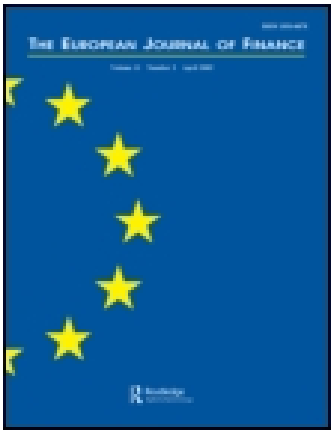

The European Journal of Finance
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rejf20

# Estimating stochastic volatility models using integrated nested Laplace approximations

Sara Martino [a], Kjersti Aas [b], Ola Lindqvist [b], Linda R. Neef [b] & Håvard Rue [a]

[a] Department of Mathematical Science, Norwegian University of Science and Technology, Alfred Getz' vei 1, 7034, Trondheim, Norway

[b] Norwegian Computing Center, Gaustadalléen 23, PO Box 114 Blindern, NO-0314, Oslo, Norway

Published online: 12 Aug 2010.

To cite this article: Sara Martino, Kjersti Aas, Ola Lindqvist, Linda R. Neef & Håvard Rue (2011) Estimating stochastic volatility models using integrated nested Laplace approximations, The European Journal of Finance, 17:7, 487-503, DOI: 10.1080/1351847X.2010.495475

To link to this article: http://dx.doi.org/10.1080/1351847X.2010.495475

*The European Journal of Finance*
Vol. 17, No. 7, August 2011, 487–503



# Estimating stochastic volatility models using integrated nested Laplace approximations

Sara Martino[a], Kjersti Aas[b]*, Ola Lindqvist[b], Linda R. Neef[b] and Håvard Rue[a]

[a]*Department of Mathematical Science, Norwegian University of Science and Technology, Alfred Getz' vei 1, 7034 Trondheim, Norway;* [b]*Norwegian Computing Center, Gaustadalléen 23, PO Box 114 Blindern, NO-0314 Oslo, Norway*



Volatility in financial time series is mainly analysed through two classes of models; the generalized autoregressive conditional heteroscedasticity (GARCH) models and the stochastic volatility (SV) ones. GARCH models are straightforward to estimate using maximum-likelihood techniques, while SV models require more complex inferential and computational tools, such as Markov Chain 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 (INLAs). INLA 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 in financial time series is mainly analysed through two classes of models; the generalized autoregressive conditional heteroscedastic (GARCH) models (Bollerslev 1986) and the stochastic volatility (SV) ones (Taylor 1982). Basically, the SV model assumes two error processes, while the GARCH model only allows for a single error term. Hence, the SV model is more flexible than the GARCH model. Carnero, Pena, and Ruiz (2004) find that in the GARCH model, the parameters explaining persistence and kurtosis are closely linked, whereas these features can be modelled independently in SV models. Hence, the latter might better represent the empirical regularities often observed in financial time series, see also Danielsson (1994) and Kim, Shepard, and Chib (1998).

However, while GARCH models, even in their most complex form, are relatively straightforward to estimate using maximum-likelihood techniques, the likelihood function in SV models does not have a closed form. Hence, these models require much more complex inferential and computational tools. Several estimation methods have been proposed in the literature, ranging from the less efficient generalized methods of moments (Andersen and Sorensen 1996), and quasi-likelihood methods (Harvey, Ruiz, and Shephard 1994) to more efficient methods such as simulated maximum likelihood (Danielsson 1994) and Markov Chain Monte Carlo (MCMC)

*Corresponding author. Email: kjersti.aas@nr.no

ISSN 1351-847X print/ISSN 1466-4364 online

DOI: 10.1080/1351847X.2010.495475
http://www.informaworld.com

(Andersen, Chung, and Sorensen 1999). For a complete review over estimation methods for SV models proposed until now, see Broto and Puiz (2004).

Since MCMC is considered one of the most efficient estimation tools (Andersen, Chung, and Sorensen 1999), much attention has been devoted to the development of efficient MCMC algorithms for SV models, e.g. Chib, Nardari, and Shephard (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, e.g. Rue and Held (2005)). Hence, due to the computational complexity of the SV models, GARCH models are still by the far the most popular in terms of real-life applications.

The purpose of this paper is to introduce a tool that may help SV models to exit the academic world and reach the financial industry. This tool, named integrated nested Laplace approximations (INLAs), was recently introduced by Rue, Martino, and Chopin (2009) as an approach for approximating posterior marginals in latent Gaussian models. 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. Moreover, INLA is easily accessible by the final user. In fact, all computations required by the INLA methodology are efficiently performed by the `inla` program, which is user friendly thanks to its `R` interface called `INLA`.

SV models belong to the class of latent Gaussian models, and in this paper we show how to adapt the theory of Rue, Martino, and Chopin (2009) to this special case. We do this not only for the basic SV model of Taylor (1982), but also for SV models with Student's $t$- and normal inverse Gaussian (NIG)-distributed noise.

The rest of the paper is organized as follows. In Section 2, we review the basic SV model and the extensions that are examined in this paper. Section 3 describes INLA and how it can be applied to the basic SV model. In Section 4, INLA is applied to fit the basic SV model with different error distributions to two different data sets. Finally, Section 5 contains a discussion and suggestions for further research.

## 2. SV models

The basic SV model introduced by Taylor (1982) is defined as

$$r_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t, \tag{1}$$

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

where $\varepsilon_t$ and $\eta_t$ are i.i.d. $\mathcal{N}(0,1)$, $r_t$ is the return and $h_t$ is the logarithm of the variance on day $t$. Hence, it 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 $\nu$, a persistence parameter $\phi \in (0,1)$ and a Gaussian noise $\eta_t$. The conditional distribution of the returns $r_t$, given $h_t$, is then Gaussian with time-varying variance.[1]

In many real data applications, the basic SV model seems to be too restrictive. Hence, several extensions can be found in the literature. For instance, heavier tails for the $\varepsilon_t$ process have been allowed by using Student's $t$-distribution (Ruiz 1994), (Harvey, Ruiz, and Shephard 1994; Chib, Nardari, and Shephard 2002) or the Generalized Error Distribution (Liesenfeld and Jung 2000)

instead of the Gaussian one. Eraker, Johannes, and Polson (2003) introduce jumps in both the return and the log-variance process. Further, Harvey, Ruiz, and Shephard (1994) and Andersson (2001) have used non-stationary SV models as alternatives to the integrated GARCH (or IGARCH) model, and Harvey, Ruiz, and Shephard (1994) also model leverage effects by letting the two noise processes, $\varepsilon_t$ and $\eta_t$, be correlated (corresponding to an asymmetric SV model). Financial assets are clearly correlated, so multivariate extensions of the SV models have lately been given much attention (see, e.g. Vol. 25 of Econometric Review).

In this paper, we concentrate on the basic SV model, but with three different distributions for $\varepsilon_t$: the Gaussian, Student's-$t$ and the NIG-distribution. To our knowledge, the basic SV model with NIG distribution for $\varepsilon_t$ has never been used before. This model must not be confused with the normal inverse Gaussian stochastic volatility (NIG-SV) model (Barndorff-Nielsen 1997). In the latter model, the conditional variance is modelled by an inverse Gaussian (IG) distribution, and the distribution of the data conditional on the volatilities by the Gaussian distribution, delivering an NIG distribution for the data. Hence, it is different from the model proposed here, which has a Gaussian distribution for $\eta_t$ and an NIG distribution for $\varepsilon_t$. Our reason for suggesting the NIG-distribution as an error distribution for the SV model is that it has previously been repeatedly used for financial applications, both as the conditional distribution of a GARCH model (Andersson 2001; Jensen and Lunde 2001; Forsberg and Bollerslev 2002; Venter and de Jongh 2002, 2004) and as the unconditional return distribution (Bølviken and Benth 2000; Eberlein and Keller 1995; Lillestøl 2000; Prause 1997; Rydberg 1997). In a study performed by Venter and de Jongh (2004), it outperformed competing skewed and heavy-tailed distributions as the conditional distribution of a one-dimensional GARCH process. For more details about the NIG distribution, see Appendix 1.



## 3. INLA for basic SV models

If we assign the common mean parameter $\nu$ in Equation (2) a Gaussian prior with mean 0 and large known variance, the basic SV model in Equations (1) and (2) may be seen as a latent Gaussian model with latent field $\boldsymbol{x}$:

$$\boldsymbol{x} = \{h_1, h_2, \ldots, \nu\} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_1)), \tag{3}$$

where $\boldsymbol{\theta}_1 = \{\sigma^2, \phi\}$. The $|\boldsymbol{x}| = n$-dimensional latent Gaussian field is partially observed through the $n_d$ conditionally independent data $\boldsymbol{r} = \{r_1, \ldots, r_{n_d}\}$ with likelihood

$$\pi(\boldsymbol{r}|\boldsymbol{x}, \boldsymbol{\theta}_2) = \prod_{t=1}^{n_d} \pi(r_t|\boldsymbol{h}, \boldsymbol{\theta}_2),$$

where $\boldsymbol{\theta}_2$ are potential parameters in the distribution for $\varepsilon_t$.

Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$. In general, the goal of inference of SV models are the marginal distributions from:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{r}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_{t=1}^{n_d} \pi(r_t|h_t, \boldsymbol{\theta}). \tag{4}$$

The main problem with inference about the posterior in Equation (4) is that it is not available in the 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 $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{r})$, they are often very slow. This is due to the generally strong correlation within the latent field $\boldsymbol{x}$ and between the latent field $\boldsymbol{x}$ and the parameters $\boldsymbol{\theta}$.

In this section, we describe a new tool for inference about $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{r})$, named INLA, which provides very accurate approximations to the marginals $\pi(h_t|\boldsymbol{r})$, $\pi(\nu|\boldsymbol{r})$ and $\pi(\theta_j|\boldsymbol{r})$ in only a fraction of the time used by clever MCMC algorithms.

### 3.1 *Fundamentals of INLA*

#### *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 $\boldsymbol{Q}$ given as the inverse of the covariance matrix, i.e. $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. In particular, it is easy to show that, for the basic SV models, $\boldsymbol{Q}$ 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). The computational efficiency of INLA 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 $\boldsymbol{x}$ is given by Equation (3), and $g_t(h_t) = \log \pi(r_t|h_t, \boldsymbol{\theta})$. The Gaussian approximation, $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{r}, \boldsymbol{\theta})$, 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

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

where $K_1$ is a normalizing 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, $\boldsymbol{Q} + \mathrm{diag}(\boldsymbol{c})$, 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 $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$ to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$.
(2) Build an approximation $\tilde{\pi}(x_t|\boldsymbol{\theta}, \boldsymbol{r})$ to $\pi(x_t|\boldsymbol{\theta}, \boldsymbol{r})$.
(3) Build an approximation $\tilde{\pi}(x_t|\boldsymbol{r})$ to $\pi(x_t|\boldsymbol{r})$ using the results from (1) and (2).

In Sections 3.2.1–3.2.3, each of the three steps are described in more detail.

#### *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 fraction with $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{r},\boldsymbol{\theta})$. Hence, the approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{r})$ is

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r}) = \left.\frac{1}{K_2}\frac{\pi(\boldsymbol{x},\boldsymbol{r},\boldsymbol{\theta})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{8}$$

where $K_2$ is a normalizing constant and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{r})$. Equation (8) is equivalent to Tierney and Kadane's (1986) Laplace approximation of a marginal posterior distribution. The approximation error is relative, and of the order $\mathcal{O}(n_d^{-3/2})$ under standard assumptions. The relative error is an important feature that allows us to accurately estimate even small quantities. In comparison, small quantities are difficult to estimate using any Monte Carlo approach, due to its additive $\mathcal{O}(N^{-1/2})$ error (here $N$ is the number of simulations). The error structure of INLA is discussed in detail in Rue, Martino, and Chopin (2009).

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

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

where $\pi_{GG}()$ is a Gaussian approximation of $\pi(\boldsymbol{x}_{-t}|x_t,\boldsymbol{\theta},\boldsymbol{r})$ and $\boldsymbol{x}^*_{-t}(x_t,\boldsymbol{\theta})$ is the modal configuration of $\pi(\boldsymbol{x}_{-t}|x_t,\boldsymbol{\theta},\boldsymbol{r})$. This approximation can be quite heavy from a computational point of view, and is therefore not convenient in practice. The solution in Rue, Martino, and Chopin (2009) is to compute a simplified version $\tilde{\pi}_{\mathrm{S}}(x_t|\boldsymbol{\theta},\boldsymbol{r})$ of $\tilde{\pi}(x_t|\boldsymbol{\theta},\boldsymbol{r})$, using a series expansion of the logarithm of $\tilde{\pi}(x_t|\boldsymbol{\theta},\boldsymbol{r})$:

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

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, Martino, and Chopin (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:

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

Here, the sum is over values of $\boldsymbol{\theta}$ with area-weights $\Delta_k$. An easy way of selecting the points $\boldsymbol{\theta}^k$ is to create a grid of points covering the area of high density for $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$. In order to correctly locate the interesting area, we compute the mode $\boldsymbol{\theta}^*$ of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$, 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, Martino, and Chopin (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 $\tilde{\pi}(x_t|\boldsymbol{r}, \boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ is the mode of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{r})$. 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*

In financial applications, one is often 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 $\pi(h_{n_d+i}|\boldsymbol{r})$, the predictive densities for the returns can be found as

$$\pi(r_{n_d+i}|\boldsymbol{r}) = \int \pi(r_{n_d+i}|h_{n_d+i})\pi(h_{n_d+i}|\boldsymbol{r})\,\mathrm{d}h_{n_d+i}. \tag{13}$$

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 the following numerical approximation of the integral in Equation (13):

$$\tilde{\pi}(r_{n_d+i}|\boldsymbol{r}) \approx \sum_j \pi(r_{n_d+i}|h^j_{n_d+i})\tilde{\pi}(h^j_{n_d+i}|\boldsymbol{r})\Delta_j. \tag{14}$$

For likelihood models which depend on additional parameters, e.g. if the error distribution is Student's-*t* or NIG, we have chosen to fix the parameter values at the posterior mode computed during the INLA process.

### 3.5 *Extensions of the basic SV model*

The INLA approximation may, without much additional effort, deal with some of the extensions of the basic SV model given by Equations (1) and (2). Estimating non-stationary SV models using INLA is straightforward. The only change from the basic stationary model is the zero structure of the prior precision matrix $\boldsymbol{Q}$ for the latent vector $\boldsymbol{x}$, which for the non-stationary model includes only the log-variances $\{h_1, h_2, \ldots\}$. All computations remain exactly the same. Non-stationary SV models can be implemented using the current version of the `R` package `INLA`.

Algorithms for SV models with leverage effects have still not been implemented in the `R` package `INLA`. However, this is work in progress, and in Appendix 2, we indicate how approximate inference using INLA can be done for SV models with correlated errors.

Multivariate extensions of the SV models have lately been given much attention (see, e.g. Vol. 25 of Econometric Review). The INLA approach has previously been applied with success

in the bivariate case, see Martino (2007) for a comparison between the performance of INLA and a Gibbs sampler algorithm applied to some bivariate SV models. The main difficulty with performing a full Bayesian analysis on multivariate SV models using INLA (but also with respect to most other estimation procedures) is the increasing number of parameters $\boldsymbol{\theta}$, which have to be numerically integrated out in order to compute an approximation to $\pi(\boldsymbol{x}_t|\boldsymbol{r})$. This is an operation which heavily suffers from the curse of dimensionality. However, in these cases, 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.

### 3.6 *Model comparison*

While the main focus of INLA 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 *a priori* equiprobable models $\mathcal{M}_1$ and $\mathcal{M}_2$ may be compared using the Bayes factor, $\mathcal{B}(1, 2)$, 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 by

$$\tilde{\pi}(\boldsymbol{r}) = \int \left. \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{r})}{\tilde{\pi}_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{r})} \right|_{\boldsymbol{x}=\boldsymbol{x}^\star(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta}. \tag{15}$$

We propose two approximations for the marginal likelihood $\pi(\boldsymbol{r})$. In the first, $\tilde{\pi}_1(\boldsymbol{r})$, the integral in Equation (15) is numerically solved using selected values $\boldsymbol{\theta}_k$ of $\boldsymbol{\theta}$. This approximation is quite precise, but expensive to compute. An alternative, cruder estimate, $\tilde{\pi}_2(\boldsymbol{r})$, of the marginal likelihood, is obtained by assuming that $\boldsymbol{\theta}|\boldsymbol{r}$ is Gaussian. Then, Equation (15) may be approximated by some known constant times $|\boldsymbol{H}|^{-1/2}$, where $\boldsymbol{H}$ is the Hessian matrix from Section 3.2.3.

Note that when computing approximations to the marginal likelihood $\pi(\boldsymbol{r})$, aiming at model comparison, it is important to include all normalizing 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, $\pi(\boldsymbol{r}|\boldsymbol{x}, \boldsymbol{\theta})$.

## 4. Examples

In this section, we apply INLA to fit the basic SV model with three different error distributions 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 INLA are efficiently implemented in the `inla` program built upon the `GMRFLib`-library (Rue and Held 2005). Both the `inla` program and the `GMRFLib`-library use the OpenMP (see `www.openmp.org`) to speed up computations for shared memory machines, i.e. multicore processors, which today are standard for new computers. A user friendly

interface to the `inla` program, the `R` package `INLA`, is available from `www.r-inla.org`. All examples in this paper have been implemented using the `INLA` package and the code can be downloaded from the above-mentioned website.

We want to compare, in terms of accuracy and the speed, the results obtained by INLA (as implemented in the `INLA` library) with those obtained by an alternative MCMC procedure. For this purpose, we have chosen to use a Gibbs sampler, as implemented in OpenBUGS (Spiegelhalter et al. 2003). OpenBUGS has been previously proposed as a tool to implement MCMC algorithms for SV models (Mayer and Yu 2000). Moreover, the programming effort needed for OpenBUGS is similar to that for the `R-INLA` package.

It would certainly be possible to obtain more efficient computing times by using different MCMC schemes (as, for instance that in Chib, Nardari, and Shephard (2002)) and programming them in C. However, a direct comparison of computing times for an MCMC algorithm and a deterministic procedure is still not possible, since there is no rule for stopping an MCMC algorithm that, in principle, could run forever. Hence, instead, we compare the order of magnitude of such computing times and the accuracy of the estimates, claiming that in the few seconds needed by INLA to obtain results, even an efficient MCMC scheme could not reach the same accuracy.

### 4.1 *SP500 index data*

Our first data set consists of 1217 daily closing quotations for the SP500 index from 2 January 2003 to 31 October 2006. Figure 1 shows the logarithmic returns for the SP500 index data.

We fit the basic SV model given by Equations (1) and (2) to the data set using the `INLA` library. We consider three different distributions for the error $\epsilon_t$; Gaussian, Student's $t$ and NIG. All error distributions are standardized to have zero mean and variance 1. The model specification is completed with the definition of prior distributions for the parameters $\boldsymbol{\theta}$. We assume the log-precision $\log 1(/\sigma^2)$ to be a vague log-Gamma, and assign

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

a Gaussian distribution, whose precision parameter is chosen such that the corresponding prior for $\phi$ 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 assign vague priors to the likelihood parameters $\boldsymbol{\theta}_2$. The Gaussian model has no parameters (i.e. $\boldsymbol{\theta}_2 = \emptyset$), the Student-$t$ model has one

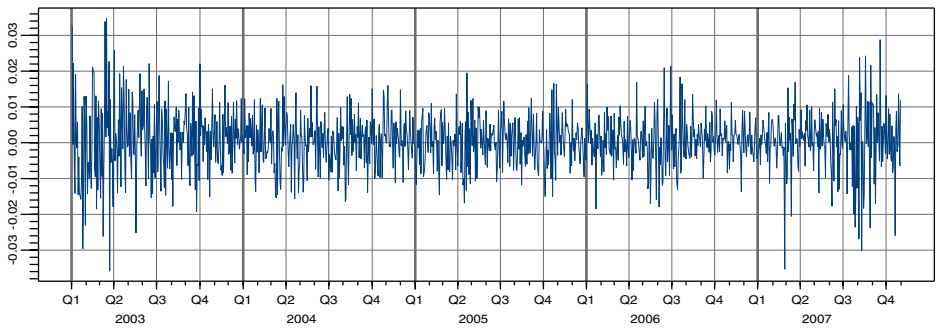


Figure 1. SP500 index: logarithmic returns from 2 January 2003 to 31 October 2006.

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

| Model | $\nu$ | $\phi$ | $\tau = 1/\sigma^2$ | d.o.f. | $\beta$ | $\psi$ |
|---|---|---|---|---|---|---|
| Gaussian | −0.52 (0.23) | 0.98 (0.006) | 100.20 (38.86) | – | – | – |
| Student's $t$ | −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 OpenBUGS.

| Model | $\nu$ | $\phi$ | $\tau = 1/\sigma^2$ | d.o.f. |
|---|---|---|---|---|
| Gaussian | −0.51 (0.27) | 0.98 (0.007) | 100.2 (37.09) | – |
| Student's $t$ | −0.48 (0.29) | 0.98 (0.005) | 142.0 (57.6) | 19.6 (15.17) |



parameter (i.e. $\boldsymbol{\theta}_2 = \xi$), and the NIG model has two parameters (i.e. $\boldsymbol{\theta}_2 = (\beta, \psi)$). Hence, the length of $\boldsymbol{\theta}$ varies from a minimum of 2 (Gaussian model for $\epsilon_t$) to a maximum of 4 (NIG model for $\epsilon_t$).

The computing time goes from 6 s for the Gaussian model to 17 s for the NIG one on a dual core 2.5 GHz laptop. The larger computing time for the NIG model is due to the larger dimension of the parameter vector $\boldsymbol{\theta}$.

Table 1 shows the estimated posterior means and standard deviations for the parameters $\boldsymbol{\theta}$ and the common mean $\nu$, as computed using `INLA`. In Table 2, the same parameters for the Gaussian and Student-$t$ models are estimated based on 50, 000 samples after convergence using OpenBUGS (the NIG model is not implemented in OpenBUGS). The estimates seem to agree. They are however not identical. We claim that such differences are due mostly to the MCMC error which, despite the large number of samples, is still present in the result of the Gibbs algorithm. This can be verified by a closer look at the Gaussian model. Figure 2 shows a comparison between the estimated values for the log variance $h_t$ at day $t$, computed using `INLA` and OpenBUGS (using 50,000 samples), respectively. Although there is a general agreement of the two estimates, differences can be noted, both in the posterior mean and the posterior 2.5% and 97.5% quantiles. We have therefore considered one specific point in time for which the two estimates seem to disagree the most ($t = 618$, indicated in Figure 2 using a vertical line). For this time point, we let

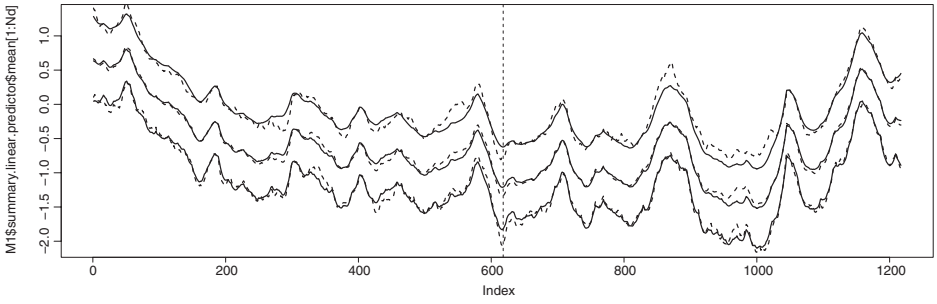


Figure 2. SP500 index: posterior mean and the 2.5% and 97.5% quantiles for the estimated log variance $h_t$ in the Gaussian model as computed by INLA (solid line) and by MCMC (broken line).

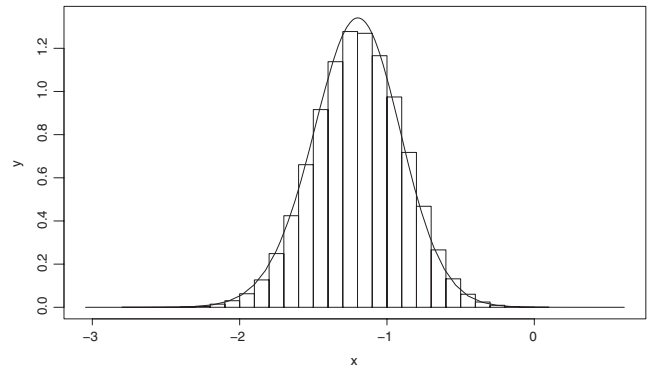


Figure 3. SP500 index: posterior marginal for $h_{618}$ in the Gaussian model computed by INLA (solid line) superimposed onto the histogram resulting from $10^6$ samples of $h_{618}$ generated using OpenBUGS.

Table 3. SP500: approximated values for the log marginal likelihood obtained using INLA.

| Model | Marginal likelihood $\log \tilde{\pi}_1(\boldsymbol{r})$ | Marginal likelihood $\log \tilde{\pi}_2(\boldsymbol{r})$ |
|---|---|---|
| Gaussian | −1391.49(*) | −1391.47(*) |
| Student's $t$ | −1393.34 | −1393.28 |
| NIG | −1396.87 | −1396.87 |

Note: The preferred model is indicated by an (*).

the Gibbs sampler produce as many as $10^6$ samples. In Figure 3, the posterior marginal $\tilde{\pi}(h_{618}|\boldsymbol{r})$ obtained by INLA is displayed together with a histogram based on the $10^6$ samples obtained using OpenBUGS. The two estimates agree almost perfectly, indicating that the difference seen in Figure 2 is due to the MCMC error. Similar results can be obtained for Student's $t$-distribution.

We have compared the fit of the three different error distributions using the two approximations of the marginal likelihoods described in Section 3.6. The results, shown in Table 3, indicate that the Gaussian distribution is preferable, but the differences between the three distributions are quite small. This is confirmed by the estimated values of the parameters in Table 1. The degrees of freedom parameter in Student's $t$-distribution is quite large, and the parameters of the NIG distribution correspond to a density which is quite similar to the Gaussian.

Figure 4(a) shows the posterior mean, and Figure 4(b) the 1% and 99% quantiles, for the estimated latent historical volatility in percent, as well as for the predicted volatility during the 60 day-period starting on 1 November 2006 for the Gaussian and the NIG model. The Gaussian, Student-$t$ and NIG densities for the predicted return (log-scale in percent) on 1 November 2006 is displayed in Figure 5, and Table 4 shows the 95% and 99% VaRs corresponding to these densities. Both the figure and the values in the table confirm that the difference between the different error distributions is small.

### 4.2 *Microsoft data*

Our second example consists of 1292 daily closing prices for the Microsoft stock from 3 January 2003 to 21 February 2008. The log-returns are shown in Figure 6. As can be seen from the figure,

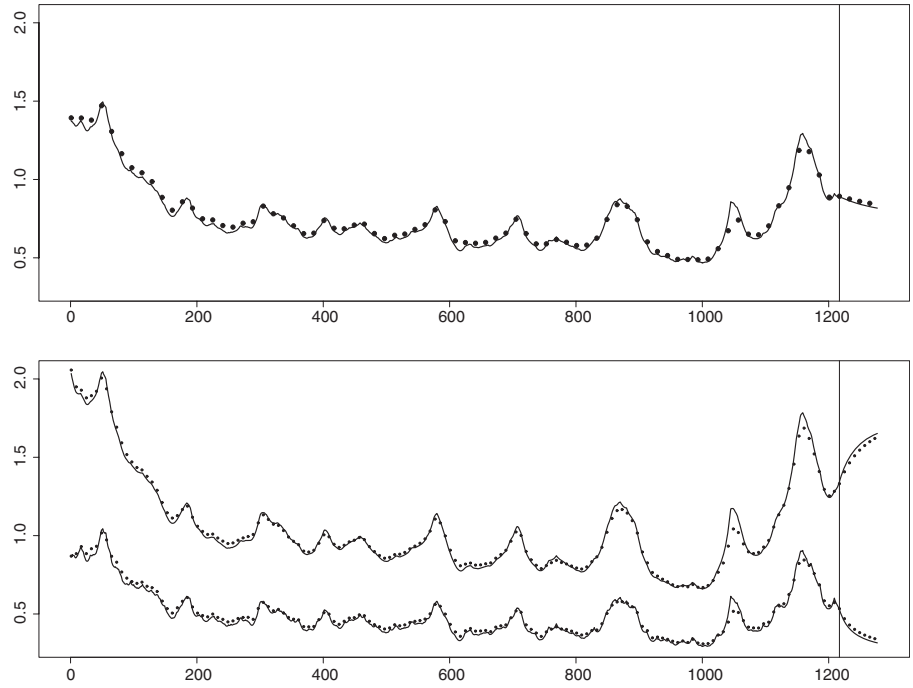


Figure 4. SP500 index: the upper panel shows posterior mean for the estimated latent historical volatility, $\exp(h_t/2)$, in percent, as well as for the predicted volatility during the 60 day-period starting on 22 February 2008 for two different models. The solid line corresponds to the Gaussian model and the dotted line to the NIG model. The vertical line corresponds to the end of the estimation period. The lower panel shows the 1% and 99% quantiles for the same quantities.

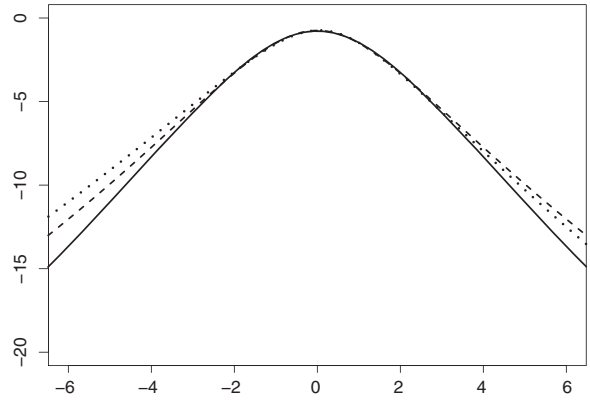


Figure 5. SP500 index: the Gaussian, Student-$t$ and NIG densities (log-scale) for the predicted return (in percent) on 1 November 2006. The solid line corresponds to the Gaussian model, the broken line to Student's $t$-model and the dotted line to the NIG model.

the volatility is larger than for the SP500 example. Moreover, the data set contains some extreme values. We fit the three SV models using the same prior specification for $\boldsymbol{\theta}$ as in the SP500 example.

Table 4. SP500: VaR on 1 November 2006 for the return in percent assuming zero mean.

| Model | $\text{VaR}_{0.95}$ | $\text{VaR}_{0.99}$ |
|---|---|---|
| Gaussian | 1.405 | 2.113 |
| Student's $t$ | 1.436 | 2.216 |
| NIG | 1.407 | 2.143 |

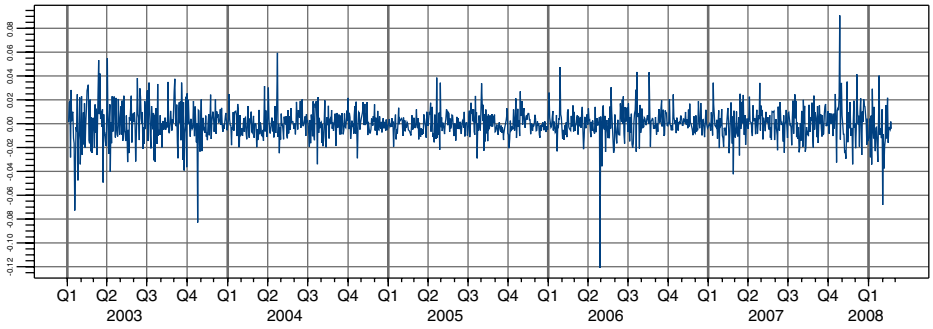


Figure 6. Microsoft stock: logarithmic returns from 3 January 2003 to 21 February 2008.

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

| Model | $\nu$ | $\phi$ | $\tau = 1/\sigma^2$ | d.o.f. | $\beta$ | $\psi$ |
|---|---|---|---|---|---|---|
| Gaussian | 0.21 (0.14) | 0.92 (0.02) | 8.98 (2.44) | – | – | – |
| Student's $t$ | 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 OpenBUGS.

| Model | $\nu$ | $\phi$ | $\tau = 1/\sigma^2$ | d.o.f. |
|---|---|---|---|---|
| AR-1 Gaussian | 0.20 (0.14) | 0.92 (0.02) | 9.1 (3.1) | – |
| AR-1 Student's $t$ | 0.45 (0.28) | 0.98 (0.01) | 67.4 (40.4) | 5.4 (0.9) |

Table 5 shows the estimated posterior mean and standard deviation for $\nu$ and $\boldsymbol{\theta}$ estimated using INLA, while Table 6 gives the corresponding OpenBUGS estimates for the Gaussian and Student's $t$-models. It can be noticed that both INLA and OpenBUGS estimate the precision for the latent vector to be considerably lower for the Gaussian model than for the two other models.

As for the SP500 example, we use the estimated marginal likelihoods to compare the fit of the models. The results are reported in Table 7. The preferred model is the one with Student's $t$-distribution for $\epsilon_t$. 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

Table 7. Microsoft: approximated values for the log marginal likelihood obtained using INLA.

| Model | Marginal likelihood $\log \tilde{\pi}_1(\boldsymbol{r})$ | Marginal likelihood $\log \tilde{\pi}_2(\boldsymbol{r})$ |
|---|---|---|
| Gaussian | −2083.39 | −2083.33 |
| Student's $t$ | −2065.87(*) | −2065.74(*) |
| NIG | −2074.06 | −2073.86 |

Note: The preferred model is indicated by an (*).

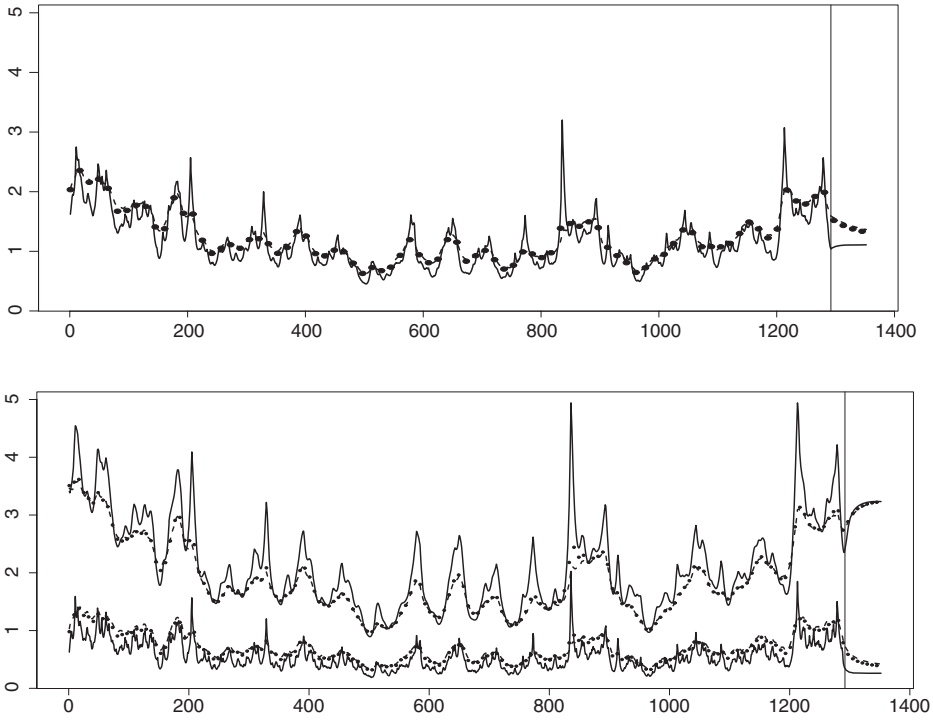


Figure 7. Microsoft stock: the upper panel shows posterior mean for the estimated latent historical volatility, $\exp(h_t/2)$, in percent, as well as for the predicted volatility during the 60 day-period starting on 22 February 2008 for two different models. The solid line corresponds to the Gaussian model and the dotted line to the NIG model. The vertical line corresponds to the end of the estimation period. The lower panel shows the 1% and 99% quantiles for the same quantities.



similar to those obtained under Student's $t$-model. Hence, it is reasonable that the model choice criteria prefers Student's $t$-model, which has one less parameter than the NIG model.

Figure 7(a) displays the posterior mean, and Figure 7(b) the 1% and 99% quantiles, for the estimated latent historical volatility in percent, as well as for the predicted volatility during the 60 day-period starting on 22 February 2008. Unlike for the SP500 example, for which all error distributions gave similar results, the Gaussian model now differs substantially from the other two, especially when dealing with the volatility prediction. The posterior mean estimated by the Gaussian model appears to be shifted, and the inter-quantile range wider, compared with the two

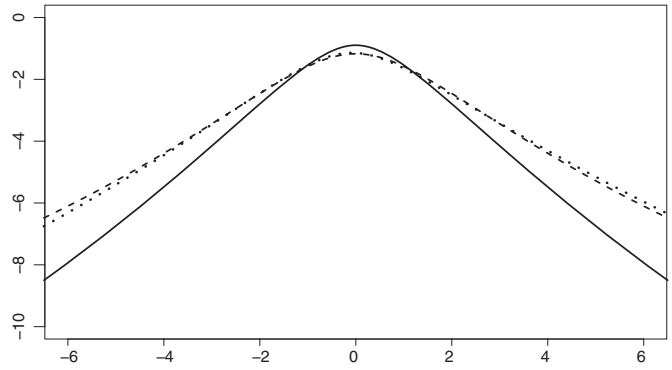


Figure 8. Microsoft stock: the Gaussian, Student-$t$ and NIG densities (log-scale) for the predicted return (in percent) on 22 February 2008. The solid line corresponds to the Gaussian model, the broken line to Student's $t$-model and the dotted line to the NIG model.

Table 8. Microsoft: VaR on 22 February 2008 for the return in percent assuming zero mean.

| Model | $\text{VaR}_{0.95}$ | $\text{VaR}_{0.99}$ |
|---|---|---|
| Gaussian | 1.731 | 2.735 |
| Student's $t$ | 2.522 | 4.161 |
| NIG | 2.588 | 4.308 |

other models. The difference is also apparent looking at the characteristics of the distributions for the predicted return (in percent), shown in Figure 8 and Table 8.

## 5. Discussion

In this paper, we have presented a new approach for performing inference on SV models named INLA. INLA computes very 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 as well as 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 the 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, $(\text{sfi})^2$. The authors thank the referee and Chris J. Adcock, the editor, for comments and suggestions that helped to improve the paper.

## Note

1. There is an alternative specification of the basic SV model in Equations (1) and (2) which reads:

$$r_t = \sigma_r \exp\left(\frac{h_t}{2}\right)\varepsilon_t, \tag{16}$$

$$h_t = \phi h_{t-1} + \sigma\eta_t. \tag{17}$$

Using this specification, the scale parameter $\sigma_r$ in Equation (16) substitutes the constant $\nu$ in Equation (2). Although it is still possible to use INLA with the parametrization given by Equations (16) and (17), its implementation is more efficient for the parametrization given by Equations (1) and (2).



## References


Andersen, T., and B. Sorensen. 1996. GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 13, no. 3: 329–52.

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, no. 1: 61–87.

Andersson, J. 2001. On the normal inverse Gaussian stochastic volatility model. *Journal of Business and Economic Statistics* 19, no. 1: 44–54.

Azzalini, A., and A. Capitanio. 1999. Statistical applications of the multivariate skew-normal distribution. *Journal of Royal Statistical Society, Series B* 61, no. 3: 579–602.

Barndorff-Nielsen, O.E. 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24, no. 1: 1–13.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, no. 3: 307–27.

Bølviken, E., and F.E. Benth. 2000. Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution. In *Proceedings of the AFIR 2000 colloquium*, eds. R. Nordberg et al., Tromsø, Norway, 87–98.

Broto, C., and E. Puiz. 2004. Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys* 18, no. 5: 613–49.

Carnero, M.A., D. Pena, and E. Ruiz. 2004. Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics* 2, no. 2: 319–42.

Chib, S., F. Nardari, and N. Shephard. 2002. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, no. 2: 281–316.

Danielsson, J. 1994. Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics* 64, no. 1: 375–400.

Eberlein, E., and U. Keller. 1995. Hyperbolic distributions in finance. *Bernoulli*, 1, no. 3: 281–99.

Eraker, B., M. Johannes, and N. Polson. 2003. The impact of jumps in volatility and returns. *The Journal of Finance* 58, no. 3: 1269–300.

Forsberg, L., and T. Bollerslev. 2002. Bridging the gap between the distribution of realized (ecu) volatility and ARCH modeling (of the euro): The GARCH-NIG model. *Journal of Applied Econometrics* 17, no. 5: 535–48.

Harvey, A.C., and N. Shephard. 1996. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, no. 3, 429–34.

Harvey, A., E. Ruiz, and N. Shephard. 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61, no. 2: 247–64.

Jeffreys, H. 1961. *The theory of probability*. New York: Oxford University Press.

Jensen, M.B., and A. Lunde. 2001. The NIG-S & ARCH model: A fat-tailed stochastic, and autoregressive conditional heteroscedastic volatility model. *Econometrics Journal* 4, no. 2, 319–42.

Kim, S., N. Shepard, and S. Chib. 1998. Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies* 65, no. 3: 361–93.

Liesenfeld, R., and R. Jung. 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15, no. 2: 137–60.

Lillestøl, J. 2000. Risk analysis and the NIG distribution. *Journal of Risk* 2, no. 4: 41–56.

Martino, S. 2007. Approximate Bayesian inference for latent Gaussian models. PhD thesis, Department 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, no. 2: 198–215.

Prause, K. 1997. Modelling financial data using generalized hyperbolic distributions. FDM preprint 48, University of Freiburg.

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

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B* 71, no. 2: 319–92.

Ruiz, R. 1994. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, no. 1: 289–306.

Rydberg, T.H. 1997. The normal inverse Gaussian Levy process: Simulation and approximation. *Communications in Statistics-Stochastic Models* 13, no. 4: 887–910.

Shephard, N., and M. Pitt. 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, no. 3: 653–67.

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. In *Time series analysis: Theory and practice 1*, ed. O.D. Anderson. Amsterdam: North Holland, 203–26.

Tierney, L., and J.B. Kadane. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of American Statistical Association* 81, no. 393: 82–86.

Venter, J.H., and P.J. de Jongh. 2002. Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk* 4, no. 2: 1–24.

Venter, J.H., and P.J. de Jongh. 2004. Selecting an innovation distribution for GARCH models to improve efficiency of risk and volatility estimation. *Journal of Risk* 6, no. 3: 27–53.

Yu, J. 2005. On leverage in a stochastic volatility models. *Journal of Econometrics* 127, no. 2: 165–78.



## Appendix 1. Using NIG as noise distribution

The NIG distribution (Barndorff-Nielsen 1997) possesses a number of attractive theoretical properties, among others is its analytical tractability. For these reasons, it has been used repeatedly for financial applications. When used as the noise distribution in an SV model, the NIG distribution has to be standardized in order to have a zero mean and a unit variance. The standardized NIG density is defined as

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

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 $\beta$ controls the skewness of the density, while $\psi$ is a shape parameter.

## Appendix 2. Asymmetric SV models

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 an 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 the MCMC procedure 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 is the Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{r})$. In order to 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

$$\pi(r_t|\boldsymbol{x},\boldsymbol{\theta}) = \pi(r_t|h_t, h_{t+1}, \boldsymbol{\theta}) = \mathcal{N}\left\{\frac{\rho}{\sigma}e^{h_t/2}[h_{t+1}-\nu+\phi(h_t-\nu)], e^{h_t}(1-\rho^2)\right\}. \tag{A1}$$

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 $\nu$. The prior distribution for the vector $\boldsymbol{x}$ is unchanged. Hence, the full conditional reads

$$\pi(\boldsymbol{x}|\boldsymbol{r},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \sum_{t=1}^{n_d} f_t(h_t, h_{t+1}, \nu)\right\}, \tag{A2}$$

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

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

where $\boldsymbol{C}_t$ is a $3 \times 3$ 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 $i$th 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

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

Here $\boldsymbol{C}$ is an $n \times n$ 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 $\nu$ 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 INLA also for SV models with correlated errors.