Preprint of the corresponding article in
Journal of Financial Econometrics (2010) **8** (4) 481-510.

MCMC estimation of the COGARCH(1,1) model

Gernot Müller*

Abstract

This paper presents a Markov chain Monte Carlo based estimation procedure for the COGARCH(1,1) model driven by a compound Poisson process. The COGARCH model is a continuous-time analogue to the discrete-time GARCH model and captures many of the stylized facts of financial time series, as has been shown in various papers. Principles for the estimation of point processes by MCMC are adapted to the special structure of the COGARCH(1,1) model. The algorithm uses discrete GARCH-type equations on a random grid which changes in each iteration of the MCMC sampler. Moreover, exact solutions of the volatility SDE of the COGARCH(1,1) model are available on this grid, so that no approximations of the COGARCH equations are necessary. The method is also applicable to irregularly spaced observations. A simulation study illustrates the quality of the MCMC estimates. Finally we fit the COGARCH(1,1) model to high-frequency data of the S&P500.

2000 MSC Subject Classifications: primary: 62M05, 62F15
secondary: 91B70, 91B84

JEL Classification: C11, C15, C22

Keywords: Bayesian inference, compound Poisson, continuous time GARCH process, Lévy process, volatility estimation

*Center for Mathematical Sciences, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany. Email: mueller@ma.tum.de.

1 Introduction

Since the seminal papers by Engle (1982) and Bollerslev (1986), many extensions of the ARCH and GARCH models have been proposed in the literature to cover different aspects of financial time series. An impressive overview over the ARCH and GARCH model family is given in Bollerslev (2008). With the increasing popularity of continuous-time models in financial econometrics - dating back to the fundamental work of Black and Scholes (1973) - also various continuous-time analogues of GARCH have been proposed over the past two decades. But what is meant exactly by a *continuous-time analogue* of GARCH? Is it a process that preserves the random recurrence volatility equation of GARCH(1,1)? Or a process that can be derived as a functional limit of GARCH? One which exhibits a GARCH-type behavior when sampled in discrete-time? Or one which is statistically equivalent to the discrete-time GARCH in the sense of Le Cam? Should such a process incorporate only one source of uncertainty as the original GARCH model? According to these and even more different ways of thinking of a continuous-time analogue of GARCH, the suggested models in the literature vary a lot in formulation and have different statistical properties. In order to understand where the COGARCH model considered in this paper stands with respect to other continuous-time GARCH models, we briefly summarize some important works dealing with continuous-time analogues of GARCH.

As one of the first dealing with this topic, Nelson (1990) derived a continuous-time limit of discrete GARCH(1,1) processes by aggregating the GARCH innovations. The corresponding limit process turned out to be a bivariate diffusion driven by two independent Brownian motions, so that the continuous-time limit incorporates two independent sources of uncertainty. Wang (2002) showed that the diffusion limit derived by Nelson is not asymptotically equivalent to the approximating GARCH sequence, in terms of Le Cam's deficiency distance (see Le Cam (1986)). One should note that it is possible to modify Nelson's approximation to obtain a limit process which is driven by a single Brownian motion only (see Corradi (2000)), but in that case the limiting volatility process is deterministic, which is an undesirable property for a model describing a price process. However, for this latter case, the sequence of discrete-time GARCH models and the continuous-time limit are equivalent, cf. Wang (2002). An extension to diffusion limits of a more general class of GARCH processes (called augmented GARCH) was obtained by Duan (1997).

Also Drost and Nijman (1993) considered temporal aggregation of volatility models. They showed that the common GARCH models of Bollerslev (1986) are not closed under temporal aggregation, which is due to the fact that semistrong ARMA models do not have this property, in contrast to weak ARMA models. Accordingly, Drost and Nijman (1993) investigated weak GARCH models, a class which is closed under temporal aggregation.

Drost and Werker (1996) introduced the corresponding class of continuous-time processes which exhibit a weak GARCH-type behavior at all discrete frequencies. They also show that the parameters of the discretized weak GARCH process correspond to certain parameters in the continuous-time weak GARCH process, so that estimation methods for the discrete-time model can be used for the continuous-time model. The class of continuous-time weak GARCH processes nests the diffusion limit by Nelson (1990), provided that it has finite fourth moment, and other models which are driven by two independent Lévy processes with finite fourth moment.

In order to avoid some limitations of the weak GARCH models Meddahi and Renault (2004) introduced the square-root stochastic autoregressive volatility (SR-SARV) models in discrete- and continuous-time, generalizing the weak GARCH models considered in Drost and Nijman (1993) and Drost and Werker (1996). The SR-SARV class takes its name from the closely related SR-SARV class discussed in Andersen (1994) and has some attractive advantages: for example, in contrast to weak GARCH models, SR-SARV allows for asymmetries, and fourth moments are not required to be finite. Moreover, the discrete-time SR-SARV models are still closed under temporal aggregation, and exact discretizations of continuous-time SR-SARV models are discrete-time SR-SARV models. For a discussion of exact discretizations of SR-SARV within the framework of irregularly spaced data see Meddahi, Renault, and Werker (2006). The continuous-time SR-SARV class is driven by a multivariate standard Brownian motion and nests several well-known models, e.g. the CEV (constant elasticity of variance) process introduced by Cox (1975), the GARCH diffusion by Nelson (1990), and even a special case of COGARCH (the COGARCH model where the driving Lévy process is a Brownian motion).

Closing our brief summary of continuous-time GARCH models, we mention three other papers where Brownian motions are used as driving processes. Aiming in option pricing applications, Kallsen and Taqqu (1998) developed a continuous-time GARCH model which is driven by a single Brownian motion. Sampled at integer times, this model follows a GARCH process. However, the volatility process is constant on intervals $[t, t+1)$ for $t \in \mathbb{N}_0$. Kallsen and Taqqu (1998) show that the model is arbitrage free and complete, and use it to derive pricing formulas for contingent claims such as European options. Also Kazmerchuk et al. (2005) use a continuous-time GARCH process which is driven by a single Brownian motion. Here the GARCH volatility process is combined with a price process defined by a stochastic delay differential equation. Lorenz (2006) got solutions of such stochastic delay differential equations as a weak limit of scaled GARCH$(p, 1)$ processes for $p \to \infty$. The continuous-time GARCH limit incorporates two independent Brownian motions, and a special case of this limiting model is Nelson's diffusion.

After this look into the history of continuous-time GARCH modelling, we now turn to the COGARCH model by Klüppelberg, Lindner, and Maller (2004) considered in this paper. The COGARCH was introduced in the spirit of (i) preserving the random recur-

rence type equation of the original GARCH model, (ii) incorporating only one (univariate) source of uncertainty, and (iii) using a general Lévy process (instead of the widely used Brownian motion) in order to be able to account for jumps. This characterization demarcates the COGARCH from all other continuous-time GARCH models mentioned above. Although the analysis for models including jumps is far less tractable, accounting for jumps seems to be important because jumps represent a significant source of non-diversifiable risk as discussed by Bollerslev et al. (2008). Meantime there is quite strong evidence for the existence of jumps in financial time series, cf. the recent studies by Barndorff-Nielsen and Shephard (2007) and Aït-Sahalia and Jacod (2009) among others.

Let us now briefly recall the definition of COGARCH. On a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ one is given a background driving Lévy process $L = (L_t)_{t\geq 0}$. Throughout it is assumed that $\mathbb{E}L_1 = 0$ and $\mathbb{E}L_1^2 = 1$. Given parameters (β, η, φ), with $\beta > 0$, $\eta > 0$, $\varphi \geq 0$, and an initial variance σ_0^2, the *integrated COGARCH(1,1) process* $G = (G_t)_{t\geq 0}$ and its *variance process* $\sigma^2 = (\sigma_t^2)_{t\geq 0}$ are defined by the two stochastic differential equations (SDEs)

$$dG_t = \sigma_t \, dL_t, \tag{1.1}$$

$$d\sigma_{t+}^2 = \beta \, dt - \eta \sigma_t^2 \, dt + \varphi \sigma_t^2 \, d[L, L]_t^{\mathfrak{d}}, \tag{1.2}$$

for $t > 0$ and with $G_0 = 0$. The process $[L, L]^{\mathfrak{d}}$ denotes the discontinuous part of the bracket process (i.e. the quadratic variation) of L, cf. Protter (2005), p.66. As an extension of this COGARCH(1,1) definition, Brockwell et al. (2006) introduced the COGARCH(p, q) models. In the present paper, however, we do not deal with these higher order models, and therefore usually just refer to COGARCH instead of COGARCH(1,1).

The similarity between the defining equations of GARCH and COGARCH is obvious. Nevertheless, a few most important questions around the COGARCH model arise immediately: Does COGARCH follow a common GARCH process when sampled in discrete time? Does it capture the *stylized facts* of financial time series? How can it be estimated?

Maller, Müller, and Szimayer (2008) proved that the COGARCH model occurs as a continuous-time limit of GARCH in a strong sense (in probability, in the Skorokhod metric). This result is no contradiction to Nelson, but due to the fact that Maller et al. (2008) used thinned, not aggregated innovations and a different scaling of the parameters. Although GARCH models can be used to approximate COGARCH, the discretely sampled COGARCH does not form an exact GARCH process. This has major implications for the estimation of COGARCH which we discuss later. Also Kallsen and Vesenmayer (2009) derived the COGARCH as a continuous-time limit of GARCH, however using a different approach than Maller et al. (2008). A recent paper by Buchmann and Müller (2009) addresses the question of statistical equivalence for the COGARCH, discussing therefore the same topic as Wang (2002) for Nelson's diffusion. The results are a bit ambiguous, but show again non-equivalence in a realistic framework.

As a byproduct of their strong convergence result, Maller et al. (2008) get a discrete-time GARCH formulation for irregularly spaced observations, with specific time-varying parameters. Such GARCH models for irregularly spaced data were also discussed in Ghysels and Jasiak (1998), Engle (2000), and Meddahi, Renault, and Werker (2006). Ghysels and Jasiak (1998) specify the total variance process as GARCH equation with time-varying parameters which mimic the functional forms derived by Drost and Werker (1996) for Nelson's GARCH diffusion approximation when the observations are equally spaced. Engle (2000) follows a more direct approach and assumes, in his simplest volatility model, that the variance per time unit follows a regular GARCH(1,1) equation with time-independent parameters. Meddahi et al. (2006) discuss the exact discretization of continuous-time stochastic volatility processes observed at irregularly spaced times (cf. the paragraph on SR-SARV models above), and compare their approach with those of Ghysels and Jasiak (1998) and Engle (2000).

The second of our questions above - whether COGARCH captures the *stylized facts* of financial time series - assesses the practical use of the COGARCH model. As other continuous-time models, COGARCH might, of course, be useful for option pricing applications, and it can easily cope with irregularly spaced or missing data. The GARCH model has been proven to be most useful for modelling financial time series, and, obviously, the feedback mechanism between mean and variance is preserved by COGARCH. A thorough theoretical investigation of many interesting properties of COGARCH, even in comparison to other continuous-time stochastic volatility models, can be found in Fasen et al. (2005). Here it is shown that the COGARCH indeed captures many of the stylized facts of financial times series. For instance, the COGARCH model, in general, exhibits regularly varying (heavy) tails, volatility jumps upwards, and clusters on high levels. More precisely, it can be shown that both the tail of the distribution of the stationary volatility and the tail of the distribution of $G(t)$ are Pareto-like under weak assumptions. It may be of interest to compare these properties of COGARCH with the corresponding properties of other continuous-time stochastic volatility models, for instance, the model by Barndorff-Nielsen and Shephard (2001). The Ornstein-Uhlenbeck process in this model has heavy tails only if the Lévy process is heavy tailed, and extreme value clusters occur only if the Lévy process has a regularly varying tail. For a detailed comparison of COGARCH to the Barndorff-Nielsen and Shephard model as well as other continuous-time models, e.g. the generalized Cox-Ingersoll-Ross model, we refer the reader to Fasen et al. (2005). However, we point out once again that the COGARCH differs from most other continuous-time stochastic volatility models by incorporating only one source of uncertainty. Moreover, the overall volatility jump sizes depend also on the actual variance, and, thus, vary over time, which is also a difference to some other models, e.g. the affine jump-diffusion model as in Duffie, Pan, and Singleton (2000).

Next we address the third question asked above, i.e. how the COGARCH model can be

fitted to a data set. A first approach to estimate the parameters of the COGARCH(1,1) model can be found in Haug, Klüppelberg, Lindner, and Zapp (2007). This paper provides a method of moment (MM) estimation procedure based on the theoretical and empirical moments and autocorrelation functions of the volatility process. For MM, there is no need to specify the driving Lévy process, however, it can only be applied to equally spaced observations and is, therefore, not applicable in many interesting situations. Moreover, it has been shown that, as for many other models, this moment method - although being consistent - is not very efficient.

Another estimation strategy was developed in Maller et al. (2008). Based on the discrete-time GARCH approximation with time-varying parameters mentioned above, a pseudo maximum likelihood (PML) procedure is derived, cf. Equations (3.2), (3.3) and (3.4) in Maller et al. (2008). In contrast to the method of moments, this method is applicable also to irregularly spaced observations. Although Maller et al. (2008) give an exact formula for the conditional variance based on the continuous-time model, in practice the volatility approximation derived in that paper has to be used. Hence, the results by Bollerslev and Wooldridge (1992) on consistency of the PML estimates (proven for semi-strong GARCH with a correct specification of the conditional mean and variance) cannot be applied, and the PML estimates are not consistent. This is in line with the results of Meddahi and Renault (2004) who considered temporally aggregated GARCH models. Table 1 therein gives strong evidence that QMLE is not consistent for weak GARCH coming even from an exact discretization of continuous-time processes. Nevertheless, simulation studies (see, for example, Table 2 in Maller et al. (2008)) have shown that PML (although being not consistent) is superior to MM (which is consistent) for not too large sample sizes, say up to about 50000 observations (superior here means that the method shows smaller mean squared errors for all parameters). Only for very large data sets, consisting of, say, 100000 observations, MM gets superior to PML. Altogether it seems that the approximation error in Maller et al. (2008) is minor compared to the inefficiency of MM, at least for not too large sample sizes. Moreover, since the PML estimates in Maller et al. (2008) (as well as the QML estimates considered in Meddahi and Renault (2004)) give reasonable approximations to the true values in simulation studies, such estimates still can serve as (quite good) starting values for other more reliable methods. That is what the PML method of Maller et al. (2008) will be used for in the present paper.

Since the method of moments turned out to be quite inefficient, although being consistent, and the (not consistent) PML method gives only slightly better estimates for realistic sample sizes, it is the goal of this paper to provide a more reliable estimation method for the COGARCH parameters. The estimation method presented here is based on a Markov chain Monte Carlo (MCMC) simulation and can also be applied to irregularly spaced observations from COGARCH. Whereas for the method of moments and the PML

method there is no need to specify the driving Lévy process, we have to do this in our Bayesian approach - aside from the specification of prior distributions. Moreover, we will assume in this paper that the driving Lévy process is a compound Poisson process, having in mind that each Lévy process (up to a possible Brownian part) can be approximated by a sequence of compound Poisson processes. For volatility modelling with COGARCH, this is not a severe restriction since we recall from Equation (1.2) that the volatility is affected only by the discontinuous part of the quadratic variation process of L. Of course, our simulation based method requires a significantly higher computation time than MM or PML. On the other hand, we also get density estimates, not only pure point estimates. In addition, as we will see, this effort is remunerated by much better estimation results than one gets from MM and PML. The MCMC algorithm presented here is based on discrete GARCH-type equations on a random grid which changes in each iteration of the MCMC sampler. Moreover, we use ideas developed by Geyer and Møller (1994) to sample from point processes. This methodology was also applied in Roberts, Papaspiliopoulos, and Dellaportas (2004) for the MCMC estimation of the model by Barndorff-Nielsen and Shephard (2001). For the COGARCH, the methodology has to be adapted, since the same driving Lévy process appears in both model equations.

Before we specify the MCMC algorithm in Section 2, let us now finally summarize some important theoretical properties of the COGARCH. They all can be found in Klüppelberg et al. (2004). As can be seen from Equation (1.1), the process G jumps at the same times as L does, and has jump sizes

$$\Delta G_t = \sigma_t \Delta L_t, \quad t \geq 0, \tag{1.3}$$

where ΔL_t denotes the jump size of L at time t. Conditions on the existence of moments and for the stationarity of the volatility process can be stated using an auxiliary process $(X_t)_{t\geq 0}$, defined by

$$X_t = \eta t - \sum_{0<s\leq t} \log[1 + \varphi(\Delta L_s)^2], \quad t \geq 0, \tag{1.4}$$

and its Laplace transform $\mathbb{E}e^{-sX_t} = e^{t\Psi(s)}$. The corresponding Laplace exponent is given by

$$\Psi(s) = -\eta s + \int_{\mathbb{R}} ((1 + \varphi x^2)^s - 1)\nu_L(dx), \quad s \geq 0. \tag{1.5}$$

The existence of a stationary distribution for the variance process σ^2 is implied by the existence of some $s > 0$ such that $\Psi(s) \leq 0$. Moreover, if $\Psi(s) < 0$ for some $s > 0$, then $\Psi(t) < 0$ for all $0 < t \leq s$. For arbitrary values of $s > 0$ it may be difficult to derive an explicit condition on η, φ and the parameters of the driving Lévy process, so that $\Psi(s) < 0$. However, for $s = 1$ Equation (1.5) writes as

$$\Psi(1) = -\eta + \varphi \int_{\mathbb{R}} x^2 \nu_L(dx). \tag{1.6}$$

If we assume that L is compound Poisson with intensity c and a jump distribution with finite second moment m_2, we immediately get $\Psi(1) = c\,\varphi\,m_2 - \eta$. If m_2 is known, it is, thus, sufficient to check

$$c\,\varphi\,m_2 - \eta \leq 0 \tag{1.7}$$

to guarantee the existence of a stationary distribution for the volatility process. Note that it may be possible for certain jump distributions to derive a corresponding condition for some $s < 1$, which would represent a weaker condition and therefore should be used then instead of condition (1.7). Finally, Corollary 4.4 in Klüppelberg et al. (2004) shows that the moment $\mathbb{E}(\sigma^{2k})$ $(k \in \mathbb{N})$ exists if and only if $\mathbb{E}L_1^{2k} < \infty$ and $\Psi(k) < 0$. If these conditions are satisfied for $k = 1, 2$, one gets

$$\mathbb{E}(\sigma_t^2) \;=\; \frac{\beta}{|\Psi(1)|} \qquad \text{and} \qquad \mathbb{E}(\sigma_t^4) \;=\; \frac{2\beta^2}{|\Psi(1)\Psi(2)|}. \tag{1.8}$$

With this theoretical background at hand, we are now ready to look into the MCMC procedure. The paper is organized as follows: In Section 2 we develop the MCMC sampler, choose prior distributions, and give some guidelines about the choice of the initial values. Section 3 assesses the performance of the algorithm in a simulation study and compares the quality of the posterior mean estimates to the quality of the PML estimates in Maller et al. (2008). In Section 4 we apply our MCMC algorithm to high-frequency data from the S&P 500 for the years 2005, 2006, and 2007. Section 5 finally contains a short summary.

2 Markov chain Monte Carlo algorithm

The fundamental idea of the MCMC algorithm is to estimate the jumps of the underlying compound Poisson process and to represent the COGARCH(1,1) model in such a form that, based on the jump times, it can be considered a *discrete GARCH model on the (random) grid of the jump times*. Therefore, the approach is different to the discrete-time GARCH formulation in Maller et al. (2008) which is defined on the grid of the *observation* times, not of the *jump* times.

Since we work with a compound Poisson process as driving Lévy process, the quadratic variation process just adds up the squared jumps of the compound Poisson process, so that Equation (1.2) can be written as

$$d\sigma_{t+}^2 = \beta\,dt - \eta\sigma_t^2\,dt + \varphi\sigma_t^2(\Delta L_t)^2. \tag{2.1}$$

Throughout we will assume that the jump distribution has finite second moment m_2 and no point mass at 0 (the latter condition is required since we cannot identify jumps of size 0), and denote the corresponding density function by $z(\cdot)$. The interarrival times between the jumps of L are exponentially distributed, a fact which also implies that almost surely no jumps occur at the observation times. Since the process G is the integrated COGARCH

process, log returns correspond to differences of G, so that L should also have negative jumps.

Let us now introduce some notations. We assume that we observe the process G on the interval $[0,T]$ at the observation times $0 < t_1 < \ldots < t_n =: T$. The set of observations is denoted by $\boldsymbol{G} := (G_{t_1}, \ldots, G_{t_n})$. The interval between the $(j-1)$th and the jth observation is denoted by I_j, i.e. $I_j := [t_{j-1}, t_j]$, and the change of G on I_j by $\Delta G_{t_j} := G_{t_j} - G_{t_{j-1}}$.

The estimated jump times of G and of L in $[0,T]$ are denoted by $0 < \tau_1 < \ldots < \tau_m < T$. Note that the index m reflects the estimated number of jumps. Therefore the value of m can vary over the iterations of the MCMC procedure. The estimated jump sizes of G at the jump times τ_i are denoted by g_i, i.e. $g_i := \Delta G_{\tau_i}$ for $i = 1, \ldots, m$. For notational convenience we set $g_0 := 0$. Further we define $\boldsymbol{g}$ to be the vector of all jump sizes g_i, i.e. $\boldsymbol{g} := (g_1, \ldots, g_m)$. The time elapsed between the $(i-1)$th and the ith estimated jump is $\Delta\tau_i := \tau_i - \tau_{i-1}$. Throughout this paper we will consider $\boldsymbol{\psi} := \{(\tau_i, g_i) | i = 1, \ldots, m\}$ as a marked point process. For notational convenience, the parameters β, η, φ, and σ_0^2 will often be collected in the vector $\boldsymbol{\theta} := (\beta, \eta, \varphi, \sigma_0^2)$, and the jump times will be combined to $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_m)$.

2.1 The volatilities at the jump times

Obviously, the COGARCH process G as well as the volatility process σ^2 does not jump between two *subsequent* jumps of the Lévy process, say during the time interval (τ_{i-1}, τ_i). Therefore, between two subsequent jumps, the process σ^2 follows an ordinary differential equation (ODE), as can be seen from Equation (2.1). Accordingly, we can apply the theory for solving ODEs to derive an exact solution of the volatility SDE (2.1) *on the grid of the jump times*. We just have to take into account, that the volatility process as defined in (1.2) and (2.1) is left-continuous. The exact solution is given as

$$\sigma^2_{\tau_i} = \frac{\beta}{\eta} + \left[\sigma^2_{\tau_{i-1}} + \varphi g^2_{i-1} - \frac{\beta}{\eta}\right] e^{-\eta\Delta\tau_i}. \tag{2.2}$$

Obviously, the volatility process has lower bound β/η, and, between the jumps, the volatility is exponentially decreasing with rate η. Now we rewrite Equation (2.2) by

$$\sigma^2_{\tau_i} = \frac{\beta}{\eta}(1 - e^{-\eta\Delta\tau_i}) + e^{-\eta\Delta\tau_i}\varphi g^2_{i-1} + e^{-\eta\Delta\tau_i}\sigma^2_{\tau_{i-1}} \tag{2.3}$$

to find a *discrete GARCH-type equation on the grid of the jump times*. In particular, we can deterministically compute an estimate of the volatility $\sigma^2_{\tau_i}$ at time τ_i when $\sigma_0^2, \beta, \eta, \varphi$, the jump sizes $g_1, \ldots, g_{i-1}$, and the jump times $\tau_1, \ldots, \tau_i$ are known or estimated. In this case we can obviously avoid the difficult update of the volatilities from their full conditionals. Our MCMC algorithm therefore estimates β, η, φ, σ_0^2, c, and the *marked point process* $\boldsymbol{\psi} = \{(\tau_i, g_i) | i = 1, \ldots, m\}$.

Note that Equations (2.2) and (2.3) give the spot variance at one time point given the jumps, not the conditional variance of a return on $[t, t+1]$ given information available at time t. In particular, Equations (2.2) and (2.3) do not play any role in the PML method which is used later to derive reasonable starting values for the MCMC chains. The PML method is based only on Equations (3.2), (3.3) and (3.4) in Maller et al. (2008). In particular, there are two important differences between (2.2) and (2.3) above and Equation (3.4) in Maller et al. (2008): First, (2.2) and (2.3) both live on the grid of the unknown (and later estimated) jump times, whereas (3.4) in Maller et al. (2008) lives on the grid of the observation times. Second, (2.2) and (2.3) are exact, whereas (3.4) in Maller et al. (2008) is an approximation. Equations (2.2) and (2.3) show how to calculate the conditional variance if samples of the jump times and sizes were available. In reality the exact jumps times and sizes will, of course, never be known, but here samples are provided by the MCMC algorithm. We emphasize once again, that between two subsequent jumps, the SDE (2.1) becomes a simple ordinary differential equation, due to the fact the the driving Lévy process is a compound Poisson process. This ODE can be solved using standard methods from analysis, giving Equation (2.2).

We point out that we do not estimate the jump sizes of the process L but those of G. The advantage of this strategy is that the restrictions which are imposed from the observations on the latent processes, are taken into account directly when the jumps of G are estimated. This is not possible for the estimation of the jumps of L, since according to Equation (1.3) the jumps of L are related to the jumps of G via the volatilities σ_t^2.

The general sampling scheme looks as follows (recall that $\boldsymbol{G} = (G_{t_1}, \ldots, G_{t_n})$ denotes the set of observations and $\boldsymbol{\theta} = (\beta, \eta, \varphi, \sigma_0^2)$ contains the COGARCH parameters):

1. Draw $\boldsymbol{\psi} = \{(\tau_i, g_i) | i = 1, \ldots, m\}$ (the jump times and sizes of G) from $f(\boldsymbol{\psi}|\boldsymbol{G}, \boldsymbol{\theta}, c)$.
2. Draw the COGARCH parameters $\beta, \eta, \varphi, \sigma_0^2$ from $f(\boldsymbol{\theta}|\boldsymbol{G}, \boldsymbol{\psi}, c)$.
3. Draw the intensity c of the compound Poisson process L from $f(c|\boldsymbol{G}, \boldsymbol{\psi}, \boldsymbol{\theta})$.

It will turn out that we have to update the latent process $\boldsymbol{\psi}$ and all other parameters by Metropolis-Hastings (MH) steps, since the exact computation of the corresponding full conditionals is analytically intractable. For MH-steps the old value will always be marked with a $^\bullet$, while the proposed value will be marked with a $^\circ$. We now consider the prior distributions and the MH-steps in detail.

2.2 Prior distributions

For the prior distributions we assume the following dependence structure:

$$f(\boldsymbol{\psi}, \varphi, \eta, c, \beta, \sigma_0^2) = f(\boldsymbol{\psi}|\beta, \eta, \varphi, \sigma_0^2, c) f(\sigma_0^2|\beta, \eta, \varphi, c) f(\varphi|\eta, c) f(\beta) f(\eta) f(c).$$

For c we take a $\Gamma(a, b)$ prior with some hyperparameters a and b and for β and η uniform priors on $(0, \infty)$. As found in simulation experiments, unstationarity of the volatility process can result in numeric instabilities of the MCMC procedure. Therefore we use unequality (1.7) and guarantee the stationarity of the volatility process by assuming that the parameter φ conditional on η and c is uniformly distributed on $[0, \eta/(cm_2)]$. The parameter σ_0^2 conditional on β, η, φ, and c has a Pareto$(x_{\min}, d)$-prior with left boundary $x_{\min} := \beta/\eta$ and parameter $d := \eta/(c\varphi m_2)$. The boundary $x_{\min}$ reflects the lower bound β/η for the stationary volatility process (cf. Klüppelberg et al. (2006), Proposition 2), whereas the parameter d is chosen so that the mean of the Pareto distribution matches the mean of the stationary distribution of the volatility process given in Equation (1.8).

Now we consider the prior distribution for the marked point process $\boldsymbol{\psi}$. This prior distribution is completely determined by the properties of the compound Poisson process L and by the implications given from the model equations. Since the jumps g_i are related to the jumps of the process L via the volatilities and therefore via the parameters β, η, φ, and σ_0^2, we assume prior dependence of $\boldsymbol{\psi}$ on the parameter vector $\boldsymbol{\theta}$ and on c. Since the jump times of G are the same as of L and the interarrival times for L are independent of each other, of the jump sizes, and of all model parameters, we assume that

$$f(\boldsymbol{\psi}|\boldsymbol{\theta}, c) = f(\boldsymbol{g}|\tau_1, \ldots, \tau_m, m, \boldsymbol{\theta}) f(\tau_1, \ldots, \tau_m|m, c) f(m|c). \tag{2.4}$$

That means, $f(\boldsymbol{\psi}|\boldsymbol{\theta}, c)$ decomposes into the distribution of the number of jumps, the distribution of the position of these jumps given the number, and the distribution of the jump sizes given the positions.

Now we write the first factor on the right-hand side of Equation (2.4) as the product

$$f(\boldsymbol{g}|\tau_1, \ldots, \tau_m, m, \boldsymbol{\theta}) \;=\; \prod_{i=1}^{m} f(g_i|g_{i-1}, \tau_1, \ldots, \tau_m, m, \boldsymbol{\theta}). \tag{2.5}$$

Following Equation (2.3) the distribution of g_i given g_{i-1}, $\tau_1, \ldots, \tau_m$, and $\boldsymbol{\theta}$ is the same as the distribution of g_i given σ_{τ_i}. Since the jumps of L have density $z(\cdot)$, we arrive, in accordance to Equation (1.3), at

$$f(\boldsymbol{g}|\tau_1, \ldots, \tau_m, m, \boldsymbol{\theta}) \;=\; \prod_{i=1}^{m} \frac{1}{\sigma_{\tau_i}} z\left(\frac{g_i}{\sigma_{\tau_i}}\right). \tag{2.6}$$

For the second factor on the right-hand side of Equation (2.4) we just refer to the basic property of Poisson processes, that, given a fixed number m of jumps in $[0, T]$ and the intensity c, the position of the jumps $\tau_1, \ldots, \tau_m$ is uniformly distributed on the set $\{0 < \tau_1 < \ldots < \tau_m < T\}$, i.e.

$$f(\tau_1, \ldots, \tau_m|m, c) \;=\; \frac{m!}{T^m} \mathbb{1}_{\{0<\tau_1<\ldots<\tau_m<T\}}. \tag{2.7}$$

Finally, the factor $f(m|c)$ in (2.4) represents the probability to have exactly m jumps in $[0,T]$ when the intensity c is known. From the properties of the compound Poisson process we know that

$$f(m|c) = e^{-cT}\frac{(cT)^m}{m!}. \tag{2.8}$$

Combining Equations (2.6), (2.7), and (2.8) we get

$$f(\boldsymbol{\psi}|\boldsymbol{\theta}, c) \;=\; e^{-cT}c^m \left[\prod_{i=1}^{m} \frac{1}{\sigma_{\tau_i}} z\left(\frac{g_i}{\sigma_{\tau_i}}\right)\right] \mathbb{1}_{\{0<\tau_1<\ldots<\tau_m<T\}}. \tag{2.9}$$

This is an exact joint density for m, $\boldsymbol{g}$ and $\boldsymbol{\tau}$ given $\boldsymbol{\theta}$ and c. However, one should always keep in mind that it contains the continuous components $\boldsymbol{g}$ and $\boldsymbol{\tau}$ as well as the discrete component m.

2.3 Sampling of the marked point process $\boldsymbol{\psi}$

The update of the marked point process $\boldsymbol{\psi}$ is central to the MCMC algorithm. Here the statistical information from the observed data is incorporated. Intuitively speaking, the hidden process $\boldsymbol{\psi}$ must be updated in such a way that it always fits together with the observations. In particular, the jump sizes g_i between two observation times have to be drawn in such a way that they sum up to the observed overall change of the process G in the corresponding observation interval. Hence, sampling of the marked point process $\boldsymbol{\psi}$ is quite involved. A general methodology for simulation procedures for spatial point processes was developed by Geyer and Møller (1994). However, their strategies are only partly applicable in our context. One main reason is that the observations $\boldsymbol{G}$ imply some restrictions on $\boldsymbol{\psi}$, so that for example a birth step as suggested in Geyer and Møller (1994) cannot be applied directly. However, like Geyer and Møller (1994) we also basically use two types of steps to draw from $\boldsymbol{\psi}$: the *displacement move* and the *birth-or-death move*. For the displacement moves, we distinguish further between a move, where the jump *sizes* of all jumps of an interval (t_{j-1}, t_j) are drawn, and a move, where the jump *times* of all jumps of such an interval are drawn. A birth move increases the number of elements in $\boldsymbol{\psi}$, and a death move decreases the number of elements in $\boldsymbol{\psi}$. In each iteration we choose one of these four moves with probability 1/4 each. More details are given in Sections 2.3.2 and 2.3.3. Let us now first investigate the full conditional distribution $f(\boldsymbol{\psi}|\boldsymbol{G}, \boldsymbol{\theta}, c)$.

2.3.1 The full conditional for $\boldsymbol{\psi}$

The observations collected in $\boldsymbol{G}$ contain important information about the process $\boldsymbol{\psi}$. Consider the interval I_j between the $(j-1)$th and the jth observation. The fact, that the jumps sizes $\{g_i|\tau_i \in I_j\}$ must add up exactly to ΔG_{t_j}, implies several restrictions on $\boldsymbol{\psi}$.

If $\Delta G_{t_j} = 0$, the process $\boldsymbol{\psi}$ does not contain any jumps in I_j. That is a direct consequence of our assumption that the jumps of L are almost surely unequal to zero.

Now, if $\boldsymbol{\psi}$ contains exactly one jump in I_j and $\Delta G_{t_j} \neq 0$, then obviously this jump must have size ΔG_{t_j}. If $\boldsymbol{\psi}$ contains k jumps in I_j with $k > 1$ and we know that $\Delta G_{t_j} \neq 0$, then only $k-1$ jumps can be chosen randomly, while the kth jump is determined by the other ones.

We will use these implications for choosing suitable proposal densities for the moves of $\boldsymbol{\psi}$. Whether the collection of jumps summarized in $\boldsymbol{\psi}$ is in accordance to the observation vector $\boldsymbol{G}$ in the sense, that the jumps in $\boldsymbol{\psi}$ imply the observation vector $\boldsymbol{G}$, can be expressed by the density $f(\boldsymbol{G}|\boldsymbol{\psi})$, which takes on only the values 0 or 1 and which therefore will also be denoted by the indicator

$$\mathbb{1}_{\{\boldsymbol{\psi} \text{ supports } \boldsymbol{G}\}} := f(\boldsymbol{G}|\boldsymbol{\psi}).$$

Now we can derive that

$$\begin{aligned} f(\boldsymbol{\psi}|\boldsymbol{G},\boldsymbol{\theta},c) &\propto f(\boldsymbol{G},\boldsymbol{\psi},\boldsymbol{\theta},c) &&= f(\boldsymbol{G}|\boldsymbol{\psi},\boldsymbol{\theta},c) f(\boldsymbol{\psi}|\boldsymbol{\theta},c) f(\boldsymbol{\theta},c) \\ &\propto f(\boldsymbol{G}|\boldsymbol{\psi}) f(\boldsymbol{\psi}|\boldsymbol{\theta},c) &&= \mathbb{1}_{\{\boldsymbol{\psi} \text{ supports } \boldsymbol{G}\}} f(\boldsymbol{\psi}|\boldsymbol{\theta},c). \end{aligned} \tag{2.10}$$

Equation (2.9) gives an explicit expression for $f(\boldsymbol{\psi}|\boldsymbol{\theta},c)$. Therefore we get

$$f(\boldsymbol{\psi}|\boldsymbol{G},\boldsymbol{\theta},c) \propto e^{-cT} c^m \left[\prod_{i=1}^{m} \frac{1}{\sigma_{\tau_i}} z\left(\frac{g_i}{\sigma_{\tau_i}}\right) \right] \mathbb{1}_{\{0<\tau_1<\ldots<\tau_m<T\}} \mathbb{1}_{\{\boldsymbol{\psi} \text{ supports } \boldsymbol{G}\}}. \tag{2.11}$$

Recall from (2.3) that the volatilities $\sigma^2_{\tau_i}$ depend on the parameters β, η, φ, and σ^2_0. Proportionality (2.11) is used for the evaluation of the acceptance probabilities for the following Metropolis-Hastings steps. Therefore it does not matter that the normalizing constant is missing.

2.3.2 Displacement Moves

For a *displacement-size-move* we choose one interval I_j randomly, with probability $(t_j - t_{j-1})/T$. Now let N_j be the current (estimated) number of jumps in this interval I_j. Let us consider the possible cases:

Of course, if $N_j = 0$, we cannot change size or position. For $N_j = 1$ we must not change the size of this only jump, because otherwise $\boldsymbol{\psi}$ would no longer support $\boldsymbol{G}$. Only in the case $N_j > 1$ we are allowed to change the jump sizes. As for all MH-steps, it is very important to choose a good proposal density $q(\boldsymbol{\psi}^\bullet, \boldsymbol{\psi}^\circ)$ to get appropriate acceptance rates. In particular, the proposal density should refer to the jump distribution of the driving compound Poisson process.

Example.

We assume just for the moment that the jumps of L are standard normally distributed. Let $g^{\bullet}_{j,1}, \ldots, g^{\bullet}_{j,N_j}$ be the current jump sizes in the interval I_j, and let $\sigma^2_{j,i}$ be the corresponding estimated volatilities. On the one hand, our proposal values $g^{\circ}_{j,1}, \ldots, g^{\circ}_{j,N_j}$ have to satisfy the equation $g^{\circ}_{j,1} + \ldots + g^{\circ}_{j,N_j} = \Delta G_{t_j}$, or, equivalently,

$$g^{\circ}_{j,N_j} = \Delta G_{t_j} - g^{\circ}_{j,1} - \ldots - g^{\circ}_{j,N_j-1}. \tag{2.12}$$

On the other hand we know that the jumps of G are normally distributed with mean 0 and variance $\sigma^2_{j,i}$. Therefore we suggest to draw the $N_j - 1$ jumps from an $(N_j - 1)$-variate normal distribution with mean vector 0 and covariance matrix $\Sigma = \mathrm{diag}(\tilde{\sigma}^2, \ldots, \tilde{\sigma}^2)$, where

$$\tilde{\sigma}^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} \sigma^2_{j,i},$$

so that $\tilde{\sigma}^2$ represents an average over the old estimated volatilities on the jump times in I_j. Taking condition (2.12) into account, this leads in the case $N_j = 2$ to sampling $g^{\circ}_{j,1}$ from $N(\Delta G_{t_j}/2, \tilde{\sigma}^2/2)$ and computing $g^{\circ}_{j,2} = \Delta G_{t_j} - g^{\circ}_{j,1}$. In the case $N_j = 3$ one has to draw $g^{\circ}_{j,1}$ and $g^{\circ}_{j,2}$ from the bivariate normal distribution

$$N\left(\begin{pmatrix} \Delta G_{t_j}/3 \\ \Delta G_{t_j}/3 \end{pmatrix}, \begin{pmatrix} 2\tilde{\sigma}^2/3 & -\tilde{\sigma}^2/3 \\ -\tilde{\sigma}^2/3 & 2\tilde{\sigma}^2/3 \end{pmatrix}\right)$$

and to compute $g^{\circ}_{j,3} = \Delta G_{t_j} - g^{\circ}_{j,1} - g^{\circ}_{j,2}$. For $N_j > 3$ the multivariate normal proposal densities can be derived analogously. Alternatively one can also choose only three jumps in I_j at random and apply a three-dimensional displacement-size-move to these jumps.

□

With regard to Equation (2.11) we note that it is usually very easy to propose new jump sizes so that $\boldsymbol{\psi}$ supports $\boldsymbol{G}$. Moreover, it is obvious that, for computing the acceptance probability, the factor $e^{-cT}c^m$ in (2.11) cancels out, since the number of elements in $\boldsymbol{\psi}$ remains unchanged under a displacement-size-move.

For the *displacement-times-move* we choose again one interval I_j randomly, with probability $(t_j - t_{j-1})/T$. Let $N_j > 0$ be the number of jumps in interval I_j (as mentioned, for $N_j = 0$ nothing happens). We draw the proposal values $\tau^{\circ}_{j,1}, \ldots, \tau^{\circ}_{j,N_j}$ uniformly on the set

$$\{(\tau_{j,1}, \ldots, \tau_{j,N_j}) \in I_j^{N_j} | \tau_{j,1} < \ldots < \tau_{j,N_j}\}.$$

Since we only draw the jump times and the jump sizes remain the same, the proposed values support $\boldsymbol{G}$. The computation of the acceptance probability is straightforward, we just note that the factor $e^{-cT}c^m$ in (2.11) cancels out again, since the number of elements in $\boldsymbol{\psi}$ remains unchanged.

2.3.3 Birth-Move and Death-Move

The aim of these types of move is to increase/decrease the number of jumps. Because of the restrictions implied by G, we use the following strategy.

For the birth-move, we first draw uniformly a time point $\tau^\circ \in [0,T]$ (i.e. we draw τ° from the density $p_{\text{create}}(\cdot) :\equiv 1/T$), search for the interval I_j such that $\tau^\circ \in I_j$, and check the current number N_j of jumps in I_j. If $N_j = 0$ we cannot add a jump in this interval. If $N_j > 0$ we add one jump at time τ° with size 0 to the current constellation in $\boldsymbol{\psi}$ and make a displacement-size-step in I_j in addition, since we do not allow for jumps of size 0.

The death-move is used to decrease the number of jumps. First we choose uniformly an index $i \in \{1,\dots,m\}$ (i.e. each index i is chosen with probability $p_{\text{remove}}(\cdot) := 1/m$) and then try to remove the jump (τ_i, g_i) from the current constellation of jumps in $\boldsymbol{\psi}$. For this we first search for the interval I_j such that $\tau_i \in I_j$, and check the current number N_j of jumps in I_j. If $N_j = 1$ we cannot remove this jump from $\boldsymbol{\psi}$ and therefore let the current constellation of jumps unchanged. If $N_j > 1$ we remove this jump from the current constellation in $\boldsymbol{\psi}$ and make a displacement-size-step in I_j in addition.

It may help to think of a complete birth as a three stage procedure: The decision to make a birth step (which happens with probability $p_{\text{birth}} := 1/4$), the creation of a new jump time using p_{create}, and a displacement-size-step with a proposal distribution q_{size}. Analogously, a complete death step is done by the following three actions: The decision to make a death step (which happens with probability $p_{\text{death}} := 1/4$), the removal of a jump using p_{remove}, and a displacement-size-step using again q_{size}.

The acceptance rate is influenced also by the proposal in the displacement-size-step, and for evaluating (2.11) one has to take into account that the number of jumps increased from m to $m+1$. Neglecting the indicator variables in Equation (2.11), the acceptance probability $\alpha = \min\{1, r\}$ for a birth-move is now determined by

$$\begin{aligned} r &= \frac{f(\boldsymbol{\psi}^\circ|\boldsymbol{G},\boldsymbol{\theta},c)}{f(\boldsymbol{\psi}^\bullet|\boldsymbol{G},\boldsymbol{\theta},c)} \frac{p_{\text{death}}}{p_{\text{birth}}} \frac{p_{\text{remove}}}{p_{\text{create}}} \frac{q_{\text{size}}(\boldsymbol{g}^\circ,\boldsymbol{g}^\bullet)}{q_{\text{size}}(\boldsymbol{g}^\bullet,\boldsymbol{g}^\circ)} \\ &= \frac{e^{-cT}c^{m+1}}{e^{-cT}c^m} \frac{f(\boldsymbol{g}^\circ|\boldsymbol{\tau}^\circ,m,\boldsymbol{\theta})}{f(\boldsymbol{g}^\bullet|\boldsymbol{\tau}^\bullet,m,\boldsymbol{\theta})} \frac{T}{(m+1)} \frac{q_{\text{size}}(\boldsymbol{g}^\circ,\boldsymbol{g}^\bullet)}{q_{\text{size}}(\boldsymbol{g}^\bullet,\boldsymbol{g}^\circ)} \\ &= \frac{cT}{m+1} \frac{f(\boldsymbol{g}^\circ|\boldsymbol{\tau}^\circ,m,\boldsymbol{\theta})}{f(\boldsymbol{g}^\bullet|\boldsymbol{\tau}^\bullet,m,\boldsymbol{\theta})} \frac{q_{\text{size}}(\boldsymbol{g}^\circ,\boldsymbol{g}^\bullet)}{q_{\text{size}}(\boldsymbol{g}^\bullet,\boldsymbol{g}^\circ)}. \end{aligned}$$

Obviously, the term $cT/(m+1)$ acts a correction factor: Since we expect cT jumps on $[0,T]$ it tends to reject a birth when $m+1 > cT$ (i.e. if we have already more jumps than expected) and tends to support a birth when $m+1 < cT$ (i.e. if we have less jumps than expected). Analogously, for the death step one gets the factor $m/(cT)$ which tends to reject a death when $m < cT$ and tends to support a death when $m > cT$.

2.4 Sampling of $\boldsymbol{\theta}$ and c

For deriving the full conditional $f(\boldsymbol{\theta}|\boldsymbol{G}, \boldsymbol{\psi}, c)$ we note that $\boldsymbol{\psi}$ contains all knowledge about $\boldsymbol{G}$. Therefore we know that

$$f(\boldsymbol{\theta}|\boldsymbol{G}, \boldsymbol{\psi}, c) \;=\; f(\boldsymbol{\theta}|\boldsymbol{\psi}, c) \;\propto\; f(\boldsymbol{\psi}|\boldsymbol{\theta}, c) f(\boldsymbol{\theta}|c) f(c) \;\propto\; f(\boldsymbol{g}|\tau_1, \ldots, \tau_m, m, \boldsymbol{\theta}) f(\boldsymbol{\theta}|c).$$

The latter proportionality can be seen from Equation (2.4).

It is quite difficult to find proposal densities for the MH step which lead to good acceptance rates. We suggest to use Equation (2.5) to get maximum likelihood estimates for the components of $\boldsymbol{\theta}$, which maximize $f(\boldsymbol{\psi}|\boldsymbol{\theta}, c)$. Recall that β, η, φ, and σ_0^2 appear in (2.9) through the volatilities $\sigma_{\tau_i}^2$, which we can compute using Equation (2.3). For all parameters contained in $\boldsymbol{\theta}$ we choose normal proposal distributions (truncated to the support of the parameter), where the mean is the corresponding ML-estimate, and the variance is determined by the behavior of the likelihood $f(\boldsymbol{\psi}|\boldsymbol{\theta}, c)$ in the closer area around the ML-estimate. A careful choice of the variance is important to achieve good acceptance rates.

For the intensity c we use again a (truncated) normal proposal. Taking the $\Gamma(a, b)$ prior for c and the conditional priors for σ_0^2 and φ into account, we get from simple calculations

$$\begin{aligned} f(c|\boldsymbol{G}, \boldsymbol{\psi}, \boldsymbol{\theta}) \;&\propto\; f(\boldsymbol{G}, \boldsymbol{\psi}, \boldsymbol{\theta}, c) \;\propto\; f(\boldsymbol{G}|\boldsymbol{\psi}) f(\boldsymbol{\psi}|\boldsymbol{\theta}, c) f(\boldsymbol{\theta}|c) f(c) \\ &\propto\; c^{m+a-1}\, e^{-c(T+b)} \left(\beta\eta^{-1}\sigma_0^{-2}\right)^{\eta/(c\varphi m_2)} \mathbb{1}_{(0,\eta/(\varphi m_2))}(c). \end{aligned}$$

Therefore the calculation of the acceptance probability is straightforward.

2.5 Starting values, update strategy, and volatility estimates

A sensible choice of the starting values is very important for a stable run of the MCMC sampler. Also the burn-in period can be shortened dramatically, when the initial constellation of the parameters is chosen in a sensible way. A simple way to get reasonable starting values for β, η, and φ is to apply the PML method by Maller et al. (2008) or, if the observations are equally spaced, the method of moments by Haug et al. (2007), and to use the corresponding estimates as starting values of the MCMC sampler. Alternatively, one can often guess rough estimates of the mean and variance of the volatility process and then choose sensible starting values using Equations (1.8) and (2.2). One should also keep in mind, that the parameter c represents not only the jump intensity of the driving Lévy process, but also plays a role in computing the mean and the variance of the volatility process $(\sigma_t^2)_{t\geq 0}$.

Another important tool to optimize the output of the MCMC sampler is the choice of a good update strategy. Since in each iteration only one jump can be added or removed, it is reasonable to update the process $\boldsymbol{\psi}$ more often than the parameters. In the following

simulation study and application, we, therefore, update the jumps times and sizes, the number of jumps, and the parameter c in each iteration, while the parameters β, η, φ and σ_0^2 are updated only each 50th iteration.

Usually, one is more interested in volatility estimates than in bare parameter estimates. In each iteration of the MCMC algorithm one can easily compute estimates for the volatility process at time t^* using the estimate for $\sigma_{\tau^*}^2$, where τ^* denotes the time of the last estimated jump before t^*. Noting that no jump occurs between time τ^* and t^* one can compute the volatility at time t^* analogously to Equation (2.3) by

$$\sigma_{t^*}^2 = \frac{\beta}{\eta}(1 - e^{-\eta(t^*-\tau^*)}) + e^{-\eta(t^*-\tau^*)}\varphi g_{\tau^*}^2 + e^{-\eta(t^*-\tau^*)}\sigma_{\tau^*}^2,$$

where $g_{\tau^*}^2$ is the estimated jump of the process G at time τ^*.

3 Simulation study

In this section we assess the quality of the MCMC estimates. We setup a simulation study, which is similar to the framework of the application in Section 4. In addition, we compare the output of the MCMC procedure with the PML method of Maller et al. (2008). In that paper it has been shown already that the PML method is usually superior to the method of moments (MM) by Haug et al. (2007). Moreover, since we deal with irregularly spaced observations MM cannot cope with, we cannot compare our method directly with MM unless restricting ourselves to the special case of equally spaced observations. To justify our estimates in Section 4, we prefer, however, to conduct a simulation study with irregularly spaced observations, while omitting another comparison with the MM method.

We start with the simulation of 50 data sets containing different realizations of the same COGARCH(1,1) process, i.e. with the same parameters and the same jump distribution and intensity of the driving compound Poisson process. This compound Poisson process is assumed to have normally distributed jumps with mean 0 and variance $m_2 = 1/20000$. We point out again, that, also under this assumption, the resulting distributions of both $G(t)$ and the stationary distribution of the volatility process has Pareto-like, i.e. heavy tails. Each of the 50 data sets contains data for $t \in [0, 1]$. To copy the framework of our application in Section 4, we observe the process at 19500 unequally spaced observation times (this corresponds to 78 observations per day over 250 trading days). We copy also the behavior of the business time scale from Section 4 for the year 2007 (cf. Table 3; for a detailed explanation of the used business time scale see Section 4). Starting from a given basic time unit u, the distance of some observations is $6.2170\,u$, whereas sometimes the process is observed again after $0.6174\,u$. Therefore, the interobservation times can differ by a factor more than 10. The simulation parameters are, for similarity with Table 4, $\beta = 0.001$, $\eta = 0.2$, $\varphi = 0.1$, $\sigma_0^2 = 0.0125$, and $c = 24\,000$.

	β		η		φ		σ_0^2	c
True	1.0000e-3		0.2000		0.1000		0.0125	24000
	PML	MCMC	PML	MCMC	PML	MCMC	MCMC	MCMC
Bias	0.0764e-3	0.0317e-3	-0.0052	0.0019	-0.0071	-0.0012	0.0002	-38.22
RMSE	0.1723e-3	0.0972e-3	0.0186	0.0123	0.0098	0.0056	0.0004	168.64

Table 1: Bias and RMSE for posterior mean estimates of β, η, φ, σ_0^2, and c, based on 50 simulated data sets. Bias and RMSE for PML estimates of β, η, and φ for comparison.

Following our own recommendation in Section 2.5, we use PML estimates of β, η, and φ as starting values for the MCMC chain to reduce the burn-in period. Based on these estimates and together with a starting value for c, an initial value for the parameter σ_0^2 is then computed as the mean of the volatility process, according to Equation (1.8). First inspections of the trace plots of our MCMC sample show that, even for quite bad starting values for c (e.g. $c = 20\,000$ or $c = 30\,000$), the chains converge within about 150 000 iterations. We emphasize once again that only the jump times and sizes, the number of jumps, and the parameter c are updated in each iteration, whereas the other four parameters are updated only each 50th iteration. Therefore, 150 000 iterations correspond to only 3 000 updates of β, η, φ, and σ_0. According to these observations, we use the first 200 000 iterations for burn-in, and then run the MCMC algorithm for additional 500 000 iterations, so that we have 10 000 samples in total to estimate the posterior distribution. The acceptance rates for the displacement and birth- and death-moves as well as for β, φ and c were around 80%, and for η and σ_0^2 at about 60%.

Table 1 shows the bias and RMSE of the posterior mean (MCMC) estimates across the 50 simulated data sets, for the parameters β, η, φ, σ_0^2, and c. For comparison, we add the corresponding values for the PML estimates of β, η, and φ. As to the bias, the posterior mean estimates for η, φ, σ_0^2, and c are quite satisfying, only the parameter β is slightly overestimated on average. Comparing MCMC and PML, we state, that for the parameter β, the PML method tends to overestimate β by more than 7%, whereas the MCMC method overestimates it just by about 3% on average. In addition, MCMC reduces the RMSE of the PML method by more than 40%. Also for η and φ, one gets the more unbiased estimates by the MCMC method. The corresponding RMSEs are again reduced by MCMC by around 30 to 40%.

Summarizing these results, the MCMC approach leads to less biased and more efficient estimates than the PML method. This improvement in the quality of the estimates is, as expected, also due to a much higher computation time for MCMC, which is, for each of the 50 data sets, about 10 hours on an Intel Core 2.66 GHz processor. In general, the computation time depends, of course, heavily on the number of observations. By comparison, the PML and the MM method have usually a computing time of a few seconds

only, however, they do not provide any information on the distribution of the parameters, and the corresponding estimates are, furthermore, not as exact as the MCMC estimates.

As already mentioned, we used the PML estimates of β, η, and φ as starting values for the MCMC chain to reduce the burn-in period. Based on this simulation study, we generally recommend for practical purposes to apply first the PML method to get good starting values. The MCMC method can than be used to make these estimates 'more exact' and to get the corresponding posterior distributions in addition.

Of course, it is not really surprising, that the MCMC method outperforms the PML method (which is, in turn, usually superior to the MM method). It has been shown also in other settings, that MCMC outperforms competing methods (for example, see Jacquier, Polson, and Rossi (1994)). However, it cannot be stated in general that each MCMC approach outperforms classical estimation methods, but it has to be checked for each new algorithm again.

4 Application to S&P500 index data

The goal of this section is to illustrate how to apply the COGARCH(1,1) model to financial data by modelling the time-varying volatility of the S&P500. The S&P500 index is one of the most widely followed measures of the US equities market, being made up of 500 of the largest stocks in the market (representing around three quarters of all US equities). In the literature there exist various papers addressing the estimation of the volatility in the S&P500. The paper by Hsieh (1991) models the S&P500 using high-frequency data. Hsieh uses 15 minute observation intervals and finds support for an EGARCH (4,4) model. Later, Andersen, Benzoni, and Lund (2002), and Eraker, Johannes, and Polson (2003), used models admitting stochastic volatility together with jumps both in returns and in volatility, for the S&P500. Lundblad (2007) estimates daily and monthly volatility of returns on the US market using data stretching over a period over more than 150 years. While he finds some evidence of a relationship between volatility and return, he is agnostic as to which of the models he uses – GARCH (1,1), TARCH (1,1), QGARCH (1,1) or EGARCH (1,1) – best fits the data.

Our analysis will use observations taken at five-minute intervals. This high-frequency structure of the data can easily be captured by the continuous-time framework of the COGARCH. Moreover, the use of COGARCH enables the analysis of irregularly spaced data, which we will be faced with after a transformation of the physical time scale to a virtual business time scale. After a brief description of the raw data on the S&P500 index, we discuss data cleaning and pre-processing. To account for intraday volatility patterns, which cannot be captured by the COGARCH model, we transform the physical time to a virtual business time using the PML method. Simultaneously, we get good starting values for our MCMC procedure. Finally we report on the results of fitting the

COGARCH model to the S&P500 by MCMC, separately for the years 2005, 2006, and 2007.

4.1 Data description and pre-processing

Based on tick-by-tick index data from the NYSE of the years 2005, 2006, and 2007, we compute 5-minutes log-returns. After that, we have, in total, 19617, 19460, and 19458 observations, for these three years respectively. Note that the number of 5-minutes log-returns per year depends not only on the number of trading days, but also on irregularities as memorials, holidays (e.g. on Christmas Eve the NYSE closes usually at 1pm), and special events (for a complete list of trading hours exceptions at NYSE since 1885 see www.nyse.com/pdfs/closings.pdf). Moreover, a few observations are missing, usually since some data is obviously reported erroneous, and has to be removed from the data set. However, taking advantage of our continuous-time approach, we do not use interpolation to fill in the missing values, but instead take the time difference to the previous available observation into account.

Before applying the COGARCH model to the data, one has to think carefully about which features of the data are to be captured by the COGARCH, and which are not. The COGARCH is designed to describe the behavior of the volatility in the data. Moreover, for long data sets, it describes the *fine* structure of the volatility, however, long-run regime switches, i.e. the *raw* structure of the volatility, cannot be detected, since we assume the COGARCH parameters β, η, and φ to be constant over time. Therefore, we must check first whether the data indeed shows a stationary volatility pattern over the whole time frame. In our data, this is definitely not the case, c.f. Figure 1, first and second row, for the data in 2007. Therefore we first pre-process the data by estimating local trends and local volatility weights. However, we must be careful not to destroy the fine structure of volatility which we want to describe by the COGARCH model. Therefore we aim at a standardization procedure which uses trends and volatility weights over longer periods such as one month, which corresponds to around 1600 observations in our setup.

In the following we denote the observed log-returns by y_i, $i = 1, \ldots, T$, whereas D denotes the number of trading days in the year under consideration (e.g. in 2007 we have $T = 19458$ and $D = 251$). Next we have to introduce a few functions, to be able to cover all aspects of the data within our subsequent formulas. First, let $d : \{1, \ldots, T\} \to \{1, \ldots, D\}$, $i \mapsto d(i)$, denote a function which returns the trading day for observation i, $A : \{1, \ldots, D\} \to \mathbb{N}$, $d \mapsto A(d)$, denote a function which returns the number of available observations on trading day d, and $M : \{1, \ldots, D\} \to \mathbb{N}$, $d \mapsto M(d)$, a function which returns the number of missing observations on trading day d. Since we take 5 minute data between 9:30 am and 4:00 pm, we usually have $A(d) = 78$ and $M(d) = 0$ on a regular trading day. Note that $M(\cdot)$ can also be 0, when the NYSE opened later or closed

earlier, since $M(\cdot)$ counts only missing values when trading really took place at that time. Although only very few observations are missing overall (less than 0.2%), we introduce, to act very precisely, the function $I : \{1, \ldots, T\} \to \mathbb{N}$, $i \mapsto I(i)$, which returns the number of 5-minute intervals elapsed before observation i. Usually $I(i) = 1$, and if, e.g. one observation is missing, $I(i) = 2$, and so on. Later, we will also need the more precise functions $I_k : \{1, \ldots, T\} \to \mathbb{N}$, $i \mapsto I_k(i)$, for $k \in \{9, 10, \ldots, 15\}$, which specify how many 5-minute intervals during trading hour k elapsed before observation i. E.g., if we have an observation at 9:55 am and the value at 10:00 am was deleted, so that the next observation i is from 10:05 am, we have $I(i) = 2$, $I_9(i) = I_{10}(i) = 1$, and $I_k(i) = 0$ for $k = 11, \ldots, 15$.

We now assume that the log-returns follow the model

$$y_i = m_{d(i)} + v_{d(i)} x_i, \qquad i = 1, \ldots, T,$$

where $m_{d(i)}$ represents a *local trend*, $v_{d(i)}$ a *local volatility weight*, and the x_i are detrended and locally reweighted log-returns. This approach takes irregularities of the stock prices

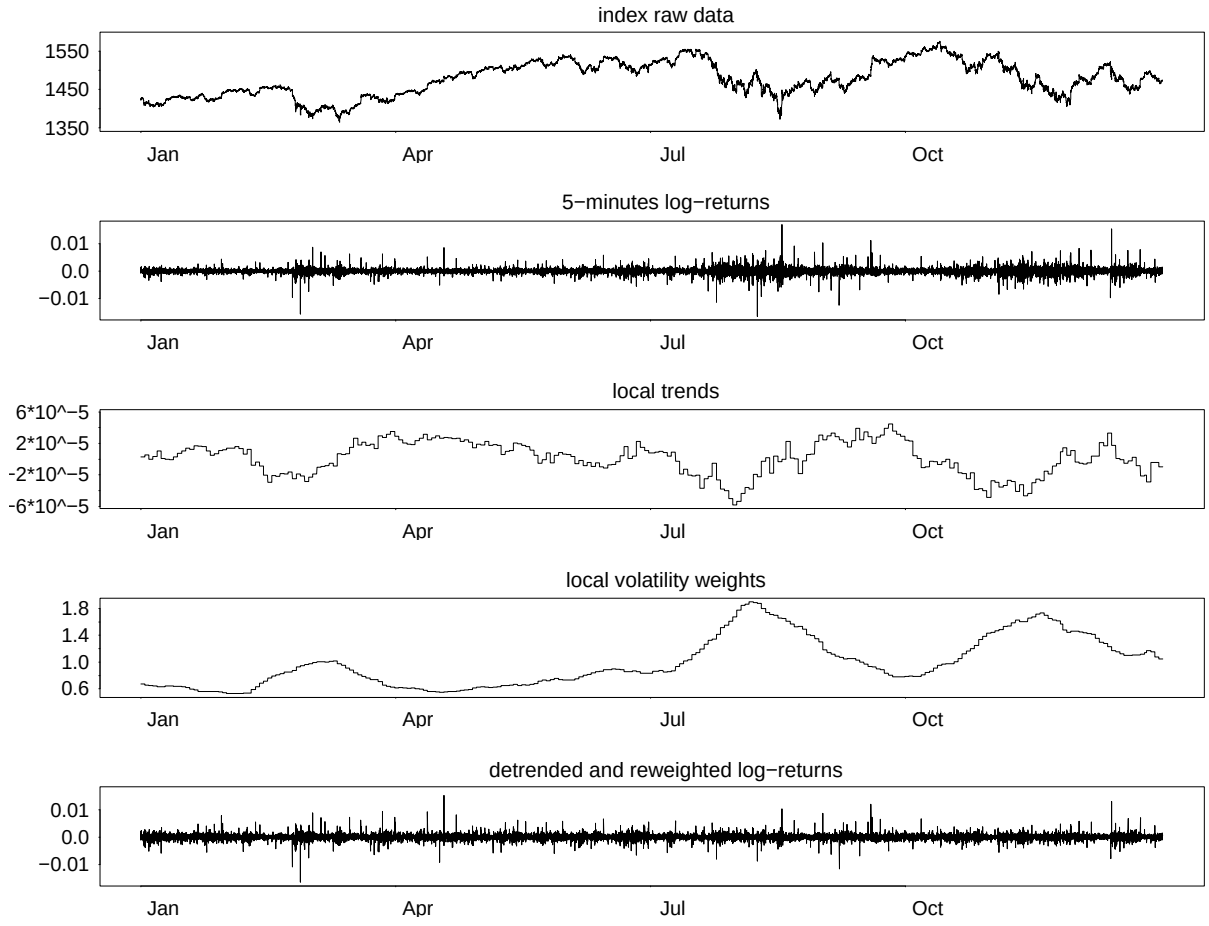

Figure 1: 5 minutely observations from S&P500 index data and corresponding log-returns, local trends, local volatility weights, and detrended and reweighted log-returns in 2007.

and index data into account, which cannot be captured by the COGARCH model. Both local trends and volatility weights are assumed to be constant over trading days, and are estimated as follows. Using a fixed $N \in \mathbb{N}$, the local trends $m_{d(i)}$ are estimated as moving averages over $2N+1$ trading days:

$$\widehat{m}_{d(i)} = \Big[\sum_{d=d(i)-N}^{d(i)+N}(A(d)+M(d))\Big]^{-1} \sum_{d=d(i)-N}^{d(i)+N} \sum_{\{j|d(j)=d\}} y_j, \qquad N < d(i) \leq D-N.$$

For the cases $d(i) \leq N$ and $d(i) > D-N$ we can again employ this formula by using data from the previous and following year, respectively, or else we can shrink the time frame of estimation. Note that the addition $A(d)+M(d)$ makes sense, since log-returns are additive.

Similarly, we estimate the volatility weights, for some $V \in \mathbb{N}$, by computing preliminary weights

$$\widehat{v}^*_{d(i)} = \Big[\sum_{d=d(i)-V}^{d(i)+V}(A(d)+M(d))\Big]^{-1} \sum_{d=d(i)-V}^{d(i)+V} \sum_{\{j|d(j)=d\}} |y_j - \widehat{m}_{d(j)}|, \quad V < d(i) \leq D-V,$$

and then by reweighting these according to

$$\widehat{v}_{d(i)} = \frac{\sum_{j=1}^{T} |y_j - \widehat{m}_{d(j)}|/\widehat{v}^*_{d(j)}}{\sum_{j=1}^{T} |y_j - \widehat{m}_{d(j)}|}\, \widehat{v}^*_{d(i)}.$$

This implies that for $\widehat{x}_i := (y_i - \widehat{m}_{d(i)})/\widehat{v}_{d(i)}$ we have $\sum_{i=1}^{T} |\widehat{x}_i| = \sum_{i=1}^{T} |y_i - \widehat{m}_{d(i)}|$, so that the magnitude of the values is preserved.

In our analysis, we set $N = V = 10$, so that we use a time frame of 21 trading days to determine the local trend and volatility weight. This seems to be a reasonable choice since we usually get larger standard errors for the COGARCH parameter estimates for very large or very small values of N and V. In Figure 1, as an example, the third and fourth row shows the estimated local trend and the estimated local volatility weight for the year 2007. The fifth row shows the detrended and reweighted log-returns. Although the S&P500 exhibits a different volatility pattern for the years 2005 and 2006, we omit the corresponding figures for these two years, since the cleaning and pre-processing procedure is the same as for 2007. We emphasize once more that both $m_{d(i)}$ and $v_{d(i)}$ depend only on the $d(i)$, so that all observations of the same day are reweighted by the same weight. This way we do not lose information about the dependence of the volatility on the exact trading time during the day. This dependence is accounted for in the following subsection.

4.2 Accounting for trading time by time transformations

To take the possible impact of trading time on the volatility into account, we first conducted a standard regression analysis for the squared log-returns to check for explanatory variables having an influence on the volatility. We used indicator variables for

the month January, for all weekdays from Monday to Friday, and for all trading hours $(9, 10, 11, \ldots, 15)$. For reasons of identifiability we had to remove one weekday and one trading hour (we chose Friday and hour 15), so that this sums up to a collection of 11 variables. The four most significant indicators always turned out to be hours 9, 12, 13, and 10, in this order. All these had p–values of less than 1%.

An easy way to account for these explanatory variables is to apply the COGARCH to a fictive business time axis. That means that we do not insert the explanatory variables into the COGARCH model itself, but first rescale the physical time axis to a business time scale. We do this using the PML method, since this way we get simultaneously good starting values for the MCMC estimation. The idea is to replace Δt_i by

$$\Delta t_i^* := \Delta t_i I(i) + h_9 I_9(i) + h_{10} I_{10}(i) + h_{12} I_{12}(i) + h_{13} I_{13}(i),$$

where h_9, h_{10}, h_{12}, and h_{13} are unknown parameters. Since we account for missing values within the functions I and I_k, respectively, Δt_i does not depend on i in our setup and serves just as the basic time unit. To get estimates on an annual basis, we choose Δt_i as $1/(\sum I(j))$, since $\sum I(j) = T + \sum M(d) = \sum A(d) + M(d)$.

Of particular interest are the quantities $f_k := (h_k + \Delta t_i)/\Delta t_i$ for $k = 9, 10, 12, 13$. These report the factors by which the basic time unit has to be rescaled during a certain trading hour to get business time. We note that there is no identifiability problem for h_9, h_{10}, h_{12}, and h_{13}, since in our model business time is the same as physical time between 11am and 12pm and after 2pm.

Table 2 and 3 contain the results of the preliminary analysis using the PML method. Table 2 shows the parameters h_9, h_{10}, h_{12}, and h_{13} which have to used for the transformation of physical time to business time. Furthermore, it contains preliminary estimates of the parameters β, η, and φ. Table 3 shows the factors which have to applied to the physical time axis to get business time. For example, the value 4.5319 of the estimate for f_9 in 2005 means that during this year business time was running at around 4.5 times faster than physical time between 9:30am and 10:00am, which reflects the high activity in the market after the opening of the exchange. Between 10am and 11am the activity decreased, but is still higher than on average. In general, during lunch time, between 12pm an 2pm, business time runs slower than physical time, usually the activity is only 60% to 75% of the average observed between 11am and 12pm as well as after 2pm.

4.3 Application of the MCMC method

To improve our estimates from the PML analysis and to derive distributions for the parameters, we now use the business time scale and employ our MCMC method to fit the COGARCH model to the data. We assume normally distributed jumps of the underlying compound Poisson process, emphasizing once more, that for this choice both the

	$h_9 \cdot 10^4$	$h_{10} \cdot 10^4$	$h_{12} \cdot 10^4$	$h_{13} \cdot 10^4$
2005	1.8003 (0.0988)	0.1923 (0.0198)	-0.1979 (0.0088)	-0.1504 (0.0106)
2006	2.1767 (0.1170)	0.2547 (0.0228)	-0.1667 (0.0103)	-0.1346 (0.0112)
2007	2.6795 (0.1377)	0.1301 (0.0193)	-0.1965 (0.0093)	-0.1766 (0.0099)

	β	η	φ
2005	0.00071658 (0.00002205)	0.2018 (0.0046)	0.1027 (0.0038)
2006	0.00069020 (0.00002447)	0.2039 (0.0054)	0.1009 (0.0049)
2007	0.00145146 (0.00008295)	0.2017 (0.0071)	0.0999 (0.0053)

Table 2: *PML estimates and corresponding approximated standard errors of the parameters h_9, h_{10}, h_{12}, h_{13}, β, η, and φ. For notational convenience, the estimates and standard errors of h_9, h_{10}, h_{12}, h_{13} have been multiplied by 10000.*

	f_9	f_{10}	f_{12}	f_{13}
2005	4.5319	1.3772	0.6117	0.7049
2006	5.2458	1.4967	0.6748	0.7374
2007	6.2170	1.2533	0.6174	0.6561

Table 3: *Factors that have to be applied to get from physical to business time scale, for trading hours 9, 10, 12, and 13, respectively.*

stationary distribution of the volatility process and the distribution of $G(t)$ have Pareto-like (heavy) tails (cf. Section 1). Also note that, in business time, our observations are irregularly spaced. We run the MCMC sampler for 700 000 iterations, as in the simulation study. Again the parameters β, η, φ, and σ_0^2 are updated only at each 50th iteration, so that, after removing the burn-in period, we have 10 000 samples to estimate the posterior density.

As starting values for β, η and φ we use the estimates obtained from the PML method. Although the variance m_2 of the jumps in the compound Poisson process can be chosen quite arbitrarily, we set $m_2 = 1/\sum I(j)$, i.e. equal to the number of observed plus missing data, in order to get the estimates nearly on an annual basis. Sometimes we report the square root of the variance estimates to give a volatility rather than a variance estimate. We saved the values of $\sigma_{\text{low}} := (\beta/\eta)^{1/2}$ and $\sigma_{\text{mean}} := (\beta/|\eta - c\varphi m_2|)^{1/2}$ in each iteration, so that we get also detailed information about the lower bound and the mean of the process (σ_t), cf. Equations (1.8) and (2.2). However, we have chosen the basic time unit in such a way that, in physical time, one year corresponds to the time interval $[0, 1]$. Since we apply the COGARCH to the business time scale, we have to rescale the estimates of

2005	β	η	φ	c	$\sigma_{\text{low,an}}$	$\sigma_{\text{mean,an}}$
post.mean	0.00068223	0.2024	0.1022	23017.37	0.0643	0.1001
cred.int.: 5% quant.	0.00065019	0.1969	0.0971	22291.81	0.0627	0.0932
95% quant.	0.00071531	0.2081	0.1069	23652.43	0.0656	0.1091
PML estimate	0.00071658	0.2018	0.1027	-	0.0660	0.0941
2006	β	η	φ	c	$\sigma_{\text{low,an}}$	$\sigma_{\text{mean,an}}$
post.mean	0.00066078	0.2041	0.1001	22825.39	0.0653	0.1005
cred.int.: 5% quant.	0.00062501	0.1972	0.0931	22167.63	0.0632	0.0926
95% quant.	0.00069723	0.2113	0.1066	23491.20	0.0669	0.1110
PML estimate	0.00069020	0.2039	0.1009	-	0.0667	0.0938
2007	β	η	φ	c	$\sigma_{\text{low,an}}$	$\sigma_{\text{mean,an}}$
post.mean	0.00131737	0.2025	0.1002	25101.72	0.0932	0.1540
cred.int.: 5% quant.	0.00119583	0.1929	0.0929	24378.09	0.0888	0.1404
95% quant.	0.00144020	0.2126	0.1074	25884.41	0.0971	0.1723
PML estimate	0.00145146	0.2017	0.0999	-	0.0977	0.1376

Table 4: Posterior mean estimates and 90% credible intervals for the parameters β, η, φ, and c, and for lower bound $\sigma_{\text{low,an}}$ and mean $\sigma_{\text{mean,an}}$ of the annualized volatility process, for years 2005, 2006, and 2007. PML estimates for comparison.

σ_{low} and σ_{mean}, in order to have them on an annual basis. The resulting estimates are called $\sigma_{\text{low,an}}$ and $\sigma_{\text{mean,an}}$ and can be derived from the original values by multiplication with $\sqrt{B}$, where B is the business time elapsed after one year. Of course, B depends on the year under consideration, and can be computed from Table 3.

4.4 Results and interpretation

Table 4 summarizes the results for the marginal posterior distributions of the parameters β, η, φ and c, as well as for the lower bound $\sigma_{\text{low,an}}$ and the long-run volatility $\sigma_{\text{mean,an}}$. For comparison we report also the PML estimates again. It is interesting, that the posterior mean estimates of η and φ are quite the same for all three years. As can be seen easily from Equation (2.2), the parameter η measures the speed of decline of a volatility burst, whereas the parameter φ measures the magnitude of such a volatility burst arising after the arrival of new information represented by a jump of the Lévy process. Therefore we conclude, that the *mechanism*, how new information influences volatility, i.e. how it changes the actual volatility and how fast this new information is absorbed by the mar-

ket, remained the same over these three years. What differs is the estimated frequency of significant news, represented by the parameter c, and the overall level of volatility, represented by the parameter β. Therefore, c carries relevant information for the volatility, however, it cannot be estimated by the PML method.

Let us now also have a closer look at the estimated lower bounds and long-run volatilities. The estimated lower bound of the volatility σ_t is 6.43% p.a. in 2005, 6.53% p.a. in 2006, and 9.32% p.a. in 2007. The corresponding PML estimates for the lower bounds are a bit higher: 6.60%, 6.67%, and 9.77%, respectively. The posterior mean estimates for the long run volatilities are 10.01% p.a. in 2005, 10.05% in 2006, and 15.40% in 2007. The corresponding PML estimates for long run volatilities are significantly lower: 9.41%, 9.38%, and 13.76%, respectively (note that in the PML method, the long run volatility (not annualized) is computed as $(\beta/(\eta-\varphi))^{1/2}$). In 2007, the PML estimate is not even contained in a 90% credible interval. By comparison, the *actual* standard deviations of the original returns were 9.33% p.a. in 2005, 9.35% p.a. in 2006, and 14.21% in 2007. Of course, these latter numbers are just rough estimates, received without taking into account a sophisticated volatility model. Nevertheless they are contained in the corresponding 90% credible intervals in Table 4.

It is always essential to get good estimates for the volatilities. However, it can be much worse to underestimate volatility than to overestimate it. Also from this point of view our Bayesian approach turns out to be very useful, since it provides information on the distribution of the estimates, as in Figure 2. It shows the estimated marginal posterior densities for the lower bound and the mean of the volatility, for the three years under consideration. The densities are a bit unsymmetric: The densities of the lower bound are slightly negatively skewed (the sample skewness is -0.65,-0.73,-0.67 for 2005, 2006, and 2007, respectively), whereas the densities for the long run volatilities are slightly positively skewed (sample skewness is 0.86, 0.87, and 0.94).

Figure 3 finally shows the estimated volatilities for February to April 2007, and, for comparison, the absolute values of the detrended and reweighted log-returns.

5 Summary

This paper provides an MCMC based estimation procedure for the COGARCH(1,1) model driven by a compound Poisson process. The basic idea of the MCMC algorithm is an exact GARCH-type solution of the volatility SDE on the grid of the jump times. The simulation study has shown, that the parameters are quite well estimated and that the estimates are more efficient than those based on the PML method. The application illustrated how to apply the COGARCH model to high-frequency financial data. We studied log-returns from the S&P500 for the years 2005, 2006, and 2007 separately. It turned out that the mechanism, how new information changes the actual volatility and how fast this

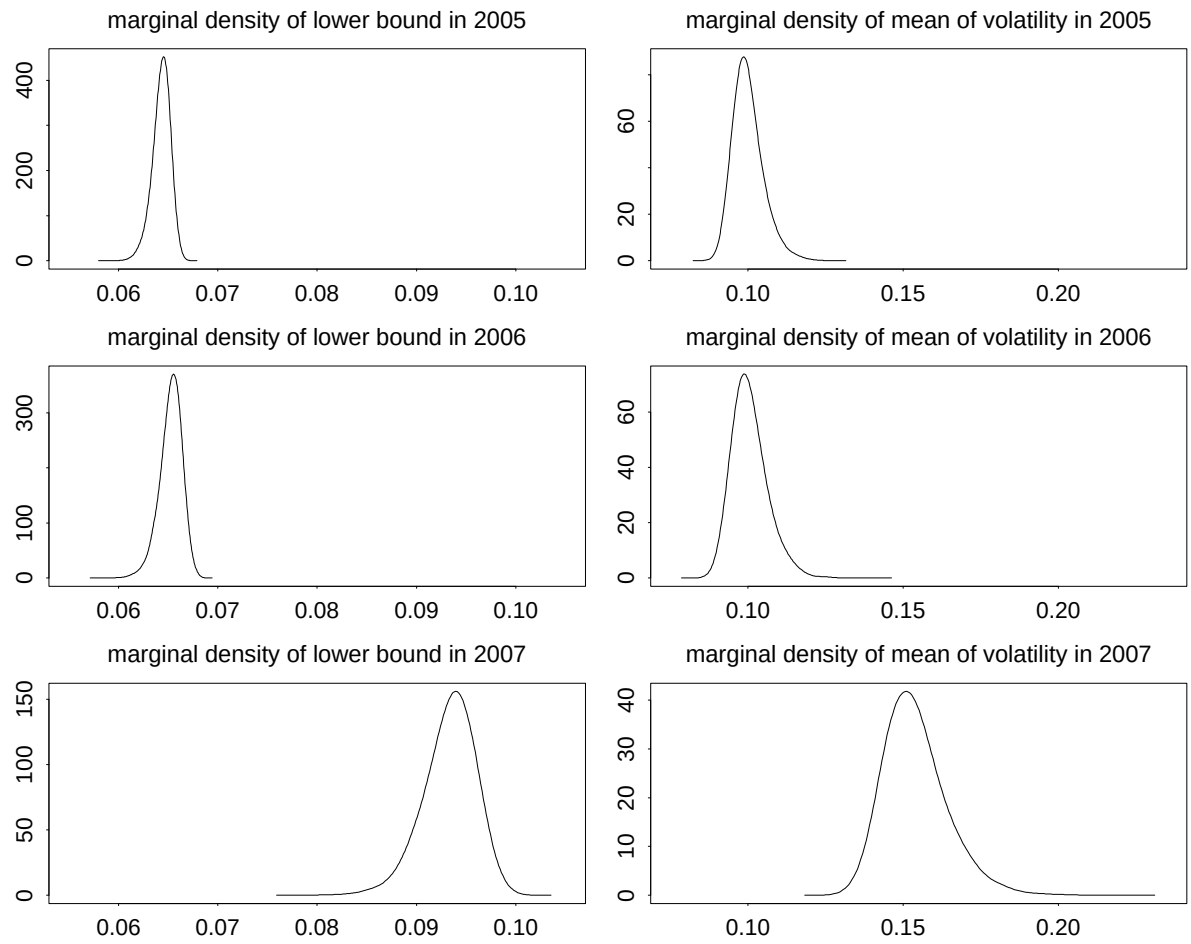

Figure 2: Estimated marginal posterior densities for the lower bound $\sigma_{\text{low,an}}$ and mean $\sigma_{\text{mean,an}}$ of the annualized volatility process, for years 2005, 2006, and 2007.

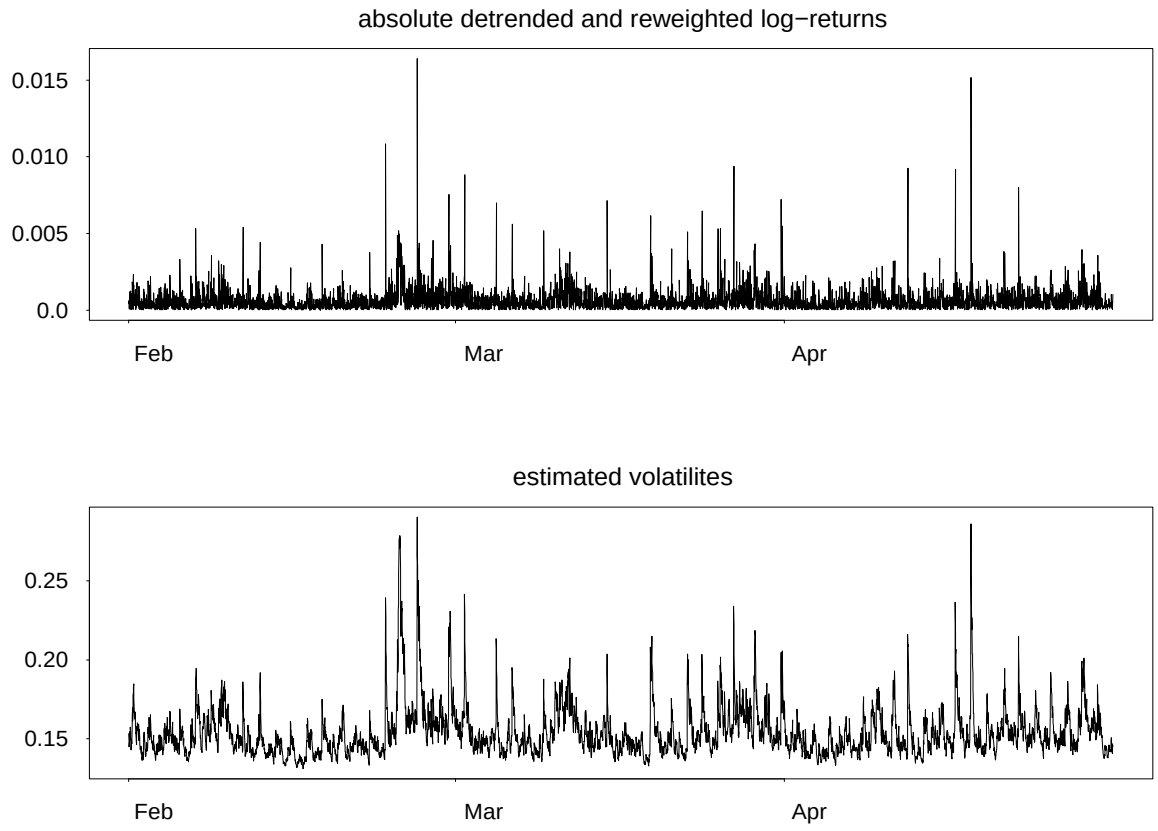

Figure 3: Top: Absolute detrended and reweighted log-returns of S&P500, February to April 2007. Bottom: Corresponding volatilities for the S&P500, estimated by MCMC.

new information is absorbed by the market, remained the same over these three years. However, the frequency of significant news as well as the overall level of volatility was different.

After considering the compound Poisson COGARCH one should now concentrate on a COGARCH which is driven by a more general class of driving Lévy processes, in particular, Lévy processes with infinite activity. For instance, one suitable choice might be the variance gamma process which is a pure jump process with finite quadratic variation. Furthermore, general model selection criteria for the COGARCH, in particular for comparison to other continuous-time models, are subject of current research.

Acknowledgement

I want to thank the *Deutsche Forschungsgemeinschaft* for the financial support via the research grant MU 2360/1-1. I am particularly grateful for the constructive and helpful comments which I received from the Editor Eric Renault. Together with the detailed remarks of two referees, they helped me a lot to improve the paper significantly. Moreover I would like to thank Claudia Czado, Claudia Klüppelberg, Ross Maller, and Alexander Szimayer for many valuable discussions.

References

Aït-Sahalia, Y. and J. Jacod (2009). Estimating the Degree of Activity of Jumps in High Frequency Data. *Annals of Statistics 37*(5), 2202–2244.

Andersen, T. (1994). Stochastic autoregressive volatility: a framework for volatility modeling. *Mathematical Finance 4*(1), 75–102.

Andersen, T., L. Benzoni, and J. Lund (2002). A Empirical Investigation of Continuous-Time Equity Return Models. *Journal of Finance 57*(3), 1239–1284.

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J.R.Stat.Soc. B 63*(2), 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2007). Variation, Jumps, Market Frictions and High Frequency Data in Financial Econometrics. *in* Blundell, R., Persson, T. and Newey, W. (Eds.)*, Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress.* Cambride University Press.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research Paper 2008-49, University of Aarhus*.

Bollerslev, T., T. Law, and G. Tauchen (2008). Risk, Jumps, and Diversification. *Journal of Econometrics 144*(1), 234–256.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*(2), 143–172.

Brockwell, P. J., E. Chadraa, and A. M. Lindner (2006). Continuous time GARCH processes. *The Annals of Applied Probability 16*(2), 790–826.

Buchmann, B. and G. Müller (2009). Limit experiments of GARCH. *Preprint, Monash University Melbourne and Technische Universität München*.

Corradi, V. (2000). Reconsidering the continuous time limit of the GARCH(1,1) process. *Journal of Econometrics 96*(1), 145–153.

Cox, J. C. (1975). Notes on option pricing I: constant elasticity of variance diffusions. Discussion Paper, Stanford University.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica 61*(4), 909–927.

Drost, F. C. and B. J. M. Werker (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics 74*(1), 31–57.

Duan, J.-C. (1997). Augmented GARCH(p,q) process and its diffusion limit. *Journal of Econometrics 79*(1), 97–127.

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*(6), 1343–1376.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica 50*(4), 987–1008.

Engle, R. F. (2000). The econometrics of ultra-high frequency data. *Econometrica 68*(1), 1–22.

Eraker, B., M. Johannes, and N. Polson (2003). The Impact of Jumps in Volatility and Returns. *Journal of Finance 58*(3), 1269–1300.

Fasen, V., C. Klüppelberg, and A. Lindner (2005). Extremal behavior of stochastic volatility models. *in* Grossinho, M.d.R., Shiryaev, A.N., Esquivel, M. and Oliviera, P.E. (Eds.)*, Stochastic Finance,* 107-155, Springer.

Geyer, C. J. and J. Møller (1994). Simulation Procedures and Likelihood Inference for Spatial Point Processes. *Scandinavian Journal of Statistics 21*(4), 359–373.

Ghysels, E. and J. Jasiak (1998). GARCH for irregularly spaced finacial data: the ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics 2*(4), 133–149.

Haug, S., C. Klüppelberg, A. Lindner, and M. Zapp (2007). Method of moment estimation in the COGARCH(1,1) model. *The Econometrics Journal 10*(2), 320–341.

Hsieh, D. (1991). Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance 46*(5), 1839–1877.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Economic and Business Statistics 12*(4), 371–389.

Kallsen, J. and M. S. Taqqu (1998). Option pricing in ARCH-type models. *Mathematical Finance 8*(1), 13–26.

Kallsen, J. and B. Vesenmayer (2009). COGARCH as a continuous-time limit of GARCH(1,1). *Stochastic Processes and their Applications 119*(1), 74–98.

Kazmerchuk, Y., A. Swishchuk, and J. Wu (2005). A continuous-time GARCH model for stochastic volatility with delay. *Canadian Applied Mathematics Quarterly 13*(2), 123–149.

Klüppelberg, C., A. Lindner, and R. Maller (2004). A continuous time GARCH process driven by a Lévy process: Stationarity and second order behaviour. *Journal of Applied Probability 41*(3), 601–622.

Klüppelberg, C., A. Lindner, and R. Maller (2006). Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. *In* Kabanov, Y., Lipster, R. and Stoyanov, J. (Eds.), *From Stochastic Calculus to Mathematical Finance. The Shiryaev Festschrift,* 393-419. Springer.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* New York: Springer.

Lorenz, R. (2006). Weak Approximation of Stochastic Delay Differential Equations with Bounded Memory by Discrete Time Series. Ph.D. thesis, Humboldt-Universität zu Berlin.

Lundblad, C. (2007). The Risk Return Tradeoff in the Long-Run: 1836-2003. *Journal of Financial Economics 85*(1), 123–150.

Maller, R., G. Müller, and A. Szimayer (2008). GARCH Modelling in Continuous Time for Irregularly Spaced Time Series Data. *Bernoulli 14*(2), 519–542.

Meddahi, N. and E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics 119*(2), 355–379.

Meddahi, N., E. Renault, and B. Werker (2006). GARCH and irregularly spaced data. *Economics Letters 90*, 200–204.

Nelson, D. (1990). ARCH models as diffusion approximations. *J. Econometrics 45*(1), 7–38.

Protter, P. (2005). *Stochastic Integration and Differential Equations.* 2nd ed. New York: Springer.

Roberts, G. O., O. Papaspiliopoulos, and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *J.R.Stat.Soc. B 66*(2), 369–393.

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *The Annals of Statistics 30*(3), 754–783.