Discussion Papers No. 601, December 2009
Statistics Norway, Research Department

Arvid Raknerud and Øivind Skare

Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes

Abstract:
This paper aims to develop new methods for statistical inference in a class of stochastic volatility models for financial data based on non-Gaussian Ornstein-Uhlenbeck (OU) processes. Our approach uses indirect inference methods: First, a quasi-likelihood for the actual data is estimated. This quasi-likelihood is based on an approximative Gaussian state space representation of the OU-based model. Next, simulations are made from the data generating OU-model for given parameter values. The indirect inference estimator is the parameter value in the OU-model which gives the best "match" between the quasi-likelihood estimator for the actual data and the quasi-likelihood estimator for the simulated data. Our method is applied to Euro/NOK and US Dollar/NOK daily exchange rates for the period 1.7.1989 until 15.12.2008. Accompanying R-package, that interfaces C++ code is documented and can be downloaded.

Keywords: stochastic volatility, financial econometrics, Ornstein-Uhlenbeck processes, indirect inference, state space models, exchange rates

JEL classification: C13, C22, C51,G10

Acknowledgement: We appreciate useful comments from Terje Skjerpen, Anders Rygh Swensen and participants at the 3rd International Conference on Computational and Financial Econometrics (CFE) held on 29-31 October 2009, in Limassol, Cyprus. Accompanying software written in C++ code (with R-interface) can be downloaded from http://folk.uio.no/skare/SV/. Financial support from the Norwegian Research Council ("Finansmarkedsfondet") is gratefully acknowledged.

Address: Arvid Raknerud, Statistics Norway, Research Department. E-mail: rak@ssb.no

Øivind Skare, Norwegian Institute of Public Health and University of Bergen, Department of Public Health and Primary Health Care. E-mail: oivind.skare@medisin.uio.no

Discussion Papers	comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

There has been an enormous research activity in the field of statistical modelling of high-frequency financial data based on non-Gaussian Ornstein-Uhlenbeck (OU) processes during the present decade. Some of the most important contributions from the point of view of this approach are three articles by Barndorff-Nielsen and Shephard (2001, 2002, 2003) (hereafter BS) and Barndorff-Nielsen et al. (2001). Overviews of recent developments in the field of financial econometrics are given in Harvey et al. (2004), Shephard et al. (2005) and Andersen et al. (2009). Traditional likelihood-based methods are generally not applicable to non-Gaussian stochastic volatility models, and we propose a new estimation method based on indirect inference (see Gourieroux et al., 1993, and Gallant and Tauchen, 1996), and apply these methods to daily Euro/NOK and US Dollar/NOK exchange rate data.

While the statistical properties of OU processes and implications for derivative pricing have been examined by BS (2001) and others (e.g. Nicolato and Venardos, 2003), many issues regarding practical implementation and estimation remain unsolved. Neither have non-Gaussian OU processes been much tested in applications. The novelty of our approach consists in a new use of indirect inference methods. In general, indirect inference combines estimation of an approximative model with simulations from an underlying "true" data generating model: First, the auxiliary model is estimated on the actual data. In our case this will be done by maximizing a Gaussian quasi-likelihood function corresponding to a linear state space representation for returns and squared returns. Then simulations are made from the underlying OU-model for given parameter values. For each simulation, the quasi-likelihood function for the simulated data is maximized. The indirect inference estimator of a parameter vector is the value of the vector in the OU-model which gives the best "match" between the quasi-likelihood estimator for the actual data and the quasi-likelihood estimator for the simulated data. Our estimation method should be seen as an alternative to the Bayesian (MCMC) approach proposed by Griffin and Steel (2006) and as complementary to pure quasi-likelihood estimation. The MCMC approach is cumbersome

for large data sets and also relies on prior distributions for all the parameters, which makes it less attractive to non-Bayesians than likelihood based methods. On the other hand, the quasi-likelihood function is constructed by means of the Kalman filter by assuming that the actual volatility process is a Gaussian latent (state) variable. Our Gaussian quasi-likelihood treats the optimal linear predictors of returns and squared returns as if they are conditional expectations, which they are not. We will investigate the consequences of this simplification for statistical inference. We also provide software written as a user friendly R-package that interfaces efficient C++ code.[1]

The applied part of this paper analyzes exchange rate volatility, using daily data from 1.7.1989 until 15.12.2008 for the Euro/NOK and US Dollar/NOK exchange rates. There exists a large literature on exchange rate dynamics, especially regarding the role of purchasing parity and uncovered interest parity. While there is some evidence that economic fundamentals may govern the behavior of exchange rates in the very long run (see MacDonald, 1999), it is now generally accepted that exchange rates at daily (or intra-daily) frequencies cannot be explained by monetary economic theory. In fact, the well-known study by Meese and Rogoff (1983) demonstrates that a wide range of exchange rate models based on economic fundamentals were unable to outperform a simple random walk model. Later work in this area, however, point out that even if a random walk is a good approximation to the conditional mean process, there is strong evidence of heteroscedasticity in the errors, in the sense that large changes tend to be followed by large changes, and small by small, leading to consecutive periods with high volatility followed by periods of relative stability (see e.g. Diebold and Nerlove, 1989). Thus, the error terms may be uncorrelated, but not independent. Generally, the modelling of the volatility of a stochastic process, which is a second order property, is much more difficult than modelling the conditional mean (a first order property). This topic is far from resolved in the econometric literature.

The rest of this paper is organized as follows: Section 2 presents the formal modelling framework, Section 3 describes the estimation method, while Section 4 discusses computational issues and presents the empirical application. Section 5 concludes.

[1] See http://folk.uio.no/skare/SV/ for software and user documentation ("How to get started").

2 Technical aspects of OU processes

Stochastic volatility models based on OU processes: In the classical contributions to modern financial theory, the *log price* or *log exchange rate*, $y^*(t)$, is modelled as a Brownian motion with drift:

$$dy^*(t) = \mu dt + \sigma dw(t),$$

where σ is the volatility parameter, μ is the drift term and $w(t)$ is a standard Brownian motion. Assume that the process is observed at discrete time points $t_n = n\Delta$, for some $\Delta > 0$, and $n = 1, 2, ..., N$. Then, integrated returns

$$y_n \equiv \int_{(n-1)\Delta}^{n\Delta} dy^*(t),\ n = 1, 2, .., N,$$

i.e., the changes in the log price over the intervals $[(n-1)\Delta, n\Delta]$, $n = 1, ..., N$, are *i.i.d.* and distributed as $\mathcal{N}(\mu\Delta, \sigma^2\Delta)$. However, there is overwhelming evidence that this model provides a poor fit to financial returns data over small to medium time intervals (see e.g. Jondeau et al., 2007, for an overview). Real time transaction data exhibit serious departure from normality and homoscedasticity and cannot be considered as independent realizations of a random variable: When Δ is small or moderate (corresponding to minutes, hours or days), the returns y_n are heavily tailed, squared returns, y_n^2, are serially correlated ("volatility clustering"), and the distribution of y_n may be is skewed. On the other hand, as Δ increases, a central limit theorem seems to be at work, so that the Gaussian model provides a better description of long-term returns.

These "stylized facts" have led to numerous attempts to build empirically more satisfactory models. A number of discrete time models (ARCH, GARCH and discrete-time stochastic variance models) have been proposed (see e.g. Engle, 1982; Diebold, 1988, Bollerslev et al., 1994; and Harvey et al., 1994). The main idea behind these models is to assume that σ^2 is a random variable which changes over time, implying that the error term in the equation for y_n is mixed Gaussian.

A starting point for diffusion-based models for stochastic volatility is the following stochastic differential equation:

$$dy^*(t) = \mu dt + \sigma(t) dw(t), \tag{1}$$

where $\sigma^2(t)$ (> 0) is a stochastic process, called *spot volatility*. In this case, $y_n|\sigma_n \sim$

$\mathcal{N}(\mu\Delta, \sigma_n^2)$, where

$$\sigma_n^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma^2(t)dt \tag{2}$$

is called *actual volatility*.

Like BS (2001, 2003), we will consider the case where $\sigma^2(t)$ is modelled as a positive non-Gaussian Ornstein-Uhlenbeck (OU) process:

$$d\sigma^2(t) = -\lambda\sigma^2(t)dt + dz(\lambda t);\ \lambda > 0,\ z(0) = 0, \tag{3}$$

where $z(t)$ is a Levy jump process with stationary, independent and positive increments (such a process is called a *subordinator*). Some important features characterize this process:

First, $\sigma^2(t)$ moves up only by jumps in $z(t)$, and then tails off exponentially at the rate λ. Thus λ determines the memory of the process: a small λ implies a long-memory volatility process, while large λ implies that past jumps are quickly discounted. The parameter λ also determines the rate at which jumps in volatility occurs.

Second, $\sigma^2(t)$ has a stationary distribution which does not depend on λ – the latter result is obtained by the peculiar timing $z(\lambda t)$. If $E(\sigma^2(t)) = \xi$ and $Var(\sigma^2(t)) = \omega^2$, it is shown in BS (2001) that

$$\sigma^2(n\Delta) = e^{-\lambda\Delta}\sigma^2((n-1)\Delta) + \eta_{2n}, \tag{4}$$

where

$$\eta_{2n} \sim i.i.d.\left(\xi(1 - e^{(-\lambda\Delta)}), \omega^2(1 - e^{(-2\lambda\Delta)})\right).$$

Thus (3) can be interpreted as a continuous time autoregressive model, where $exp(-\lambda\Delta)$ is the autoregressive parameter in the corresponding (exact) discrete-time transition equation for $\sigma^2(n\Delta)$.

Finally, many analytical results about the distribution and dependence structure of integrated returns, y_n, and integrated volatility, σ_n^2, are available. For example, as $t_N \to \infty$,

$$t_N^{-1}\sum_{n=1}^{N}\sigma_n^2 \to \xi \text{ and } t_N^{-1/2}(\sum_{n=1}^{N} y_n - \mu t_N) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \xi). \tag{5}$$

Thus, non-normality vanishes under temporal aggregation. This result also corresponds to stylized facts about financial returns data. A similar result holds for the popular class of

ARCH models (see Diebold, 1988). However, an advantage of OU based stochastic volatility models compared to more traditional discrete-time approaches is that they generate many closed form solutions under temporal aggregation. Moreover, because estimation of the model at different time frequencies is just a matter of choosing a different Δ, the parameters of the model are (trivially) invariant under temporal aggregation. These results are important both in order to study volatility, to price derivatives and to estimate empirical models. In contrast, if we formulate a GARCH model for a given time frequency (e.g., daily) and then decide to estimate the model on another frequency (e.g., weekly), the latter model is no longer a GARCH model; GARCH processes are generally not closed under aggregation (see Drost and Nijman, 1993).

Modelling approaches: A Levy process with stationary, independent and positive increments is characterized by the Levy measure, W, on the positive half line (see Jondeau et al., 2007, Ch. 17). Subject to regularity conditions, W determines the cumulant generating function $k(\theta)$ of $z(1)$ through the relation:

$$k(\theta) \equiv \ln\left(E\left[e^{-\theta z(1)}\right]\right) = -\int_{R_+}\left[1-e^{-\theta x}\right]dW(x)$$

Furthermore, the cumulant generating function $k^*(\theta)$ of $\sigma^2(t)$ is determined from $k(\theta)$ through the equation:

$$\begin{aligned} k^*(\theta) &\equiv \ln E(e^{-\theta\sigma^2(t)}) \\ &= \int_0^\infty k(\theta e^{-s})ds. \end{aligned}$$

Hence, the marginal distribution of $\sigma^2(t)$ is fully specified given $W(x)$. The upper tail mass function

$$W^+(x) = \int_x^\infty dW(u)$$

and its inverse

$$W^{-1}(x) = \inf\{y > 0 : W^+(y) \leq x\},$$

play a large role in *simulations* of OU processes through the following relation (see BS, 2001):

$$\int_0^\Delta f(\lambda s)dz(\lambda s) \stackrel{D}{=} \sum_{i=1}^\infty W^{-1}(a_i/\Delta\lambda)f(\lambda\Delta r_i), \tag{6}$$

where $f(\cdot)$ is an arbitrary function and $\{a_i\}$ and $\{r_i\}$ are two mutually independent sequences of random variables: The r_i are independently and uniformly distributed on $[0,1]$ and $a_1 < a_2 < \cdots$ are arrival times of a Poisson process with intensity 1. Equation (6) can be used to simulate realizations of $z(\lambda t)$ and $\sigma^2(t)$ using the recursive relations

$$\begin{bmatrix} z(\lambda n\Delta) \\ \sigma^2(n\Delta) \end{bmatrix} = \begin{bmatrix} z(\lambda(n-1)\Delta) \\ e^{-\lambda\Delta}\sigma^2((n-1)\Delta) \end{bmatrix} + \begin{bmatrix} \eta_{1n} \\ \eta_{2n} \end{bmatrix}, \tag{7}$$

where

$$\begin{bmatrix} \eta_{1n} \\ \eta_{2n} \end{bmatrix} \stackrel{D}{=} \begin{bmatrix} \int_0^\Delta dz(\lambda t) \\ e^{-\lambda\Delta}\int_0^\Delta e^{\lambda t} dz(\lambda t) \end{bmatrix}, \tag{8}$$

i.e., (6) can be used to simulate realizations of the stochastic integrals appearing in (8). Examples of two realizations of actual volatility σ_n^2 with $\lambda = 0.15$ and $\lambda = 0.6$, respectively, and $\Delta = 1$, $\xi = 0.2$ and $\omega^2 = 0.3$ are shown in Figure 1. We see that for small λ, the series jumps infrequently and then tails off very slowly. In the limit when $\lambda = 0$, we obtain the constant volatility model: $\sigma_n^2 = \Delta\xi$. On the other hand, with $\lambda = 0.6$, jumps occur frequently but tail off very quickly, leading to an erratic volatility series.

A natural modelling approach is to start by choosing a parametric family for the (marginal) distribution of $\sigma^2(t)$. Obviously, not all distributions on the positive half line are consistent with the OU assumption. In fact, the family of distributions which is consistent with this assumption is the class of self-decomposable distributions on R_+ (see BS, 2001). In general, a random variable x (not necessarily restricted to R_+) is self-decomposable if, for any $c \in (0,1)$, there exists a random variable x_c, independent of x, such that

$$x \stackrel{\mathcal{D}}{=} cx + x_c.$$

A prime example of a self-decomposable distribution is the stationary AR(1) model: $x_n = \rho x_{n-1} + \varepsilon_n$, with $|\rho| < 1$, $x_0 = \sum_{s=0}^{\infty} \rho^s \varepsilon_{-s}$, with ε_n *i.i.d.* white noise. In this case x_n has the same distribution as x_{n-1}. As pointed out above, an OU process can be seen as a continuous time AR(1) process. Other examples of self-decomposable distributions on R_+ are the lognormal distribution (see Bondesson, 2002, and BS, 2003), which was advocated by Andersen et al. (2001) to model actual volatility in the context for exchange rate data, and the generalized inverse Gaussian distribution, which contains the inverse Gaussian, inverse χ^2 and Gamma distribution as special cases. If $\sigma^2(t)$ has an inverse Gaussian distribution, then y_n is distributed as $\mu\Delta + \sigma\varepsilon$, where σ^2 has an inverse Gaussian

distribution and ε has a standard normal distribution. This is known as the generalized hyperbolic distribution. Special cases include the normal inverse Gaussian and the Student t distribution. The former has been applied to daily Norwegian stock returns data by Bølviken and Benth (2000).

In the empirical part of this paper, we follow Griffin and Steel (2006) by only considering the Gamma marginal distribution for $\sigma^2(t)$: $\sigma^2(t) \sim \text{Gamma}(\nu, \alpha)$, where $\nu > 0$ is the scale parameter and α is the precision parameter. In particular, $E(\sigma^2(t)) \equiv \xi = \nu/\alpha$ and $Var(\sigma^2(t)) \equiv \omega^2 = v/\alpha^2$. For this distribution

$$W^{-1}(a_i/\lambda\Delta) = \max\left\{0, \frac{1}{\alpha}\ln(\frac{\nu\lambda\Delta}{a_i})\right\}, \tag{9}$$

which is zero for $a_i \geq \nu\lambda\Delta$ (see BS, 2001). Hence, simulation of (7) is almost trivial to carry out, since the infinite sum in (6) can be replaced by a finite one. An important feature of the representation (6) is that the simulations of $\{a_i\}$ and $\{r_i\}$ do not depend on unknown parameters. Thus estimation of the model based on simulations can be done by keeping the simulated draws of $\{a_i\}$ and $\{r_i\}$ unchanged, as the parameters are varied during the estimation algorithm. An algorithm for exact simulations from Inverse Gaussian-OU processes, is given by Zhang and Zhang (2008).

Greater flexibility within the framework of OU processes can be achieved – without sacrificing analytical tractability – by superposition. That is, by replacing $\sigma^2(t)$ with a sum of m independent OU processes:

$$\sigma^2(t) = \sum_{j=1}^{m} \sigma_j^2(t), \tag{10}$$

where the $\sigma_j^2(t)$ are independent OU processes, with mean, variance and autocorrelation function ξ_j, ω_j^2 and $r_j(s) = \exp(-\lambda_j|s|)$. Let $r(s)$ denote the autocorrelation function of $\sigma^2(t)$. Then

$$\begin{aligned} E(\sigma^2(t)) &= \sum_{j=1}^{m} \xi_j \equiv \xi,\ Var(\sigma^2(t)) = \sum_{j=1}^{m} \omega_j^2, \\ r(s) &= \sum_{j=1}^{m} w_j e^{-\lambda_j|s|}, \text{ with } w_j = \omega_j^2 / \sum_{j=1}^{m} \omega_j^2. \end{aligned} \tag{11}$$

Let $z_j(t)$ denote the Levy-process corresponding to $\sigma_j^2(t)$ and define

$$\sigma_{jn}^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma_j^2(t)dt. \tag{12}$$

Indeed, as shown in our empirical application to exchange rate data, superposition is essential for obtaining a good fit to the data. It is also possible to extend (1) to incorporate leverage effects, i.e., a negative correlation between returns and changes in actual volatility, but we do not consider this extension in this paper.

3 Estimation

BS (2001) give an approximate state space representation of the OU model for volatility discussed above using the first and second order properties of y_n and y_n^2. In particular, they show that:

$$\begin{aligned} y_n &= \mu\Delta + u_{1n} \\ y_n^2 &= \mu^2\Delta^2 + \sigma_n^2 + u_{2n}, \end{aligned} \tag{13}$$

where $E(u_{i1}|\sigma_n) = 0$ for $i = 1, 2$. In Section 3.2 we extend the state space representation proposed by BS (2001) to the case with superposition (10)-(11), with

$$\sigma_n^2 = \sum_{j=1}^{m} \sigma_{jn}^2, \tag{14}$$

cf. (12). The state space form allows us to formulate a Gaussian quasi-likelihood function, to make inference about realized volatility, σ_n^2, and to estimate the parameters $\mu, \xi,\ \lambda_1, .., \lambda_m, \omega_1^2, ..., \omega_m^2$. However, because the background driving Levy process is a jump process, and therefore far from normally distributed, this approach is not efficient. Moreover, quasi-likelihood estimators are not generally consistent. These concerns have motivated researchers to investigate other possible estimators. In this paper we shall explore an approach that uses indirect inference methods and computer simulations.

3.1 Indirect inference

The idea of combining a quasi-likelihood function (or an approximate model) with simulations from an underlying "true" model is called indirect inference; see Gourieroux et al. (1993) and Heggland and Frigessi (2004) – who apply this method to queue models with partially observed data. This method seems appropriate in our situation, where computing the exact likelihood is infeasible because the σ_n^2 must be "integrated out" of the conditional density of $y_1, ..., y_N$ given $\sigma_1^2, ..., \sigma_N^2$ to obtain the likelihood function. A

procedure for Bayesian inference based on Markov chain Monte Carlo (MCMC) methods is outlined by BS (2001): A prior distribution is placed on the parameters, and both the parameters and the latent volatility process are updated in a Metropolis-Hastings algorithm. Due to possible high correlation between the latent variables and the parameters, convergence may be very slow, although Griffin and Steel (2006) show that convergence can be considerably faster for the particular case of the Gamma-OU volatility process.

We shall now discuss our proposed indirect inference method in some more detail. For concreteness, consider the situation where the marginal distribution of $\sigma^2(t)$ is the Gamma distribution: $\sigma^2(t) \sim \text{Gamma}(\nu, \alpha)$. Let θ^0 denote the vector of the true parameter values of the underlying data generating model: $\theta^0 = (\omega^0, \xi^0, \lambda^0, \mu^0)$ (with $\omega^0 = \sqrt{\nu^0}/\alpha^0$ and $\xi^0 = \nu^0/a^0$ for Gamma-OU processes), while ψ^*= (ω^*,ξ^*, λ^*, μ^*) is the pseudo-true parameters in the quasi-likelihood, i.e., the probability limit of the quasi-likelihood estimator

$$\widehat{\psi}_N = \arg\max_{\psi} N^{-1} L(\psi; \overrightarrow{y}_N), \tag{15}$$

where $L(\psi; \overrightarrow{y}_N)$ is the Gaussian quasi log-likelihood function based on the actual data $\overrightarrow{y}_N = \{y_n\}_{n=1}^N$. This function can be decomposed sequentially as

$$N^{-1} L(\psi; y_N) = \frac{1}{N} \sum_{n=1}^{N} \ln f(y_n | \overrightarrow{y}_{n-1}; \psi), \tag{16}$$

where $f(y_n | \overrightarrow{y}_{n-1}; \psi)$ is the conditional density of y_n given $\overrightarrow{y}_{n-1}$ under the quasi-likelihood assumptions. Note that if the quasi-likelihood estimate of, say λ, is inconsistent, then λ^* will differ from λ^0, where the difference equals the asymptotic bias of the quasi-likelihood estimator. Moreover, under the conditions of Gourieroux et al. (1993),

$$\sqrt{N}(\widehat{\psi}_N - \psi^*) \overset{D}{\Rightarrow} \mathcal{N}(0, J^{-1} I J^{-1}), \tag{17}$$

where

$$\begin{aligned} I(\psi) &= \lim_N Var(N^{-1/2} \frac{\partial}{\partial \psi} L(\psi; \overrightarrow{y}_N)) \\ J(\psi) &= -\text{p}\lim N^{-1} \frac{\partial^2}{\partial \psi \partial \psi'} L(\psi; \overrightarrow{y}_N)) \\ J &= J(\psi^*) \text{ and } I = I(\psi^*). \end{aligned}$$

The purpose of simulations in indirect inference is to establish a link function between θ and ψ, which will enable us to estimate θ^0 from $\widehat{\psi}_N$. For a given $\theta = (\omega, \xi, \lambda, \mu)$ we

can simulate a sequence of the σ_n^2, denoted $\{\sigma_n^{2(s)}\}_{n=1}^N$, and a sequence $\{\varepsilon_n^{(s)}\}_{n=1}^N$, where $\varepsilon_n^{(s)} \sim \mathcal{IN}(0,1)$. Then we obtain a simulated sequence $\overrightarrow{y}_N^{(s)}(\theta) = \{y_n^{(s)}\}_{n=1}^N$, given θ, where

$$y_n^{(s)} = \mu\Delta + \sigma_n^{(s)}\varepsilon_n^{(s)}.$$

In the Gamma-OU-case we can simulate σ_n^2 using (6), (7) and (9). We can then write

$$\overrightarrow{y}_N^{(s)} = \varphi(\overrightarrow{e}_N^{(s)}, \theta), \tag{18}$$

where $\varphi(\cdot,\cdot)$ is a continuous function in θ and $\overrightarrow{e}_N^{(s)} = \{\varepsilon_n^{(s)}, a_n^{(s)}, r_n^{(s)}\}_{n=1}^N$ is the s'th simulated sequence of $\{\varepsilon_n,\, a_n, r_n\}_{n=1}^N$. To denote $\overrightarrow{y}_N^{(s)}$ as a function of θ for *fixed* $\overrightarrow{e}_N^{(s)}$, we use the notation $\overrightarrow{y}_N^{(s)}(\theta)$.

We obtain a quasi-likelihood estimator $\widehat{\psi}^{(s)}(\theta)$ for the simulated data as follows:

$$\widehat{\psi}_N^{(s)}(\theta) = \arg\max_{\psi} L(\psi;\, \overrightarrow{y}_N^{(s)}(\theta)). \tag{19}$$

Since $\overrightarrow{y}_N^{(s)}(\theta)$ is a continuous function of θ according to (18), $\widehat{\psi}^{(s)}(\theta)$ is also continuous in θ. We will refer to (19) as the *inner* optimization. Under certain regularity conditions (see Gourieroux et al., 1993),

$$\widehat{\psi}_N^{(s)}(\theta) \xrightarrow{P} b(\theta),$$

the so-called binding function, as the number of observations N tends to infinity, where $b(\theta)$ is continuously differentiable and defined through

$$b(\theta) = \arg\max_{\psi} \lim_{n} E_\theta(\ln f(y_n|\overrightarrow{y}_{n-1}; \psi)). \tag{20}$$

Thus $\psi^* = b(\theta^0)$. If $b(\cdot)$ was known, a consistent estimator of θ would be $\widehat{\theta}_N = b^{-1}(\widehat{\psi}_N)$. Kuk (1995) utilizes this relation to obtain finite-sample corrections of estimators which are known to be consistent.

We follow here the approach of Gourieroux et al. (1993) and obtain $\widehat{\theta}_N^{(s)}$ by minimizing, with respect to θ, the distance between $\widehat{\psi}_N^{(s)}(\theta)$ and $\widehat{\psi}_N$ (the quasi-likelihood estimate of ψ based on the actual data, $\overrightarrow{y}_N$) in a weighted mean squared error sense:

$$\begin{aligned} \widehat{\theta}_N^{(s)} &= \arg\min_{\theta} (\widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta))' \Omega_N (\widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta)) \\ &\equiv \arg\min_{\theta} \left\| \widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta) \right\|_{\Omega_N}. \end{aligned} \tag{21}$$

Asymptotically, $||\widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\widehat{\theta}_N^{(s)})|| \xrightarrow{P} 0$, and if

$$\widehat{\psi}_N = \widehat{\psi}_N^{(s)}(\widehat{\theta}_N^{(s)}), \tag{22}$$

the choice of Ω_N is immaterial for the distribution of $\widehat{\theta}_N^{(s)}$. In practice, the solution of (21) is subject to numerical optimization errors, and an exact solution satisfying (22) is not feasible. In our application we accept solutions with $||\widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta)||_{\Omega_N} < \varepsilon$, for a given tolerance level ε, where

$$\Omega_N = \frac{\partial^2}{\partial\psi\partial\psi'} L(\psi; \overrightarrow{y}_N). \tag{23}$$

Note that Ω_N is a by-product of any quasi-Newton routine for maximizing the quasi-likelihood function. Intuitively, when estimating θ, most weight is given to deviations between components of $\widehat{\psi}_N$ and $\widehat{\psi}_N^{(s)}(\theta)$ which are most accurately identified in the quasi-likelihood estimation of ψ.

To reduce estimation uncertainty due to simulations, Gourieroux et al. (1993) consider different alternatives, including replacing $\widehat{\psi}_N^{(s)}(\theta)$ in (21) by the average $S^{-1}\sum_{s=1}^{S}\widehat{\psi}_N^{(s)}(\theta)$. However, due to round-off errors in the Kalman filter, the *outer* minimization (21) is not feasible in that case when SN is very large. As our final indirect inference estimator, we propose instead the average across S indirect inference estimators:

$$\widehat{\theta}_N^S = \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_N^{(s)},$$

where S is chosen so as to keep the estimation uncertainty due to simulations (i.e., the Monte Carlo standard error) below a desired tolerance level. The next proposition shows that $\widehat{\theta}_N^S$ has the same asymptotic distribution as the indirect inference estimators considered by Gourieroux et al. (1993). Moreover, the conditional independence property of $\widehat{\theta}_N^{(s)}$ given $\overrightarrow{y}_N$ can be utilized to obtain a simple estimator of the variance of $\widehat{\theta}_N^S$, based on the sample variance of $\widehat{\theta}_N^{(s)}$.

Proposition 1 *Under the regularity conditions of Gourieroux et al. (1993), given that* $\widehat{\theta}_N^{(s)}$ *for* $s = 1, ..., S$ *satisfies (22), then as* $N \to \infty$

$$\sqrt{N}(\widehat{\theta}_N^S - \theta_0) \overset{D}{=} \left[\frac{\partial b(\theta^0)}{\partial\theta'}\right]^{-1} J^{-1} \times \tag{24}$$
$$\left(N^{-1/2}\frac{\partial}{\partial\psi}L(\psi^*; \overrightarrow{y}_N) - S^{-1}\sum_{s=1}^{S} N^{-1/2}\frac{\partial}{\partial\psi}L(\psi^*; \varphi(\overrightarrow{e}_N^{(s)}, \theta^0))\right)$$

with

$$\begin{aligned} Var(\widehat{\theta}_N^S) &\simeq N^{-1}(1+\frac{1}{S})\left[\frac{\partial b(\theta^0)}{\partial \theta'}\right]^{-1} J^{-1} I J^{-1}\left[\frac{\partial b(\theta^0)}{\partial \theta'}\right]^{-1\prime} \\ &\simeq (1+\frac{1}{S}) Var(\widehat{\theta}_N^{(s)} | \overrightarrow{y}_N). \end{aligned} \tag{25}$$

The proof is given in Appendix A. Note that (25) is very useful as it gives a non-parametric estimate of $Var(\widehat{\theta}_N^S)$ from the sample variance of $\widehat{\theta}_N^{(s)}$ for $s = 1, ..., S$ (which are conditionally independent estimators, given $\overrightarrow{y}_N$). It also allows us to monitor the convergence of the indirect inference estimator by recursively calculating $\widehat{\theta}_N^S$ for increasing S.

3.2 The quasi-likelihood function based on a Gaussian state space model

The indirect inference method outlined above requires a large number of quasi-likelihood estimates. These estimates must therefore be evaluated rapidly, even for large N. However, latent variable models are estimated by means of the EM or ECM algorithm, which are notoriously slow. In a similar way as Raknerud et al. (2010), we combine features of the EM algorithm with an efficient quasi-Newton algorithm. In the EM algorithm, the log-likelihood function, $L(\psi)$ (which in our case will be a quasi log-likelihood function), is decomposed as:

$$L(\psi) = M(\psi|\psi') - H(\psi|\,\psi'), \tag{26}$$

where $M(\psi|\psi')$ is maximized iteratively with respect to ψ to update ψ'. Importantly, the function $M(\psi|\psi')$ has the following property:

$$\left.\frac{\partial L(\psi)}{\partial \psi}\right|_{\psi=\psi'} = \left.\frac{\partial M(\psi|\psi')}{\partial \psi}\right|_{\psi=\psi'}, \tag{27}$$

which follows from the fact that ψ' is the maximizer of $H(\psi|\psi')$, and hence a stationary point. Hence the derivatives $\partial L(\psi')/\partial\psi$ can easily be obtained by *analytic* differentiation of $M(\psi|\psi')$. These derivatives can then be used as inputs in a fast quasi-Newton algorithm, where the log-likelihood can be calculated from the state space form using conventional methods (see e.g. Harvey, 1989). In our experience, the convergence of the algorithm is extremely fast compared to the EM algorithm, and ideally suited when estimation must be repeated many times. Some background for (26) and (27) is given in Dempster et al.

(1977), with discussions. See also Fahrmeir and Tutz (1994). An explicit derivation of $M(\psi|\psi^0)$ and $\partial M(\psi|\psi^0)/\partial\psi$ is given in Appendix B for the state space model presented below.

A linear state space representation Consider the measurement equations (13). We have

$$\begin{aligned} u_{1n} &= \int_{(n-1)\Delta}^{n\Delta} \sigma(u)dW(u), \\ u_{2n} &= u_{1n}^2 + 2\mu\Delta u_{1n} - \sigma_n^2. \end{aligned} \tag{28}$$

Thus

$$u_{1n}|\sigma_n \sim N(0, \sigma_n^2), \tag{29}$$

and we can write

$$\begin{aligned} u_{1n} &= \sigma_n\varepsilon_n;\ \varepsilon_n \sim \mathcal{IN}(0,1) \\ u_{2n} &= \sigma_n^2(\varepsilon_n^2 - 1) + 2\mu\Delta\sigma_n\varepsilon_n. \end{aligned} \tag{30}$$

Let $u_n = (u_{1n}, u_{2n})'$ and consider the case of superposition (10). Then we have the following result: $Var(u_n) = \Sigma$, where

$$\Sigma = \begin{bmatrix} \xi\Delta & 2\mu\Delta^2\xi \\ 2\mu\Delta^2\xi & \sum_{j=1}^m \left[4\omega_j^2\lambda_j^{-2}\{e^{-\lambda_j\Delta} - 1 + \lambda_j\Delta\}\right] + 2\xi^2\Delta^2 + 4\mu^2\Delta^3\xi \end{bmatrix}. \tag{31}$$

A detailed derivation is given in Appendix A.

Let $blockdiag(A_1, ..., A_m)$ denotes the blockdiagonal matrix with i'th block equal to A_i. In Appendix A we derive the following state space representation:

Proposition 2 *Assume that $\sigma^2(t)$, $\sigma_j^2(t)$ and σ_{jn}^2 are given by (10)-(12). Let $\widetilde{\alpha}_n = [\sigma_{1n}^2 - \xi_1\Delta, \sigma_1^2(n\Delta) - \xi_1, ..., \sigma_{mn}^2 - \xi_m\Delta, \sigma_m^2(n\Delta) - \xi_m]'$ be the state vector, and let $Y_n = [y_n, y_n^2]'$ be the observation vector. Then*

$$\begin{aligned} Y_n &= \tau + G\widetilde{\alpha}_n + u_n \\ \widetilde{\alpha}_n &= \boldsymbol{F}\ \widetilde{\alpha}_{n-1} + \widetilde{\eta}_n \end{aligned} \quad ; \quad n = 1, ..., N, \tag{32}$$

where $E(\widetilde{\eta}_n) = E(u_n) = 0$, $Var(u_n) = \Sigma$ is given in (31), $Var(\widetilde{\eta}_n) \equiv Q = blockdiag(Q_1, ..., Q_m)$, with

$$Q_j = 2\omega_j^2 \begin{bmatrix} \lambda_j^{-2}\left(-\frac{3}{2} - \frac{1}{2}e^{-2\lambda_j\Delta} + 2e^{-\lambda_j\Delta} + \lambda_j\Delta\right) & \lambda_j^{-1}\left(1 - e^{-\lambda_j\Delta} - \frac{1}{2}(1 - e^{(-2\lambda_j\Delta)}\right) \\ \lambda_j^{-1}\left(1 - e^{-\lambda_j\Delta} - \frac{1}{2}(1 - e^{(-2\lambda_j\Delta)}\right) & \frac{1}{2}(1 - e^{(-2\lambda_j\Delta)}) \end{bmatrix},$$

$F = blockdiag(F_1, ..., F_m)$, *with*

$$F_j = \begin{bmatrix} 0 & \frac{1-e^{-\lambda_j \Delta}}{\lambda_j} \\ 0 & e^{-\lambda_j \Delta} \end{bmatrix},$$

$G = \begin{bmatrix} G_1 & \cdots & G_m \end{bmatrix}$, *with*

$$G_j = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

and

$$\tau = \begin{bmatrix} \mu\Delta \\ \mu^2\Delta^2 + \xi\Delta \end{bmatrix}. \tag{33}$$

It is clear from the above representation that one cannot identify $\xi_1, ..., \xi_m$ separately, only the sum $\xi = \sum_{j=1}^{m} \xi_j$. Furthermore, we note that

$$Q_j = 2\omega_j^2 \begin{bmatrix} 0 & 0 \\ 0 & \lambda_j\Delta \end{bmatrix} + O((\lambda_j\Delta)^2) \tag{34}$$

and thus becomes singular when $(\lambda_j\Delta)^2 \approx 0$. Approximating Q_j with the first term in (34), is equivalent to the approximation

$$\sigma_{jn}^2 \simeq \Delta\sigma_j^2((n-1)\Delta). \tag{35}$$

Therefore, for components $\sigma_j^2(t)$ with small $\lambda_j\Delta$, the approximation (35) can be used to eliminate redundant (i.e., almost linearly dependent) state variables, i.e. we do not need to include $\sigma_{jn}^2 - \xi_j\Delta$ in the state vector.

To calculate the corresponding quasi-likelihood function, let $a_{t|s} = E(\widetilde{\alpha}_t|Y_1, ..., Y_s)$ and $V_{t|s} = Var(\widetilde{\alpha}_t|Y_1, ..., Y_s)$, which are easily computed by means of the Kalman filter and smoother (see Appendix B), under assumption of joint normality of all random variables. Furthermore, let ψ denote the vector with the unknown parameters. The (quasi) log-likelihood function based on the Gaussian state space model then takes the standard form:

$$L(\psi) = -\frac{1}{2}\sum_{n=1}^{N} \ln|D_n| + \left(Y_n - \tau - Ga_{n|n-1}\right)' {D_n}^{-1} \left(Y_n - \tau - Ga_{n|n-1}\right).$$

See Appendix B for more details and computation of derivatives.

4 Application to exchange rate data

We use $N = 5000$ daily returns data from 1.7.1989 until 15.12.2008 for the Euro/NOK and US Dollar/NOK exchange rates, i.e., the daily changes in the log price of Euro and Dollar, respectively, measured in Norwegian kroner. Thus $\Delta = 1$ corresponds to one day. The return series are depicted in Figure 2. The financial crisis that broke out in September 2008 is clearly visible, leading to large spikes in the figure. Kernel smoothed density estimates of the two returns series are displayed in Figure 3. Descriptive statistics calculated from these empirical densities are shown in Table 1, including measures of skewness and kurtosis for daily returns, y_n, $n = 1, \ldots, 5000$; and scaled 5-days returns: $\sqrt{5^{-1}} \sum_{i=1}^{5} y_{5(m-1)+i}$, $m = 1, ..., 1000$, cf. (5). Table 1 shows that the empirical coefficient of skewness is zero for all practical purposes, which is common for exchange rate data. For the daily returns, we find excess kurtosis (above 3) for both Euro and Dollar, but less so for Dollar (4.47) than for Euro (6.18). As predicted by (5), both coefficients of kurtosis are closer to 3 for the 5-days returns than for the daily returns. The raw data have coefficients of kurtosis that are considerably larger than those reported in Table 1, but these are extremely vulnerable to outliers and therefore not very informative.

Computational issues The *inner* optimization of the quasi-likelihood $L(\psi; \overrightarrow{y}_N^{(s)}(\theta))$ with respect to ψ (for given θ) is carried out by means of a quasi-Newton algorithm that incorporates Fletcher's line search sub-algorithm (Fletcher, 1987, p. 34). Fast convergence is facilitated by good starting values. During the *outer* optimization, small steps are made in the θ–space. Since $\overrightarrow{y}_N^{(s)}(\theta)$ is continuous in θ, the subsequent inner maximizers with respect to ψ are close to each other. Therefore, the previous inner maximizer is typically a very good starting value for the next inner optimization problem. We consider the inner optimization as having converged when the gradient vector has no components exceeding 0.001 in absolute value.

To take restrictions on the parameters into account, these are reparametrized as follows:

$$\begin{aligned} \lambda_1 &= \frac{\lambda_{\max}}{1 + e^{-c_1}}, \ \lambda_j = \frac{\lambda_{j-1}}{1 + e^{-c_j}} \text{ for } j = 2, .., m, \\ \omega_j &= e^{c_{m+j}} \text{ for } j = 1, \ldots, m, \text{ and } \xi = e^{c_{2m+1}}, \end{aligned} \tag{36}$$

where $\lambda_{\max}$ is a pre-specified upper bound on λ_1, and $c_1, \ldots, c_{2m+1}$ are unrestricted parameters. Note that $\lambda_1 > \lambda_2 > ... > \lambda_m$.

While the *inner* maximization (with respect to ψ) is relatively straightforward, the *outer* maximization (21) (with respect to θ) is much more complicated. First, the function $\widehat{\psi}_N^{(s)}(\theta)$ can only be evaluated numerically. Second, even if $\overrightarrow{y}_N^{(s)}(\theta)$ is continuous in θ, it is not a continuously differentiable function. The reason is that $W^{-1}(a_i/\lambda)$ has kinks at $a_i = \nu\lambda$, see (9), forcing us to use derivative-free methods instead of numerical differentiation. This also means that the simpler minimum chi-squared estimator proposed, in a similar context, by Gallant and Long (1997) or the EMM method used by Andersen et al. (1999) are not applicable in our situation. Instead we apply a (slightly simplified) version of the conjugate direction method due to Brent (1973, Ch. 7), in combination with the derivative-free line search algorithm from the Numerical Recipes library (Press et al., 1994, p. 419). Using the convergence criterion

$$\left\| \widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta) \right\|_{\Omega_N} < 0.1, \tag{37}$$

where Ω_N is given by (23). When $m = 1$, to fulfil (37) requires about 150 *inner* optimizations, each of which requires – on average – 12 inner function evaluations (i.e., evaluations of $L(\psi)$ and its derivatives). The corresponding numbers when $m = 2$ are 535 and 13, respectively. Implementing our algorithm in C++, running a HP xw6600 workstation with an Intel(R) Xeon(R) E5420 2.50GHz processor, each evaluation of $L(\psi)$ (including its gradient) takes about 0.05 seconds when $m = 1$ and 0.06 seconds when $m = 2$. Thus evaluation of the outer function $||\widehat{\psi}_N - \widehat{\psi}_N^{(s)}(\theta)||_{\Omega_N}$ takes typically less time than one second.

Sometimes our outer optimization algorithm converges to a point where (37) is not fulfilled. This is unusual for $m = 1$, but when $m = 2$, ω_1 or ω_2 may converge towards zero, implying that there is only one volatility component. This indicates that either $\widehat{\psi}_N^{(s)}(\theta)$ does not span $\widehat{\psi}_N$, or that the numerical optimization fails. Numerical optimization methods may, of course, be sensitive to starting values and there is no guarantee of convergence towards a global optimum. The imposition of the rather strong criterion (37) is a way of ensuring that an optimal point is, indeed, found. Thus, simulations where (37) are not met are disregarded when forming our final indirect inference estimator $\theta_N^S = S^{-1} \sum \theta_N^{(s)}$. The effect of this selection on the finite sample properties of the estimator is

unclear and must be addressed by simulation studies. In our application we use $S = 100$, which means that the standard error due to the simulations (finite S) contributes to less than 1/10 of the total standard error of θ_N^S, cf. (25).

Estimation results Results from quasi-likelihood estimation of the model without superposition, $m = 1$, and with superposition of $m = 2$ volatility processes are shown in Table 3. When estimating models with $m = 3$, we obtain indistinguishable estimates of λ_1 and λ_2. Thus $m = 2$ seems to be adequate. Griffin and Steel (2006) came to the same conclusion using daily U.S. stock returns data.

For both exchange rates, the smallest λ (λ_2) is estimated to around 0.015 and the largest λ to around 0.45. Figure 4 depicts the empirical versus estimated (model-based) autocorrelation functions (ACFs) of squared returns, y_n^2, and actual (integrated) volatility, σ_n^2. The figure shows a good agreement between the data and the model. In the beginning, the estimated ACF tail off quickly (the effect of λ_1) and then very slowly after 5–10 days (due to λ_2). The model without superposition is not able to pick up the slowly decaying empirical autocorrelation pattern for lags exceeding 5-10 days. We also see from the estimates of ω_1^2 and ω_2^2 in Table 2, that the ACF for Dollar has relatively more weight on the lowest λ compared to Euro, leading to a more slowly decaying pattern.

We note that the estimated average spot volatility $E(\sigma^2(t)) = \xi$ is much higher for Dollar/NOK (0.44) than for Euro/NOK (0.12). The spot volatility of the dollar/NOK rate also fluctuates much more over time: $Var(\sigma^2(t)) = \omega_1^2 + \omega_2^2$ is estimated to 0.34 for Dollar/NOK, but only 0.12 for Euro/NOK. That the Dollar-volatility is much larger than the Euro-volatility is also evident from Figure 5, which shows the predicted values of actual volatility, σ_n^2, obtained from the Kalman smoother. We see that the actual Dollar-volatility is almost uniformly higher than the Euro-volatility over the sample period.

The indirect inference estimators obtained by averaging $\widehat{\theta}_N^{(s)}$ for $s = 1, ..., S$; and with $S = 100$ simulated sequences $\{y_n\}_{n=1}^N$, are shown in Table 3. We first note that all the parameter estimates are almost identical to the quasi-likelihood estimates reported in Table 2. When considering that the Monte Carlo standard error is 1/10 of the standard errors reported in the parentheses of Table 3, we conclude that the quasi-likelihood and indirect inference estimates are not significantly different. On the other hand, the standard errors and confidence intervals generated by the two methods differ substantially with

respect to several of the parameters. Confidence intervals are obtained by transforming symmetric intervals for the c_j-parameters in (36) back to the original parameters. Most notably, the "sandwich" matrix estimator $J^{-1}IJ^{-1}$ in (17), gives larger standard errors and wider confidence intervals for λ_1 and λ_2 than do the non-parametric standard errors and corresponding confidence intervals calculated from the sample variance of $\widehat{\theta}_N^{(s)}$; see (25). The results with regard to the estimation uncertainty of ω_1^2 and ω_2^2 go in the opposite direction, with much narrower confidence intervals emerging from the quasi-likelihood-based method.

To evaluate the properties of the two methods more systematically, we have carried out two small simulation studies, reported in tables 4 and 5. In Table 4, we simulate data from a model with $m = 1$ and $\lambda = 0.1$, using $S = 10$ replications to obtain $\widehat{\theta}_N^S$. In Table 5, we simulate from a model with $\lambda = 0.5$. All other parameters are equal in the two simulation studies. The results confirm that the two methods give almost identical point estimates and both have almost the same level of precision (with $S = 10$). The studies also show that both methods give, for all practical purposes, unbiased estimators. On the other hand, there appears to be a systematic difference with respect to assessment of estimation uncertainty, confirming the differences noted above. First, concentrating on the 95 percent confidence intervals, we see that the actual coverage of the indirect inference-based intervals are close to 95 percent; varying between 91 and 95 percent across the different parameters and simulations. On the other hand, the quasi-likelihood method gives confidence intervals for λ with a coverage of 96-99 percent, whereas the coverage for ω^2 is much too small: only 70 percent in both tables. These differences with respect to λ and ω^2 are significant, and they confirm that the quasi-likelihood based standard errors are upward biased with regard to λ, but downward biased with regard to ω^2.

5 Conclusions

In this paper we have developed a new method for indirect inference for a class of stochastic volatility models for financial data based on non-Gaussian Ornstein-Uhlenbeck (OU) processes that were originally proposed in the context of financial econometrics by Barndorff-Nielsen and Shephard (2001). The volatility in this class of models is driven by Levy jump processes. Many analytical results about the distribution and dependence

structure of integrated volatility are available, leading to exact discrete time transition equations that can be formulated on a state space form, with white noise error terms. By assuming that these error terms have a multivariate normal distribution (which is contrary to the model assumptions), we obtain an approximative Gaussian state space representation of the OU-based model, which can be estimated on data by means of the Kalman filter and smoother. The resulting estimator is a quasi-likelihood estimator. By combining quasi-likelihood estimation with simulations from the data generating OU-model for given parameter values, we successfully implement a procedure for indirect inference. The indirect inference estimator is the parameter value in the OU-model which gives the best "match" between the quasi-likelihood estimator for the actual data and the quasi-likelihood estimator for the simulated data. Accompanying software written in C++ code is documented and can be downloaded.

In an application using 5000 daily exchange rate observations from 1.7.1989 until 15.12.2008 for the Euro/NOK and US Dollar/NOK exchange rates, we demonstrated that our estimation algorithm is feasible with large data sets and have good convergence properties. The indirect inference and quasi-likelihood estimator gave almost identical point estimates, but the two methods led to significantly different answers when it came to assessing estimation uncertainty (standard errors and confidence intervals). In a simulation study, we found that the actual coverage of 95 percent confidence intervals were close to 95 percent using indirect inference, but could be as low as 70 percent when the intervals were based on the classical formula for estimating standard errors in misspecified models (i.e., the "sandwich" matrix estimator). There are several related topics that we will address in future research, e.g., allowing leverage effects in the model, in the sense that positive shocks to volatility is associated with lower expected returns, as well as multivariate extension.

References

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001). "The distribution of realized exchange rate volatility." *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T.G., H.-J. Chung and B.E. Sørensen (1999). "Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study." *Journal of Econometrics*, 91, 61-87.

Andersen, T.G., R.A. Davis, J.-P. Kreiss and T. Mikosch (Eds.) (2009). *Handbook of Financial Time Series.* Berlin: Springer.

Barndorff-Nielsen O.E. and D.R. Cox (1989). *Asymptotic techniques for use in statistics.* London: Chapman and Hall.

Barndorff-Nielsen O.E., T. Mikosch and S.I. Resnick (Eds.) (2001). *Lévy processes : theory and applications.* Boston: Birkhäuser.

Barndorff-Nielsen O.E. and N. Shephard (2001). "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics." *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 63, 167-207 Part 2.

Barndorff-Nielsen O.E. and N. Shephard (2002). "Econometric analysis of realized volatility and its use in estimating stochastic volatility models." *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 64, 253-280 Part 2.

Barndorff-Nielsen O.E. and N. Shephard (2003). "Integrated OU processes and non-Gaussian OU-based stochastic volatility models." *Scandinavian Journal of Statistics*, 30, 277-295.

Bondesson, L. (2002). "On the Levy measure of the Lognormal and the LogCauchy distributions." *Methodology and Computing in Applied Probability*, 4, 243-256.

Bølviken, E. and F. E. Benth (2000). "Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution." *Proceedings AFIR 2000 Colloquium*, Tromsø, Norway, pp. 87-98.

Bollerslev T., R.F. Engle and D.B. Nelson (1994). "ARCH models," in *Handbook of Econometrics* (Eds.: R.F. Engle and D.L. McFadden), Vol. 4, pp. 2959-3038. Amsterdam: North-Holland.

Brent, R.P. (1973). textitAlgorithms for minimization without derivatives. Englewood Cliffs: Prentice-Hall.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Diebold, F.X. (1988). *Empirical modeling of exchange rate dynamics.* New York: Springer.

Diebold, F.X. and M. Nerlove (1989). "The dynamics of exchange rate volatility: A multivariate latent factor ARCH model." *Journal of Applied Econometrics*, 4, 1-21.

Drost, F.C. and T.E. Nijman (1993). "Temporal aggregation of GARCH processes." textitEconometrica, 61, 909-927.

Engle R.F. (1982)."Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation." *Econometrica*, 50, 987-1007.

Fahrmeier, L. and G. Tutz (1994). *Multivariate statistical modelling based on generalized linear models.* New York: Springer.

Fletcher, R. (1987). *Practical methods of optimization.* New York: John Wiley

Gallant A.R. and G. Tauchen (1996). "Which moments to match?" *Econometric Theory*, 12, 657-681.

Gallant A.R. and J.R. Long (1997). "Estimating stochastic differential equations efficiently by minimum chi-square." *Biometrika*, 84, 125-141.

Gourieroux C., A. Monfort and E. Renault (1993). "Indirect inference." *Journal of Applied Econometrics*, 8, Supplement, S85-S118.

Griffin, J.E. and M.F.J. Steel (2006). "Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility." *Journal Econometrics*, 134, 605-644.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge. Cambridge University Press.

Harvey, A.C., S.J. Koopman and N. Shephard (Eds.) (2004). *State space and unobserved component models: Theory and applications.* Cambridge. Cambridge University Press.

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

Heggland, K. and A. Frigessi (2004). "Estimating functions in indirect inference." *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 66, 447-462.

Jondeau, E., S.-H. Poon and M. Rockinger (2007). *Financial Modeling under Gaussian Distributions.* New York: Springer.

Kuk, A.Y.C. (1995). "Asymptotically unbiased estimation in generalized linear models with random effects". *Journal of the Royal Statistical Society Series B-Statistical Methodology.* 57, 395-407.

MacDonald, R. (1999). "Exchange rate behaviour: Are fundamentals important?" *Economic Journal*, 109, 673-691.

Meese, R. and K. Rogoff (1983). "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" In *Exchange Rates and International Macroeconomics* (Ed.: J.A. Frenkel), National Bureau of Economic Research, pp. 67-112.

Nicolato, E. and E. Venardos (2003). "Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type." *Mathematical Finance*, 13, 445-466.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1994). *Numerical recipes in C.* Cambridge: Cambridge University Press.

Raknerud, A., T. Skjerpen and A.R. Swensen (2010). "Forecasting key macroeconomic variables from a large number of predictors: A state space approach", *Forthcoming, Journal of Forecasting.*

Shephard, N. (Ed.) (2005). *Stochastic volatility: selected readings,* Oxford: Oxford University Press.

Zhang, S. and X. Zhang (2008). "Exact simulation of IG-OU processes." *Methodology and computing in applied probability*, 10, 337-355.

Appendix A: Proofs

Proof of Proposition 1 Averaging over $s = 1, \ldots, S$, (24) follows directly from the expansion of $\widehat{\psi}_N^{(s)}(\theta^0)$ in Gourieroux et al. (1993) (who use the notation $\beta_T^h(\theta^0)$) and the relation (22). From (24),

$$\begin{aligned} Var(\widehat{\theta}_N^S|y_N) &\simeq \frac{1}{S}\left[\frac{\partial b(\theta^0)}{\partial\theta}\right]^{-1} J^{-1}Var(N^{-1/2}\frac{\partial}{\partial\psi}L(\psi^*;\varphi(\overrightarrow{e}_N^{(s)},\theta^0)))J^{-1}\left[\frac{\partial b(\theta^0)}{\partial\theta}\right]^{-1\prime} \\ &= \frac{1}{S}Var(\widehat{\theta}_N^{(s)}|y_N), \end{aligned}$$

where $\simeq$ means asymptotically equivalent. Since $N^{-1/2}\frac{\partial}{\partial\psi}L(\psi^*;\varphi(\overrightarrow{e}_N^{(s)},\theta^0))$ (for $s = 1, ..., S$) and $N^{-1/2}\frac{\partial}{\partial\psi}L(\psi^*;y_N)$ are independent and have the same distribution (the latter is a "simulation of the nature"), we obtain

$$Var(\widehat{\theta}_N^S|y_N) \simeq (1+\frac{1}{S})Var(\widehat{\theta}_N^{(s)}|y_N).$$

Q.E.D.

Proof of (31) From (30) and the rule of double expectation

$$\begin{aligned} Var(u_{1n}) &= E(\sigma_n^2) = \xi\Delta \\ Var(u_{2n}) &= 2E(\sigma_n^4) + 4\mu^2\Delta^2E(\sigma_n^2) \\ &= 2\{Var(\sigma_n^2) + E(\sigma_n^2)^2\} + 4\mu^2\Delta^3\xi \\ &= \sum_{j=1}^{m}\left[4\omega_j^2\lambda_j^{-2}\{e^{-\lambda_j\Delta} - 1 + \lambda_j\Delta\}\right] + 2\xi^2\Delta^2 + 4\mu^2\Delta^3\xi \\ E(u_{1n}u_{2n}) &= 2\mu\Delta E(\sigma_n^2) = 2\mu\Delta^2\xi, \end{aligned}$$

where we have used that

$$E(\sigma_n^4) = Var(\sigma_n^2) + \xi^2\Delta^2,$$

with

$$Var(\sigma_n^2) = \sum_{j=1}^{m}Var(\sigma_{jn}^2) = \sum_{j=1}^{m}\left[4\omega_j^2\lambda_j^{-2}\{e^{-\lambda_j\Delta} - 1 + \lambda_j\Delta\}\right]$$

for independent OU processes $\sigma_j^2(t)$. Q.E.D.

Proof of Proposition 2 First, let us examine the case *without* superposition. Then, from (2)-(3),

$$\sigma_n^2 = \lambda^{-1}\left[z(\lambda n\Delta) - z(\lambda(n-1)\Delta) - \sigma^2(n\Delta) + \sigma^2((n-1)\Delta)\right]. \tag{38}$$

Using (7) and (38), we obtain

$$\begin{bmatrix} \lambda\sigma_n^2 + \sigma^2(n\Delta) \\ \sigma^2(n\Delta) \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-\lambda\Delta} \end{bmatrix} \sigma^2((n-1)\Delta) + \begin{pmatrix} \eta_{1n} \\ \eta_{2n} \end{pmatrix}. \tag{39}$$

Setting $\eta_n = (\eta_{1n}, \eta_{2n})'$, we have

$$\begin{aligned} E(\eta_n) &= \xi \begin{bmatrix} \lambda\Delta \\ 1 - e^{-\lambda\Delta} \end{bmatrix} \\ Var(\eta_n) &= 2\omega^2 \begin{bmatrix} \lambda\Delta & 1 - e^{-\lambda\Delta} \\ 1 - e^{-\lambda\Delta} & \frac{1}{2}(1 - e^{-2\lambda\Delta}) \end{bmatrix}, \end{aligned} \tag{40}$$

see BS (2001). Setting $\alpha_n = [\lambda\sigma_n^2, \sigma^2(n\Delta)]'$, we obtain from (39)

$$\alpha_n = \begin{bmatrix} 0 & 1 - e^{-\lambda\Delta} \\ 0 & e^{-\lambda\Delta} \end{bmatrix} \alpha_{n-1} + \begin{bmatrix} \eta_{1n} - \eta_{2n} \\ \eta_{2n} \end{bmatrix},$$

where

$$E\left(\begin{bmatrix} \eta_{1n} - \eta_{2n} \\ \eta_{2n} \end{bmatrix}\right) = \xi \begin{bmatrix} \lambda\Delta - 1 + e^{-\lambda\Delta} \\ 1 - e^{-\lambda\Delta} \end{bmatrix}.$$

It is convenient to redefine the state vector so that it becomes mean zero. We find $E(\alpha_n)$ by solving:

$$E(\alpha_n) = \begin{bmatrix} 0 & 1 - e^{-\lambda\Delta} \\ 0 & e^{-\lambda\Delta} \end{bmatrix} E(\alpha_n) + \xi \begin{bmatrix} \lambda\Delta - 1 + e^{-\lambda\Delta} \\ 1 - e^{-\lambda\Delta} \end{bmatrix}.$$

Hence

$$E(\alpha_n) = \xi \begin{bmatrix} \lambda\Delta \\ 1 \end{bmatrix}.$$

Define

$$\widetilde{\alpha}_n = \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \left(\alpha_n - \xi \begin{bmatrix} \lambda\Delta \\ 1 \end{bmatrix}\right).$$

Then $\widetilde{\alpha}_n = [\sigma_n^2 - \xi\Delta, \sigma^2(n\Delta) - \xi]'$, as in Proposition 2 (in the special case without superposition), and, from (13),

$$\begin{aligned} \begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda^{-1} & 0 \end{pmatrix} \alpha_n + \begin{pmatrix} u_{1n} \\ u_{2n} \end{pmatrix} \\ &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \alpha_n + \begin{pmatrix} u_{1n} \\ u_{2n} \end{pmatrix} \\ &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 + \xi\Delta \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \widetilde{\alpha}_n + \begin{pmatrix} u_{1n} \\ u_{2n} \end{pmatrix}. \end{aligned} \tag{41}$$

Moreover,

$$\begin{aligned}\widetilde{\alpha}_n &= \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} 0 & 1-e^{-\lambda\Delta} \\ 0 & e^{-\lambda\Delta} \end{bmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \widetilde{\alpha}_{n-1} + \widetilde{\eta}_n \\ &= \begin{bmatrix} 0 & \frac{1-e^{-\lambda\Delta}}{\lambda} \\ 0 & e^{-\lambda\Delta} \end{bmatrix} \widetilde{\alpha}_{n-1} + \widetilde{\eta}_n, \end{aligned} \tag{42}$$

where

$$\widetilde{\eta}_n = \begin{bmatrix} \lambda^{-1}\left[(\eta_{1n} - \eta_{2n}) - \xi(\lambda\Delta - 1 + e^{-\lambda\Delta})\right] \\ \eta_{2n} - \xi(1 - e^{-\lambda\Delta}) \end{bmatrix}.$$

Thus

$$E(\widetilde{\eta}_n) = 0$$

and

$$\begin{aligned} Var(\widetilde{\eta}_n) &= \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} Var(\eta_n) \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & 1 \end{pmatrix} \\ &= Q. \end{aligned} \tag{43}$$

In the general case with superposition, the expression for Q_j follows directly if we replace Q by Q_j, λ by λ_j and ω by ω_j in (40) and (43). The expression for F_j follows from (42), replacing $\widetilde{\alpha}_n$ by $\widetilde{\alpha}_{jn} = \left[\lambda\sigma^2_{jn}, \sigma^2_j(n\Delta)\right]'$ and λ by λ_j. Since the $\sigma^2_j(t)$ are independent, Q and F become blockdiagonal matrices. The remaining part of the proposition follows directly, since by (13) and (14),

$$\begin{aligned}\begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \sum_{j=1}^{m} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \left(\widetilde{\alpha}_{jn} + \xi_j \begin{bmatrix} \Delta \\ 1 \end{bmatrix} \right) + \begin{pmatrix} u_{1n} \\ u_{2n} \end{pmatrix} \\ &= \tau + G\widetilde{\alpha}_n + u_n. \end{aligned}$$

(recall that $\xi = \sum_{j=1}^{m} \xi_j$). Q.E.D.

Appendix B: Derivation of $M(\psi|\psi^0)$ and its derivatives

We derive these expression for the state space model with superposition (32)-(33).

The Kalman filter and smoother: $a_{t|s}$ and $V_{t|s}$ We assume $\widetilde{\alpha}_1 = 0$, since we may ignore the initial value problem because N is large. Then

Kalman filtering:

$$
\begin{aligned}
a_{0|0} &= 0, V_{0|0} = 0 \\
\text{For } n &= 1, ..., N\text{:} \\
a_{n|n-1} &= F a_{n-1|n-1} \\
V_{n|n-1} &= F V_{n-1|n-1} F' + Q \\
D_n &= G V_{n|n-1} G' + \Sigma \\
K_n &= V_{n|n-1} G' D_n^{-1} \\
a_{n|n} &= a_{n|n-1} + K_n \left(Y_n - \tau - G a_{n|n-1} \right) \\
V_{n|n} &= V_{n|n-1} - K_n G V_{n|n-1}.
\end{aligned}
$$

The required conditional expectations $a_{n|N}$ and variances $V_{n|N}$ are obtained in subsequent backward smoothing recursions (see Fahrmeir and Tutz, 1994, p. 265):

Kalman smoothing:

$$
\begin{aligned}
\text{For } n &= N, ..., 2\text{:} \\
a_{n-1|N} &= a_{n-1|n-1} + B_n (a_{n|N} - a_{n|n-1}) \\
V_{n-1|N} &= V_{n-1|n-1} + B_n (V_{n|N} - V_{n|n-1}) B_n',
\end{aligned}
$$

where

$$
B_n = V_{n-1|n-1} F' V_{n|n-1}^{-1}. \tag{44}
$$

Expressions for $M(\psi|\psi^0)$ and $\partial M(\psi|\psi^0)/\partial\psi$ First

$$
M(\psi|\psi^0) = E\left\{ \ln f(Y, \alpha; \psi) |\, Y; \psi^0 \right\}, \tag{45}
$$

where $f(Y,\alpha;\psi)$ is generic notation for the joint normal density function of (Y,α) given the parameter vector ψ, where $Y=\{Y_n\}_{n=1}^N$ is the observed Y_n-vectors, $\alpha=\{\widetilde{\alpha}_n\}_{n=1}^N$ are the latent variables and $E\left\{\cdot|\,Y;\psi^0\right\}$ denotes the condition expectation given Y and evaluated at $\psi=\psi^0$. We then need to evaluate

$$M(\psi|\,\psi^0)=M^{(1)}(\Sigma,\tau,G|\psi^0)+M^{(2)}(\digamma,Q|\psi^0), \tag{46}$$

where

$$\begin{aligned}M^{(1)}(\Sigma,\tau,G|\psi^0)=&-\frac{N}{2}\ln|\Sigma|\\&-\frac{1}{2}\sum_{n=1}^{N}\left[tr\left\{\Sigma^{-1}(Y_n-\tau-Ga_{n|N})(Y_n-\tau-Ga_{n|N})'\right\}+tr\left\{\Sigma^{-1}GV_{n|N}G'\right\}\right]\end{aligned} \tag{47}$$

and

$$\begin{aligned}M^{(2)}(\digamma,Q|\psi^0)=&\sum_{i=1}^{m}\Bigg(-\frac{N}{2}\ln|Q_i|\\&-\frac{1}{2}\sum_{n=1}^{N}\left[tr\left\{Q_i^{-1}(a_{n|N}^{(i)}-F_ia_{n|N}^{(i)})(a_{n|N}^{(i)}-F_ia_{n|N}^{(i)})'\right\}\right].\\&+tr\left\{Q_i^{-1}\left(V_{n|N}^{(i,i)}-(V_{n|N}'B_n')^{(i,i)}F_i'-F_i\left(B_nV_{n|N}\right)^{(i,i)}\right)+F_iV_{n|N}^{(i,i)}F_i'\right\}\Bigg)\end{aligned} \tag{48}$$

where for a general $2m$ dimensional vector a, $a^{(i)}(i=1,2,\ldots,m)$ is a 2-dimensional vector defined by the partition

$$a=\begin{bmatrix}a^{(1)}\\ \\ a^{(m)}\end{bmatrix},$$

and for a general $2m\times 2m$ matrix A, $A^{(i,j)}(i,j=1,2,\ldots,m)$ is a 2×2 matrix defined by the partition

$$A=\begin{bmatrix}A^{(1,1)} & A^{(1,2)} & ... & A^{(1,m)}\\ A^{(2,1)} & A^{(2,2)} & ... & A^{(2,m)}\\ \\ A^{(m,1)} & A^{(m,2)} & & A^{(m,m)}\end{bmatrix},$$

B_n is defined in (44), and we have utilized that

$$\mathsf{E}(\alpha_n\alpha_{n-1}'|\,Y;\psi)=a_{n|N}a_{n-1|N}{}'+V_{n|N}'B_n'.$$

The partial derivatives are then given by:

$$
\begin{aligned}
\frac{\partial M^{(1)}(\Sigma,\tau,G|\psi^0)}{\partial G} &= \sum_{n=1}^{N}\Sigma^{-1}\left(\left(Y_n-\tau-Ga_{n|N}\right)a'_{n|N}-GV_{n|N}\right)\\
\frac{\partial M^{(1)}(\Sigma,\tau,G|\psi^0)}{\partial \tau} &= \sum_{n=1}^{N}\Sigma^{-1}\left(Y_n-\tau-Ga_{n|N}\right)\\
\frac{\partial M^{(1)}(\Sigma,\tau,G|\psi^0)}{\partial vec(\Sigma)} &= -\frac{N}{2}vec(\Sigma^{-1})+\frac{1}{2}(\Sigma^{-1}\otimes\Sigma^{-1})\sum_{n=1}^{N}vec\left[\left(Y_n-\tau-Ga_{n|N}\right)\left(Y_n-\tau-Ga_{n|N}\right)'\right.\\
&\quad \left.+GV_{n|N}G'\right]\\
\frac{M^{(2)}(\boldsymbol{F},Q|\psi^0)}{\partial vec(Q_i)} &= -\frac{N}{2}vec(Q_i^{-1})+\frac{1}{2}(Q_i^{-1}\otimes Q_i^{-1})\sum_{n=1}^{N}vec\left[(a^{(i)}_{n|N}-F_ia^{(i)}_{n-1|N})(a^{(i)}_{n|N}-F_ia^{(i)}_{n-1|N})'\right.\\
&\quad \left.+V^{(i,i)}_{n|N}-F_i(B_nV_{n|N})^{(i,i)}-(V'_{n|N}B'_n)^{(i,i)}{F_i}'+F_i(V_{n-1|N})^{(i,i)}{F_i}'\right]\\
\frac{\partial M^{(2)}(\boldsymbol{F},Q|\psi^0)}{\partial F_i} &= Q_i^{-1}\left[\left(\sum_{n=1}^{N}a^{(i)}_{n|N}a^{(i)}_{n-1|N}{}'+(V'_{n|N}B'_n)^{(i,i)}\right)-\Phi\left(\sum_{n=1}^{N}a^{(i)}_{n-1|N}a^{(i)}_{n-1|N}{}'+V^{(i,i)}_{n-1|N}\right)\right].
\end{aligned}
\tag{49}
$$

Finally, $(\Sigma,\tau,G,\boldsymbol{F},Q)$ are functions of the free parameters ψ, and the partial derivatives with respect to ψ are trivially obtained by using the chain rule on (49) .

Figures and tables

Table 1: **Descriptive statistics based on kernel-smoothed density estimates.** 5000 daily and 1000 (scaled) 5-day returns

	Euro/NOK		**Dollar/NOK**	
	$\Delta = 1$	$\Delta = 5$	$\Delta = 1$	$\Delta = 5$
Mean	0.001	0.006	0.0002	-0.0005
Variance	0.10	0.11	0.46	0.48
Skewness	0.17	0.29	0.15	0.07
Kurtosis	6.18	4.52	4.47	3.33

Table 2: **Quasi-likelihood estimates.** Euro/NOK and Dollar/NOK exchange rates (1.7.1989-15.12.2008)

	Euro/NOK				**Dollar/NOK**			
	$m = 1$		$m = 2$		$m = 1$		$m = 2$	
	estimate*	95% CI**	estimate	95% CI	estimate	95% CI	estimate	95% CI
μ	.016 (.005)	[.006, .025]	.016 (.005)	[.006, .025]	.008 (.01)	[−.012, .030]	.008 (.01)	[−.005, .022]
λ_1	.23 (.07)	[.12, .40]	.39 (.22)	[.16, 1.08]	.090 (.05)	[.02, .30]	.46 (.18)	[.22, .93]
λ_2	–	–	.014 (.013)	[.002, .06]	–	–	.018 (.007)	[.008, .035]
ξ	.11 (.003)	[.11, .12]	.12 (.003)	[.11, .12]	.47 (.01)	[.44, .50]	.46 (.014)	[.43, 0, 49]
ω_1^2	.09 (.002)	[.09, .10]	.08 (.006)	[.07, .09]	.31 (.01)	[.29, .34]	.20 (.058)	[.11, .36]
ω_2^2	–	–	.02 (.007)	[.01, .04]	–	–	.12 (.04)	[.06, .24]

*Standard errors in parentheses based on (estimated) asymptotic covariance matrix $J^{-1}IJ$

**95 % Confidence Intervals (CI) are transformed CI of unrestricted parameters $c_1, ..., c_{2m+1}$

Table 3: **Indirect inference estimates.** Models with m=1 and m=2 volatility components

	Euro/NOK				**Dollar/NOK**			
	$m=1$		$m=2$		$m=1$		$m=2$	
	estimate*	95% CI**	estimate	95% CI	estimate	95% CI	estimate	95% CI
μ	.014 (.016)	$[-.018, .047]$	.017 (.02)	$[-.012, .030]$	.007 (.03)	$[-.043, .058]$	.011 (.03)	$[-.002, .049]$
λ_1	.24 (.05)	[.15, .35]	.45 (.12)	[.27, .77]	.091 (.02)	[.058, .140]	.44 (.12)	[.25, .75]
λ_2	–	–	.015 (.008)	[.004, .04]	–	–	.013 (.005)	[.005, .027]
ξ	.11 (.016)	[.08, .15]	.12 (.02)	[.07, .18]	.47 (.04)	[.38, .57]	.44 (.06)	[.34, .59]
ω_1^2	.10 (.038)	[.04, .21]	.10 (.06)	[.03, .33]	.33 (.08)	[.21, .55]	.24 (.06)	[.14, .40]
ω_2^2	–	–	.02 (.013)	[.001, .07]	–	–	.10 (.05)	[.05, .42]

*Standard errors in parenthesis based on sample variance of $\theta_N^{(s)}$ for $s = 1, ..., S$, with $S = 100$ replicated simulations

**95 % Confidence Intervals (CIs) for λ_j, ξ and ω_j^2 are transformed CIs from the unrestricted parameters $c_1, .., c_{2m+1}$

Table 4: **Monte Carlo simulation I.** Model with m=1 volatility component and 100 simulations

		Quasi-likelihood estimator				**Indirect inference estimator with** $S = 10$			
	true value	mean*	actual SE	estimated SE**	coverage of 95% CI***	mean*	actual SE	estimated SE**	coverage of 95% CI***
μ	0	.002 (0.002)	.022	.023	96	.002 (.002)	.022	.023	95
λ	0.1	.103 (0.002)	.024	.038	99	.103 (.002)	.026	.022	95
ξ	.5	.497 (0.004)	.038	.025	84	.499 (.004)	.040	.039	94
ω^2	.3	.296 (0.006)	.062	.042	70	.309 (.008)	.082	.072	92

*Standard errors in parenthesis based on sample variance of corresponding estimator in 100 Monte Carlo simulations

**Square root of the mean estimated variances across Monte Carlo simulations

*** Actual coverage of nominal 95% CIs

Table 5: **Monte Carlo simulation II.** Model with m=1 volatility component and 100 simulations

		Quasi-likelihood estimator				**Indirect inference estimator with** $S = 10$			
	true value	mean*	actual SE	estimated SE**	coverage of 95% CI***	mean*	actual SE	estimated SE**	coverage of 95% CI***
μ	0	.004 (0.002)	.024	.023	91	.002 (.002)	.025	.025	91
λ	.5	.51 (0.01)	.104	.150	96	.53 (.013)	.127	.143	94
ξ	.5	.50 (0.002)	.021	.026	98	.50 (.002)	.022	.022	91
ω^2	.3	.30 (0.005)	.056	.031	70	.31 (.006)	.062	.057	90

*Standard errors in parenthesis based on sample variance of corresponding estimator in 100 Monte Carlo simulations

**Square root of mean estimated variances across Monte Carlo simulations

*** Actual coverage of nominal 95% CIs

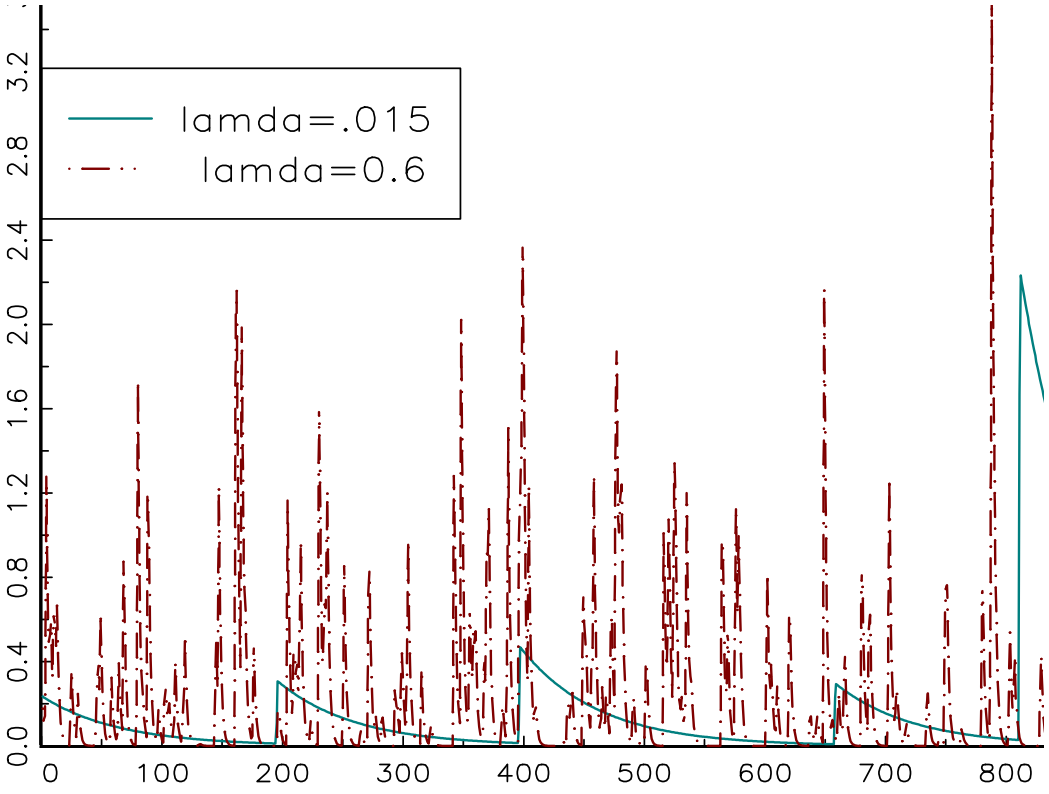

Figure 1: Two simulated volatility series σ_n^2

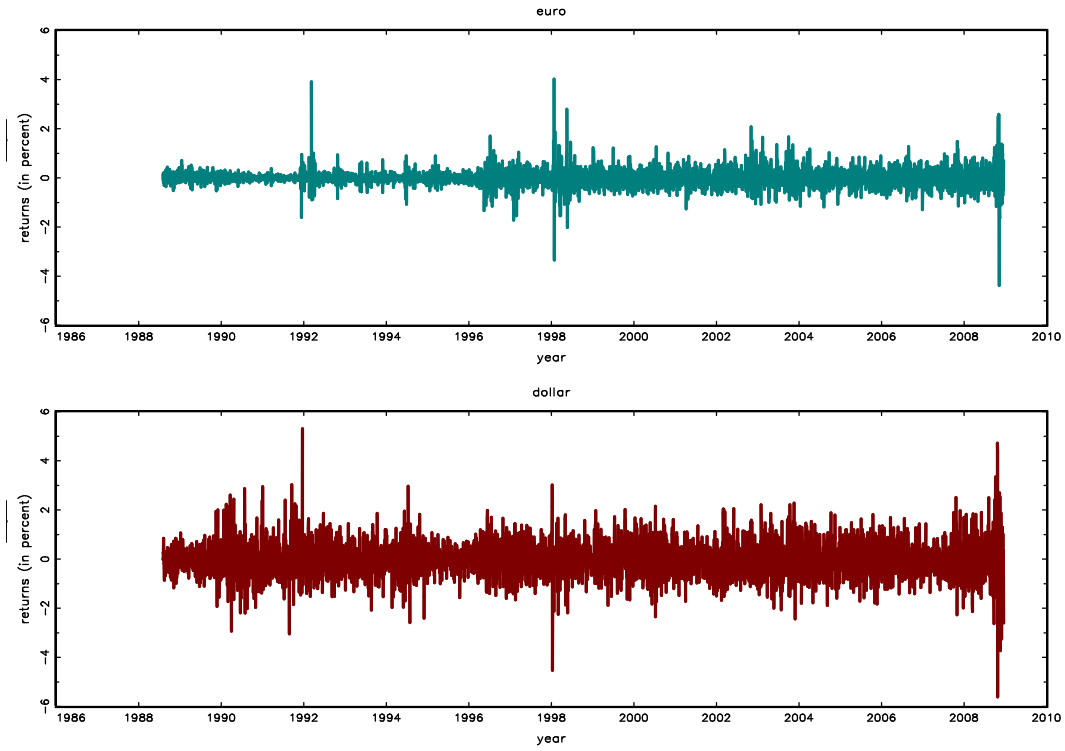

Figure 2: Returns: 1.7.1989 until 15.12.2008

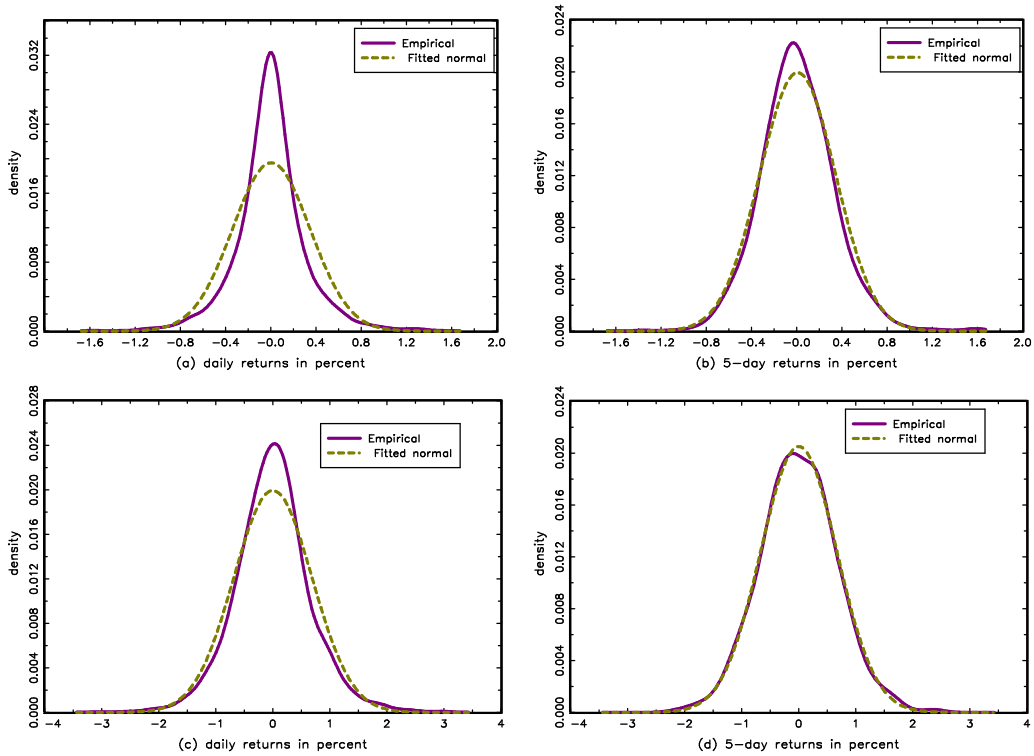

Figure 3: Kernel density estimates of returns. Euro in (a) and (b), Dollar in (c) and (d)

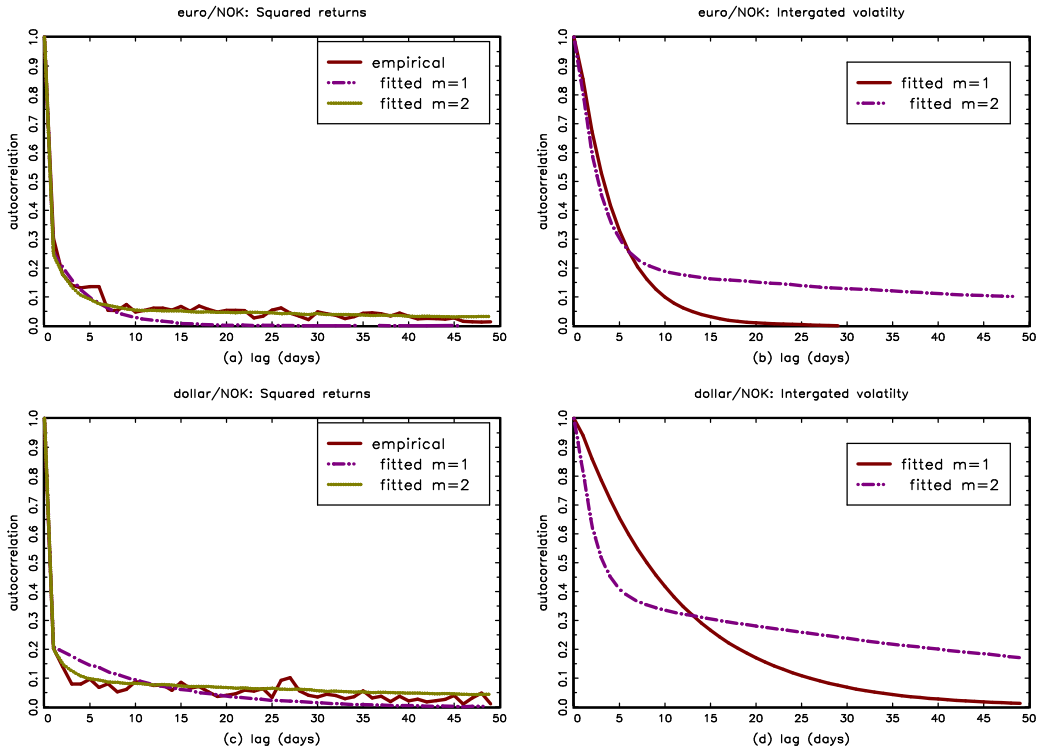

Figure 4: Autocorrelation functions

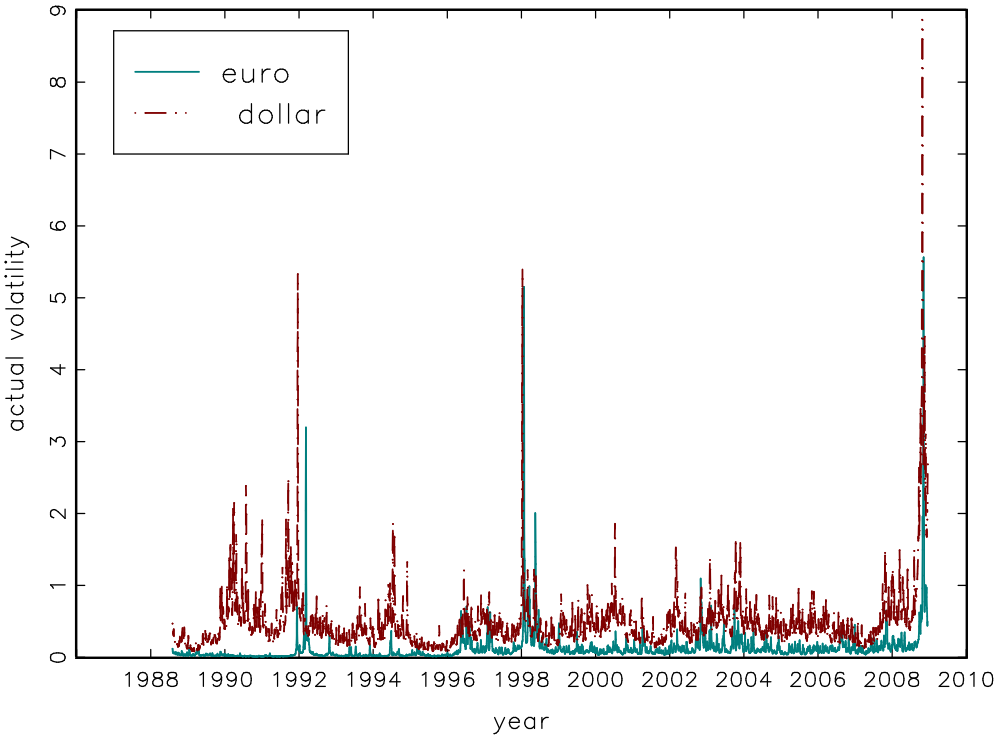

Figure 5: Actual volatility estimates