \int \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) f(y_t) dy_t$$

$$\overset{Y_t \sim iid}{\Rightarrow} \quad \pi(\mathbf{p}, \rho|\mathbf{N}, \mathbf{l}) = C \prod_{t=1}^{n} P_{\mathbf{p},\rho}^{F}(L_{1,t} = l_{1,t}, ..., L_{K,t} = l_{K,t})$$

by Fubini's Theorem and Y_t iid. All conditions of Proposition 1.1 are satisfied, because $|Bin(\cdot)| \leq 1$. Sampling from π can be performed by Gibbs Sampling. Changing the prior does not matter as long as all prior densities are differentiable.

Next note, in practical implementation the algorithm seems to improve if the threshold D is sampled instead of p. This works as follows. Assume $F = G$ and F is stable under convolution, e.g $F = G = \Phi$. So the distribution of Y stays in the same class and can be calculated say $\tilde{F}$. We have $p = \tilde{F}(D)$, e.g. $p = \Phi(D)$. To obtain the same result we have to change the prior correctly. By the transformation theorem and because the prior density of p, f_p, is uniform and independent over all rating classes, in the Gaussian case we obtain the following prior distribution of D for every $j = 1, ..., K$:

$$f_D(D_j) = f_p(\Phi(D_j)) \left| \frac{d\Phi}{dD_j} \right| = 1\phi(D_j).$$

In general we have to replace Φ by $\tilde{F}$ and ϕ by the derivative of $\tilde{F}$, provided it exists. Finally, sampling from the posterior

$$\pi(\mathbf{D}, \rho, \mathbf{y}|\mathbf{N}, \mathbf{l}) \propto \prod_{t=1}^{n} P(\mathbf{L}_t = l_t|N_t, \mathbf{D}, \rho, y_t) p(\mathbf{D}) p(\rho) f(\mathbf{y}) \tag{5}$$

with likelihood

$$P_{D,\rho}(\mathbf{L}_t = l_t|y_t) = \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}),\ p_{j|y_t} = G\left(\frac{D_j - \sqrt{\rho}y_t}{\sqrt{1-\rho}}\right), \quad (6)$$

priors

$$D_{R_j} \overset{iid}{\sim} \tilde{F},\ \text{e.g. } \mathbf{N}(0,1),\ j = 1, ..., K,\ y_t \overset{iid}{\sim} F,\ \rho \sim \mathbf{U}(0,1),$$

and setting $p = \tilde{F}(D)$ gives the same result as before. $\mathbf{U}(0,1)$ denotes the the uniform distribution and note that $\mathbf{U}(0,1) = \mathbf{B}(1,1)$, where $\mathbf{B}(a,b)$ denotes the beta distribution.

Roughly speaking, priors leading to an ML-estimate are called "uniformative priors". In fact, one has no idea where the parameter might be in the parameter space. The probability of the parameter ρ or p_{R_j} is set uniform over the parameter space $(0,1)$. We know however that p_{R_j} is greater than p_{R_k} when R_k is the better rating. So taking in mind that modes of the resulting Markov chains are no longer ML-estimates, Bayesian methodology further gives the possibility to include an "informative" prior belief into the estimation:

$$\begin{aligned} p_{R_j} &\overset{iid}{\sim} \mathbf{B}(a_j, b_j) \text{ resp. } D_{R_j} \overset{iid}{\sim} f_D(a_j, b_j, \tilde{F}) \\ &\text{or} \quad D_{R_j} \overset{iid}{\sim} \mathbf{N}(\mu_j, \sigma_j^2),\ j = 1, ..., K \\ y_t &\overset{iid}{\sim} F \text{ (e.g. } \mathcal{N}(0,1)/NIG(\psi)),\ t = 1, ..., n, \\ \rho &\sim \mathbf{B}(a, b) \end{aligned}$$

where $f_D(a_j, b_j, \tilde{F}) = f_{\mathbf{B}(a_j,b_j)}(\tilde{F}(D))|\tilde{f}(D)|$. a_j and b_j are supposed to be implied by a certain preinformation on p_j. A potential method to obtain this information is discussed next.

A natural idea to impose such a preinformation is to derive default probabilities from equity time series by applying an asset value model, see e.g. Bluhm et al. (2002). Moody's KMV[1] (henceforth KMV) exactly does that by issuing a number called *expected default frequency* (EDF). Although the exact way of calculation is not published by KMV, theoretically KMV-EDFs can be calculated every day from 1981-2005 as in S&P's default study. Whereas default rates provide a single data point per year only, equity time series offer large data sets over the last years. In fact, apart from potential problems concerning the availability of those KMV-EDFs, KMV-EDFs from 1981-2005 would just present a competitive method to calculate default probabilities for this period. So imposing a preinformation more related to

[1]Credit Monitor, CreditEdge, CreditEdge Plus, CreditMark, DealAnalyzer, EDFCalc, Private Firm Model, Portfolio Preprocessor, GCorr, the Moody's KMV logo, Moody's KMV Financial Analyst, Moody's KMV LossCalc, Moody's KMV Portfolio Manager, Moody's KMV Risk Advisor, Moody's KMV RiskCalc, RiskAnalyst, Expected Default Frequency, and EDF are trademarks owned by of MIS Quality Management Corp. and used under license by Moody's KMV Company.

short-term-information seems to be more interesting. A prior belief over the current development of selected companies can be included provided KMV-EDFs exist. In particular, this is possible for traded companies only.

In detail, suppose a bank might have bought KMV-EDFs for their listed companies weekly x years before the day they like to analyze their portfolio. A calibration of a_j and b_j to these KMV-EDFs, then includes an own portfolio information from the latest x years into the estimation of the default probability p_j.

Finally, to actually obtain a_j and b_j, $j = 1, .., K$, we randomly choose a sufficiently large number of companies from every rating bucket (AAA-CCC) over a time horizon of three years (here weekly in the years 2004-2006) and build the average μ_j and the standard deviation σ_j of KMV-EDFs in every rating class.[2] Afterwards the implied parameters a_j and b_j are established as follows:

We know that

$$\mu := E(\check{X}) = \frac{a_j}{a_j + b_j},\ \sigma_j^2 := V(\check{X}) = \frac{a_j b_j}{(a_j + b_j)^2(a_j + b_j + 1)}, \tag{7}$$

where

$$\check{X} \sim \mathbf{B}(a_j, b_j),\ a_j > 0,\ b_j > 0 \tag{8}$$

and so

$$a_j = \frac{(1-\mu_j)\mu_j - \sigma_j^2}{\sigma_j^2 + \sigma_j^2(\frac{1-\mu_j}{\mu_j})},\ b_j = \frac{1-\mu_j}{\mu_j} a_j. \tag{9}$$

To guarantee positivity of a_j and b_j, we assume $\sigma_j^2 < \mu_j(1-\mu_j)$. This is satisfied in all cases, we consider, see Table 1.

Rating class	μ	σ^2	$\mu(1-\mu)$	a	b
AAA	0.0002	3.99E-08	2.00E-04	1.0020	5009
AA	0.0007	1.93E-06	7.00E-04	0.2531	361.3
A	0.001	4.98E-06	9.99E-04	0.2000	199.8
BBB	0.0014	6.295E-06	0.0014	0.3100	221.1
BB	0.0066	1.40E-04	0.0066	0.3043	45.79
B	0.0242	8.79E-04	0.0236	0.6419	25.88
CCC	0.0779	5.57E-03	0.0718	1.0115	11.97

TABLE 1. Priors via KMV-EDFs

Interestingly the obtained "pre"-default probabilities are a lot tighter than usual. This might depend on the chosen period and/or the random choice of the considered companies. As company names cannot be stated here, we refrain from a further interpretation.

[2]As a KMV-EDF is a commercially distributed product, the applied procedure is therefore chosen such that no link to an actual KMV-EDF of a particular company can be drawn. Further details about the considered portfolio cannot be published here.

1.2.3. *Bayesian Estimation: Binomial Gibbs Sampler.* For a better understanding of the algorithm, we present a simple example, how estimation works in McNeil and Wendin (2006)/Wendin (2006):

(1) $\theta = (D_A, D_{BB}, \rho, y_1, ..., y_5)$, set $i = 0$
(2) Set $D^{(0)} = (0.5\, 0.5)'$, $\rho^{(0)} = 0.1$, $(y_1, ..., y_5)^{(0)} = (0\, 1\, 2\, -0.5\, -1)$
(3) Simulate

$$\begin{aligned} D_A^{(i+1)} &\quad \text{from} \quad \pi(D_A | D_{BB}^{(i)}, \rho^{(i)}, y_1^{(i)}, .., , y_5^{(i)}, \mathbf{l}, \mathbf{N}) \\ D_{BB}^{(i+1)} &\quad \text{from} \quad \pi(D_{BB} | D_A^{(i+1)} \rho^{(i)}, y_1^{(i)}, ..., y_5^{(i)}, \mathbf{l}, \mathbf{N}) \\ &\quad \vdots \\ y_5^{(i+1)} &\quad \text{from} \quad \pi(y_5 | D_A^{(i+1)} \rho^{(i+1)}, y_1^{(i+1)}, .., , y_4^{(i+1)}, \mathbf{l}, \mathbf{N}) \end{aligned}$$

(4) set i=i+1, if $i + 1 < I$, go to step 3, where I is the number of iterations.

For i large enough ($i \geq \bar{b}$ =burn in), $\theta^{(i)}$ is approximately a sample of the posterior function $\pi(\mathbf{D}, \rho, \mathbf{y} | \mathbf{l}, \mathbf{N})$. Full conditional distributions of π are derived as follows. As a function of y_s, (e.g. y_1) given $\mathbf{D}, \rho, \mathbf{N}, \mathbf{l}$:

$$\begin{aligned} \pi(y_s | \mathbf{D}, \rho, \mathbf{N}, \mathbf{l}) &\propto \prod_{t=1}^{n} \prod_{j=1}^{K} Bin(N_{R_j,t}, p_{j|y_t}, l_{R_j,t}) p(\mathbf{D}) p(\rho) f(\mathbf{y}) \\ &\propto \prod_{j=1}^{K} Bin(N_{R_j,s}, p_{j|y_s,\rho,p_{R_j}}, l_{R_j,s}) f(y_s) \end{aligned} \tag{10}$$

where $f(y_s) = \phi(y_s)$ and

$$p_{j|y}(y_s) = \Phi\left(\frac{D_j - \sqrt{\rho} y}{\sqrt{1-\rho}} \right)(y_s) =: p_{j|y_s,\rho,D_j}. \tag{11}$$

Samples from the non-standardized density $\pi(y_s | \mathbf{D}, \rho, \mathbf{N}, \mathbf{l})$ can be drawn by ARMS-Algorithm, see also McNeil and Wendin (2006)/Wendin (2006) and Gilks et al. (1995). Analogous derivations yield full conditionals for p, ρ, and all other y_t. In a general framework, we increase 5 to n, and include the remaining rating classes.

1.2.4. *Comparison in a Gaussian One factor model.* Suppose we stay in a Gaussian framework. We can then summarize that we have already discussed three different estimators. The classical ML-estimator is derived from the likelihood (6) by applying Markov Chain Monte Carlo methods, denoted by ML-MCMC and the two estimators discussed in Niethammer and Overbeck (2008): in an infinite granular model, the ML-estimator (ML) can be solved analytically; the moment estimator (BO) can be derived semianalytically (we need to solve an equation for ρ). For better comparability, we calculate the ML-estimator in an infinite granular model also with fixed ρ

determined by the moment estimation:

$$\hat{p}_k = \Phi\left(\frac{\sqrt{1-\hat{\rho}_k}}{n}\sum_t \Phi^{-1}\left(\frac{l_{k,t}}{N_{k,t}}\right)\right),$$

where $\hat{\rho} = \hat{\rho}_{BO}$. This estimator is denoted by MLρ. In addition, we can change priors of e.g. $p(\rho) = \mathbf{B}(9, 90)$ (i.e. the distribution is quite symmetric and the mean of the prior is 0.091). We introduce B as an acronym.

Estimators obtained from S&P default study (2006) are shown in Table 2. ML-MCMC and B are calculated as the mode of the Markov Chain obtained from the above Gibbs-sampler. For a comparison, we also display the mean for the ML-estimator (ML-MCMC).

θ	mean(MCMC)	ML-MCMC	B	BO	MLρ	ML
p_{CCC}	0.2754	0.2686	0.2681	0.2292	0.2073	0.2334
p_B	0.0589	0.0555	0.0561	0.0512	0.0526	0.0519
p_{BB}	0.0134	0.0111	0.0123	0.0117	0.0099	0.0134
p_{BBB}	0.0036	0.0029	0.0031	0.0027	0.0015	0.0033
p_A	0.0006	0.001	0.001	0.0004	0.0004	0.0004
ρ	0.0955	0.1012	0.1012	0.0966*	0.0966*	0.1805*

Sample size=20000, burn in=1000, autocorrelation=a=2.
*=this number is the average of the ρ over all ratings classes.

TABLE 2. A comparison of estimators in the Gaussian one factor model

Finally, Figure 1 presents a comparison of the Bayesian estimator B and the moment estimator BO w.r.t. the common factor estimation. For the Bayesian estimator we take the mean over the posterior of $\mathbf{y}$. The graph for the ML-estimator (ML-MCMC, $a = 1, b = 1$) is quite similar to the Bayesian estimator (B, $a = 9$, $b = 90$). The common factor for moment estimator $\hat{y}$ is obtained from the estimator $(\hat{p}_1, ..., \hat{p}_K, \hat{\rho}_1, ..., \hat{\rho}_K)$ as in Bluhm and Overbeck (2003), with $g_k(s_t) = \Phi\left(\frac{\Phi^{-1}(\hat{p}_k)-\sqrt{\hat{\rho}_k}s_t}{\sqrt{1-\hat{\rho}_k}}\right)$:

$$(\hat{y}_1, ..., \hat{y}_n) = \arg\min_{\psi, s_1, ..., s_T} \sum_{t=1}^{n}\sum_{k=1}^{K}\left|\frac{l_{t,k}}{N_{t,k}} - g_k(s_t)\right|^2$$

1.3. **Bayesian Extensions: Random Factor Loading and Stochastic Volatility.** In a Bayesian framework it is now easy to build more complex models. In particular, we want to discuss models including time dependent correlation driven by the common factor y. McNeil and Wendin (2006)/Wendin (2006) discusses ARMA processes (autoregressive moving average, $y_t = \alpha y_{t-1} + \epsilon_t$) and the influence of macroeconomic factors (CFNAI[3], an American activity index $y_t = \beta CFNAI_t + \epsilon_t$) on loss rates. We

[3]CFNAI= Chigago Fed National Activity Index can be downloaded from http://www.chicagofed.org/economic_research_and_data/cfnai.cfm

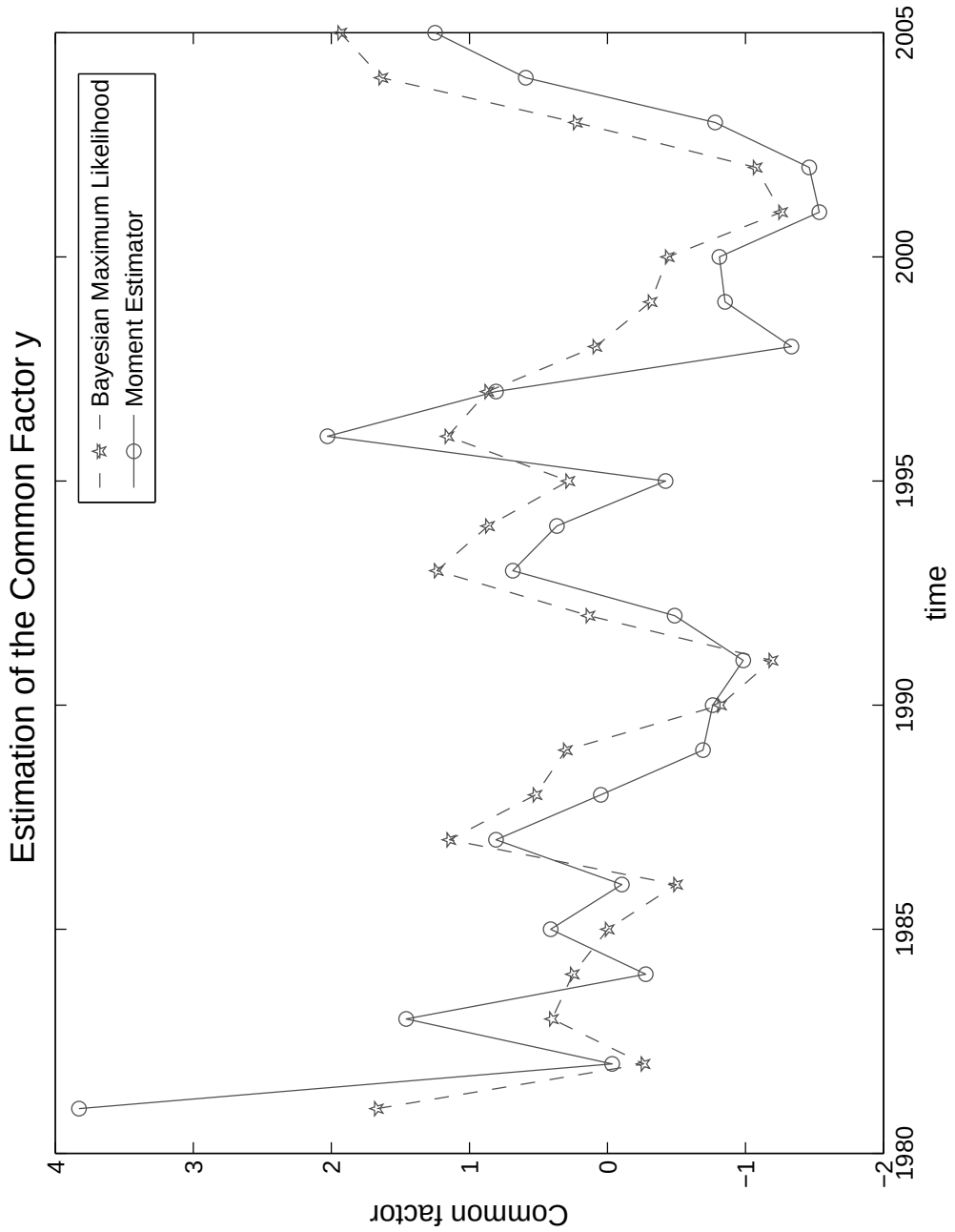


FIGURE 1. Comparison of latent factors: Moment Estimation (BO) vs. Bayesian "Maximum Likelihood" (B, mean)

also implemented them, but there is no qualitative difference on our correlation study, i.e. on the fluctuation of ρ. A further discussion including industry indeces and equity time series is postponed to Section 4.

We restrict our consideration upon the Random factor loading, see below, and the following Stochastic volatility model:

$$y_t|h_t \overset{i.i.d.}{\sim} \mathcal{N}(0, e^{h_t}),\ h_t - \mu_h = \phi_h(h_{t-1} - \mu_h) + \sigma_h u_t,\ u_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

where $\tilde{\psi} = (\mu_h, \phi_h, \sigma_h)$. In principle, we just have to exchange $f(y_t)$ by $p(y_t|h_t)p(h_t|\tilde{\psi})$. Several methods to sample h_t given y_t are described in Kim et al. (1998). For sake of completeness, one can additionally set $Z_{i,t}|h_{2t} \sim \mathcal{N}(0, e^{h_{2t}})$, where h_{2t} is independent of $h_{1t} := h_t$. A sampling method (Kalman filter and simulation smoother of de Jong and Shephard (1995)) is described in Chib et al. (2002) and Chib et al. (2006) in the multidimensional case. In their multi factor model, h_{2t} drives stochastic volatility whereas h_{1t} drives stochastic correlation.

We think, the latter model is too overloaded for the considered model. Not least as we assume that our variance is standardized to 1. We therefore stick to a time dependent variance of y only. We postpone the stochastic volatility model to Section 2, where we also introduce a new (asset value) algorithm. This algorithm is more stable and therefore yields more reliable results. Although random factor loading works well with the Binomial algorithm:

1.3.1. *Random Factor Loading: Two State Time Dependent Correlation?* We start with the model. Correlation is set dependent on the common factor

$$\rho_t|y_t = \begin{cases} \rho_L & \text{if} \quad y_t > \text{level} \\ \rho_H & \text{if} \quad y_t \leq \text{level} \end{cases} \tag{12}$$

with full conditionals:

$$\begin{aligned} p(\rho_H|\mathbf{D}, \mathbf{y}, \mathbf{N}, \mathbf{l}) &\propto \prod_{t=1}^{n}\prod_{j=1}^{K} Bin(N_{j,t}, p_{j|y_t,\rho_t}, l_{j,t})p(\mathbf{D})p(\rho_t)p(\mathbf{y}) \\ &\propto \prod_{t:\ y_t \leq \text{level}}\prod_{j=1}^{K} Bin(N_{j,t}, p_{j|y_t,\rho_H,D_j}, l_{j,t})p(\rho_H) \end{aligned}$$

similar for ρ_L and so

$$\rho_t^{(i)} = \mathtt{1}_{y_t \leq level}\rho_H^{(i)} + (1 - \mathtt{1}_{y_t \leq level})\rho_L^{(i)},$$

where i denotes the ith iteration.

In the following S&P loss rate example we set level$= -1$. That ensures that enough $y_t^{(i)}$ are below the level to update ρ_H. We perform three simulation studies (2 x data, simulation: 1 x 1-to-1 test). We set $I = 100000$. The prior on p is chosen to be uniform over the interval $[0, 1]$. Moreover, we face the following set up:

(1) Data: S&P default study

Model: two regimes of ρ, (ρ_L, ρ_H), $B(1.5, 6)$ prior on ρ_H with mean 0.2, and iid Gaussian prior on y_t. Firstly we take the same prior on ρ_L as on ρ_H. Secondly, we choose $B(9, 90)$ with mean 0.0909.

(2) Data: S&P default study

Model: two regimes of ρ, (ρ_L, ρ_H), $B(9, 90)$ prior on ρ_H and ρ_L with mean 0.09, and iid Gaussian prior on y_t.

Optionally an observed factor CFNAI or/and an AR(1) on y are included.

(3) Data: simulated data with $\rho_L = 0.08$ and $\rho_H = 0.25$.

Model: model and priors are chosen as in 2. Optionally, we simulate an observed factor saison=4/AR(1) process, otherwise y_t is set iid normal.

We obtain the following results. By plotting the chains of the parameters, convergence is approximately reached after 4000 iterations. Convergence of ρ_H takes a bit longer. All included options give no significant different results. In detail, we get:

ρ	prior mean	prior sd.	post. mean	post. sd.	tendency
ρ_L (1a. S&P)	0.200	0.1372	0.09433	0.03354	ρ_H <prior
ρ_H (1a. S&P)	0.200	0.1372	0.09350	0.06103	ρ_H <prior
ρ_L (1b. S&P)	0.091	0.0939	0.08735	0.02156	ρ_H <prior
ρ_H (1b. S&P)	0.200	0.1372	0.09321	0.05673	ρ_H <prior
ρ_L (2.S&P)	0.091	0.0939	0.08472	0.02072	ρ_H <prior
ρ_H (2.S&P)	0.091	0.0939	0.08352	0.02388	ρ_H <prior
ρ_L (3.Sim)	0.091	0.0939	0.07770	0.01480	ρ_H <prior
ρ_H (3.Sim)	0.091	0.0939	0.20310	0.02540	ρ_H >prior

In all studies we impose a normal prior $N(0, 1)$. No observed saisonal factor or AR-process is included.

TABLE 3. Prior analysis: a comparison

In **study 1**, $p_{|y_t}$ (here the mean of $p_{|y_t^{(i)}}$ is meant see Remark 1.1) fits l_t/N_t reasonable, but a bit worse than in study 2. In addition, data show a tendency to a decreasing ρ_H. The mean of ρ_H suggested by the prior is not confirmed, it decreases from 0.2 to 0.09. There is no much impact if the prior on ρ_L changes. The standard deviation using the second prior is naturally smaller as the first prior. It seems that the latter prior fits the truth better than the first one.

In **study 2**, the mean of $p_{|y_t^{(i)}}$ fits l_t/N_t well. ρ_H stays around the prior belief (no tendency to a higher ρ_H than imposed by the mean of the prior distribution). There is no significant difference to ρ_L. If y has a AR(1) prior

or if CFNAI is excluded, we get the same stylized fact as for an iid prior, i.e. $\alpha = 0$ in $y_t = \alpha y_{t-1} + \epsilon_t$.

Our **simulation study**, (3), confirms the results of study 2. All parameters are recovered, independently if (prior of) y is an ARMA processes or driven by a seasonal component. Loss rates are fitted very well. But ρ_H and ρ_L are close to 0.25 and 0.08 (level -1). The deviation from 0.25 is due to the prior belief. The standard deviation is quite small and the mean 0.0909 is quite far away from 0.25. This fact is only completely ruled out when n goes to infinity!

Summarizing, if there were a random factor loading in the data, the algorithm would find two different ρ, although a prior belief suggest that ρ are equal, see Table 3!

2. Asset Value Algorithm

This section is devoted to a new Markov Chain Monte Carlo algorithm. It is based on the same Gaussian one factor model as before, but also includes stochastic volatility e^h of the factor y, i.e. $y|h \sim \mathcal{N}(0, e^h)$. A sensitivity analysis shows that the algorithm is a lot more stable than the already discussed Binomial approach. Stochastic volatility in the common factor can be examined and is applied to S&P default study (2006). An idea how the algorithm can be extended to a multi factor setting is presented.

2.1. AVA: Why Implementing another Algorithm?

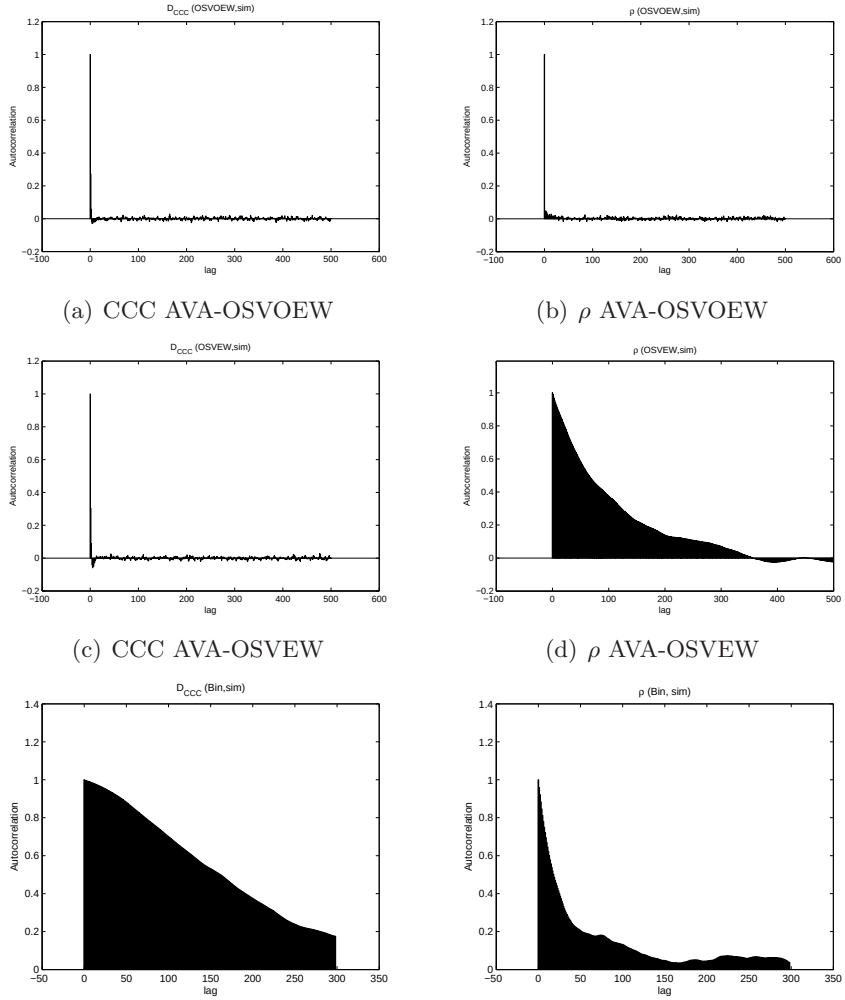
If the model becomes complex, in the above Binomial approach we face the following problems:

- Sensitivity to initial values (convergence takes very long, ARMS does not perform well) and to the transformation $p = \Phi(D)$
- High autocorrelation within chains
- Stochastic volatility does not work well

We therefore develop a new Gibbs Sampler, called Asset Value Algorithm (AVA). Autocorrelation within the chains is reduced essentially, see Figure 2. The algorithm is stable to examine cases with a complex time dependent structure in loss rates, as stochastic volatility models:

2.2. Asset Value Algorithm (AVA).

The essential device of AVA is to include the unknown asset value process X in our considerations. The algorithm is then more natural as e.g. equity and asset value process can be compared, see Figure 12. As in Example 1.2, the main trick is to sample X given ρ, y, and the data $\mathbf{N}, \mathbf{l}$. Knowing X then carries sufficient information to sample all other parameters p, y, D via a usual regression. In this step knowledge of $\mathbf{N}, \mathbf{l}$ is not needed as we condition on X. In principle, we apply the idea explained in Example 1.2 by setting $X = \breve{X}$ and $y = \tilde{y}$. However, we build groups implying a switch from a Bernoulli to a Binomial experiment. Furthermore, y is not known, it also has to be sampled. That is no problem,



(a) CCC AVA-OSVOEW (b) ρ AVA-OSVOEW

(c) CCC AVA-OSVEW (d) ρ AVA-OSVEW

(e) CCC BIN (f) ρ BIN

FIGURE 2. Autocorrelation within Markov Chains - a comparison
All variants are sampled without taking stochastic volatility into account, denoted by (OSV). BIN is an acronym for the Binomial Gibbs-Sampler. AVA-OSVOEW and AVA-OSVEW are the two model variants of AVA as described in (14) and (15).

because in every step of the sampler we condition on the other parameters. Technically, given the grouped X, the regression equation then exactly fits into the multidimensional model given in Chib et al. (2006). Full conditionals can be calculated as described there, similar but a bit more general

as in our example. We start defining a "rating class index"

$$\tilde{X}_{kt} = -D_k + \sqrt{\rho} y_t + \sqrt{1-\rho}\frac{1}{N_{kt}}\sum_{i=1}^{N_{kt}} Z_{i,t},\ k = A, ..., CCC$$

and a "world index"'

$$X_t = \frac{1}{N_t}\sum_{k=1}^{K}(N_{tk}\tilde{X}_{kt} + N_{tk}D_k) \Rightarrow X_t = \sqrt{\rho} y_t + \sqrt{1-\rho}\bar{Z}_t \qquad (13)$$

where $N_t = \sum_{k=1}^{K} N_{tk}$, $\bar{Z}_t \sim \mathcal{N}(0, 1/N_t)$. The algorithm then works as follows:

(1) Set starting values and number of iterations I, $i = 0$
(2) Simulation of $\rho^{(i)}$ given $X^{(i-1)}, h^{(i-1)}, \mathbf{N}$ (or also $y^{(i-1)}$)
(3) Simulation of $y^{(i)}$ given $\rho^{(i)}, h^{(i-1)}, \mathbf{N}, X^{(i-1)}$
(4) Generate $h^{(i)}$ given $y^{(i)}$
(5) Simulation of $D_k^{(i)}$ given $\tilde{X}_k^{(i-1)}, \rho^{(i)}, h^{(i)}, \mathbf{N}$
(6) Simulation of $\tilde{X}_k^{(i)}$ given $\mathbf{N}, \mathbf{l}, y^{(i)}, \rho^{(i)}, h^{(i)}, D_k^{(i)}$ and calculate $X^{(i)}$
(7) Set $i = i + 1$, go to 2 if $i \leq I$

In detail, with $f_{\mathcal{N}}(X|\mu, \sigma^2)$ denoting the Gaussian density with mean μ and variance σ^2 we follow the following algorithm:

(1) Set starting value of $y_t^{(0)}, h_t^{(0)}, \tilde{X}_{tk}^{(0)}$, $D_k^{(0)}, k = 1, ..., K$ for $t = 1, ..., n$, set $i = 1$, set maximum number of iteration I
(2) Sample $\rho^{(i)}$ using ARMS with full conditional distribution, similar to Chib et al. (2006). We propose two alternatives:

variant 1 (OEW):

$$p(\rho^{(i)}|X) \propto p(\rho)\prod_{t=1}^{n} f_{\mathcal{N}}(X_t^{(i-1)}|0, \rho e^{h_{1t}^{(i-1)}} + (1-\rho)e^{h_{2t}^{(i-1)}}\frac{1}{N_t}) \qquad (14)$$

or when the distribution of y_t is purely prior, variant 2 (EW):

$$p(\rho^{(i)}|X) \propto p(\rho)\prod_{t=1}^{n} f_{\mathcal{N}}(X_t^{(i-1)}|\sqrt{\rho} y_t^{(i-1)}, (1-\rho)e^{h_{2t}^{(i-1)}}\frac{1}{N_t}) \qquad (15)$$

(3) Sample $y_t^{(i)} \sim \mathcal{N}(\mu_y^{(i)}, (\sigma_y^2)^{(i)})$ where

$$\mu_y^{(i)} = (\sigma_y^2)^{(i)}\sqrt{\rho}(1-\rho)^{-1}N_t e^{-h_{2t}^{(i-1)}} X_t^{(i-1)}, \qquad (16)$$

$$(\sigma_y^2)^{(i)} = (\rho(1-\rho)^{-1}N_t e^{-h_{2t}^{(i-1)}} + e^{-h_{1t}^{(i-1)}})^{-1} \qquad (17)$$

with $\rho = \rho^{(i)}$. This works as likelihood and prior are normal and therefore conjugate.
(4) Generate $h_t^{(i)}$ given $y_t^{(i)}$, $t = 1, ..., n$ using Kim et al. (1998) (for $h_{2t} \equiv 0$ one could use Chib et al. (2002) but that is not done in our study).

(5) Sample $D_k^{(i)}, k = 1, ..., K$ as follows

$$p(D_k^{(i)}|X) \propto p(D_k) \prod_{t=1}^{T} f_{\mathcal{N}}(\tilde{X}_{tk}^{(i-1)}| - D_k, \rho e^{h_{1t}^{(i)}} + (1-\rho)e^{h_{2t}^{(i)}} \frac{1}{N_{tk}})$$

with $\rho = \rho^{(i)}$. If $p(D_k) = f_{\mathcal{N}}(\mu_k, \sigma_k^2)$, sampling without ARMS is possible (we face conjugate distributions!).

(6) Generate $\tilde{X}_{k,t}^{(i)} = \frac{1}{N_{k,t}}(\sum_{j=1}^{l_{k,t}} X_{i,t}^{D_k} + \sum_{j=1}^{N_{k,t}-l_{k,t}} X_{i,t}^{ND_k})$ where

$$\begin{aligned} X_{i,t}^{D_k}|h_t, \rho, y_t, D_k &\sim T\mathcal{N}_{(-\infty,0]}(-D_k + \sqrt{\rho}y_t, e^{h_{2t}}(1-\rho)) \\ X_{i,t}^{ND_k}|h_t, \rho, y_t, D_k &\sim T\mathcal{N}_{(0,\infty)}(-D_k + \sqrt{\rho}y_t, e^{h_{2t}}(1-\rho)) \end{aligned}$$

with $h_t = h_t^{(i)}, \rho = \rho^{(i)}, y_t = y_t^{(i)}, D_k = D_k^{(i)}$, and calculate $X_t^{(i)}$ given in (13).

(7) Set $i = i + 1$, if $i \leq I$ go to step (2)

Note, in (2) the second method (OEW) performs best, correlation between the chains is lowest. Derivation of the full conditional distribution in step 2 and 5 are straightforward as in the Binomial approach (BIN) before. In step 3, the world index is known, we face a usual one dimensional regression model. The full conditional distribution can be deduced from the multidimensional regression model in Chib et al. (2006) or Chib and Greenberg (1996, Example 3.1). Step 4 is described in Kim et al. (1998), we use (SV) as a short hand if h is included, otherwise we set (OSV). Step 6 then works as in Example 1.2.

Autocorrelation presented in Figure 2 does not change when including stochastic volatility in AVA. However, for the Binomial approach we do not obtain reasonable results. The increase in autocorrelation of AVA is negligable whereas BIN shows an extreme increase in autocorrelation, when including stochastic volatility.

2.3. **Time dependent Stochastic Correlation? An Empirical Study.** We perform a second simulation study. We examine if the common factor is generated by a time dependent variance.

2.3.1. *Set up.* Recall,

$$h_t = \phi_h h_{t-1} + \sigma_h u_t, \; u_t \overset{iid}{\sim} \mathcal{N}(0,1).$$

We consider three studies. In the first study, we simulate a process representing the common factor including stochastic volatility. The second study solely simulates a standard normal common factor. Both studies are supposed to be compared with our data:

(1) Simulation of h with $\phi_h = 0.5, \; \sigma_h = 1, \; y_t|h_t \sim \mathcal{N}(0, e^{h_t})$
(2) Simulation of h with $\phi_h = 0, \; \sigma_h = 0$, so $y \sim \mathcal{N}(0,1)$, $h_t = 0$
(3) S&P Default study

The aim is to find out, if (3) is closer to (1) or (2). (1) shall also show that stochastic volatility is actually fitted. Clearly, as u is random and $y_t|h_t \sim (0, e^{h_t})$, a smoothed fit of h is more than we can expect. Further recall that $n = 25$:

2.3.2. *Results.* We set the number of iterations equal to 20000. A minor autocorrelation adjustment is allowed for, depending on the single parameter. A burn in of 10000 and an increase of the iterations brought no significant improvement. Furthermore, omitting the autocorrelation adjustment shows no major difference except in the Binomial approach. As parameter distributions are quite symmetric (see Figure 5), we present parameter means:

Sim/Model	CCC	B	BB	BBB	A	ρ
S/real	-0.4959	-1.4466	-2.0749	-2.5121	-3.0902	0.08
S(h=0)/SVEW	-0.4829	-1.4472	-2.0774	-2.5154	-3.0916	0.0552
S(h=0)/SVOEW	-0.4864	-1.4411	-2.0727	-2.5088	-3.0797	0.0681
S/BIN (chain 1)	-0.4646	-1.4110	-2.0330	-2.4714	-3.0253	0.0820
S/BIN (chain 2)	-0.4951	-1.4118	-2.0645	-2.5031	-3.0576	0.0802
S/OSVOEW	-0.5044	-1.4535	-2.0837	-2.5250	-3.1147	0.0769
S/OSVEW	-0.5066	-1.4565	-2.0865	-2.5294	-3.1263	0.0723
S/SVOEW	-0.5210	-1.4731	-2.0884	-2.5585	-3.1498	0.0853
S/SVEW	-0.5222	-1.4770	-2.0919	-2.5656	-3.1490	0.0793
D/BIN (chain 1)	-0.6018	-1.5759	-2.2334	-2.7175	-3.2580	0.0871
D/BIN (chain 2)	-0.5981	-1.5726	-2.2294	-2.7139	-3.2553	0.0874
D/SVOEW	-0.5863	-1.5324	-2.1801	-2.6512	-3.3829	0.0848
D/SVEW	-0.5898	-1.5360	-2.1826	-2.6594	-3.3982	0.0772
D/OSVOEW	-0.5930	-1.5340	-2.1847	-2.6624	-3.3955	0.0827
D/OSVEW	-0.5943	-1.5372	-2.1875	-2.6629	-3.3989	0.0785

SV=stochastic volatility, OSV=no SV, OEW=(14), EW=(15), D=data, S=Simulation, BIN= Binomial (OSV)

TABLE 4. Threshold and correlation estimates for different models

In **study 1**, we obtain a quite good fit of $\frac{l_{tk}}{N_{kt}} \approx p_{k|y_t}$, see Figure 6. The tendency of h is fitted well, but a bit too smooth (very stable for variant 1), see Figures 8(b) and 9. In view of the fact that we fit in "second order", i.e. the variance, this is a quite good result. D, ρ are recovered pretty sound as well as σ. ϕ is a bit too high because h is too smooth, see Figure 4.

In **study 2**, we again observe a quite good fit of $\frac{l_{tk}}{N_{kt}} \approx p_{k|y_t}$. Means (over iterations) of h_t oscillates closely around zero, hence the densities of h_t are approximately centered around 0. $\sigma \approx 0.21$ for several simulations. Note, σ cannot be zero because our prior does not contain 0! ϕ fluctuates around zero, but quite unstable, the variance is quite high. For our analysis this does not really matter. Important is that h is close to zero and fits well.

AVA-SVOEW	CCC	B	BB	BBB	A	ρ
mean	-0.5863	-1.5324	-2.1801	-2.6512	-3.3829	0.0844
median	-0.5862	-1.5327	-2.1790	-2.6512	-3.3803	0.0817
std	0.0539	0.0580	0.0680	0.0780	0.1110	0.0233
min	-0.7860	-1.7757	-2.4737	-3.0641	-4.1438	0.0252
max	0.5016	0.5012	0.5000	0.5000	-1.2937	0.3759
autocov	0.0029	0.0034	0.0046	0.0061	0.0123	0.0005
mode	-0.5744	-1.5812	-2.2235	-2.6144	-3.3362	0.0738
25% quantile	-0.6867	-1.6320	-2.3023	-2.7902	-3.6000	0.0469
75% quantile	-0.4867	-1.4336	-2.0600	-2.5140	-3.1774	0.1355
$p = \Phi(D)$:						
75% quantile	0.3132	0.0758	0.0197	0.0060	0.0007	
75%q of $m_{\tilde{L}}$	0.3263	0.0735	0.0134	0.0038	0.0009	

TABLE 5. Results S&P Default Study (AVA-SVOEW)

In **study 3**, the fit of $\frac{l_{tk}}{N_{kt}} \approx p_{k|y_t}$ and parameters are quite well, see Figure 5 and 7. Of course worse than in 1. and 2., because it cannot be expected that the considered loss rates actually follow a one factor model. Remarkable: h and ϕ behave similar as in 2.! $\sigma \approx 0.21$, means of h are all very close to zero suggesting a constant variance of y over time, see Figure 4, 10, and 11. We therefore suggest a constant asset correlation:

$$\rho_t|h_t := Corr(X_{i,t}|h_t, X_{j,t}|h_t) = (\rho e^{h_t} + (1-\rho))^{-1} e^{h_t}\rho \approx \rho.$$

Further graphs are given in the end of the chapter. Figure 12 shows the obtained world index plotted against an appropriately standardized Dow Jones (yearly log returns times 100) and the common factor y. Finally note, ϕ maps into $(-1, 1)$, often it is useful to transform values of ϕ by means of a Fisher's Z transform ($tanh^{-1}(\phi)$). In particular, this is done when the parameter is close to the boundary. In our case, we detect no significant difference.

2.4. **Sensitivity Analysis.** When modeling stochastic volatility, the Binomial approach is quite unstable to initial values, the one to one test fails, i.e. parameters are barely recovered. Whereas in AVA-SVOEW, after a very small burn in, a difference between two chains cannot be detected, see Table 4. This also holds if very extreme initial values are taken, see Figure 8. Variant 2 (SVEW) shows a very small deviation of chains, see Figure 9 (in all studies initial values are chosen very different). This fact is again negligible when comparing with the difference obtained by the Binomial algorithm see Figure 3.

Clearly, on the one hand AVA takes longer than the Binomial algorithm because X has to be sampled. On the other hand, efficiency is gained by not applying ARMS in every step. Convergence is approached much faster. Finally, in AVA-SVOEW almost every iteration can be taken into account

for our parameter analysis whereas in the Binomial approach, reasonable would be to take at least every 100 element of the sampled Markov chains as the autocorrelation is high.

One can now claim that SVOEW mixes up model and prior information of y (in the derivation of the full conditionals y is not assumed to be purely prior). As in classical statistics, this produces no major problem as long as y is actually normal. If that is not the case, SVEW is more appropriate and furthermore an adjustment when sampling D has to be made. This apprehension is fortunately not confirmed. A fit of the data shows no difference between obtained parameters of both algorithms, see Table 4. Furthermore, we can question, if we actually make a big mistake when y has fatter tails e.g. if y is t-distributed. We perform a simulation study, where y is simulated as a t-distributed random variable with d.f. equal to 3. In fact, SVEW recovers peaks of y better than SVOEW. However, SVOEW still performs very well. In particular, when comparing with the Binomial approach, see Figure 3. In a Binomial approach it is easy to replace the normal prior by a t-prior. But still a t-prior brought no improvement. y is not fitted in more complex models!

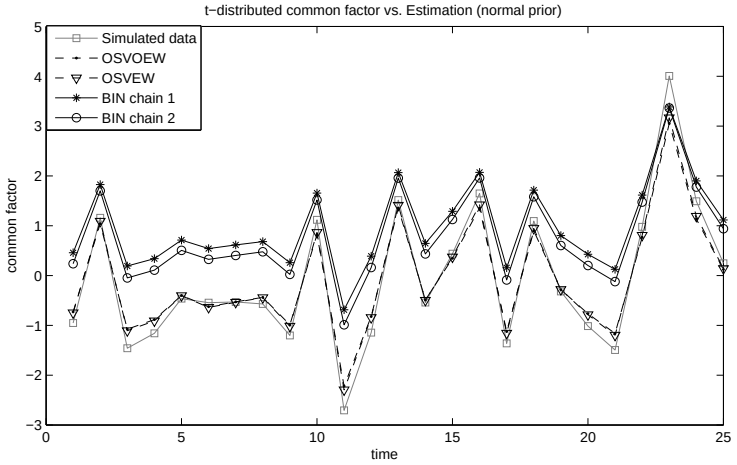


FIGURE 3. t-distributed common factor, but normal prior

Finally, note McNeil and Wendin (2006)/Wendin (2006) uses a logit model instead of probit model. Furthermore, the model is slightly different, ρ is not directly sampled. Our studies cannot be fully compared to the one in (Wendin, 2006).

2.5. **Multi Factor Model.** Last but not least, we sketch how to extend the one factor model to a model which can handle a whole correlation matrix, e.g. caused by different industries or regions. We assume that the multi factor stochastic volatility model introduced in Chib et al. (2006) is a model

for our asset value process. The difference is that the asset value process is not observable. We can apply the same trick as in the one dimensional case:

2.5.1. *Model.* We assume that the (return) asset value process of firm j in group g has the following form:

$$X_{j,t} = B_g f_t + \tilde{K}_{g,t} q_{g,t} + \lambda_{gt}^{-\frac{1}{2}} \tilde{\sigma}_{g,t} \epsilon_{j,t}, \tag{18}$$

where $\epsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. $B_g f_t$ drives stochastic correlation and $\tilde{\sigma}_{g,t}$ stochastic volatility, both driven by a $(\dim f + \#\text{groups})$-dimensional process h, $\lambda_{jt}^{-1/2} \epsilon_{j,t} \sim t_{(\nu_j)}$. $\tilde{K}_{g,t} q_{g,t}$ model shocks to the groups. For an exact definition see Chib et al. (2006).

With $\epsilon_{g,t} = {n_{tg}}^{-\frac{1}{2}} \frac{1}{n_{tg}} \sum_j^{n_{tg}} \epsilon_{j,t} \sim \mathcal{N}(0,1)$ a group index can be determined by

$$\overline{X}_{g,t} = B_g f_t + \tilde{K}_{g,t} q_{g,t} + {n_{tg}}^{\frac{1}{2}} \lambda_{gt}^{-\frac{1}{2}} \tilde{\sigma}_{g,t} \epsilon_{g,t}. \tag{19}$$

(18) is exactly the model given in Chib et al. (2006). So sampling could be performed if $X_{j,t}$ were known. The trick is again to sample $X_{j,t}$ as in a probit model, thresholds as in the one factor stochastic vola model:

2.5.2. *Multi Factor Models: Full Algorithm.*

(1) Set starting value of $\overline{X}_{g,t}^{(0)}$, $g = 1, ..., \#\text{groups}$, for $t = 1, ..., n$, set $i = 1$, set maximum number of iteration I
(2) Sample $B^{(i)}, f_t^{(i)}, \tilde{K}_t^{(i)}, q_t^{(i)}, h_t^{(i)}, \lambda_t^{(i)}$ using $\overline{X}^{(i-1)}, \mathbf{N}$ via Gibbs Sampling described in Chib et al. (2006).
(3) Sample $D_k^{(i)}, k = 1, ..., K$ given $B^{(i)}, f_t^{(i)}, \tilde{K}_t^{(i)}, q_t^{(i)}, \lambda_t^{(i)}, \mathbf{N}$
(4) Generate $\overline{X}_t^{(i)}$ from $B^{(i)}, f_t^{(i)}, \tilde{K}_t^{(i)}, q_t^{(i)}, h_t^{(i)}, \lambda_t^{(i)}, D_t^{(i)}, \mathbf{N}, \mathbf{l}$ for $t = 1, ..., n$
(5) set $i = i + 1$, if $i \leq I$ go to step (2)

In 3, one derives full conditionals given f, i.e. f is purely prior analogously to AVA-SVEW. Independently on f, one would have to form $g \times d_f$-groups, where d_f is the dimension of f.

3. Risk Figures

In this section, we want to examine risk figures obtained from our Markov Chain Monte Carlo method. We stick to real world default time series and analyze one year expected loss of a CDO tranche. As mentioned before the obtained expected loss therefore does not directly drive the price of the CDO. CDO tranches are compared to the corresponding numbers obtained from our moment estimators in Niethammer and Overbeck (2008). Both estimation techniques are different. The classical moment estimation already assumes that an infinite number of firms is considered, whereas in the Bayesian approach this assumption is not made. Further, in our MCMC approach we estimate a single ρ only. In contrary, moment estimators produce a different ρ_k for every rating class and any interconnection between the rating classes is missing. Furthermore, in the classical procedure intervals of

the risk figures are determined asymptotically. In the Bayesian approach, we have I values of D_k and ρ, where I is the number of considered iterations (potentially subtracted by a burn in and an autocorrelation adjustment). We plug them into the risk figures and obtain a distribution of the risk figure without any asymptotic theory, see Table 6/7/8. It is therefore not surprising that we have already examined differences in the parameters estimates for different estimation techniques. This fact is naturally carried forward to the risk figures. In view of the quite small data basis results are not so far away.

$EL^{K2\%}_{K1\%}$	K1	K2	$a^{classic}_{2.5\%}$	EL	$b^{classic}_{97.5\%}$	$a^{MCMC}_{2.5\%}$	EL	$b^{MCMC}_{97.5\%}$
CCC	14	29	0.3335	0.4888	0.6441	0.5382	0.6556	0.7816
B	3	6	0.3534	0.5156	0.6777	0.3934	0.5555	0.7066
BB	0	3	0.2445	0.3608	0.4772	0.3197	0.4149	0.5311
BBB	0	3	0.0514	0.0889	0.1264	0.0819	0.1275	0.1793
A	0	3	0.0041	0.0131	0.0221	0.0059	0.0124	0.0209

TABLE 6. Expected loss K1%-K2 tranche%

$EL^{6\%}_{3\%}$	$a^{classic}_{2.5\%}$	$EL^{classic}$	$b^{classic}_{97.5\%}$	$a^{MCMC}_{2.5\%}$	EL^{MCMC}	$b^{MCMC}_{97.5\%}$
CCC	0.9194	0.9757	1.0000	0.9251	0.9894	1.000
B	0.3534	0.5156	0.6778	0.3934	0.5555	0.7066
BB	0	0.0247	0.0500	0.0118	0.05777	0.1204
BBB	8.30E-05	1.105E-04	1.38E-04	2.21E-04	3.09E-03	1.69E-02
A	5.39E-08	5.75E-08	6.11E-08	2.43E-09	5.25E-05	3.36E-04

TABLE 7. Expected loss 3%-6% tranche classical vs. SVOEW

$EL^{3\%}_{0\%}$	$a^{MCMC}_{2.5\%}$	EL^{MCMC}	$b^{MCMC}_{97.5\%}$
CCC	0.9822	0.9981	1.0000
B	0.6955	0.8843	0.9638
BB	0.3197	0.4149	0.5311
BBB	0.0819	0.1275	0.1793
A	0.0059	0.0124	0.0209

TABLE 8. Expected loss 0%-3% tranche

Table 6 shows tranches more appropriate to the loss profile of the rating classes. Attachment and detachment point are mapped to the default probability of the rating class as in the classical approach via a moneyness mapping. The CCC tranche is mapped to the same moneyness as the B tranche, but with default probabilities of the classical approach to guarantee comparability. Tranche sizes in Table 6 are though adjusted to the default

$EL^{K2\%}_{K1\%}$	K1	K2	$b^{classic}_{97.5\%} - a^{classic}_{2.5\%}$	$b^{mix}_{97.5\%} - a^{mix}_{2.5\%}$	$b^{MCMC}_{97.5\%} - a^{MCMC}_{2.5\%}$
CCC	14	29	0.3116	0.4299	0.2434
B	3	6	0.3244	0.4553	0.3132
BB	0	3	0.2326	0.2857	0.2114
BBB	0	3	0.0750	0.1122	0.0974
A	0	3	0.0180	0.0225	0.0150

TABLE 9. Expected loss K1%-K2 tranche%: comparison of quantile range

probability. A default event of the considered tranche has got a reasonable probability. The length of the confidence intervals are quite similar for both approaches, see Table 9. Confidence intervals of our MCMC method (also assuming iid common factor) are closer to confidence intervals not including a time dependence $a^{classic}$. If an adjustment of the tranche size is not made, probabilities of hitting a tranche become quite small and almost hit the boundary zero. Estimation of the confidence intervals in the classical approach is not accurate enough. Confidence intervals of the moment estimator are very small for good ratings or when the hitting probability of a tranche is small (see Table 7 and 8), too small as we have seen in Niethammer and Overbeck (2008). This is confirmed here, the distribution of our parameters and risk figures from our Bayesian analysis show much wider and realistic confidence intervals in these cases. Furthermore, confidence intervals are not necessarily symmetric, not least because loss distributions are usually not symmetric. So it is not probable that estimation errors are carried forward symmetrically. Finally if ρ is sampled for every rating class separately, confidence intervals of moment estimations for ρ are a lot smaller for good ratings! The Bayesian confidence intervals can be seen as a quite good benchmark for our moment estimation as it is neither assumed that we have large number of companies nor that the number of time periods goes to infinity. Hence, our findings in Niethammer and Overbeck (2008) can be regarded as confirmed.

4. Concluding Remarks

Bayesian statistics possesses many valuable properties. We can nicely treat rare data as uncertainty is always taken into account in the distribution of the parameter. Parameter confidence intervals are given automatically. Since the establishment of Markov chain Monte Carlo methods (MCMC) and fast computing possibilities Bayesian analysis provides a method to flexibly solve all kinds of problems in statistics. The paper is therefore devoted to Bayesian analysis and among that we dealt with its relationship to classical statistics. We start comparing different approaches to estimate the main parameters in a Gaussian factor model based on S&P default study (2006) providing loss rates in five rating classes. Bayesian concepts

are introduced to smartly solve numerical integration problems caused by the ML-estimator. Moreover, it offers a flexible methodology to handle complex models as e.g. replacing the Gaussian by a general distribution. Strong methods to perform a thorough correlation study are given.

In our first correlation study, we find no evidence for two stages of correlation depending on the performance of our economy (random factor loading). Estimation is performed applying a Binomial type approach proposed by McNeil and Wendin (2006)/Wendin (2006)/Gössl (2005). Unfortunately, the algorithm gets unstable for a stochastic volatility model on the common factor. This problem is addressed by a new asset value algorithm, which reduces autocorrelations in the Markov chains enormously. In addition, it delivers an asset value index for every rating class. But again, time dependence in a stochastic volatility model cannot be detected. Further, we suggest a multi factor model to account for industry and country specification.

Finally, we take MCMC estimates and its confidence intervals as a benchmark to judge the asymptotic confidence intervals derived in Niethammer and Overbeck (2008). In view of a slightly different model specification our comparison of risk figures in classical and Bayesian statistics shows that an asymptotic analysis is not that bad as long as default and hitting probabilities of a tranche are not too close to 0 or 1.

All over one should have in mind, that we are working on a data set of loss rates. This data set is quite small in terms of observed periods. Further, assertions from equity data or additional financial statement information on the individual firms can look quite different, see e.g. Buraschi et al. (2006); Das et al. (2005). Except from (Das et al. , 2005), all authors directly consider equity data and correlations between the time series. Das et al. (2005) build their analysis on estimated individual firms' probabilities of default (henceforth EPDs) provided by Moody's Investors Services. Such EPDs are derived from a distance to default measure as described in Merton (1974) and so also from equity data. Some additional financial information is also included. A descriptive analysis is performed grouped in industry sectors or rating buckets. Data are on a monthly basis (1/87-10/00), but grouped in four periods of about three years. Significantly different correlations between EPDs within the periods are detected in their descriptive analysis as well as in a regime switching model (a test for equal correlations in different regimes in a hidden Markov chain type model is rejected). This is not surprising as these EPDs are driven by equity time series. Historical loss rates are not included. Moreover, observations are taken on a monthly basis. We conclude that the model, the actual data source, and time intervals play a major role when interpreting results on correlations. This does not change the observation that applying estimation methods to the natural default data source - yearly loss rates - show no evidence of time dependence of the correlation in several model variants of the Gaussian one factor model. In particular, there is no support for a random factor loading

from historical loss rates. Nevertheless one should note that in our study we exclusively work on an historical data set. An assessment of the market is not included. Market data as credit spreads or tranche spreads are not considered. So when pricing products, a calibration to market data is more important than an actually historical evidence for a model. A random factor loading (two stage correlation model) can be still applicable.

References

Ahluwalia, R., McGinty, L., Beinstein, E., and Watts, M. (2004), Introducing base correlations, JPMorgan Research, http://mm.jpmorgan.com.

Andersen, L. and Sidenius, J., (2004), Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings, Journal of Credit Risk, 1, 1, 29-70.

Bluhm, C. and Overbeck, L. (2003), Estimating Systematic Risk in Uniform Credit Portfolios, Credit Risk. Hrsg.: G.Bol et al. Contributions to Economics Physica-Verlag, Heidelberg.

Bluhm, C., Overbeck, L., and Wagner, C.K.J. (2002), An Introduction to Credit Risk Modeling, Financial Mathematics Series, Chapman & Hall/CRC, London.

Buraschi, A., Porchia, P., and Trojani, F. (2006), Correlation Risk and Optimal Portfolio Choice, preprint, Imperical College/University of St. Gallen.

Chib, S., Nardari, F., Shephard, N. (2002), Markov Chain Monte Carlo Methods for Stochastic Volatility Models, Journal of Econometrics, 108, 281-316.

Chib, S. , Nardari, F., Shephard, N. (2006), Analysis of High Dimensional Multivariate Stochastic Volatility Models, Journal of Econometrics, 134,2, 341-371.

Chib, S. and Greenberg, E. (1995), Understanding the Metropolis-Hastings Algorithm, The American Statistician, 49, 4, 327-335.

Chib, S. and Greenberg (1996), Markov Chain Monte Carlo in Econometrics, Econometric Theory, 12, 409-431.

Das, S.R., Freed, L., and Geng, G. (2005), Correlated Default Risk, preprint, http://www.gloriamundi.org/picsresources/dgfk.pdf.

Gamerman, D. and Lopez, H. F. (2006), Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Texts in Statistical Science, Chapman & Hall/CRC.

Gilks, W. R., Best, N.G., and Tan K.K.C. (1995), Adaptive Rejection Metropolis Sampling with Gibbs Sampling, Applied Statistics, 4, 455-475.

Gössl, C. (2005), Predictions based on certian uncertainties - a bayesian credit portfolio approach, HypoVereinsbank AG, London, preprint.

Jones, G. L. (2004), On the Markov Chain Central Limit Theorem, Probability Surveys, 1, 299-320.

de Jong, P. and Shephard, N., (1995) The Simulation Smoother for Time Series Models, Biometrika, 82, 2, 339-350.

Kalemanova, A., Schmid, B., and Werner, R. (2005), The Normal Inverse Gaussian Distribution for Synthetic CDO pricing, www.default-risk.com.

Kim, S., Shephard, N., and Chib, S. (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, Review of Economic Studies, 65, 361-393.

Lancaster, T. (2004), An Introduction to Modern Bayesian Econometrics, Blackwill Publishers.

McNeil, A. and Wendin, J. P. (2007), Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk, Journal of Empirical Finance, 14, 2, 134-149.

Merton, R. C. (1974), On the pricing of corporate debt: the risk structure of interest rates, Journal of Finance, 29, pp. 449-470.

Niethammer, C. R. (2008), Default Correlations and the Effect of Estimation Errors on Risk Figures, preprint.

S& P default study, Annual 2005 Global Corporate Default Study and Rating Transitions, Standard& Poor's, Global Fixed Income Research, January 2006.

Wendin, J. E. P. Bayesian Methods in Portfolio Credit Risk Management, dissertation ETH Zurich No. 16481.

Willemann, S. (2004), An Evaluation of the Base Correlation Framework for Synthetic CDOs, preprint, Available at SSRN: http://ssrn.com/abstract=637944.

E-mail address: christina.niethammer@uni-konstanz.de (University of Konstanz/UniCredit Markets & Investment Banking)

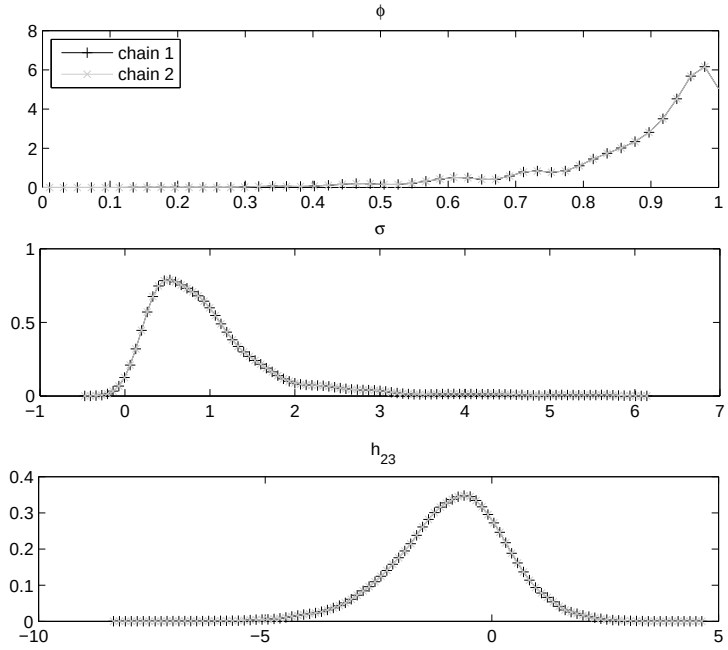


FIGURE 4. Parameter simulation study SV.1

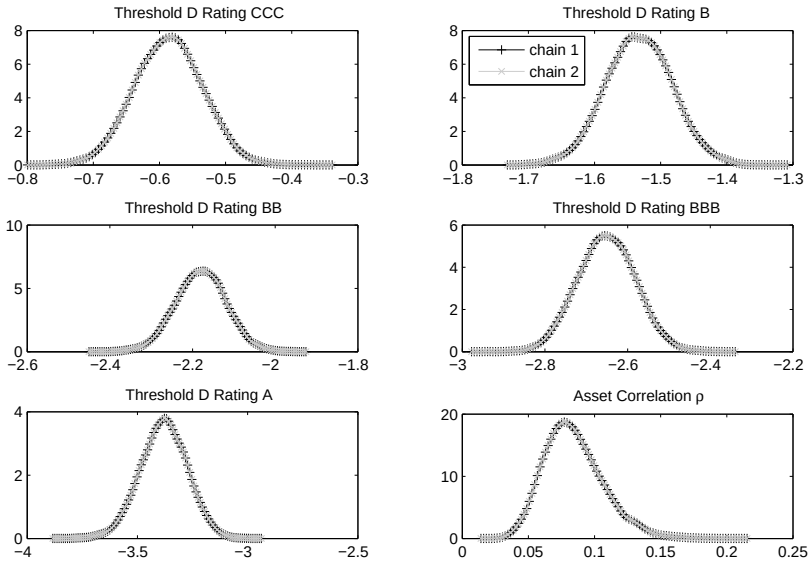


FIGURE 5. Parameter estimates (S&P default study, SVOEW)

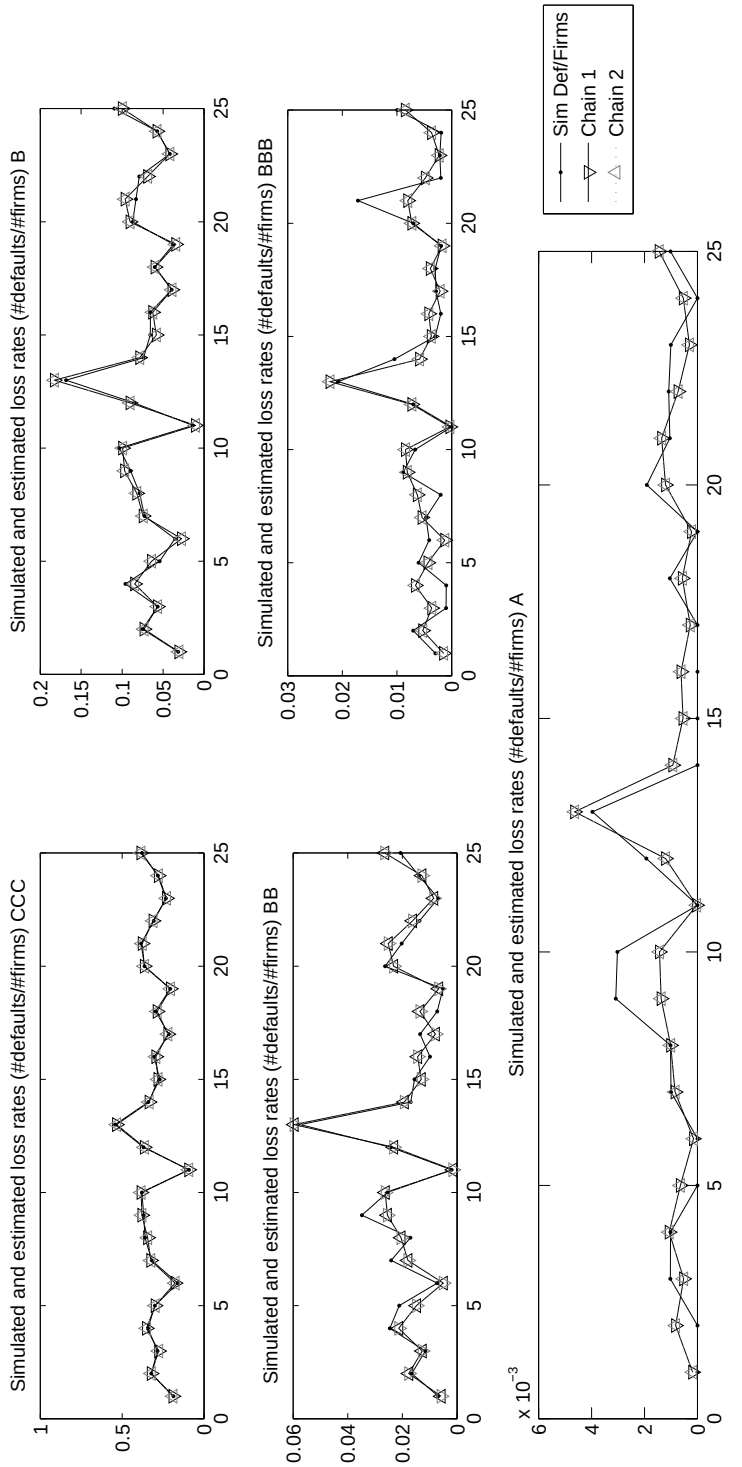


FIGURE 6. Fit of default time series $p_{|y_t} \approx \frac{l_t}{N_t}$ (Simulated data)

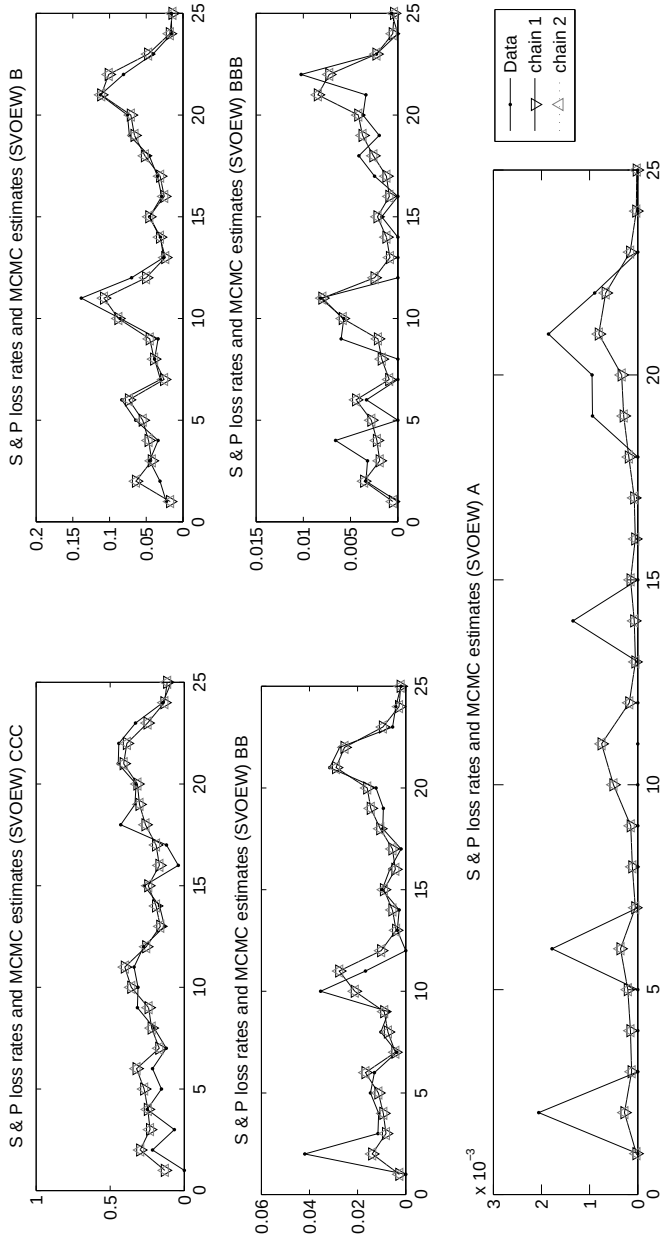
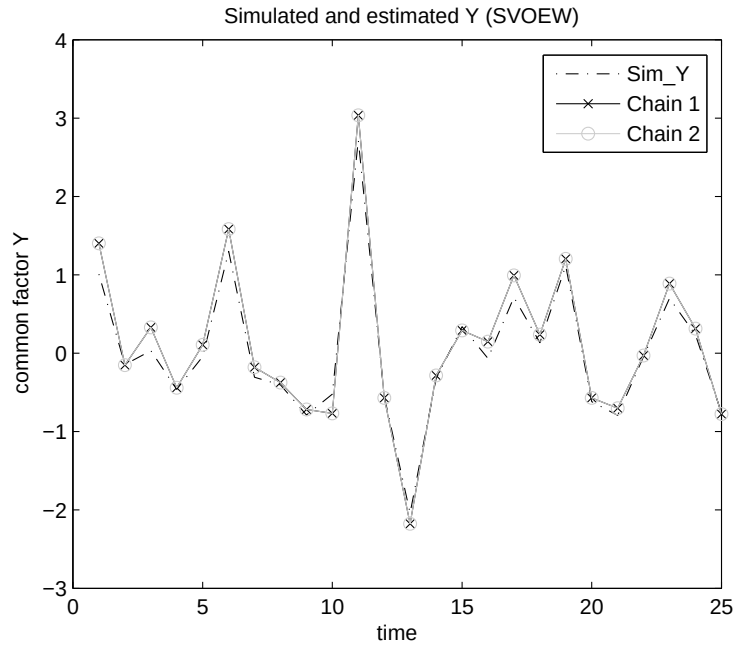
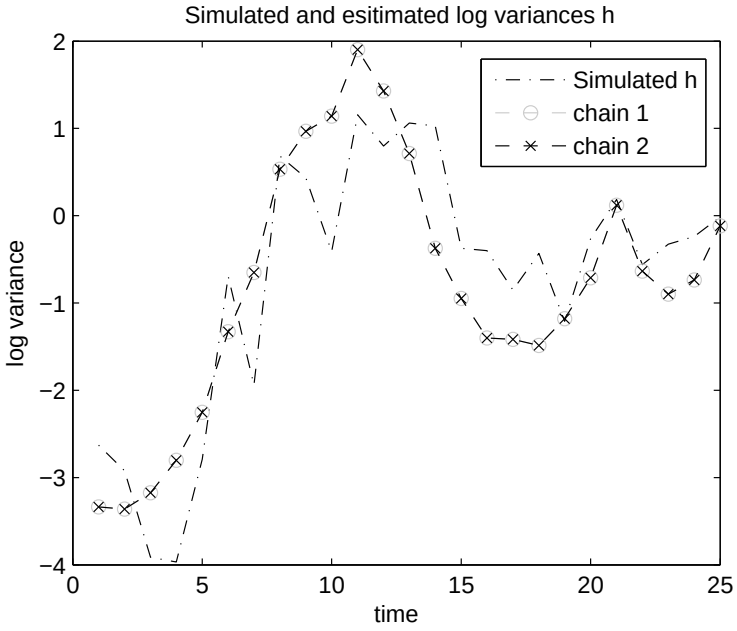


FIGURE 7. Fit of default time series $p_{|y_t} \approx \frac{l_t}{N_t}$ (S&P)



(a) Fit of common factor



(b) Fit of log variances

FIGURE 8. Fit of common factor and its log variance (SVOEW)

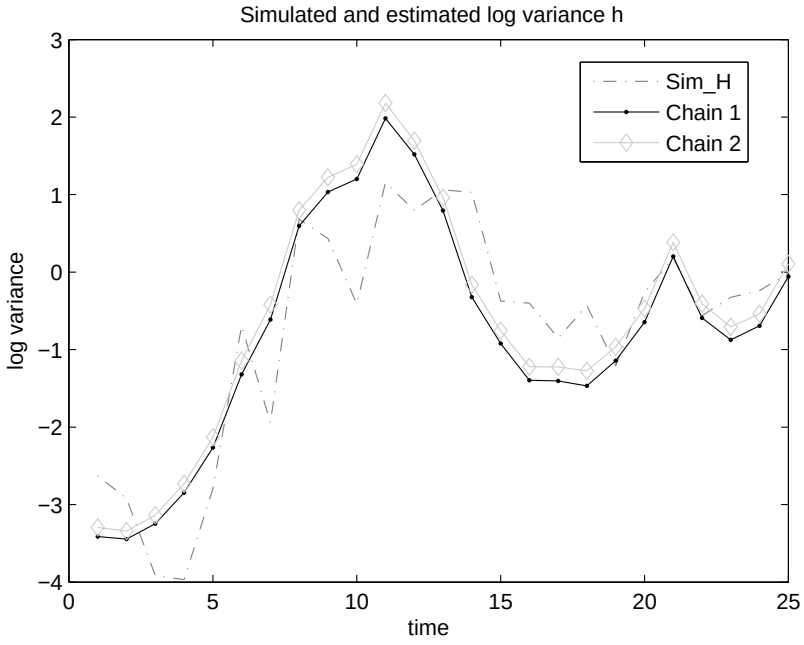


FIGURE 9. Fitted log variance in model variant SVEW

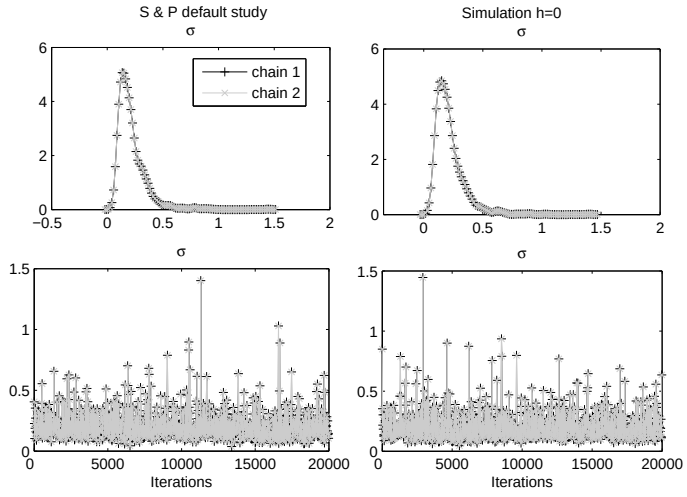
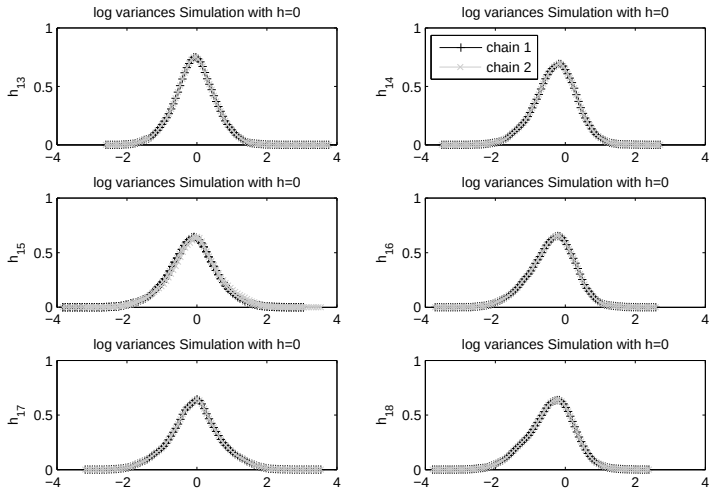
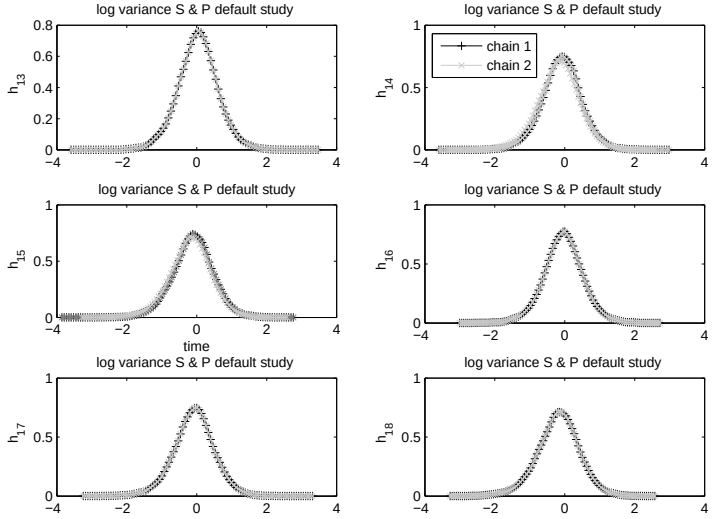


FIGURE 10. σ of $h_t = \phi h_{t-1} + \sigma u_t$ data vs. simulation $h = 0$



(a) Simulation study $h = 0$



(b) S&P default study

FIGURE 11. Estimated densities of log variances

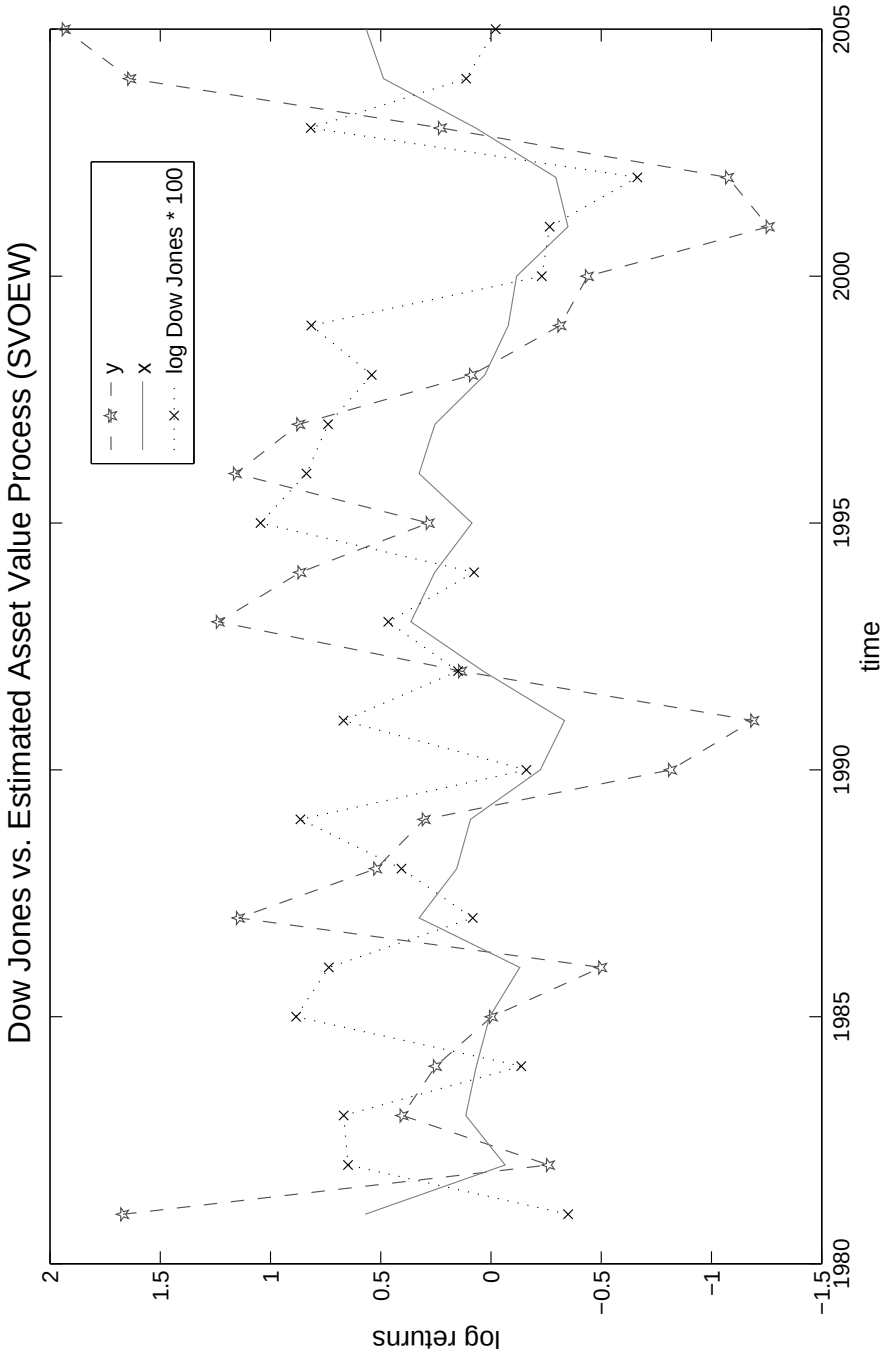


FIGURE 12. Estimated asset value process vs. Dow Jones