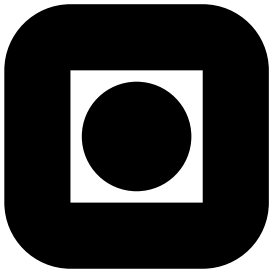
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Estimating Stochastic Volatility Models Using Integrated Nested Laplace Approximations

by

Sara Martino, Kjersti Aas, Ola Lindqvist , Linda Reiersølmoen Neef &
Håvard Rue

PREPRINT
STATISTICS NO. 8/2008



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2008/S8-2008.ps
Sara Martino has homepage: http://www.math.ntnu.no/∼martino
E-mail: martino@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and
Technology, N-7491 Trondheim, Norway.

Estimating Stochastic Volatility Models Using Integrated Nested Laplace Approximations

Sara Martino[(1)], Kjersti Aas[(2)], Ola Lindqvist[(2)] ,
Linda Reiersølmoen Neef[(2)] & Håvard Rue[(1)]

[(1)] *Department of Mathematical Science, NTNU, Norway*
[(2)] *Norwegian Computing Center, Oslo, Norway*

Abstract

Volatility in financial time series is mainly analysed through two classes of models; the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models and the Stochastic Volatility (SV) ones. GARCH models are straight-forward to estimate using maximum likelihood techniques, while SV models require more complex inferential and computational tools, like Markov Chains Monte Carlo (MCMC). Hence, although provided with a series of theoretical advantages, SV models are in practice much less popular than GARCH ones.

In this paper we solve the problem of inference for some SV models by applying a new inferential tool, Integrated Nested Laplace Approximations (INLA), which substitutes MCMC simulations with accurate deterministic approximations, making a full Bayesian analysis of many kinds of SV models extremely fast and accurate. Our hope is that the use of INLA will help SV models to become more appealing to the financial industry where, due to their complexity, they are rarely used in practice.

Keywords

Approximate Bayesian Inference, Laplace Approximation, Latent Gaussian Models, Stochastic Volatility Model.

1 Introduction

Time varying volatility is one of the main features of financial time series, like asset prices and exchange rates. It is mainly analysed through two classes of models; the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models and the Stochastic Volatility (SV) ones. The GARCH models were introduced by Engle (1982) and Bollerslev (1986). In these models, the conditional variance is assumed to be a function of previous observations and variances. Several different versions of the GARCH model have been proposed in order to accommodate for observed features like heavy tails and leverage effects (see for example Glosten et al. (1993); Nelson (1990); Engle et al. (1987) and Franses et al. (2008)).

In the SV models, the variance is specified to follow some latent stochastic process. SV models were introduced in their basic form by Taylor (1982). They can be written as

$$r_t = \exp(h_t/2)\varepsilon_t, \quad (1)$$

$$h_t = \nu + \phi(h_{t-1} - \nu) + \sigma\eta_t, \quad (2)$$

where ε_t and η_t are i.i.d. $\mathcal{N}(0,1)$, r_t is the return and h_t is the logarithm of the variance on day t. Several extensions to the original SV model can be found. For example, heavier tails in the returns have been allowed by using the Student-t distribution (Ruiz, 1994; Harvey et al., 1994) or the Generalised Error Distribution (Liesenfeld and Jung, 2000) to model the random noise, ε_t, in Equation (1). Eraker et al. (2003) introduce jumps in both the return and the log-variance process. Harvey et al. (1994) model leverage effects by letting the two noise processes, ε_t and η_t, be correlated.

At present, GARCH models are by far the most popular in terms of real life applications. However, this seems mostly to be due to computational convenience (Fleming and Kirby, 2003). The superior performance of SV models over GARCH-type models in terms of in-sample fitting is generally accepted in the literature (Danielsson, 1994; Geweke, 1994; Kim et al., 1998), while there seems to be some debate about out-of-sample performance (see for example Bluhm and Yu (2001), Yu (2002), Sadorsky (2004), and Pederzoli (2006)). It should be noticed, however, that because of the computational complexity of SV models, it is often only the basic SV model that is compared to the GARCH alternatives. In general, SV models are recognised to be more flexible and realistic than GARCH models (Kim et al., 1998). Moreover, they represent the natural discrete time version of the continuous time models upon which much of modern finance theory has been developed.

The main difficulty related to SV models is computational. In fact, while GARCH-type models, even in their most complex form, are relatively straightforward to estimate using maximum likelihood optimisation techniques, SV models require a much more involved statistical and computational implementation. In fact, because of the unobserved conditional variance, the likelihood function in SV models does not have a closed form. The challenge is even bigger if we leave the basic SV model for more complex versions. Several estimation methods have been proposed in the literature, ranging from the less efficient generalised methods of moments (Andersen and Sorensen, 1996), and quasi-likelihood methods (Harvey et al., 1994) to more efficient methods such as simulated maximum likelihood (Danielsson, 1994) and Markov Chain Monte Carlo (MCMC) (Andersen et al., 1999). For a complete review over estimation methods for stochastic volatility models proposed until now, see Broto and Puiz (2004).

Since MCMC is considered one of the most efficient estimation tools (Andersen et al.,

1999), much attention has been devoted to the development of efficient MCMC algorithms for SV models, e.g. Chib et al. (2002), and Shephard and Pitt (1997). However, creating fast and efficient MCMC algorithms is far from easy. In fact, the large dimension of the latent log-variance vector and the strong correlation structures which are often found in SV models, make even well constructed MCMC algorithms slow, and their convergence dubious to assure (see for example Rue and Held (2005)).

In this we paper solve the inferential problem for SV models by Integrated Nested Laplace Approximations (INLA). INLA was introduced by Rue et al. (2009). It is a new approach for inference on latent Gaussian models (which most SV models can be thought to be part of). It substitutes MCMC sampling with a series of numerical approximations, providing very accurate estimates for the posterior marginals and the parameters of the model in only a fraction of the time needed by MCMC algorithms. We show that, using INLA, a complete Bayesian analysis of SV models is possible with largely reduced computational costs, even when we depart from the basic model. In particular, the model for ϵ_t in Equation (1) can be changed without large impacts on the estimation procedure.

Moreover, INLA makes it possible to compute routinely, and at a very low additional cost, other quantities of interest, such as marginal likelihoods. These can be used as a tool for model comparison. Many different models can then be tested against each other, without waiting for long computations.

A thorough comparison between results from INLA and MCMC has been performed in Rue et al. (2009). They show that INLA outperforms even smart MCMC schemes in terms of accuracy. Even extremely long MCMC runs could not detect any error in the approximation produced by INLA, despite the computational time for INLA being only a small fraction of the time needed by MCMC.

The aim of this paper is not to introduce new SV models or to assess which model best fits a certain data set, but to show how existing models, which are considered 'hard' to solve can be computed quickly and in a user friendly way using INLA. Our hope is that the use of INLA will help SV models to exit the academic world and reach the financial industry, where, due to the estimation complexity, few actors use them in practice. With almost instant inference, the estimation can be done in real time by an active trader or risk manager operating in the market.

The rest of the paper is organised as follows. In Section 2, we introduce the basic SV model and the extensions that are examined in this paper. Section 3 describes the basics of INLA and how it can be applied to SV models. In Section 4, INLA is applied to fit different SV models to two different data sets, while Section 5 discusses how INLA could be applied to generalisations of the basic SV model. Finally, we end with a discussion in Section 6.

2 Stochastic Volatility Models

2.1 Model description

The basic SV model, given by Equations (1) and (2), assumes that the latent log-variance vector $\boldsymbol{h} = \{h_1, \ldots, h_n\}$ follows a stationary autoregressive model of order 1 (AR1) with a common mean ν, a persistence parameter $\phi \in (0, 1)$ and a Gaussian noise η_t. The conditional distribution of the returns r_t, given h_t, is Gaussian with time varying variance.

In many real data applications, a Gaussian distribution for ϵ_t seems to be too restrictive. Financial time series, in fact, tend to show heavier tails than those of a Gaussian

distribution. In this paper, we consider three different models for ϵ_t: the Gaussian, the Student-t and the normal inverse Gaussian (NIG) distribution. All three distributions are standardised in order to have zero mean and unit variance. The standardised NIG density is defined as

$$f(y;\boldsymbol{\theta}_1) = \frac{\gamma\psi}{\pi}\sqrt{\frac{\beta^2+\psi^2}{(\gamma x+\beta)^2+\psi^2}}\ \exp\left(\psi^2+\beta(\gamma x+\beta)\right)\ K_1\left(\sqrt{(\beta^2+\psi^2)\left((\gamma x+\beta)^2+\psi^2\right)}\right),$$

where $K_1(\cdot)$ is the modified Bessel function of the third kind of order 1 and $\gamma^2 = 1+\beta^2/\psi^2$. The parameter β controls the skewness of the density, while ψ is a shape parameter. The NIG model is widely used in finance. In the context of SV models it has been used to define a model, alternative to that in Equations (1) and (2), in which an explicit formulation for the likelihood function exists, see for example Barndorff-Nielsen (1997) and Andersson (2001). However, to our knowledge, the NIG distribution has never been used to model the random noise ϵ_t in Equation (1). The Student-t distribution, on the other hand, is a quite popular choice for modelling ϵ_t, see for example Chib et al. (2002).

The model specification is completed with the definition of prior distributions for all the parameters of the model. For the common mean ν in Equation (2) we assume a Gaussian prior with a large known variance and mean 0. The joint density for the vector $\boldsymbol{x} = \{h_1, h_2, \ldots, \nu\}$, with length $|\boldsymbol{x}| = n$, will then be Gaussian with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ governed by the parameters $\boldsymbol{\theta}_1 = \{\phi, \sigma^2\}$. To the variance σ^2, we assign a vague inverse Gamma prior. As for the persistence parameter, ϕ, we assign the parameter

$$\phi^* = \text{logit}\frac{\phi+1}{2}$$

a Gaussian distribution whose precision parameter is chosen such that the corresponding prior for ϕ is roughly uniform in $(0,1)$. This is done to have all parameters defined over the whole real line, which gives computational simplifications. However, other priors could be defined without much influence on the estimation procedure. Finally, we need to assign priors to the parameters in the distribution for ϵ_t. We indicate these generically with $\boldsymbol{\theta}_2$: the Gaussian model has no parameters ($\boldsymbol{\theta}_2 = \emptyset$), the Student-$t$ model has one parameter ξ ($\boldsymbol{\theta}_2 = \xi$), and the NIG model has two parameters ($\boldsymbol{\theta}_2 = (\beta, \psi)$). We assign vague priors to all these parameters. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ be the vector of all model parameters, its lenght M then varies from a minimum of 2 (Gaussian model for ϵ_t) to a maximum of 4 (NIG model for ϵ_t).

2.2 Inferential goals

We assume that, in general, the goal of inference for the models presented in Section 2.1, is the marginals from the posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ given the n_d observed returns $\boldsymbol{r}$. The posterior distribution is proportional to the product of the priors and the likelihood, i.e.:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{r}) \ \propto\ \pi(\boldsymbol{\theta})\ \pi(\boldsymbol{x} \mid \boldsymbol{\theta})\ \prod_{t=1}^{n_d} \pi(r_t \mid h_t, \boldsymbol{\theta}). \tag{3}$$

The marginals $\pi(h_{n_d+i}|\boldsymbol{r})$ and $\pi(h_t|\boldsymbol{r})$ from the posterior density in Equation (3) can be used for prediction and smoothing of the latent log-volatility vector. Further, the marginals $\pi(\theta_j|\boldsymbol{r})$ and $\pi(\nu|\boldsymbol{r})$ can be used for parameter estimation.

One is often also interested in predicting future returns through the density $\pi(r_{n_d+i}|\boldsymbol{r})$. If we can compute marginals for the latent volatility vector, the predictive densities for the returns can be found as

$$\pi(r_{n_d+i}|\boldsymbol{r}) = \int \pi(r_{n_d+i}|h_{n_d+i})\pi(h_{n_d+i}|\boldsymbol{r})\mathrm{d}h_{n_d+i}. \tag{4}$$

This is a one dimensional integral, which is not difficult to solve numerically. If the likelihood density $\pi(r_{n_d+i}|h_{n_d+i})$ depends on some parameters $\boldsymbol{\theta}_2$ (e.g. in the case of ϵ_t being Student-t or NIG-distributed), a common practice is to fix the parameters at some convenient value; for example the posterior mean, and consider $\pi(r_{n_d+i}|\boldsymbol{r}, \hat{\boldsymbol{\theta}}_2)$. ¿From the density in Equation (4) it is possible to numerically compute the Value at Risk (VaR). The VaR is essentially a quantile of the return distribution, widely used as a measure of risk in the financial industry.

3 Integrated Nested Laplace Approximation

The main problem with inference about Equation (3) is that it is not available in closed form, since the likelihood (as a function of $\boldsymbol{x}$) is far from Gaussian. Although it is possible to implement MCMC algorithms to explore $\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{r})$, they are often very slow. We introduce a new tool for inference about marginals from $\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{r})$, named Integrated Nested Laplace Approximations (INLA), which provide very accurate approximations to $\pi(h_t \mid \boldsymbol{r})$, $\pi(\nu \mid \boldsymbol{r})$, and $\pi(\theta_j \mid \boldsymbol{r})$ in only a fraction of the time used by clever MCMC algorithms. Before describing the INLA scheme in details in Section 3.2, we present the main building blocks of this approach in Section 3.1. Section 3.3 discusses the speed-up of INLA using an Empirical Bayes approach, while Sections 3.4 and 3.5 describe how prediction and model comparison may be conducted using INLA.

3.1 Fundamentals of the INLA approach

3.1.1 Sparse matrix computations

The Gaussian vector $\boldsymbol{x}$ exhibits a particular conditional dependence (or Markov) structure which is reflected in its precision matrix (the inverse of the covariance matrix) $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. In particular, it is easy to show that, for the SV models presented here, $\boldsymbol{Q}$ is a tridiagonal matrix, where in addition also the last row and column are non-zero (this row and column are due to the common mean term). Most of the entries in the precision matrix are zero, which means that $\boldsymbol{Q}$ is sparse. The computational efficiency of the INLA approach relies on the sparseness of the precision matrix. All matrix operations, like solving systems and determinant computations, can be solved much faster for sparse matrices than for dense ones, see Rue and Held (2005).

3.1.2 The Gaussian approximation

The core of the INLA procedure is a Gaussian approximation to densities of the form

$$\pi(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum g_t(h_t)\right\}, \tag{5}$$

where $\boldsymbol{x} = \{\boldsymbol{h}, \nu\}$ as previously stated, and $g_t(h_t) = \log \pi(r_t|h_t, \boldsymbol{\theta})$. The Gaussian approximation, $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta})$, is found by matching the mode and the curvature at the mode. The

mode is computed iteratively using a Newton-Raphson algorithm. The approximation is

$$\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{r},\boldsymbol{\theta}) = K_1 \ \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q}+\text{diag}(\boldsymbol{c}))(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{6}$$

where K_1 is a normalising constant, $\boldsymbol{\mu}$ is the modal value of $\pi(\boldsymbol{x}|\boldsymbol{r},\boldsymbol{\theta})$ and the vector $\boldsymbol{c}$ is given by second order terms in the Taylor expansion of $\sum g_t(h_t)$ at the modal value $\boldsymbol{\mu}$. Note that the new precision matrix, $\boldsymbol{Q}+\text{diag}(\boldsymbol{c})$, will have exactly the same non-zero structure as the precision matrix $\boldsymbol{Q}$.

3.2 The INLA scheme

The INLA algorithm builds approximations to $\pi(x_t|\boldsymbol{r})$, $t = 1,\ldots,n$ using the following scheme:

1. Build an approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$.
2. Build an approximation to $\pi(x_t|\boldsymbol{\theta},\boldsymbol{r})$.
3. Compute an approximation to $\pi(x_t|\boldsymbol{r})$ using the approximations from steps 1 and 2.

In Sections 3.2.1-3.2.3 each of the three steps are described in more details.

3.2.1 Approximating $\pi(\boldsymbol{\theta}|\boldsymbol{r})$

An approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ is built starting from the identity

$$\pi(\boldsymbol{\theta}|\boldsymbol{r}) = \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{r})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})} \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{r})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})}, \tag{7}$$

and then approximating the denominator of the rightmost part of Equation (7) via the Gaussian approximation described in Section 3.1.2. Hence, the approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ is:

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r}) = K_2 \ \left.\frac{\pi(\boldsymbol{x},\boldsymbol{r},\boldsymbol{\theta})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{8}$$

where K_2 is the normalising constant, $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})$ is the Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})$. Equation (8) is equivalent to Tierney and Kadane (1986)'s Laplace approximation of a marginal posterior distribution, suggesting that the approximation error is of order $\mathcal{O}(n_d^{-3/2})$.

3.2.2 Approximating $\pi(x_t|\boldsymbol{\theta},\boldsymbol{r})$

Approximating $\pi(x_t|\boldsymbol{\theta},\boldsymbol{r})$ is the most challenging part of the INLA procedure. The starting point is the identity

$$\pi(x_t|\boldsymbol{\theta},\boldsymbol{r}) = \frac{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})}{\pi(\boldsymbol{x}_{-t}|x_t,\boldsymbol{\theta},\boldsymbol{r})} \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{r})}{\pi(\boldsymbol{x}_{-t}|x_t,\boldsymbol{\theta},\boldsymbol{r})}, \tag{9}$$

where $\boldsymbol{x}_{-t}$ indicates that element t of the vector has been removed. The denominator of the rightmost fraction is approximated using the results from Section 3.1.2, giving the following approximation to $\pi(x_t|\boldsymbol{\theta}, \boldsymbol{r})$:

$$\widetilde{\pi}(x_t|\boldsymbol{\theta}, \boldsymbol{r}) \propto \left. \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{r})}{\widetilde{\pi}_{GG}(\boldsymbol{x}_{-t}|x_t, \boldsymbol{\theta}, \boldsymbol{r})} \right|_{\boldsymbol{x}_{-t}=\boldsymbol{x}^*_{-t}(x_t,\boldsymbol{\theta})}. \tag{10}$$

This approximation can be quite heavy from a computational point of view, and is therefore not convenient in practice. The solution in Rue et al. (2009) is to compute a simplified version $\widetilde{\pi}_{\mathrm{S}}(x_t|\boldsymbol{\theta}, \boldsymbol{r})$ of $\widetilde{\pi}(x_t|\boldsymbol{\theta}, \boldsymbol{r})$, using a series expansion of the logarithm of $\widetilde{\pi}(x_t|\boldsymbol{\theta}, \boldsymbol{r})$:

$$\log \widetilde{\pi}_{\mathrm{S}}(x_t|\boldsymbol{\theta}, \boldsymbol{r}) = \text{constant} - \frac{1}{2}x_t^2 + \gamma_t^{(1)}(\boldsymbol{\theta})x_t + \frac{1}{6}x_t^3\gamma_t^{(3)}(\boldsymbol{\theta}) + \cdots . \tag{11}$$

Here $\gamma_t^{(1)}$ and $\gamma_t^{(3)}$ are the terms in the Taylor expansion, which can be derived re-using already performed computations, resulting in a very fast algorithm.

Since the third order term is unbounded, Equation (11) does not define a proper density. We therefore fit the logarithm of a Skew-Normal distribution (Azzalini and Capitanio, 1999) to the Taylor Expansion. The fit is done such that the third derivative at the mode is $\gamma_t^{(3)}$, the mean is $\gamma_t^{(1)}$ and the variance is 1. In this way, $\gamma_t^{(3)}$ only contributes to the skewness, whereas the adjustment in the mean comes from $\gamma_t^{(1)}$. For details on how to derive the series expansion in Equation (11) and on how to fit the Skew-Normal density to Equation (11), see Rue et al. (2009).

3.2.3 Approximating $\pi(x_t|\boldsymbol{r})$

Once approximations to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ and $\pi(x_t|\boldsymbol{\theta}, \boldsymbol{r})$ are available, we compute an approximation to $\pi(x_t|\boldsymbol{r})$ using the numerical integration scheme:

$$\widetilde{\pi}(x_t|\boldsymbol{r}) = \sum_k \widetilde{\pi}(x_t|\boldsymbol{\theta}^k, \boldsymbol{r})\widetilde{\pi}(\boldsymbol{\theta}^k|\boldsymbol{r})\Delta_k, \tag{12}$$

for some selected values $\boldsymbol{\theta}^k$ of $\boldsymbol{\theta}$. An easy way to select the points $\boldsymbol{\theta}^k$, is to create a grid of points covering the area of high density for $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$. In order to correctly locate the interesting area, we compute the mode $\boldsymbol{\theta}^*$ of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$, and its Hessian $H(\boldsymbol{\theta}^*)$ at the mode, and use those as guidelines. It turns out that as long as the points are well located in the $\boldsymbol{\theta}$-space, the number of points necessary to obtain a good accuracy of Equation (12), does not need to be high. For more details about the grid construction and alternative ways to select points in the $\boldsymbol{\theta}$-space, see Rue et al. (2009).

3.3 Empirical Bayes scheme

The scheme described in Section 3.2 performs a full Bayesian analysis. A faster alternative would be to assume an Empirical Bayes approach and approximate $\pi(x_t|\boldsymbol{r})$ with $\widetilde{\pi}(x_t|\boldsymbol{r}, \boldsymbol{\theta}^*)$ where $\boldsymbol{\theta}^*$ is the mode of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$. One then avoids the numerical integration described in Section 3.2.3. The Empirical Bayes is very accurate in cases where the posterior distribution of $\boldsymbol{\theta}|\boldsymbol{r}$ is reasonably regular.

3.4 Predicting future return values

As stated in Section 2.2 it is often of interest to predict future return values, r_{n_d+i}. The INLA approach does not compute the marginals, $\pi(r_{n_d+i}|\boldsymbol{r})$, directly. However, an approximation for such densities can easily be found via numerical approximation of the integral in Equation (4) as

$$\widetilde{\pi}(r_{n_d+i}|\boldsymbol{r}) \approx \sum_j \pi(r_{n_d+i}|h^j_{n_d+i})\widetilde{\pi}(h^j_{n_d+i}|\boldsymbol{r})\Delta_j, \tag{13}$$

where $\widetilde{\pi}(h^j_{n_d+i}|\boldsymbol{r})$ is the approximation of the posterior marginal for $h^j_{n_d+i}|\boldsymbol{r}$ computed by INLA. For likelihood models which depend on additional parameters, e.g. where the error distribution is Student-t or NIG, we have chosen to fix the parameter values at the posterior mode computed during the INLA process.

3.5 Model comparison

While the main focus of the INLA approach is to compute posterior marginals, it is also possible to compute other interesting quantities with low additional costs. In particular, marginal likelihoods $\pi(\boldsymbol{r})$ are useful quantities for model comparison. For instance, two competing and apriori equiprobable models $\mathcal{M}_1$ and $\mathcal{M}_2$ may be compared using the Bayes factor, defined as the ratio between the corresponding marginal likelihoods:

$$\mathcal{B}(1,2) = \frac{\pi(\boldsymbol{r}|\mathcal{M}_1)}{\pi(\boldsymbol{r}|\mathcal{M}_2)}.$$

Jeffreys (1961) provides a scale for the interpretation of $\log \mathcal{B}(1,2)$. Model comparison becomes particularly interesting when a fast inference procedure like INLA makes it possible to fit more models to the same data set.

In the INLA framework, the marginal likelihood, $\pi(\boldsymbol{r})$, can be approximated as the normalising constant of $\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$:

$$\widetilde{\pi}(\boldsymbol{r}) = \int \left.\frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{r})}{\widetilde{\pi}_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})}\right|_{\boldsymbol{x}=\boldsymbol{x}^\star(\boldsymbol{\theta})} d\boldsymbol{\theta}. \tag{14}$$

We propose two approximations for the marginal likelihood $\pi(\boldsymbol{r})$. The first, cruder one, $\widetilde{\pi}_1(\boldsymbol{r})$, is based on approximating the density of $\boldsymbol{\theta}|\boldsymbol{r}$ with a Gaussian distribution

$$\widetilde{\pi}_G(\boldsymbol{\theta}|\boldsymbol{r}) = \mathcal{N}(\boldsymbol{\theta}^*, \boldsymbol{H}^{-1}).$$

Here $\boldsymbol{\theta}^*$ is the modal configuration of $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ and $\boldsymbol{H}$ is the negative Hessian matrix computed at the mode. This gives

$$\widetilde{\pi}_1(\boldsymbol{r}) = (2\pi)^{M/2}|\boldsymbol{H}|^{-1/2}. \tag{15}$$

The second approximation, $\widetilde{\pi}_2(\boldsymbol{r})$, is more precise, but also more expensive to compute. It assumes no parametric form of the density of $\boldsymbol{\theta}|\boldsymbol{r}$. Instead it is computed by solving the integral in Equation (14) numerically using selected values $\boldsymbol{\theta}_k$ of $\boldsymbol{\theta}$. This approximation allows departures from Gaussianity, a feature often encountered in real applications.

Note that when computing an approximation to the marginal likelihood $\pi(\boldsymbol{r})$, aiming to use it for model comparison, it is important to include all normalising constants that appear in the priors for the hyperparameters, $\pi(\boldsymbol{\theta})$, and the latent vector, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$. Moreover, one has to include the constants in the likelihood term, $\pi(\boldsymbol{r}|\boldsymbol{x},\boldsymbol{\theta})$.

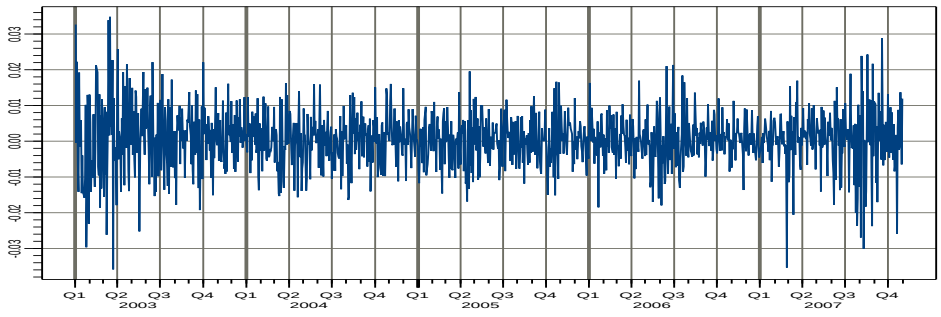


Figure 1: **SP500 index:** Logarithmic returns from 02.01.2003 to 31.10.2006.

4 Examples

In this section we apply the INLA approach to fit the models presented in Section 2.1 to two different data sets. We will emphasize that the aim of this paper is rather to present some of the possibilities offered by the INLA approach than to claim which model is the best for each data set.

All algorithms necessary for the INLA approach are efficiently implemented in the `inla` program, built upon the `GMRFLib`-library (Rue and Held, 2005, Appendix), open-source and freely available from `http://www.math.ntnu.no/~hrue/`. Both the `GMRFLib`-library and the `inla` program use the Open-MP API (see `www.openmp.org`) to speed up computation on multi-core processors. A user friendly interface to the `GMRFLib`-library and `inla` program, the `R` package `INLA`, is also available from the same web page. All examples in this paper have been implemented using the `R` interface.

Detailed comparisons between INLA and MCMC results can be found in Rue et al. (2009). In this paper, we have compared the results and computing times using INLA to those using a Gibbs sampler, as implemented in WinBUGS/OpenBUGS (Spiegelhalter et al., 2003).Our choice of WinBUGS/OpenBUGS is due to two reasons. First, it has been proposed as a tool to implement MCMC algorithms on SV models by Mayer and Yu (2000). Second, the programming effort the user has to put in WinBUGS/OpenBUGS is similar to that he has to put in the `R` package `INLA`.

4.1 SP500 index data

Our first data set consists of 1217 daily closing quotations of the SP500 index from 02.01.2003 to 31.10.2006. Figure 1 shows the logarithmic returns for the SP500 index data. We fit the models described in Section 2.1 to the data set. The main benefit of the INLA approach is its speed. The computing time goes from 5 seconds for the Gaussian model, to 10 seconds for the NIG one on a quad core 2.6GHz CPU. Table 1 shows the estimated posterior mean and standard deviation for the parameters as computed from the approximated posterior marginals. In Table 2 the same parameters for the Gaussian and Student-t models are estimated based on 10^6 samples using WinBUGS/OpenBUGS (the NIG model is not implemented in WinBUGS/OpenBUGS). Despite the large number of samples, trace plots (not included here) still show signs of the "stickiness" of the MCMC algorithm. The time used by WinBUGS/OpenBUGS to produce 10^6 samples was close to two hours.

Table 1: **SP500:** Posterior means and standard deviations for the parameters in the SV models obtained using **INLA**.

Model	ν	ϕ	$\tau = 1/\sigma^2$	d.o.f.	β	ψ
Gaussian	-0.52(0.23)	0.98(0.006)	95.15(35.66)	-	-	-
Student-t	-0.50(0.25)	0.98(0.005)	123.44(50.75)	17.14(8.58)	-	-
NIG	-0.48(0.25)	0.99(0.005)	135.11(56.14)	-	-0.32(0.19)	2.39(0.64)

Table 2: **SP500:** Posterior means and standard deviations for the parameters in the SV models fit to the SP500 index data set obtained using **WinBUGS/OpenBUGS**.

Model	ν	ϕ	$\tau = 1/\sigma^2$	d.o.f.
Gaussian	-0.52(0.25)	0.98(0.007)	96.6(41.09)	-
Student-t	-0.48(0.29)	0.98(0.005)	142.0(57.6)	19.6(15.17)

The three models may be compared using marginal likelihoods. We compute both approximations described in Section 3.5. The results, shown in Table 3 indicate that the Gaussian distribution for ϵ_t is preferable. This is not strange if we study the estimated parameters in Table 1. The degrees of freedom parameter in the Student-t model is quite large, and the parameters of the NIG distribution correspond to a density which is quite similar to the Gaussian.

Figure 2 shows the posterior mean and the 25 and 97,5% quantiles for the estimated latent historical volatility, $\exp(h_t/2)$, in percent, as well as for the predicted volatility during the 60 day-period starting in 01.11.2006. Figure 3 displays the Gaussian, Student-t and NIG densities for the predicted return (in percent) at 01.11.2006, and Table 4 shows the 95 and 99% VaRs corresponding to these densities. Both the figures and the values in the table indicate that the differences between the models is small.

4.2 Microsoft data

Our second example consists of 1292 daily closing prices of the Microsoft stock from 03.01.2003 to 21.02.2008. The log-returns are shown in Figure 4. They are much more volatile than the SP500 index returns, and also contain some extreme values.

Table 5 shows the estimated posterior mean and standard deviation for the parameters estimated using INLA, and Table 6 reports the WinBUGS/OpenBUGS estimates for the parameters in the Gaussian and Student-t models. It can be noticed that, both for INLA and WinBUGS, the estimated precision for the latent vector is considerably lower for the Gaussian model than for the two others.

Table 3: **SP500:** Approximated values for the log marginal likelihood obtained using INLA. The the preferred model is indicated by a $(*)$.

Model	Marg. Likelihood $\log \widetilde{\pi}_1(\boldsymbol{r})$	Marg. Likelihood $\log \widetilde{\pi}_2(\boldsymbol{r})$
Gaussian	-1391.47(*)	-1391.49(*)
Student-t	-1393.28	-1393.34
NIG	-1396.87	-1396.87

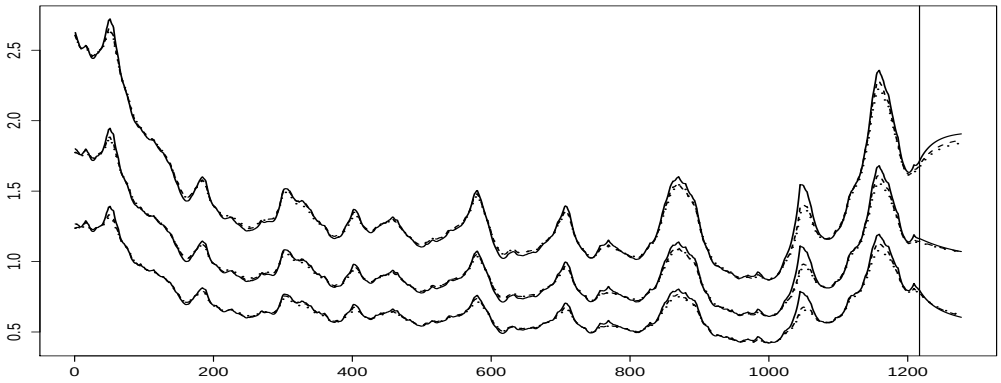


Figure 2: **SP500 index:** Posterior mean and the 25 and 97,5% quantiles for the estimated latent historical volatility, $\exp(h_t/2)$, in percent, as well as for the predicted volatility during the 60 day-period starting in 01.11.2006 for the three different models. The solid line corresponds to the Gaussian model, the broken line to the Student-t model and the dotted line to the NIG model. The vertical line corresponds to the end of the estimation period.

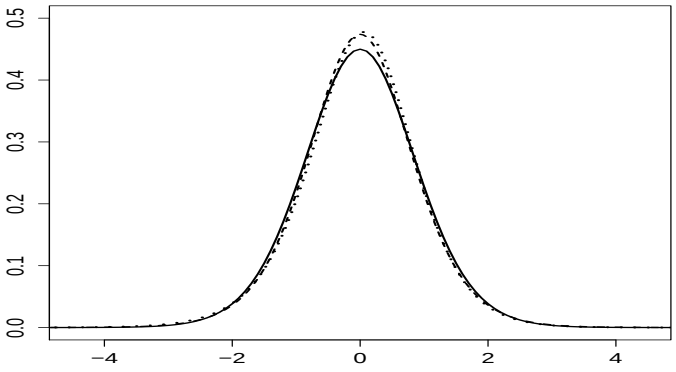


Figure 3: **SP500 index:** The Gaussian, Student-t and NIG densities for the predicted return (in percent) at 01.11.2006. The solid line corresponds to the Gaussian model, the broken line to the Student-t model, and the dotted line to the NIG model.

Table 4: **SP500:** VaR at 01.11.2006 for the return in percent assuming zero mean.

Model	$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$
Gaussian	1.405	2.113
Student-t	1.436	2.216
NIG	1.407	2.143

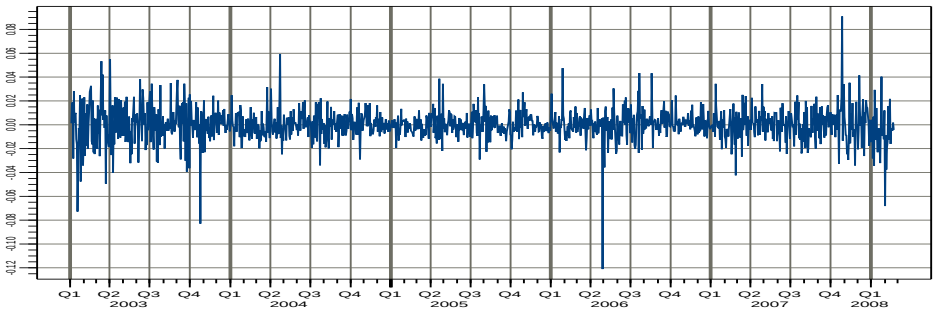


Figure 4: **Microsoft stock:** Logarithmic returns from 03.01.2003 to 21.02.2008.

Table 5: **Microsoft:**Posterior means and standard deviations for the parameters in the SV models obtained using **INLA**.

Model	ν	ϕ	$\tau = 1/\sigma^2$	d.o.f.	β	ψ
Gaussian	0.21(0.14)	0.92(0.02)	8.98(2.44)	-	-	-
Student-t	0.45(0.25)	0.98(0.01)	61.55(27.61)	5.48(0.89)	-	-
NIG	0.41(0.23)	0.98(0.01)	48.33(20.89)	-	0.11(0.08)	1.28(0.15)

Table 6: **Microsoft:** Posterior means and standard deviations for the parameters in the SV models obtained using **WinBUGS/OpenBUGS**.

Model	ν	ϕ	$\tau = 1/\sigma^2$	d.o.f.
AR-1 Gaussian	0.20(0.14)	0.92(0.02)	9.1(3.1)	-
AR-1 Student-t	0.45(0.28)	0.98(0.01)	67.4(40.4)	5.4(0.9)

Table 7: **Microsoft:** Approximated values for the log marginal likelihood obtained using INLA. The the preferred model is indicated by a $(*)$.

Model	Marg. Likelihood $\log \widetilde{\pi}_1(\boldsymbol{r})$	Marg. Likelihood $log\widetilde{\pi}_2(\boldsymbol{r})$
Gaussian	-2083.33	-2083.39
Student-t	-2065.74(*)	-2065.87(*)
NIG	-2073.86	-2074.06

Table 8: **Microsoft:** VaR at 22.02.08 for the return in percent assuming zero mean.

Model	$\text{VaR}_{0.95}$	$\text{VaR}_{0.99}$
Gaussian	1.842	3.054
Student-t	2.593	4.315
NIG	2.532	4.172

As in the SP500 example, we use the estimated marginal likelihoods to compare the models. The results are reported in Table 7.

The preferred model is the one with Student-t distribution for ϵ_t. This agrees well with the fact that the Microsoft data presents extreme values that would not be allowed under a Gaussian model. The results for the NIG distribution are very similar to those obtained under the Student-t model. Hence, it is reasonable that the model choice criteria prefers the Student-t model with only one parameter in favour of the NIG model with two.

Figure 5(a) displays the posterior mean, and Figure 5(b) the 25 and 97,5% quantiles, for the estimated latent historical volatility in percent, as well as for the predicted volatility during the 60 day-period starting in 22.02.2008. On the contrary to the SP500 example, for which all models gave similar results, the Gaussian model now differs quite much from the other two. This is especially the case for the volatility prediction. The posterior mean appears to be shifted, and the inter-quantile range wider, compared to the two other models. The difference is also apparent looking at the characteristics of the distributions for the predicted return (in percent), shown in Figure 6 and Table 8.

5 Extensions

The INLA approximation may, without much additional effort, deal with some of the extensions of the SV model most commonly proposed in the literature. In this section we describe how non-stationarity, leverage effects and several dimensions may be modelled using INLA.

5.1 Non-stationary models

For real financial time series, the mean-reverting parameter ϕ in Equation (2) is often estimated to be very close to one, indicating an almost non-stationary model. This might create problems in the estimation of the model parameters, since the stationary AR1 model assumes a constant mean level for the volatility, while the actual mean could be slowly changing over time.

The non-stationary random walk model of order 1 (RW1) is an alternative to the AR1 model for the latent log-variance. Non-stationary SV models have been used as

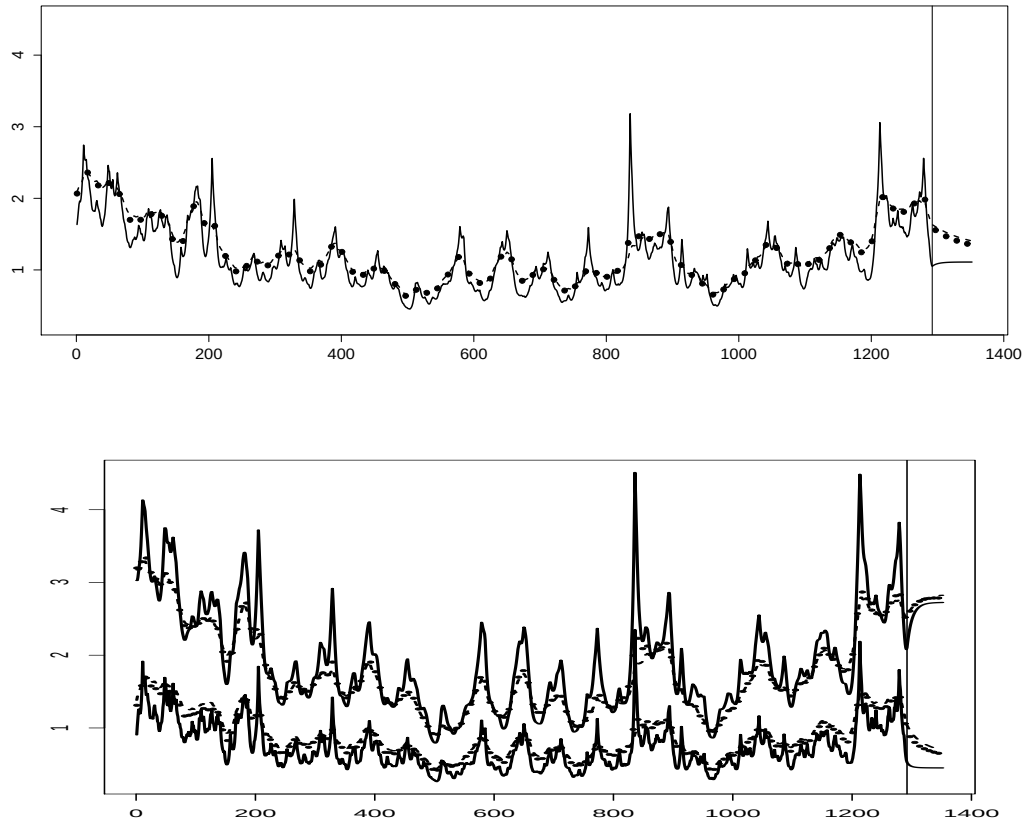


Figure 5: **Microsoft stock:** The upper panel shows posterior mean for the estimated latent historical volatility, $\exp(h_t/2)$, in percent, as well as for the predicted volatility during the 60 day-period starting in 22.02.2008 for the three different models. The lower panel shows the 25 and 97,5% quantiles for the same quantities. The solid line corresponds to the Gaussian model, the broken line to the Student-t model and the dotted line to the NIG model. The vertical line corresponds to the end of the estimation period.

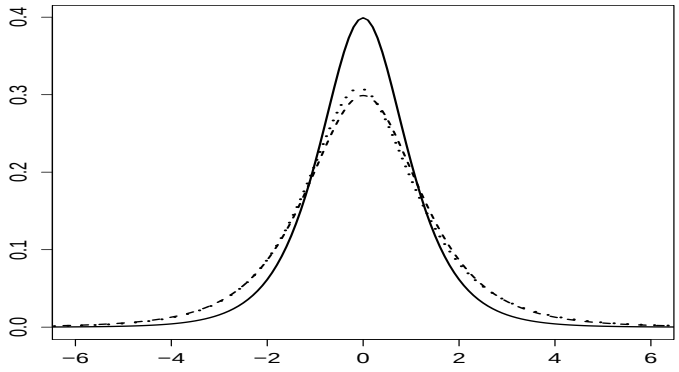


Figure 6: **Microsoft stock:** The Gaussian, Student-t and NIG densities for the predicted return (in percent) at 22.02.2008. The solid line corresponds to the Gaussian model, the broken line to the Student-t model and the dotted line to the NIG model.

alternatives to the integrated GARCH (or IGARCH) model, see Harvey et al. (1994) and Andersson (2001). In a non-stationary SV model, Equation (1) is unchanged, while Equation (2) becomes

$$h_t = h_{t-1} + \sigma\eta_t, \tag{16}$$

where η_t is standardised Gaussian noise.

Estimating this kind of models using the INLA approach is straightforward. The only change from the stationary model is the zero structure of the prior precision matrix $\boldsymbol{Q}$ for the latent vector $\boldsymbol{x}$, which now includes only the log-variances $\{h_1, h_2, \dots\}$. All computations remain exactly the same. Non-stationary SV models can be implemented using the current version of the `R` package `INLA`.

5.2 Asymmetric SV models

One feature often observed in financial studies is that volatility responds asymmetrically to positive and negative return shocks. Several explanations have been proposed in the literature to explain the presence of such an asymmetric relationship between volatility and returns. Two of the most widely cited, Black (1976) and Christie (1982), suggest that the asymmetry reflects a change in financial leverage. In particular, the argument is that, when an asset experiences a positive (negative) return, it becomes less (more) risky, thus decreasing (increasing) its volatility. In other words, there is a negative correlation between returns and volatility. This is known as the *leverage* effect.

A univariate SV model with leverage effects was first introduced by Harvey and Shephard (1996), letting the two error processes in Equations (1) and (2) be negatively correlated. Formally, $\text{Corr}(\epsilon_t, \eta_{t+1}) = \rho$, with $\rho < 0$. Note that we prefer to model $\text{Corr}(\epsilon_t, \eta_{t+1})$, and not $\text{Corr}(\epsilon_t, \eta_t)$, because the former is more logically appealing both from a theoretical and a empirical point of view, see Yu (2005).

The SV model with leverage effects is estimated by the quasi-likelihood method in Harvey and Shephard (1996), and by MCMC in Mayer and Yu (2000). We have not yet implemented the algorithms for this kind of models in the `R` package `INLA`. However, it can be done, and in this section we describe how we can perform approximate inference using INLA for SV models with correlated errors.

The core of the INLA approach is the Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{r})$. In order determine this approximation for models with leverage effects, we need to have an expression for the likelihood of each data point, $\pi(r_t|\boldsymbol{x}, \boldsymbol{\theta})$. With some algebra, it can be shown that

$$\pi(r_t|\boldsymbol{x}, \boldsymbol{\theta}) = \pi(r_t|h_t, h_{t+1}, \boldsymbol{\theta}) = \mathcal{N}\left\{\frac{\rho}{\sigma}e^{h_t/2}[h_{t+1} - \nu + \phi(h_t - \nu)], e^{h_t}(1-\rho^2)\right\}. \tag{17}$$

Note that, unlike for the models analysed before, each data point, r_t, depends on three nodes of the latent vector, namely, h_t, h_{t+1} and the common mean ν. The prior distribution for the vector $\boldsymbol{x}$ is unchanged. Hence, the full conditional reads

$$\pi(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_{t=1}^{n_d} f_t(h_t, h_{t+1}, \nu)\right\}, \tag{18}$$

where $f_t(h_t, h_{t+1}, \nu) = \log \pi(r_t|h_t, h_{t+1}, \nu, \boldsymbol{\theta})$. We can expand $f_t(h_t, h_{t+1}, \nu)$ around the point $(h_t^0, h_{t+1}^0, \nu^0)$ obtaining

$$f_t(h_t, h_{t+1}, \nu) \approx \text{Const} + (h_t, h_{t+1}, \nu)\boldsymbol{b}_t - \frac{1}{2}(h_t, h_{t+1}, \nu)\boldsymbol{C}_t(h_t, h_{t+1}, \nu)^T,$$

where $\boldsymbol{C}_t$ is a 3×3 symmetric matrix and $\boldsymbol{b_t}$ a column vector of dimension 3. Both $\boldsymbol{b}_t$ and $\boldsymbol{C}_t$ are functions of the gradient and the Hessian matrix of $f_t(h_t, h_{t+1}, \nu)$, computed at $(h_t^0, h_{t+1}^0, \nu^0)$, and depend on the value of the hyperparameter vector $\boldsymbol{\theta}$. Let c_{ij}^t indicate element i, j of the matrix $\boldsymbol{C}_t$ and b_i^t indicate the ith element of vector $\boldsymbol{b}_t$, where $i, j = 1, 2, 3$. Moreover, let

$$\boldsymbol{C} = \begin{bmatrix} c_{11}^1 & c_{12}^1 & 0 & \dots & 0 & c_{13}^1 \\ c_{21}^1 & c_{22}^1 + c_{11}^2 & c_{12}^2 & \dots & 0 & c_{23}^1 + c_{13}^2 \\ 0 & c_{21}^2 & c_{22}^2 + c_{11}^3 & \dots & 0 & c_{23}^2 + c_{13}^3 \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & & & \dots & c_{22}^{n_d-1} + c_{11}^{n_d} & c_{23}^{n_d-1} + c_{13}^{n_d} \\ c_{31}^1 & c_{32}^1 + c_{31}^2 & c_{32}^2 + c_{31}^3 & \dots & c_{32}^{n_d-1} + c_{31}^{n_d} & \sum_{j=1}^{n_d} c_{33}^j \end{bmatrix},$$

and

$$\boldsymbol{b}^T = \left[b_1^1, b_2^1 + b_1^2, b_2^2 + b_1^3, \dots, \sum_{j=1}^{n_d} b_3^j \right].$$

Here $\boldsymbol{C}$ is a $n \times n$ matrix, where n is the dimension of $\boldsymbol{x}$, and $\boldsymbol{b}$ is a vector of length n. We can build a Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta})$ with precision matrix $\boldsymbol{Q} + \boldsymbol{C}$, and mean given by the solution of $(\boldsymbol{Q} + \boldsymbol{C})\boldsymbol{x}^* = \boldsymbol{b}$, where $\boldsymbol{x}^*$ is the modal configuration of $\pi(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta})$. Note that since h_t, h_{t+1} and ν are neighbours in the prior model for $\boldsymbol{x}$, the Gaussian approximation is a Gaussian Markov random vector with respect to the same graph, and therefore preserves the Markov properties of the prior distribution of $\boldsymbol{x}$.

Starting from the Gaussian approximation described above, it is possible to derive all the other algorithms necessary to implement the INLA approach also for SV models with correlated errors.

5.3 Multivariate SV models

Multivariate extensions of the SV models have lately been given much attention (see for example Vol. 25 of Econometric Review) There are several reasons, both economical and econometric, why multivariate volatility models are important. Financial assets are clearly correlated. Hence, the knowledge of the correlation structures is vital in many financial applications such as asset pricing, optimal portfolio risk management, and asset allocation. Compared to their univariate counterparts, multivariate models for financial assets have to be able to capture more features of the data: both returns and volatility may be cross-dependent. Moreover, volatility can spill over from one market to another, such that the knowledge about one asset can help predicting another. Here, we describe how INLA can be applied to bivariate SV models. The same ideas can be applied to higher dimensions.

A quite generic bivariate SV model can be written as

$$\begin{aligned} \boldsymbol{r}_t &= \boldsymbol{\Omega}_t \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon), \\ \boldsymbol{h}_t &= \boldsymbol{\nu} + \boldsymbol{\Phi}(\boldsymbol{h}_{t-1} - \boldsymbol{\nu}) + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\eta), \end{aligned} \tag{19}$$

where $\boldsymbol{r}_t = \{r_{t1}, r_{t2}\}$ are the observed log-returns at time t, $\boldsymbol{\epsilon}_t = \{\epsilon_{t1}, \epsilon_{t2}\}$ and $\boldsymbol{\eta}_t = \{\eta_{t1}, \eta_{t2}\}$ are two independent bivariate noise terms, and $\boldsymbol{h}_t = \{h_{t1}, h_{t2}\}$ the latent volatil-

ity. Moreover,

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_\epsilon = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_\eta = \begin{pmatrix} 1/\tau_{\eta_1} & \rho_\eta/\sqrt{\tau_{\eta_1}\tau_{\eta_2}} \\ \rho_\eta/\sqrt{\tau_{\eta_1}\tau_{\eta_2}} & 1/\tau_{\eta_2} \end{pmatrix}, \quad \boldsymbol{\Omega}_t = \begin{pmatrix} \exp(h_{1t}/2) & 0 \\ 0 & \exp(h_{2t}/2) \end{pmatrix}.$$

By specialising the different matrices in the model, different features of the data can be captured (see for example Yu and Mayer (2006)).

The core of the INLA approach is the Gaussian approximation to the full conditional of the latent vector, presented in Section 3.1.2. For the model in Equation (19), the full conditional reads

$$\pi(\boldsymbol{x}|\boldsymbol{r},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \sum_{t=1}^{n_d} g_t(\boldsymbol{h}_t)\right\}, \tag{20}$$

where $\boldsymbol{x} = (\boldsymbol{h}^T, \boldsymbol{\nu}^T)$ and $g_t(\boldsymbol{h}_t) = \log \pi(\boldsymbol{r}_t|\boldsymbol{h}_t, \rho_\epsilon)$. We can derive the Gaussian approximation to Equation (20) using the procedure described in Section 3.1.2. The only difference is that the likelihood term, $g_t(\boldsymbol{h}_t)$, depends on a bivariate vector instead of a scalar. The Gaussian approximation then reads

$$\widetilde{\pi}_G(\boldsymbol{x}) \propto \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T(\boldsymbol{Q} + \mathrm{diag}(\boldsymbol{C}))(\boldsymbol{x}-\boldsymbol{\mu})\}, \tag{21}$$

where $\mathrm{diag}(\boldsymbol{C})$ is a matrix of the form

$$\begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{0} & & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \vdots & & & & \\ \boldsymbol{0} & & \dots & \boldsymbol{C}_n & \boldsymbol{0} \\ \boldsymbol{0} & & & \dots & \boldsymbol{0} \end{bmatrix}. \tag{22}$$

Each term $\boldsymbol{C}_t$ is defined as

$$\boldsymbol{C}_t = -\begin{bmatrix} \frac{\partial^2 g_t(\boldsymbol{x}_t)}{\partial x_{t1}^2} & \frac{\partial^2 g_t(\boldsymbol{x}_t)}{\partial x_{t1}\partial x_{t2}} \\ \frac{\partial^2 g_t(\boldsymbol{x}_t)}{\partial x_{t1}\partial x_{t2}} & \frac{\partial^2 g_t(\boldsymbol{x}_t)}{\partial x_{t2}^2} \end{bmatrix}_{\boldsymbol{x}_t=\boldsymbol{\nu}}.$$

Starting from the Gaussian approximation above, it is possible to derive all other algorithms necessary to implement the INLA approach also for the bivariate SV models. For a comparison between the performance of the INLA approach and a Gibbs sampler algorithm applied to some bivariate SV models, see Martino (2007).

The main difficulty with performing a full Bayesian analysis on multivariate SV models using the INLA approach (but also with respect to most other estimation procedures) is the number of hyperparameters in the model. Remember, in fact, that in order to compute an approximation to $\pi(\boldsymbol{x}_t|\boldsymbol{r})$, we have to numerically integrate out the hyperparameters, an operation which heavily suffers from the curse of dimensionality. The Empirical Bayes approach described in Section 3.3 might be sufficient in practice. The investigation of this will be the scope of further research.

6 Discussion

In this paper we have presented a new approach to do inference on SV models named Integrated Nested Laplace Approximations (INLA). INLA computes extremely accurate approximations to the posterior marginals of latent log-variances and parameters in SV models. Such posterior marginals constitute the basis for inference on the model parameters and for prediction of future volatility and returns.

The main advantage of INLA is its speed: it provides answers in only a small fraction of the time used by a well designed MCMC algorithm. Near instant inference makes it possible for the user to fit several models to the same data set, or to analyse a large number of data sets in a reasonable time.

An additional advantage of INLA is that it can be used almost as a black box so that the programming effort of the final user is reduced to minimum. The `GMRFLib`-library and the `inla` program provide efficient implementation of all the algorithms needed, while the `R` package `INLA` makes these tools easily available for the final user.

Our hope is that almost instant inference, together with user-friendly implementation tools, will help SV models to exit the academic world and make them more appealing for the financial industry.

Acknowledgements

This work is sponsored by Statistics for Innovation, $(\text{sfi})^2$.

References

Andersen, T., H. Chung, and B. Sorensen (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics 91*, 61–87.

Andersen, T. and B. Sorensen (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics 13*, 329–352.

Andersson, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business and Economic Statistics 19*, 44–54.

Azzalini, A. and A. Capitanio (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of Royal Statistical Society, Series B 61*, 579:602.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*, 1–13.

Black, F. (1976). Studies of stock market volatility changes. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bluhm, H. and J. Yu (2001). Forecasting Volatility: Evidence from the German Stock Market. In *Modelling and Forecasting Financial Volatility*, pp. 173–193.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Broto, C. and E. Puiz (2004). Estimation Methods for Stochastic Volatility Models: A Survey. *Journal of Economic Surveys 18*, 613–649.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*, 281–316.

Christie, A. (1982). The stochastic behaviour of common stock variances: values leverage and interest rates effects. *Journal of Financial Economics 10*, 407–432.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics 64*, 375–400.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica 50*, 987–1008.

Engle, R., D. Lilien, and R. Robins (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica 55*, 391–407.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance 58*, 1269–1300.

Fleming, J. and C. Kirby (2003). A closer look at the relation between GARCH and Stochastic Autoregressive Volatility. *Journal of financial econometrics 1*, 365–419.

Franses, P., M. van der Leij, and R. Paap (2008). A simple test for GARCH against a Stochastic Volatility model. *Journal of Financial Econometrics 6*, 291–306.

Geweke, E. (1994). Bayesian comparison of econometric models. *Econometrica 57*, 1317–1340.

Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the normal excess return on stocks. *Journal of Finance 48*, 1779–1801.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics 14*, 429–434.

Jeffreys, H. (1961). *The theory of probability*. Oxford press.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Liesenfeld, R. and R. Jung (2000). Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics 15*(2), 137–160.

Martino, S. (2007). *Approximate Bayesian Inference for Latent Gaussian Models*. Ph. D. thesis, Dep. of Mathematics, Norwegian University of Science and Technology.

Mayer, R. and J. Yu (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal 3*, 198–215.

Nelson, D. (1990). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

Pederzoli, C. (2006). Stochastic volatility and GARCH: A comparison based on UK stock data. *The European Journal of Finance 12*, 41–59.

Rue, H. and L. Held (2005). *Gaussian markov random fields: Theory and applications*, Volume 104 of Monographs on statistics and applied probability. Chapman & Hall, London.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Accepted as a discussion paper in JRSS, Series B.

Ruiz, R. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics 63*, 289–306.

Sadorsky, P. (2004). A comparison of some alternative volatility forecasting models for risk management. Proceedings of the second IASTED international conference: Financial Engineering and Applications, November.

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

Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Wilks (2003). *WinBUGS User Manual (Version 1.4)*. Cambridge, UK: MRC Biostatistics Unit.

Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices 1961-1979. *Time Series Analysis: Theory and Practice 1*, 203–226.

Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of American Statistical Association 81*, 82–86.

Yu, J. (2002). Forecasting volatility in the New Zealand stock market. *Applied Financial Economics 12*, 193–202.

Yu, J. (2005). On leverage in a stochastic volatility models. *Journal of Econometrics 127*, 165–178.

Yu, J. and R. Mayer (2006). Multivariate stochastic volatility models: Bayesian estimation and models comparison. *Econometric Reviews 25*, 361–384.