Chapter 1

Introduction

This theses is about nonlinear, non-Gaussian state space models and their application in finance. In finance many of the variables of interest such as default probability and volatility are inherently latent and we have only noisy measurements about these variables. State space models provide a natural modelling framework. The term state space model refers to a class of probabilistic graphical models (see Koller and Friedman (2009)) that describes the joint distribution of the latent state variable and the observed measurement. State space models present a flexible modelling tool as they can capture complex features of the data.

We discuss different applications of state space models in finance such as dynamic credit risk models and different stochastic volatility models. The Bayesian estimation of these models is challenging as we have to integrate out the latent state from the joint distribution of parameters and states to arrive at the likelihood. The dimension of this integral is equal to the number of observations in our time series, which means that we have to solve an integral where we have to integrate over potentially several thousands of variables or in a high frequency return context millions. We use Monte Carlo integration methods which are well suited for large dimensional problems and not directly effected by the dimension of the integrals as opposed to standard grid
based numerical integration techniques. We will see that Monte Carlo methods are especially useful when we want to sequentially update the posterior distribution of the parameters as new information arrives.

In this chapter I give a short introduction into latent variable models and Bayesian inference as these topics are at the core of this thesis. The first section shows how conditional independence assumptions can help to specify statistical models such as latent variable models. The second section reviews the basics of Bayesian statistics putting emphasis on Monte Carlo methods as these are unavoidable tools to carry out modern Bayesian inference. Finally, the last section discusses the outline of the thesis and gives a short summary of the topics and contributions in the thesis.

1.1 Latent variable models

Modelling a large number of variables such as a financial time series is challenging, because we have to specify a high dimensional joint distribution. Imagine that we want to model a series of random variables \( y_{1:T} = (y_1, y_2, \ldots, y_T) \). An obvious first attempt would be to model the returns with a joint normal distribution. This approach is naive because returns can have fatter tails than the normal distribution. Moreover it is infeasible for larger \( T \), because we have to specify the \( T \times 1 \) mean vector and a \( T \times T \) covariance matrix, which has \( T(T + 1)/2 \) free parameters. Therefore, we need more assumptions to have a grip on this problem.

The usual solution to this problem is to use conditional independence assumptions. These assumptions allow us to factorize the joint distribution and specify lower dimensional distributions. Conditional independence assumptions can be represented by a directed acyclic graph (DAG) whose nodes represent random variables and an edge pointing from a parent node to a child node encodes a direct influence of the parent variable on the child variable. If a model can be represented by a DAG then we call it
1.1 Latent variable models

a Bayesian network or belief network and the joint distribution can be factorized as

\[ p(y_1, y_2, \ldots, y_T) = \prod_{t=1}^{T} p(y_t | \text{pa}(y_t)), \]  

(1.1)

where \( \text{pa}(y_t) \) denotes the parent nodes of \( y_t \). One can read more about conditional independence assumptions and probabilistic graphical models in Koller and Friedman (2009).

Returning to the problem of modelling a time series, probably the simplest meaningful assumption for a time series model is the Markov assumption which states that \( y_t \) given \( y_{t-1} \) is conditionally independent of \( y_i, i < t - 1 \) or more formally

\[ p(y_t | y_{t-1}, y_{t-2}, \ldots) = p(y_t | y_{t-1}) \]  

(1.2)

for every \( t \), where \( p(y_t | y_{t-1}) \) is the conditional density function of \( y_t \) conditionally \( y_{t-1} \). This means that only \( y_{t-1} \) has a direct influence on \( y_t \) and the parent of \( y_t \), \( \text{pa}(y_t) \) is \( y_{t-1} \). The corresponding DAG representation is depicted on Figure 1.1

\[ \ldots \rightarrow y_{t-1} \rightarrow y_t \rightarrow y_{t+1} \rightarrow \ldots \]

Fig. 1.1 DAG representation of a Markov model

Equation (1.1) implies that we can factor the joint distribution as

\[ p(y_1, y_2, \ldots, y_T) = \prod_{t=1}^{T} p(y_t | y_{t-1}). \]  

(1.3)

If we are willing to assume that

\[ p(y_t | y_{t-1}) = N \left( \phi y_{t-1}, \sigma^2 \right), \]  

(1.4)
we can reduce the number of parameters to two and we end up with a zero mean autoregressive process of order one. This shows that, by assuming some structure on the joint distribution in terms of conditional independence assumptions, we can greatly reduce the number of parameters.

Throughout this thesis I concentrate on latent variable models that are also known as state space models or hidden Markov models. Hidden Markov models are simple special cases of Bayesian networks, which are applicable to modelling time series data. These models are successfully applied in different fields of science e.g., engineering, biology, finance, etc. Hidden Markov models assume that there is an unobserved time series $x_{1:T}$ which is Markovian and that we observe a time series $y_{1:T}$ which depends on $x_{1:T}$ such that $y_t$ given $x_t$ is conditionally independent from all the other $y_i, i \neq t$. Figure 1.2 depicts the DAG representation of the hidden Markov models.

Fig. 1.2 DAG representation of a hidden Markov model

The general formulation of hidden Markov models is given by

\begin{align*}
y_t & \sim p(y_t | x_t) \\
x_t & \sim p(x_t | x_{t-1}) \\
x_1 & \sim p(x_1),
\end{align*}

\begin{align*}
(1.5) & \\
(1.6) & \\
(1.7) & 
\end{align*}
where \( p(y_t|x_t) \) is the observation density, \( p(x_t|x_{t-1}) \) is the transition density and \( p(x_1) \) is the initial density.

Hidden Markov models are widespread in financial econometrics, because they are flexible and they encompass a large number of models, even if the conditional independence assumptions might sound restrictive at first. Moreover, they have an intuitive appeal and a tractable factorization.

In the next few paragraphs we discuss some of the models that we cover in more details in the thesis. These models illustrate the flexibility of nonlinear, non-Gaussian state space models.

### 1.1.1 Stochastic intensity model for corporate defaults

The fist model is a dynamic credit risk model based on ? and Duffie et al. (2009a). We observe a portfolio of firms over time and some firm specific and macro variables. We would like to explain the dynamics of the default probability based on observable variables and a latent frailty factor. We link the explanatory variables and the default probabilities through the default intensity \( \lambda_{ti} \) of firm \( i \) at time \( t \),

\[
\lambda_{ti} = \lim_{\Delta t \downarrow 0} \frac{P[t < T_i \leq t + \Delta t | T_i > t]}{\Delta t},
\]

where \( T_i \) is the default time of firm \( i \). We model the default intensity for firm \( i \) at time \( t \) as a function of firm specific and macro fundamentals collected in a vector \( c_{it} \) and a latent frailty factor \( f_t \)

\[
\log \lambda_{ti} = \mu + \beta c_{it} + \gamma f_t, \tag{1.9}
\]

\[
f_{t+1} = \phi f_t + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, 1 - \phi^2), \tag{1.10}
\]

\[
f_1 \sim N(0, 1).
\]
Introduction

Given the piece-wise constant default intensity specification we can write the probability of default and survive period \( t \) as

\[
\prod_{i=1}^{N_t} \exp \left( D_{it} \log \lambda_{it} - S_{it} \lambda_{it} \right), \tag{1.11}
\]

where \( N_t \) is the number of firms in period \( t \), \( D_{it} \) is a default indicator which is one if firm \( i \) defaults in period \( t \) and \( S_{it} \) is the time that firm \( i \) spent in the portfolio in period \( t \). Equation (1.10) give the transition density while equations (1.9) and (1.11) determine the observation density.

1.1.2 Gamma Ornstein–Uhlenbeck stochastic volatility model

The second example is a Gamma Ornstein–Uhlenbeck stochastic volatility model for stock returns. The model was introduced in Barndorff-Nielsen and Shephard (2001) and Barndorff-Nielsen and Shephard (2002). In this model conditional on the volatility the returns follow a normal distribution

\[
y_n = \mu + \beta \sigma_n^2 + \sigma_n \varepsilon_n, \quad \varepsilon_n \sim N(0, 1), \tag{1.12}
\]

while the variance follows a Gamma Ornstein–Uhlenbeck process with unconditional mean \( \xi \), variance \( \omega^2 \), and persistence parameter \( \lambda \). The transition density of the Gamma Ornstein–Uhlenbeck process is not available in close form. However we can simulate from the process as follows

\[
k \sim \text{Poi}(\lambda \xi^2/\omega^2), \quad c_i \sim \text{U}(n, n+1), \quad e_i \sim \text{Exp}(\xi/\omega^2), \quad i = 1, \ldots, k; \tag{1.13}
\]

\[
z_{n+1} = e^{-\lambda} z_n + \sum_{j=1}^{k} e^{-\lambda(n+1-c_j)} e_j, \quad \sigma_n^2 = \frac{1}{\lambda} \left[ z_n - z_{n+1} + \sum_{j=1}^{k} e_j \right]. \tag{1.14}
\]
1.1 Latent variable models

Equations (1.12) and (1.14) define a nonlinear, non-Gaussian state space model. The inference of this model is challenging, because we cannot even write down the transition density in closed form.

1.1.3 Integer valued stochastic volatility model

The final example is an integer valued stochastic volatility model for high frequency trade by trade stock price differences. We model the changes in prices expressed in the number of ticks with a negative binomial difference distribution ($\Delta$NB), which is defined as the difference of two negative binomial random variables. We assume that the log variance can be decomposed into a slowly moving daily pattern $s_t$ and a persistent local volatility component $x_t$, which captures the variation in volatility on top of the daily seasonality pattern. Conditionally on the variance $\lambda_t$ the integer returns follow a zero inflated $\Delta$NB distribution

$$ y_t = \begin{cases} r_t & \text{with} \ (1 - \gamma)f_{\Delta\text{NB}}(r_t; \lambda_t, \nu), \\ 0 & \text{with} \ \gamma + (1 - \gamma)f_{\Delta\text{NB}}(0; \lambda_t, \nu), \end{cases} \quad (1.15) $$

$$ \log \lambda_t = h_t = \mu_h + s_t + x_t, \quad (1.16) $$

$$ s_t = w_t \beta, \quad (1.17) $$

where $f_{\Delta\text{NB}}(r_t; \lambda_t, \nu)$ is the probability mass function of the $\Delta$NB distribution, $\gamma$ is the zero inflation parameter that puts some extra probability weight on zero. The local volatility component $x_t$ is modelled by a zero mean AR(1) process, with the transition equation

$$ x_{t+1} = \phi x_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta). \quad (1.18) $$
1.2 Bayesian inference and Monte Carlo methods

Bayesian inference is concerned with updating the subjective prior distribution $p(\theta)$ of some model parameter $\theta$ based on the observation of some data $y_{1:T}$ to arrive at the subjective conditional distribution of the parameter $p(\theta|y_{1:T})$ which is called the posterior distribution. Bayes rule gives a simple updating rule in the form of

$$p(\theta|y_{1:T}) = \frac{p(\theta|y_{1:T})p(\theta)}{p(y_{1:T})} = \frac{p(\theta|y_{1:T})p(\theta)}{\int_\theta p(\theta|y_{1:T})p(\theta)d\theta}.$$  \hfill (1.19)

This formula may seem simple, however the evaluation of the integral in the denominator can become problematic when the dimension of $\theta$ gets large.

In latent variable models we are not only interested in the posterior distribution of the parameters but we also would like to make inference on the latent variables $x_{1:t}$. We are interested in the distribution of the latent state $x_t$ given different subsets of the data. The smoothing distribution is the posterior distribution of the $x_t$ state at time $t$ conditional on all the available data $y_{1:T}$ where $t \leq T$

$$p(x_t|y_{1:T}).$$  \hfill (1.20)

A second distribution is the filtering distribution

$$p(x_t|y_{1:t}),$$  \hfill (1.21)

which gives the distribution of the state at time $t$ given all the available data at time $t$.

We can factorize the smoothing density in the following way:

$$p(x_t|y_{1:T}) = \frac{p(x_t, y_{1:T})}{p(y_{1:T})} = \frac{p(y_{t+1:T}|x_t) p(x_t, y_{1:t})}{p(y_{1:T})} = \frac{p(y_{t+1:T}|x_t)}{p(y_{t+1:T}|y_{1:t})} \frac{p(x_t|y_{1:T})}{p(x_t|y_{1:T})}.$$  \hfill (1.22)
Notice that the second factor is the filtering density. We can further decompose the filtering density as

\[
p(x_t | y_{1:t}) = \frac{p(x_t, y_{1:t})}{p(y_{1:t})} = \int \frac{p(x_t, x_{t-1}, y_{1:t})}{p(y_{1:t})} dx_{t-1}
\]

\[
= \int \frac{p(y_t | x_t)p(x_t | x_{t-1})p(x_{t-1}, y_{1:t-1})}{p(y_{1:t-1})p(y_{1:t})} dx_{t-1}
\]

\[
= \int \frac{p(y_t | x_t)p(x_t | x_{t-1})}{p(y_t | y_{1:t-1})} p(x_{t-1} | y_{1:t-1}) dx_{t-1}.
\] (1.23)

Moreover, we note that

\[
p(y_{t+1:T} | x_t) = \int \int p(y_{t+1:T}, x_{t+1} | x_t) dx_{t+1} = \int \int p(y_{t+1:T}, x_{t+1} | x_t) dx_{t+1} dx_{t-1}
\]

\[
= \int \int p(y_{t+2:T} | x_{t+1})p(y_{t+1} | x_{t+1})p(x_{t+1} | x_t) dx_{t+1}.
\] (1.24)

Equations (1.23) and (1.24) gives the so called forward filtering and backward smoothing recursions, which facilitate the calculation of the smoothing and filtering distributions. However, these recursions involve integrals and \(p(y_t | y_{1:t-1})\) the likelihood contribution of \(y_t\) which can be expressed as the following double integral

\[
p(y_t | y_{1:t-1}) = \int \int p(y_t, x_t, x_{t-1} | y_{1:t-1}) dx_{t-1} dx_t
\]

\[
= \int \int p(y_t | x_t)p(x_t | x_{t-1})p(x_{t-1} | y_{1:t-1}) dx_{t-1} dx_t.
\] (1.25)

The usefulness of the recursions in (1.23) and (1.24) depends on whether we can efficiently evaluate these integrals. If the state is finite and small this can be done, because the integrals reduce to summations. In the special case of a linear Gaussian
state space models the integrals are available in closed form and the recursions are called the Kalman filtering and smoothing recursions. However, for general hidden Markov models the filtering and smoothing recursions are not available in closed form.

In this section we have already seen that efficiently solving integrals in large dimensions is crucial for Bayesian statistics. In the rest of this section I discuss two integration methods that are particularly useful in evaluating high dimensional integrals namely Monte Carlo integration and Markov Chain Monte Carlo methods.

All the above mentioned integrals, which we are interested in, can be written in the following form

\[ E_p(f) = \int f(u)p(u)du, \]  

(1.26)

where \( f \) is a square integrable function

\[ \int |f(u)|^2 p(u)du < \infty, \]  

(1.27)

and \( p(u) \) is a probability density function. We can approximate the integral using the strong law of large numbers and calculating

\[ E_{p,N}^{MC}(f) = \frac{1}{N} \sum_{i=1}^{N} f(u^i), \]  

(1.28)

where \( u^1, u^2, \ldots, u^N \) are i.i.d draws from the distribution \( p(u) \).

To see why Monte Carlo integration is useful in solving large dimension integrals compared to numerical integration methods, we have to check the precision of our Monte Carlo integral estimate. Fortunately, the central limit theorem shows that

\[ \sqrt{N} \left[ E_{p,N}^{MC}(f) - E_p(f) \right] \]  

(1.29)
1.2 Bayesian inference and Monte Carlo methods

has an asymptotically normal distribution which can be used to construct asymptotic confidence bound for $E_p(f)$. In case of real valued $f$, we get the

$$\left[ E_{p,N}^{MC}(f) - \frac{1}{\sqrt{N}} c_\alpha \sigma_{N}^{MC}(f) , \ E_{p,N}^{MC}(f) + \frac{1}{\sqrt{N}} c_\alpha \sigma_{N}^{MC}(f) \right]$$

(1.30)

as a confidence interval for $E_p(f)$, where

$$\left[ \sigma_{N}^{MC}(f) \right]^2 = \text{Var} \left[ E_{p,N}^{MC}(f) \right] = \frac{1}{N} \sum_{i=1}^{N} \left[ f(u^i) - E_{p,N}^{MC}(f) \right]^2 .$$

(1.31)

The above equations show that by increasing the number of draws $N$, we can get any desired precision and the convergence rate is $\sqrt{N}$, which means that for one digit of extra accuracy, we have to increase the number of replications 100 times. Moreover, and maybe more importantly the precision does not depend on the dimension of $u$, which is the reason why Monte Carlo methods are so popular.

However, in practice there is a problem with Monte Carlo integration as described in the previous paragraph, namely we might not be able to draw from $p(u)$, as is usually the case for the integrals in Bayesian statics. In this case the above Monte Carlo integration is infeasible. We highlight two solutions for this problem. One of them uses Markov Chain Monte Carlo (MCMC) integration. The idea of MCMC is that, instead of drawing an i.i.d. sequence one may consider a Markov-dependent sequence which has $p(u)$ as its stationary distribution

$$E_{p,N}^{MCMC}(f) = \frac{1}{N} \sum_{i=1}^{N} f(u^i) ,$$

(1.32)

where $u^1, u^2, \ldots, u^N$ is a Markov Chain with stationary distribution $p(u)$. The construction of the Markov-dependent sequence with invariant distribution $p(u)$ is feasible in general, but might require case by case design and careful tuning.
Introduction

An other way of turning an infeasible Monte Carlo integration into a practical algorithm is to draw $u^1, u^2, \ldots, u^N$ from a distribution $q(u)$ from which we can sample and correct the Monte Carlo estimator of the integral such that it takes into account the difference between $p(u)$ and $q(u)$. This method is called importance sampling.

To understand importance sampling, first note the following identity,

$$E_p(f) = \int f(u)p(u)du = \int f(u)\frac{p(u)}{q(u)}q(u)du = E_q \left(f \frac{p}{q}\right), \quad (1.33)$$

which suggests the following importance sampling estimate of the integral

$$E_{IS}^{q,N}(f) = \frac{1}{N} \sum_{i=1}^{N} f(u^i) \frac{p(u^i)}{q(u^i)}, \quad (1.34)$$

where $u^1, u^2, \ldots, u^N$ are i.i.d draws from the distribution $q(u)$. The variance of the important sampling estimator is

$$\left[\sigma_{IS}^{q,N}(f)\right]^2 = \text{Var} \left[E_{IS}^{q,N}(f)\right] = \frac{1}{N} \sum_{i=1}^{N} \left[f(u^i)\frac{p(u^i)}{q(u^i)} - E_{IS}^{q,N}(f)\right]^2, \quad (1.35)$$

which can be smaller than the variance of the Monte Carlo estimate given that $q(u)$ is close to $f(u)p(u)$. The design of $q$ needs careful consideration. More information about Monte Carlo methods can be found in Robert and Casella (2004) and Cappé et al. (2005).

1.3 Thesis outline

In Chapter 2 we propose a new Bayesian methodology for designing flexible proposal densities for the estimation of parameters in a nonlinear, non-Gaussian state space model. The proposal is a mixture of Student’s $t$ distributions, which minimizes the Kullback-Leibler divergence between the posterior distribution and the mixture of Student’s $t$ proposals. We show that a highly efficient Bayesian procedure emerges when these proposal densities are used in an independent Metropolis-Hastings algorithm.
or in importance sampling. We present extensive simulation evidence for stochastic intensity and stochastic volatility models based on Ornstein–Uhlenbeck processes. Our method provides a computationally more efficient alternative to several recently proposed algorithms such as the SMC² method by Chopin et al. (2013a) and the density tempered marginalized sequential Monte Carlo estimation procedure by Duan and Fülöp (2015). For our empirical study, we analyse the performance of our methods for corporate default panel data and stock index returns.

In Chapter 3 we show how sequential Bayesian estimation can help to explain the variation in corporate credit spreads. If there is an unobserved component in corporate default intensities, then part of the fluctuation in corporate bond prices can be attributed to the variation in beliefs about this latent factor. We show evidence of a latent frailty process in the default intensities in US corporate defaults even after including macro and firm specific variables by comparing Bayes factors of models with and without frailty factors. We proxy for changes in agents’ beliefs about the unobserved factor by calculating the changes in conditional expectations of the frailty level, persistence and volatility. We demonstrate that changes in frailty related variables help to explain the variation in US corporate credit spreads.

In Chapter 4 we discuss stochastic volatility models of high frequency price changes. Trade by trade high frequency price changes are affected by the fact that on most stock exchanges prices are defined on a discrete grid. We check different integer valued distributions to account for the afore mentioned discreteness. Moreover, we propose an integer valued stochastic volatility model which exhibits fat tails. Our model explicitly takes into account the discreteness of the observed prices changes, fat tails, and intraday patterns of volatility. We propose a Markov chain Monte Carlo estimation method, which takes advantage of data augmentation and an auxiliary mixture sampling. Using our model we are able to decompose the volatility into a persistent daily seasonality pattern and an autoregressive component which captures the deviation of volatility from the daily pattern. We illustrate our methodology using tick by tick data of
several stocks from the NYSE in different periods. Using predictive likelihoods we find evidence in favour of the dynamic $\Delta NB$ model.

Chapter 5 summarizes the main findings of the thesis.