Czado, Haug:

Quasi maximum likelihood estimation and prediction in the compound Poisson ECOGARCH(1,1) model

Sonderforschungsbereich 386, Paper 516 (2006)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Quasi maximum likelihood estimation and prediction in the compound Poisson ECOGARCH$(1,1)$ model

Claudia Czado
Technische Universität München

Stephan Haug
Technische Universität München

Abstract

This paper deals with the problem of estimation and prediction in a compound Poisson ECOGARCH$(1,1)$ model. For this we construct a quasi maximum likelihood estimator under the assumption that all jumps of the log-price process are observable. Since these jumps occur at unequally spaced time points, it is clear that the estimator has to be computed for irregularly spaced data. Assuming normally distributed jumps and a recursion to estimate the volatility allows to define and compute a quasi-likelihood function, which is maximised numerically. The small sample behaviour of the estimator is analysed in a small simulation study. Based on the recursion for the volatility process a one-step ahead prediction of the volatility is defined as well as a prediction interval for the log-price process. Finally the model is fitted to tick-by-tick data of the New York Stock Exchange. (JEL: C32, C51, C53)

Keywords: ECOGARCH; QMLE; Lévy process; leverage effect; prediction

1 Introduction

Discrete time GARCH type processes have become very popular in financial econometrics to model returns of stocks, exchange rates and other series observed at equidistant time points. They have been designed (see Engle (1982) and Bollerslev (1986)) to capture so-called *stylised facts* of such data, which are e.g. volatility clustering, dependence without correlation and tail heaviness. A further characteristic is that stock returns are negatively correlated with changes in the volatility, i.e. that volatility tends to increase after negative shocks and to fall after positive ones. This effect is called *leverage effect* and can not be modeled by a GARCH type process without further extensions. This

[0]Address for correspondence: Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, D-85748 Garching, Germany; Tel.: ++49 (0)89 28917040; email: haug@ma.tum.de

finding led Nelson (1991) to introduce the exponential GARCH(p,q) process, which is able to model this asymmetry in stock returns. This process X is defined as

$$\begin{aligned} X_i &= \sigma_i \epsilon_i \\ \log(\sigma_i^2) &= \mu + \sum_{k=1}^{p} \beta_k f(\epsilon_{i-k}) + \sum_{k=1}^{q} \alpha_k \log(\sigma_{i-k}^2)\,, \end{aligned}$$

where $p, q \in \mathbb{N}, \mu, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p \in \mathbb{R}$ and $\alpha_q \neq 0\,, \beta_p \neq 0$. For stationarity it is assumed that the *autoregressive polynomial* $\alpha(z) := 1 - \alpha_1 z - \cdots - \alpha_q z^q$ and the *moving average polynomial* $\beta(z) := \beta_1 + \beta_2 z + \cdots + \beta_p z^{p-1}$ have no common zeros and that $\alpha(z) \neq 0$ on $\{z \in \mathbb{C} \mid |z| \leq 1\}$. The errors $(\epsilon_i)_{i\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}(\epsilon_1) = 0$ and $\mathbb{V}\mathrm{ar}(\epsilon_1) = 1$. Finally $f(\cdot)$ is a known function such that $\mathbb{E}(|f(\epsilon_i)|) < \infty$ and $\mathbb{V}\mathrm{ar}(f(\epsilon_i)) < \infty$. To achieve an asymmetric relation between the stock return X_i and the future volatility σ_{i+1}^2, $f(\epsilon_i)$ must be a function of the magnitude and the sign of ϵ_i as noted by Nelson (1991). Therefore he proposed the following function:

$$f(\epsilon_i) := \theta\epsilon_i + \gamma[|\epsilon_i| - \mathbb{E}(|\epsilon_i|)]\,,$$

with real coefficients θ and γ.

The availability of high frequency financial data and the need to analyse such irregularly measured time series calls for stochastic models in continuous time which mimic the behaviour of financial time series models in discrete time. Different kinds of continuous time models have been proposed to model the stylised facts of financial data and which can be extended to describe a leverage effect, as e.g. the stochastic volatility model of Barndorff-Nielsen and Shepard (2001) or the continuous time GARCH$(1,1)$ process defined in Klüppelberg, Lindner and Maller (2004). Only recently has been defined a continuous time model by Haug and Czado (2007), the ECOGARCH(p,q) process, which aims at being the natural continuous time analogue of the discrete time EGARCH model. The only driving source of this process is a zero mean Lévy process, similar to the COGARCH$(1,1)$ process of Klüppelberg, Lindner and Maller (2004). Recall that any Lévy process $L := (L_t)_{t\geq 0}$ on $\mathbb{R}$ has a characteristic function of the form $\mathbb{E}(e^{iuL_t}) = \exp\{t\psi_L(u)\}\,, \quad t \geq 0$, with

$$\psi_L(u) := i\gamma_L u - \frac{\tau_L^2}{2}u^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iux\chi_{(-1,1)}(x))\nu_L(dx)\,, \quad u \in \mathbb{R},$$

where $\tau_L^2 \geq 0,\ \gamma_L \in \mathbb{R}$, the measure ν_L satisfies $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}} \min(x^2, 1)\nu_L(dx) < \infty$ and $\chi_A(\cdot)$ denotes the indicator function of the set $A \subset \mathbb{R}$. The measure ν_L is called the *Lévy measure* of L. For more details on Lévy processes we refer to Sato (1999) or Applbaum (2004).

Haug and Czado (2007) consider a zero mean Lévy processes L defined on a probability space $(\Omega, \mathcal{F}, P)$ with jumps $\Delta L_t := L_t - L_{t-}$, which implies that the Lévy-Itô decomposition (see e.g. Theorem 2.4.16 in Applbaum (2004)) of L is given by

$$L_t = B_t + \int_{\mathbb{R}\setminus\{0\}} x\tilde{N}_L(t, dx)\,, \qquad t \geq 0,$$

where B is a Brownian motion with variance τ_L^2 and $\tilde{N}_L(t,dx) = N_L(t,dx) - t\nu_L(dx)$, $t \geq 0$, is the compensated random measure associated to the Poisson random measure

$$N_L(t,A) = \#\{0 \leq s < t; \Delta L_s \in A\} = \sum_{0<s\leq t} \chi_A(\Delta L_s), \qquad A \in \mathcal{B}(\mathbb{R} \setminus \{0\}),$$

on $\mathbb{R}_+ \times \mathbb{R} \setminus \{0\}$, which is independent of B.

The *exponential continuous time* GARCH$(1,1)$ *process* is then defined as follows:

Let $L = (L_t)_{t\geq 0}$ *be a zero mean Lévy process with Lévy measure* ν_L *such that* $\int_{|x|\geq 1} x^2 \nu_L(dx) < \infty$. *Then the* exponential COGARCH$(1,1)$ *process* G, *abbreviated to* ECOGARCH$(1,1)$, *is defined as the stochastic process satisfying,*

$$dG_t \;:=\; \sigma_{t-} dL_t, \quad t > 0, \quad G_0 = 0, \tag{1}$$

where the log-volatility process $\log(\sigma^2) = (\log(\sigma_t^2))_{t\geq 0}$ *is the* CARMA$(1,0)$ *process (see e.g. Brockwell (2001)), with mean* $\mu \in \mathbb{R}$ *and state space representation*

$$\begin{aligned} \log(\sigma_t^2) &:= \mu + b_1 X_t, \quad t \geq 0\,, \\ dX_t &= -a_1 X_t dt + dM_t\,, \quad t > 0\,, \end{aligned}$$

with $a_1 \in \mathbb{R}_+$ *and* $b_1 \in \mathbb{R}$. *Here* $X_0 \in \mathbb{R}$ *is independent of the driving Lévy process* L *and*

$$M_t := \int_{\mathbb{R}\setminus\{0\}} h(x)\tilde{N}_L(t,dx)\,, \qquad t > 0, \tag{2}$$

is a zero mean Lévy process with $h(x) := \theta x + \gamma|x|$ *and parameters* $(\theta,\gamma) \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$.

While it is of high theoretical interest to be able to design a continuous time process which possesses the desired stylised facts including a leverage effect, it is of paramount practical interest to be able to fit such a model to real data and to facilitate prediction of future volatility and log returns. To be able to perform such parameter estimation and prediction we restrict in this paper to the case of a compound Poisson process as driving Lévy process for a ECOGARCH$(1,1)$ process. We follow a quasi maximum likelihood approach for parameter estimation. The quasi maximum likelihood estimator (QMLE) will be derived under the assumption of full observations of the sample path, i.e. we assume that we are able to observe every jump. Since the jump points are a series of unequally spaced time points it is clear that the estimation can and in fact should be done for irregularly spaced data.
The paper is now organized as follows. In Section 2 we introduce the quasi maximum likelihood estimator for the parameters of the process. Section 3 presents a simulation study to investigate the small sample behavior of the QMLE. In Section 4 we show how to compute one-step ahead predictions for the volatility and a prediction interval for the one-step ahead log-price process. An analysis of General Motors stock log-prices is done in Section 5. Concluding remarks are made in Section 6.

2 Quasi MLE in the compound Poisson ECOGARCH$(1,1)$ model

We consider an ECOGARCH$(1,1)$ process driven by a compound Poisson process L with Lévy symbol

$$\psi_L(u) = \int_{\mathbb{R}} (e^{iux} - 1)\lambda F_{0,1/\lambda}(dx)\,,$$

where $F_{0,1/\lambda}(\cdot)$ is a distribution function with mean 0 and variance $1/\lambda$. Hence the mean and variance of L are independent of λ and given by $\mathbb{E}(L_t) = 0$ and $\mathbb{V}\mathrm{ar}(L_t) = t$, respectively. This means L has representation $L_t = \sum_{k=1}^{J_t} Z_k$, $t > 0$, $L_0 = 0$, where $(J_t)_{t\geq 0}$ is an independent Poisson process with intensity $\lambda > 0$ and jump times $(t_k)_{k\in\mathbb{N}}$. The Poisson process J is also independent from the i.i.d. sequence of jump sizes $(Z_k)_{k\in\mathbb{N}}$, with distribution function $F_{0,1/\lambda}$. The Lévy process M, defined in (2), is in this case also a compound Poisson process and given by the following expression

$$M_t = \sum_{k=1}^{J_t} [\theta Z_k + \gamma|Z_k|] - \gamma\lambda K t\,, \quad t > 0,$$

with $K = \int_{\mathbb{R}} |x| F_{0,1/\lambda}(dx)$ (cf. also Example 3.5 in Haug and Czado (2007)). The log-volatility process at time $t \geq 0$ is then of the form

$$\begin{aligned}
\log(\sigma_t^2) &= \mu + b_1 X_t = \mu + b_1 e^{-a_1 t} X_0 + \int_0^t b_1 e^{-a_1(t-s)} dM_s \\
&= \mu + b_1 e^{-a_1 t} X_0 + \sum_{k=1}^{J_t} b_1 e^{-a_1(t-t_k)} [\theta Z_k + \gamma|Z_k|] - \gamma\lambda K \frac{b_1}{a_1}(1 - e^{-a_1 t})
\end{aligned}$$

and from (1) it follows that the log-price process is given by

$$G_t = \sum_{k=1}^{J_t} \sigma_{t_k -} Z_k\,, \quad t > 0, \quad G_0 = 0\,,$$

with jump times t_k, $k \in \mathbb{N}$.

We assume now that we observe G at n consecutive jump times $0 = t_0 < t_1 < \cdots < t_n < T$, $n \in \mathbb{N}$, over the time interval $[0, T]$. The state process X has then the following autoregressive representation

$$\begin{aligned}
b_1 X_{t_i} &= b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} + \sum_{k=J_{t_{i-1}}+1}^{J_{t_i}} b_1 e^{-a_1(t_i - t_k)} [\theta Z_k + \gamma|Z_k|] \\
&\quad - \gamma\lambda \int_{t_{i-1}}^{t_i} b_1 e^{-a_1(t_i - s)} K ds \\
&= b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} + b_1 \theta Z_i + b_1 \gamma \left(|Z_i| - \frac{\lambda K}{a_1}(1 - e^{-a_1 \Delta t_i}) \right), \qquad (3)
\end{aligned}$$

where $\Delta t_i := t_i - t_{i-1}$, $i = 1, \ldots, n$. Here we used $J_{t_{i-1}} + 1 = J_{t_i} = i$. This implies that the left-hand limit of the log-volatility process at the jump times $0 = t_0 < t_1 < \cdots < t_n$ is given by

$$\log(\sigma^2_{t_i-}) = \mu + b_1 e^{-a_1 \Delta t_i} X_{t_{i-1}} - b_1 \gamma \frac{\lambda K}{a_1}(1 - e^{-a_1 \Delta t_i}) \,. \tag{4}$$

Since b_1 serves only as a scaling coefficient it will be set equal to one for identifiability reasons from now on. The observations of the log-price process are given by

$$G_{t_i} = \sum_{k=1}^{J_{t_i}} \sigma_{t_k-} Z_k = G_{t_{i-1}} + \sigma_{t_i-} Z_i \,. \tag{5}$$

Hence the return at time t_i is equal to $G^{\Delta t_i}_{t_i} := G_{t_i} - G_{t_{i-1}} = \sigma_{t_i-} Z_i$.

The parameter estimation is done in two steps. The rate λ of the Poisson process J can be estimated given only the jump times t_i, therefore this is done in a first step. Since we observe the total number n of jumps for the Poisson process J over T intervals of length one the MLE of λ is given by

$$\widehat{\lambda}_n := \frac{n}{T} \,.$$

To estimate the remaining parameters $\boldsymbol{\vartheta} := (a_1, \theta, \gamma, \mu)$ we use similar ideas as in the discrete time case to solve the parameter estimation problem. Quasi maximum likelihood estimation in discrete time conditionally heteroscedastic time series models is e.g. explained in Straumann (2005).

Consider the following decomposition of the conditional log-likelihood given the initial value X_0

$$\log \rho_{\boldsymbol{\vartheta}}(G^{\Delta t_1}_{t_1}, \ldots, G^{\Delta t_n}_{t_n} | X_0) \;=\; \sum_{i=1}^{n} \log \rho_{\boldsymbol{\vartheta}}(G^{\Delta t_i}_{t_i} | G^{\Delta t_{i-1}}_{t_{i-1}}, \ldots, G^{\Delta t_1}_{t_1}, X_0) \,,$$

where we assume that $G^{\Delta t_i}_{t_i}$ given $G^{\Delta t_{i-1}}_{t_{i-1}}, \ldots, G^{\Delta t_1}_{t_1}, X_0$ is conditionally normal distributed with mean zero and variance $\sigma^2_{t_i-}/\lambda$. This implies that the conditional log-likelihood has the representation

$$\log \rho_{\boldsymbol{\vartheta}}(G^{\Delta t_1}_{t_1}, \ldots, G^{\Delta t_n}_{t_n} | X_0) \;=\; -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left(\log(\sigma^2_{t_i-}/\lambda) - \frac{(G^{\Delta t_i}_{t_i})^2}{\sigma^2_{t_i-}/\lambda}\right). \tag{6}$$

Since the volatility is unobservable, (6) can not be evaluated numerically. Therefore we need an approximation of the state process X, which together with (4) gives estimates of $\sigma^2_{t_1-}, \ldots, \sigma^2_{t_n-}$. Given parameters $\boldsymbol{\vartheta}$ and λ, an approximation of the recursion (3) is used to compute estimates of the state process X by

$$\widehat{X}_{t_i}(\boldsymbol{\vartheta}, \lambda) \;=\; e^{-a_1 \Delta t_i} \widehat{X}_{t_{i-1}}(\boldsymbol{\vartheta}, \lambda) + \theta \frac{G^{\Delta t_i}_{t_i}}{\widehat{\sigma}_{t_i-}(\boldsymbol{\vartheta}, \lambda)} + \gamma\left(\frac{|G^{\Delta t_i}_{t_i}|}{\widehat{\sigma}_{t_i-}(\boldsymbol{\vartheta}, \lambda)} - \lambda \widehat{K} \Delta t_i\right), \tag{7}$$

$i = 1, \ldots, n$, where $\widehat{K} := \sqrt{\frac{2}{\pi\lambda}} = \mathbb{E}(|W|)$, $W \sim N(0, 1/\lambda)$.

Here we used $(1 - e^{-z}) \approx z$ for z small and $\frac{G_{t_i}^{\Delta t_i}}{\widehat{\sigma}_{t_i-}(\boldsymbol{\vartheta},\lambda)}$ approximates the innovation Z_i. The recursion needs a starting value $\widehat{X}_0$. We set X_0 equal to the mean value of the stationary distribution of X, which is zero. (7) can be understood as the log-volatility description in a discrete time EGARCH(1, 1) model for irregularly spaced data. We work with this approximation since in our numerical experiences it provided better results than (3). This is due to the independence of a_1 from the part compensating the absolute jumps, which is approximated by the Gaussianity assumption. If we would use (3) the optimisation with respect to a_1 tries to account for the approximation, which results in convergence problems and biased estimates of the autoregressive parameter e^{-a_1}.

Recursion (7) together with expression (4) provides then estimates of the volatility given by

$$\widehat{\sigma}_{t_i-}^2(\boldsymbol{\vartheta}, \lambda) := \exp(\mu + e^{-a_1 \Delta t_i} \widehat{X}_{t_{i-1}}(\boldsymbol{\vartheta}, \lambda) - \gamma\lambda\widehat{K}\Delta t_i)\,, \quad i = 1, \ldots, n.$$

Based on the approximation of the volatility we define the quasi log-likelihood function for $\boldsymbol{\vartheta}$ given the data $\mathbf{G}^\Delta := (G_{t_1}^{\Delta t_1}, \ldots, G_{t_n}^{\Delta t_n})$ and the MLE $\widehat{\lambda}_n$ by

$$L(\boldsymbol{\vartheta}|\mathbf{G}^\Delta, \widehat{\lambda}_n) := -\frac{1}{2}\sum_{i=1}^{n} \log(\widehat{\sigma}_{t_i-}^2(\boldsymbol{\vartheta}, \widehat{\lambda}_n)) - \frac{1}{2}\sum_{i=1}^{n} \frac{(G_{t_i}^{\Delta t_i})^2}{\widehat{\sigma}_{t_i-}^2(\boldsymbol{\vartheta}, \widehat{\lambda}_n)/\widehat{\lambda}_n}\,. \tag{8}$$

Observe that $L(\boldsymbol{\vartheta}|\mathbf{G}^\Delta, \widehat{\lambda}_n)$ does not contain the constant $\frac{n}{2}\log(\widehat{\lambda}_n/2\pi)$, since it does not affect the optimisation with respect to $\boldsymbol{\vartheta}$.

Maximising the log-likelihood function (8) with respect to $\boldsymbol{\vartheta}$ over the parameter space $\Theta := \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}^2$ yields QML estimates

$$\widehat{\boldsymbol{\vartheta}}_n := \arg\max_{\boldsymbol{\vartheta}\in\Theta} L(\boldsymbol{\vartheta}|\mathbf{G}^\Delta, \widehat{\lambda}_n) \tag{9}$$

of $\boldsymbol{\vartheta}$. As a byproduct we get a parametric estimator of the volatility. If we first determine the QMLE $\widehat{\boldsymbol{\vartheta}}_n$ in (9) then we can substitute $\widehat{\boldsymbol{\vartheta}}_n$ into (7) and get estimates

$$\widehat{\sigma}_{t_i-}^2(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n) := \exp(\widehat{\mu}_n + e^{-\widehat{a}_1^n \Delta t_i} \widehat{X}_{t_{i-1}}(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n) - \widehat{\gamma}_n\widehat{\lambda}_n\widehat{K}\Delta t_i) \tag{10}$$

of the volatility at the jump times $t_1, \ldots, t_n$ based on $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{a}_1^n, \widehat{\theta}_n, \widehat{\gamma}_n, \widehat{\mu}_n)$.

3 Simulation study

The performance of the QMLE for small samples will be investigated in a simulation study. In all of the following simulation cases we will consider a compound Poisson ECOGARCH(1, 1) observed at all jump times t_i over the time interval $[0, 1\,500]$. The estimates will be computed for 1 000 independent replications in each case.

In all of the following cases we have taken the parameter a_1 equal to 0.1, the intensity λ equal to 2 and the mean μ of the log-volatility process will be equal to -3. This implies that we expect per replication 3 000 observations. The leverage parameter θ and γ will vary over the examples. In most of the cases θ will be negative and γ positive, i.e. we model the leverage effect as observed in stock price data. If $-\gamma < \theta < 0$, this corresponds to the case where a positive shock in the return data increases the log-volatility process less than a negative one of the same magnitude. For $\theta < -\gamma < 0$ a positive shock decreases the log-volatility, whereas a negative one increases it. The last example will illustrate the case where a positive shock in the log-price process increases the log-volatility process more than a negative one of the same size and we denote it as non leverage case. For a more detailed discussion of the leverage effect see also Section 3.1 in Haug and Czado (2007). For the distribution of the innovations Z_i we will consider two different cases.

3.1 Leverage case with Gaussian jump distribution

First the innovations Z_i are normally distributed with mean 0 and variance $1/\lambda$. We computed the empirical mean ($\widehat{\text{mean}}$), relative bias ($\widehat{\text{rbias}}$), and mean squared error ($\widehat{\text{MSE}}$) for all parameter estimates based on 1 000 independent replications. The corresponding results are summarised in Table 1.

In the leverage case we observe a satisfying performance of the QMLE. The relative bias of $\widehat{a}_1$, $\widehat{\theta}$ and $\widehat{\gamma}$ varies between -0.0361 and 0.0211 over the different parameter sets. For the mean $\widehat{\mu}$ of the log-volatility process a larger relative bias is observed. It also increases for larger values of γ. Moreover we seem to underestimate μ consistently, shown by a negative bias in each case. The quality of the separately estimated $\widehat{\lambda}$ remains of course unchanged over all parameter settings.

The goodness of fit of our estimation method is further investigated by an analysis of the fitted innovations for the case, where $\theta = -0.3$, $\gamma = 0.4$ and the other parameters remain fixed. The fitted innovations are given by $\widehat{Z}_i := G_{t_i}^{\Delta t_i}/\widehat{\sigma}_{t_i-}, \quad i = 1, \ldots, n$. Since our innovations were normally distributed with mean zero and standard deviation equal to $1/\sqrt{2}$, we expect the average $\overline{Z} := \frac{1}{n}\sum_{i=1}^{n} \widehat{Z}_i$ of the fitted innovations for one replication close to zero, their empirical standard deviation $\left(\frac{1}{n-1}\sum_{i=1}^{n}(\widehat{Z}_i - \overline{Z})^2\right)^{1/2}$ close to $1/\sqrt{2} \approx 0.70711$ and their empirical skewness close to zero. For all three quantities we computed $\widehat{\text{mean}}$ and $\widehat{\text{MSE}}$ over all 1 000 replications. The results are reported in Table 2 and indicate a good fit.

Under the assumption of a correctly estimated volatility the fitted innovations are a white noise series, in particular the innovations and also the squared innovations should be uncorrelated. The correlation of the squared innovations was checked by performing a Ljung-Box test (cf. Ljung and Box (1978)). The test statistic is given by

$$Q = n(n+2)\sum_{k=1}^{m} \frac{\widehat{\rho}_{\widehat{Z}^2}(k)^2}{n-k},$$

where $\widehat{\rho}_{\widehat{Z}^2}(k)$ is the empirical autocorrelation function of the the squared fitted innova-

Gaussian	$\widehat{a}_1$	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1021	-0.1007	0.1927	-2.9058	2.0008
	$(5{\cdot}10^{-4})$	$(5{\cdot}10^{-4})$	$(6{\cdot}10^{-4})$	$(3{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0211	0.0077	-0.0361	-0.0313	0.0004
$\widehat{\text{MSE}}$	0.0003	0.0003	0.0004	0.0161	0.0014
	$(1{\cdot}10^{-5})$	$(1{\cdot}10^{-5})$	$(2{\cdot}10^{-5})$	$(6{\cdot}10^{-4})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1800	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0979	-0.1797	0.1938	-2.8980	2.0001
	$(7{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(6{\cdot}10^{-4})$	$(4{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	-0.0202	-0.0016	-0.0310	-0.0339	$2{\cdot}10^{-5}$
$\widehat{\text{MSE}}$	0.0005	0.0014	0.0004	0.0274	0.0012
	$(6{\cdot}10^{-5})$	$(4{\cdot}10^{-4})$	$(2{\cdot}10^{-5})$	$(4{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.3000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1004	-0.3027	0.1930	-2.8982	2.0000
	$(3{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(6{\cdot}10^{-4})$	$(4{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0038	0.0091	-0.0349	-0.0339	$1{\cdot}10^{-5}$
$\widehat{\text{MSE}}$	0.0001	0.0013	0.0004	0.0290	0.0012
	$(1{\cdot}10^{-5})$	$(5{\cdot}10^{-4})$	$(5{\cdot}10^{-4})$	$(1{\cdot}10^{-2})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1017	-0.0996	0.3877	-2.8028	2.0001
	$(4{\cdot}10^{-4})$	$(6{\cdot}10^{-4})$	$(7{\cdot}10^{-4})$	$(4{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0170	-0.0037	-0.0307	-0.0657	0.0005
$\widehat{\text{MSE}}$	0.0001	0.0004	0.0006	0.0562	0.0013
	$(7{\cdot}10^{-6})$	$(2{\cdot}10^{-5})$	$(3{\cdot}10^{-5})$	$(2{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1800	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1007	-0.1792	0.3864	-2.8122	1.9985
	$(4{\cdot}10^{-4})$	$(7{\cdot}10^{-4})$	$(7{\cdot}10^{-4})$	$(4{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0071	-0.0044	-0.0338	-0.0625	-0.0007
$\widehat{\text{MSE}}$	0.0001	0.0005	0.0007	0.0526	0.0013
	$(6{\cdot}10^{-6})$	$(2{\cdot}10^{-5})$	$(3{\cdot}10^{-5})$	$(2{\cdot}10^{-3})$	$(5{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.3000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1010	-0.3004	0.3880	-2.7998	2.0023
	$(3{\cdot}10^{-4})$	$(7{\cdot}10^{-4})$	$(7{\cdot}10^{-4})$	$(4{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0102	0.0014	-0.0298	-0.0667	0.0011
$\widehat{\text{MSE}}$	0.0001	0.0005	0.0006	0.0595	0.0014
	$(4{\cdot}10^{-6})$	$(2{\cdot}10^{-5})$	$(2{\cdot}10^{-5})$	$(2{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$

Table 1: Estimated mean and MSE for $\widehat{a_1}, \widehat{\theta}, \widehat{\gamma}, \widehat{\mu}$ and $\widehat{\lambda}$ with corresponding estimated standard deviations in brackets based on 1 000 replications with normally distributed jump sizes. In the third row of each case the relative bias is shown.

Gaussian	mean$(\widehat{Z}_i)$	std$(\widehat{Z}_i)$	skewness$(\widehat{Z}_i)$
$\widehat{\text{mean}}$	-0.00004 ($4{\cdot}10^{-4}$)	0.70743 ($2{\cdot}10^{-4}$)	0.00062 ($1{\cdot}10^{-3}$)
$\widehat{\text{MSE}}$	0.00017 ($8{\cdot}10^{-6}$)	0.00004 ($2{\cdot}10^{-6}$)	0.00209 ($9{\cdot}10^{-5}$)

Table 2: Estimated mean and MSE for the mean, standard deviation and skewness of the fitted innovations with corresponding estimated standard deviations in brackets for normally distributed jumps based on 1 000 replications.

tions for one replication, and asymptotically χ^2-distributed with m degrees of freedom under the null hypothesis of no correlation. The number of lags m taken into account to compute the statistic was set equal to $\sqrt{n}$ (cf. Section 9.4 in Brockwell and Davis (1987)). The null hypothesis of no correlation was rejected 49 times out of 1 000 simulations at the 0.05 level. The empirical mean of the 1 000 p-values was equal to 0.52, which shows that a majority of the test statistics has a rather large p-value confirming the hypothesis of no correlation.

3.2 Leverage case with student-t jump distribution

In all the previous examples the jump distribution was Gaussian. Now we want to compute the QMLE under the assumption of sampling innovations from a student-t distribution. We will consider a t-distribution with 6 degrees of freedom. Since we assume $\mathbb{E}(Z_1) = 0$ and $\mathbb{V}\text{ar}(Z_1) = 1/\lambda$, we have to scale the innovations in an appropriate way. The intensity λ will again be equal to 2 and μ remains at -3. The parameters a_1, θ and γ taken on the same values as in the Gaussian case. The results are reported in Table 3.

The quality of the estimators for a_1, θ, γ and μ is reduced due to the model misspecification. The relative bias and MSE have increased for almost all parameter settings. Concerning μ the relative bias for example has doubled compared to results for normally distributed jumps. But overall the results are still satisfying. Indicating that the QMLE provides reasonable values even if the true distribution of the returns is much heavier tailed than the assumed one.

Analogously to the Gaussian case we investigated the goodness of fit by an analysis of the fitted innovations for the case, where $\theta = -0.3$ and $\gamma = 0.4$. Since we scaled the innovations such that they have mean zero and a standard deviation equal to $1/\sqrt{2}$, we expect the empirical mean of the estimated innovations close to zero, their empirical standard deviation close to 0.70711 and their empirical skewness close to zero. For all three quantities we computed $\widehat{\text{mean}}$ and $\widehat{\text{MSE}}$ over all 1 000 replications. The results are reported in Table 4 and indicate a reasonable fit. The null hypothesis of the Ljung-Box test was rejected 84 times out of 1 000 simulations at the 0.05 level. The empirical mean of the 1 000 p-values was 0.59.

student-t	$\widehat{a}_1$	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0978	-0.0995	0.1917	-2.7748	2.0020
	$(1{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$	$(9{\cdot}10^{-4})$	$(6{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	-0.0213	-0.0046	-0.0415	-0.0750	0.0010
$\widehat{\text{MSE}}$	0.0012	0.0017	0.0010	0.0873	0.0014
	$(8{\cdot}10^{-5})$	$(6{\cdot}10^{-4})$	$(5{\cdot}10^{-5})$	$(9{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1800	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1007	-0.1829	0.1934	-2.7776	2.0001
	$(8{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(9{\cdot}10^{-4})$	$(5{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0077	0.0165	-0.0331	-0.0741	$4{\cdot}10^{-5}$
$\widehat{\text{MSE}}$	0.0007	0.0020	0.0008	0.0757	0.0013
	$(6{\cdot}10^{-5})$	$(6{\cdot}10^{-4})$	$(5{\cdot}10^{-5})$	$(5{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.3000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1013	-0.3034	0.1926	2.0012	-2.7738
	$(5{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(9{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(4{\cdot}10^{-4})$
$\widehat{\text{rbias}}$	0.0131	0.0115	-0.0369	0.0006	-0.0754
$\widehat{\text{MSE}}$	0.0002	0.0015	0.0008	0.0014	0.0668
	$(2{\cdot}10^{-5})$	$(5{\cdot}10^{-4})$	$(6{\cdot}10^{-5})$	$(6{\cdot}10^{-5})$	$(3{\cdot}10^{-3})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1022	-0.1010	0.3884	-2.5405	2.0025
	$(5{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$	$(6{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0221	0.0101	-0.0290	-0.1531	0.0012
$\widehat{\text{MSE}}$	0.0003	0.0009	0.0012	0.2441	0.0014
	$(2{\cdot}10^{-5})$	$(4{\cdot}10^{-5})$	$(6{\cdot}10^{-5})$	$(6{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.1800	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1015	-0.1819	0.3861	-2.5518	2.0004
	$(5{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$	$(6{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0151	0.0109	-0.0347	-0.1493	0.0002
$\widehat{\text{MSE}}$	0.0003	0.0011	0.0013	0.2346	0.0013
	$(1{\cdot}10^{-5})$	$(2{\cdot}10^{-4})$	$(5{\cdot}10^{-5})$	$(5{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$
$\boldsymbol{\vartheta}$, λ	0.1000	-0.3000	0.4000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.1020	-0.3030	0.3876	-2.5481	1.9997
	$(4{\cdot}10^{-4})$	$(1{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$	$(6{\cdot}10^{-3})$	$(1{\cdot}10^{-3})$
$\widehat{\text{rbias}}$	0.0199	0.0099	-0.0309	-0.1506	-0.0002
$\widehat{\text{MSE}}$	0.0013	0.0009	0.0010	0.0789	0.0014
	$(8{\cdot}10^{-5})$	$(4{\cdot}10^{-4})$	$(5{\cdot}10^{-5})$	$(4{\cdot}10^{-3})$	$(6{\cdot}10^{-5})$

Table 3: Estimated mean and MSE for $\widehat{a_1}, \widehat{\theta}$, $\widehat{\gamma}, \widehat{\mu}$ and $\widehat{\lambda}$ with corresponding estimated standard deviations in brackets based on 1 000 replications with t-distributed jump sizes. In the third row of each case the relative bias is shown.

$\widehat{\text{mean}}$	-0.00007 ($4{\cdot}10^{-4}$)	0.70763 ($2{\cdot}10^{-4}$)	-0.00295 ($6{\cdot}10^{-3}$)
$\widehat{\text{MSE}}$	0.00016 ($7{\cdot}10^{-6}$)	0.00006 ($7{\cdot}10^{-6}$)	0.03846 ($5{\cdot}10^{-3}$)

Table 4: Estimated mean and MSE for the mean, standard deviation and skewness of the fitted innovations with corresponding estimated standard deviations in brackets for student-t distributed jumps based on 1 000 replications.

3.3 Non-leverage case

So far we have only considered the leverage case. In this last example we will have the following relation: $0 < \theta < \gamma$. This means that a positive shock in the return data increases the log-volatility process more than a negative one. For normally distributed jumps the results in the non-leverage case are shown in the upper section of Table 5. One can observe an increased relative bias for a_1, θ and γ compared to the leverage case. But the results are still acceptable. For t-distributed innovations the results are shown in the lower section of Table 5. The conclusions which can be made are similarly to the Gaussian case. It is interesting however that the leverage parameter θ is estimated more accurately for t-distributed jumps. It can be concluded that the influence of the sign of the leverage parameter θ is not important for the performance of the QMLE.

Gaussian	$\widehat{a}_1$	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
$\boldsymbol{\vartheta}$, λ	0.1000	0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0938	0.0980	0.1911	-2.8977	2.0024
	($1{\cdot}10^{-3}$)	($1{\cdot}10^{-3}$)	($8{\cdot}10^{-4}$)	($7{\cdot}10^{-3}$)	($1{\cdot}10^{-3}$)
$\widehat{\text{rbias}}$	-0.0614	-0.0201	-0.0447	-0.0341	0.0012
$\widehat{\text{MSE}}$	0.0012	0.0015	0.0007	0.0595	0.0014
	($9{\cdot}10^{-5}$)	($5{\cdot}10^{-4}$)	($1{\cdot}10^{-4}$)	($1{\cdot}10^{-2}$)	($6{\cdot}10^{-5}$)
student-t	$\widehat{a}_1$	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
$\boldsymbol{\vartheta}$, λ	0.1000	0.1000	0.2000	-3.0000	2.0000
$\widehat{\text{mean}}$	0.0976	0.0988	0.1913	-2.7873	2.0031
	($1{\cdot}10^{-3}$)	($1{\cdot}10^{-3}$)	($1{\cdot}10^{-3}$)	($6{\cdot}10^{-3}$)	($1{\cdot}10^{-3}$)
$\widehat{\text{rbias}}$	-0.0235	-0.0117	-0.0432	-0.0709	0.0015
$\widehat{\text{MSE}}$	0.0013	0.0009	0.0009	0.0789	0.0013
	($8{\cdot}10^{-5}$)	($4{\cdot}10^{-5}$)	($5{\cdot}10^{-5}$)	($4{\cdot}10^{-3}$)	($6{\cdot}10^{-5}$)

Table 5: Estimated mean and MSE for $\widehat{a_1}, \widehat{\theta}, \widehat{\gamma}, \widehat{\lambda}$ and $\widehat{\mu}$ with corresponding estimated standard deviations in brackets based on 1 000 replications with normally distributed (*top*) and t-distributed (*bottom*) jump sizes. In the third row of each case the relative bias is shown.

3.4 Empirical characteristics

Finally some empirical characteristics of the volatility process will be presented. In particular we computed for each set of parameters the empirical mean $\overline{\sigma^2}$, variance $s^2(\sigma^2)$ and 99% quantile of the volatility process. Further we estimated the correlation $\mathbb{C}\text{orr}(\Delta G_{t_i}, \sigma^2_{t_i})$, which should be negative in the leverage case (see also Section 3.1 in Haug and Czado (2007) for details). This is done for Gaussian as well as t-distributed jumps. The results are shown in Table 6.

Gaussian	θ, γ	$\overline{\sigma^2}$	$s^2(\sigma^2)$	$\widehat{q}_{0.99}$	$\widehat{\mathbb{C}\text{orr}}(\Delta G_{t_i}, \sigma^2_{t_i})$
	-0.10, 0.2	0.0633	0.0015	0.2084	-0.1441
	-0.18, 0.2	0.0680	0.0031	0.2822	-0.2085
	-0.30, 0.2	0.0815	0.0116	0.4952	-0.2597
	-0.10, 0.4	0.0999	0.0244	0.7055	-0.0902
	-0.18, 0.4	0.1083	0.0433	0.8415	-0.1459
	-0.30, 0.4	0.1368	0.2569	1.3444	-0.2073
	0.10, 0.2	0.0633	0.0015	0.2076	0.1449
student-t	θ, γ	$\overline{\sigma^2}$	$s^2(\sigma^2)$	$\widehat{q}_{0.99}$	$\widehat{\mathbb{C}\text{orr}}(\Delta G_{t_i}, \sigma^2_{t_i})$
	-0.10, 0.2	0.0636	0.0163	0.2193	-0.1485
	-0.18, 0.2	0.0681	0.0044	0.3005	-0.2120
	-0.30, 0.2	0.0831	0.1933	0.5338	-0.2615
	-0.10, 0.4	0.1022	0.0594	0.7977	-0.0937
	-0.18, 0.4	0.1109	0.0849	0.9216	-0.1569
	-0.30, 0.4	0.1442	0.9888	1.5152	-0.2151
	0.10, 0.2	0.0632	0.0017	0.2157	0.1476

Table 6: Empirical mean, variance and 99% quantile of the volatility process with parameters $a_1 = 0.1, \mu = -3$ and $\lambda = 2$. The empirical correlation between the current jump in the log-price process and future volatility.

The estimated correlation is negative in all of the leverage cases and positive for the non-leverage case ($\theta = 0.1, \gamma = 0.2$). We further observe a slightly increased variance and greater quantiles for t-distributed jumps, which seems reasonable.

4 Prediction

The aim of this section is to show how prediction can be done in this framework. In particular we will construct a one-step ahead prediction of the volatility process and also derive the prediction density in that case. Since future innovations of the log-price process are positive or negative with probability 0.5, it is not possible to define a sensible one-step ahead point-prediction of the log-price process. But we can construct a prediction interval for the next observation of the log price, thus taking care about the size of the innovation.

4.1 One-step ahead prediction of the volatility

In this section we want to show how to compute a prediction $P_r(\sigma^2_{t_{n+1}})$ of the volatility $\sigma^2_{t_{n+1}}$, which means the volatility right after the next jump, conditional on the information $\widehat{X}_{t_n}$, $\widehat{\boldsymbol{\vartheta}}_n$, $\widehat{\lambda}_n$ and Δt_{n+1}. Recall that the parameters $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{a}_1^n, \widehat{\theta}_n, \widehat{\gamma}_n, \widehat{\mu}_n)$ and $\widehat{\lambda}_n$ are estimated based on the observations $G_{t_1}, \ldots, G_{t_n}$. As a first step we will need a prediction of the log-volatility process at time t_{n+1}, which is the $n+1$-th jumptime of the log-volatility process. Since we estimated the parameters by maximising a Gaussian likelihood it seems reasonable to simulate the future jump $\widehat{Z}_{n+1}$ as a normally distributed random variable with mean 0 and variance $1/\widehat{\lambda}_n$. If we substitute in the recursion (7) for $i = n+1$ the quantity $G_{t_{n+1}}^{\Delta t_{n+1}}/\widehat{\sigma}_{t_{n+1}-}(\boldsymbol{\vartheta}, \lambda)$ by $\widehat{Z}_{n+1}$, we get

$$\widehat{X}_{t_{n+1}}(\boldsymbol{\vartheta}, \lambda) \;=\; e^{-a_1 \Delta t_{n+1}} \widehat{X}_{t_n}(\boldsymbol{\vartheta}, \lambda) + \theta \widehat{Z}_{n+1} + \gamma \left(|\widehat{Z}_{n+1}| - \lambda \widehat{K} \Delta t_{n+1} \right) .$$

Hence we will get a prediction of $\log(\sigma^2_{t_{n+1}})$ the log-volatility process at time t_{n+1}, which is denoted by $P_r(\log(\sigma^2_{t_{n+1}}))$, by the following equation

$$P_r(\log(\sigma^2_{t_{n+1}})) := \widehat{\mu}_n + e^{-\widehat{a}_1^n \Delta t_{n+1}} \widehat{X}_{t_n}(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n) + \widehat{\theta}_n \widehat{Z}_{n+1} + \widehat{\gamma}_n \left(|\widehat{Z}_{n+1}| - \widehat{\lambda}_n \widehat{K} \Delta t_{n+1} \right) .$$

The one-step ahead prediction of $\sigma^2_{t_{n+1}}$ the volatility process at time t_{n+1}, which is denoted by $P_r(\sigma^2_{t_{n+1}})$, is then defined by applying the exponential function to the prediction $P_r(\log(\sigma^2_{t_{n+1}}))$, i.e.

$$P_r(\sigma^2_{t_{n+1}}) := \exp(P_r(\log(\sigma^2_{t_{n+1}}))) . \tag{11}$$

For ease of notation we will omit in the following the dependence of $\widehat{X}_{t_n}$ on $\widehat{\boldsymbol{\vartheta}}_n$ and $\widehat{\lambda}_n$. Given $\widehat{X}_{t_n}$, $\widehat{\boldsymbol{\vartheta}}_n$, $\widehat{\lambda}_n$ and Δt_{n+1} $P_r(\sigma^2_{t_{n+1}})$ is a monotone function of $\widehat{Z}_{n+1}$. Hence we can in addition easily derive a prediction density of $\sigma^2_{t_{n+1}}$, which is denoted by $p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1})$. The form of the density depends on the sign and size of $\widehat{\theta}_n$ and $\widehat{\gamma}_n$. If e.g. $\widehat{\theta}_n < -\widehat{\gamma}_n < 0$, then the prediction density for $\sigma^2_{t_{n+1}}$ is given by

$$\begin{aligned}
&p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}) = \\
&\quad \frac{1}{s} \Bigg[\chi_{(-\infty, \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})]}(\log(s) - \widehat{\mu}_n) \left| \frac{1}{\widehat{\theta}_n + \widehat{\gamma}_n} \right| \\
&\qquad\qquad \times \sqrt{\frac{\widehat{\lambda}_n}{2\pi}} \exp\left\{ -\frac{\widehat{\lambda}_n}{2} \left(\frac{\log(s) - \widehat{\mu}_n - \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})}{\widehat{\theta}_n + \widehat{\gamma}_n} \right)^2 \right\} \\
&\quad + \chi_{(\varphi(\widehat{X}_{t_n}, \Delta t_{n+1}), \infty)}(\log(s) - \widehat{\mu}_n) \left| \frac{1}{\widehat{\theta}_n - \widehat{\gamma}_n} \right| \\
&\qquad\qquad \times \sqrt{\frac{\widehat{\lambda}_n}{2\pi}} \exp\left\{ -\frac{\widehat{\lambda}_n}{2} \left(\frac{\log(s) - \widehat{\mu}_n - \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})}{\widehat{\theta}_n - \widehat{\gamma}_n} \right)^2 \right\} \Bigg] ,
\end{aligned} \tag{12}$$

for each $s > 0$, where $\varphi(\widehat{X}_{t_n}, \Delta t_{n+1}) := e^{-\widehat{a}_1^n \Delta t_{n+1}} \widehat{X}_{t_n} - \widehat{\gamma}_n \widehat{\lambda}_n \widehat{K} \Delta t_{n+1}$.

A different shape of the prediction density we get for $-\widehat{\gamma}_n < \widehat{\theta}_n < 0$:

$$\begin{aligned} p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}) &= \frac{1}{s} \chi_{(\varphi(\widehat{X}_{t_n}, \Delta t_{n+1}), \infty)}(\log(s) - \widehat{\mu}_n) \left| \frac{1}{\widehat{\theta}_n - \widehat{\gamma}_n} \right| \\ &\quad \times \sqrt{\frac{\widehat{\lambda}_n}{2\pi}} \exp\left\{ -\frac{\widehat{\lambda}_n}{2} \left(\frac{\log(s) - \widehat{\mu}_n - \varphi(\widehat{X}_{t_n}, \Delta t_{n+1})}{\widehat{\theta}_n - \widehat{\gamma}_n} \right)^2 \right\}, \end{aligned}$$

for each $s > 0$.

Next we illustrate how the shape of the prediction density for $\sigma^2_{t_{n+1}}$ depends on the estimated parameters. Therefore we consider again 1 000 samples of a compound Poisson ECOGARCH(1, 1) process with parameters $\mu = 0, a_1 = 0.1, \theta = -0.3, \gamma = 0.2$ and $\lambda = 2$. The sample paths are over the time interval $[0, 1\,000]$ yielding on average 2 000 observations. First we estimate the parameters, as explained in Section 2, based on the first $n = 1900$ observations. To compute a prediction density for $\sigma^2_{t_{n+1}}$ by (12) we will further need $\widehat{X}_{t_n}$ and Δt_{n+1}. Since both of them will be different for each sample, we set them equal to 1 and 0.5, respectively, to make prediction densities comparable. For each sample the prediction density was then computed over the same grid on the interval $[1.5, 4.7]$. Everything was done for two different jump distributions. In the first case the jumps Z_i were normally distributed with mean 0 and variance $1/\lambda$, whereas in the second example scaled t-distributed with 6 degrees of freedom and same mean and variance as in the normal case. In the first row of Figure 1 one can see on the left hand side six replications of prediction densities $p(\cdot|1, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, 0.5)$ for normally distributed jumps and on the right hand side for the t-distributed ones. In both cases we also computed a mean prediction density

$$\overline{p}(s|1, \widehat{\boldsymbol{\vartheta}}, \widehat{\lambda}_n, 0.5) := \frac{1}{N} \sum_{i=1}^{N} p^i(s|1, \widehat{\boldsymbol{\vartheta}}^i_n, \widehat{\lambda}^i_n, 0.5) \,,$$

for s lying on the grid and $N = 1\,000$. The results are shown in the second row of Figure 1 together with true prediction density $p(\cdot|1, \boldsymbol{\vartheta}, \lambda, 0.5)$. One can observe that despite the fact that we have a larger bias in the estimates $\widehat{\boldsymbol{\vartheta}}_n$ for t-distributed jumps compared to normally distributed ones, the prediction densities show similar behaviour.

For assessing the quality of the forecasts we will consider a scoring rule, which assigns a numerical score based on the prediction density and the value actually observed. We will work with the logarithmic score, for other scoring rules see Gneiting and Raftery (2007). Assume we have N observations of the volatility $\sigma^2_{t^1_{n+1}}, \ldots, \sigma^2_{t^N_{n+1}}$. For each of them we compute a prediction density $p^i(\cdot|\widehat{X}_{t^i_n}, \widehat{\boldsymbol{\vartheta}}^i_n, \widehat{\lambda}^i_n, \Delta t^i_{n+1})$, $i = 1, \ldots, N$, which will be evaluated at the observation point. The logarithmic score for the i-th sample is then defined as

$$LS_i := \log\left(p^i(\sigma^2_{t^i_{n+1}}|\widehat{X}_{t^i_n}, \widehat{\boldsymbol{\vartheta}}^i_n, \widehat{\lambda}^i_n, \Delta t^i_{n+1}) \right) .$$

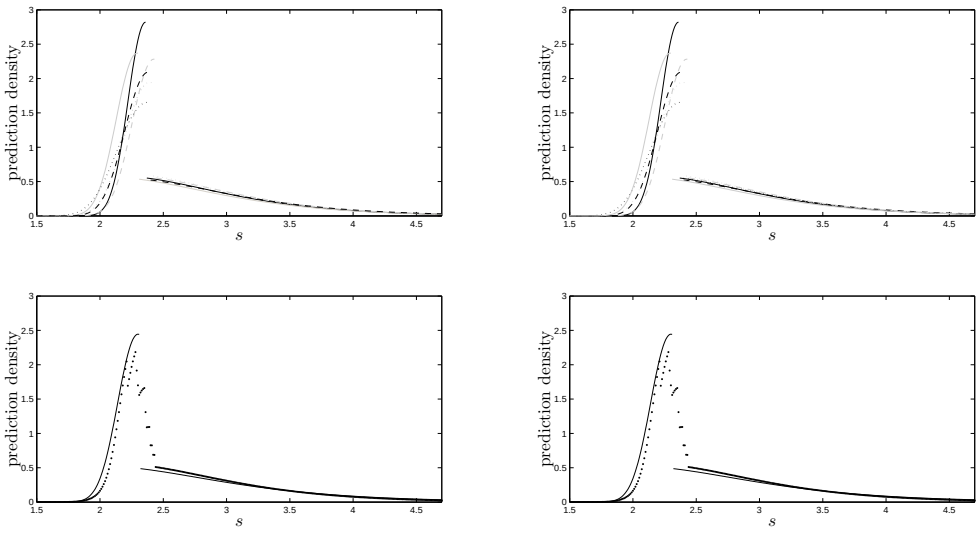

Figure 1: Prediction densities $p^i(\cdot|1, \widehat{\boldsymbol{\vartheta}}_n^i, \widehat{\lambda}_n^i, 0.5)$, $i - 1, \ldots, 6$, for the first 6 replication (*top row*), the mean prediction density $\overline{p}(s|1, \widehat{\boldsymbol{\vartheta}}, \widehat{\lambda}_n, 0.5)$ over the 1 000 replications and true (*solid line*) prediction density $p(\cdot|1, \boldsymbol{\vartheta}_n, \lambda_n, 0.5)$(*bottom row*) for normally (*left*) and t-distributed (*right*) jumps.

Observe that the higher LS is, the better the prediction will be. Note that this evaluation is only feasible in a simulation set up since we do not observe the volatility process in real data. To associate a numerical score to the all N samples we further calculate the mean logarithmic score MLS defined as

$$MLS := \frac{1}{N} \sum_{i=1}^{N} LS_i.$$

For our previous considered simulation examples, where we had $\mu = 0, a_1 = 0.1, \theta = -0.3, \gamma = 0.2$ and $N = 1\,000$ replications, we get MLS's of -0.0716 and -0.8285 for Gaussian and t-distributed jumps, respectively. This result is not surprising, since we assumed normally distributed jumps to compute the prediction density. Thus we would expect a better performance of the prediction, if the jumps are really normally distributed. One also has to mention that the logarithmic score is rather sensitive to outliers which produce large negative values.

Given the prediction density we are able to define a second one-step ahead prediction of the volatility. More precisely we take the mode of the corresponding prediction distribution as prediction of $\sigma^2_{t_{n+1}}$, defined by

$$P_m(\sigma^2_{t_{n+1}}) := \arg\max_{s>0} \; p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1}) \, .$$

Both predictions P_r and P_m are compared for the two simulation cases by computing the empirical relative bias $\widehat{\text{rbias}}(P_k) := \frac{1}{N} \sum_{i=1}^{N} \left(\frac{P_k(\sigma^2_{t^i_{n+1}{}^+})}{\sigma^2_{t^i_{n+1}{}^+}} - 1 \right)$ and the empirical relative

mean squared error $\widehat{\text{RMSE}}(P_k) := \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_k(\sigma^2_{t^i_{n+1}+})}{\sigma^2_{t^i_{n+1}+}} - 1\right)^2$ for $k = r, m$. Remember that we have $N = 1\,000$ replications in each case. The results are reported in Table 7.

	$\widehat{\text{rbias}}(P_r)$	$\widehat{\text{RMSE}}(P_r)$	$\widehat{\text{rbias}}(P_m)$	$\widehat{\text{RMSE}}(P_m)$
Gaussian	0.0364	0.1301	-0.0895	0.0428
student-t	0.0399	0.1383	-0.0974	0.0517

Table 7: Relative bias and MSE for the recursive and mode prediction in case Gaussian and student-t distributed jumps.

Concluding one can say that despite the reduced performance of the QMLE in the non-normal case, the one-step ahead prediction ability is acceptable compared to the normal case. The results for the recursive prediction P_r are actually very similar. For the mode prediction greater differences can be observed. This is also rather obvious, since it predicts the mode of a distribution, which is more suitable if the innovations Z are not so heavy tailed. Concerning the RMSE the mode prediction P_m is preferable compared to P_r in both cases, but one has to deal with a larger relative bias of P_m.

Finally we want to analyse how the accuracy of the prediction density depends on the number of observations. For the 1 000 sample paths in each case, we estimated the parameters based on all observations $G_{t_1}, \ldots, G_{t_n}$, the last 1 500 $G_{t_{n-1499}}, \ldots, G_{t_n}$, the last 1 000 $G_{t_{n-999}}, \ldots, G_{t_n}$ and finally the last 500 $G_{t_{n-499}}, \ldots, G_{t_n}$. For all four scenarios we computed the prediction density and logarithmic score, which was then averaged over all 1 000 samples. The resulting MLS's are summarised in Table 8.

Recall that the average number of observations over the 1 000 samples is 2 000. Comparing the MLS's one can recognise that at least 1 500 observations should be taken into account to estimate the parameters. If not, the prediction quality will reduced considerably. In our real data example considered in Section 5, 1 500 observations correspond to roughly three quarters of a trading day.

4.2 Prediction interval for the log-price

If we consider again the log-price process G defined in (5) one observes that the future return $G^{\Delta t_{i+1}}_{t_{i+1}}$ has variance $\sigma^2_{t_{i+1}-}/\lambda$ under the assumption $Z_{i+1} \sim N(0, 1/\lambda)$. From (4) it follows that given X_{t_i} and Δt_{i+1} the volatility $\sigma^2_{t_{i+1}-}$ is known up to parameters $\boldsymbol{\vartheta}$ and λ and can be calculated by (10) given $\widehat{\boldsymbol{\theta}}$ and $\widehat{\lambda}$. This enables us to construct a 95% prediction interval for the one-step ahead log-price $G_{t_{i+1}}$. It is given by the following

Number of observations	n	1 500	1 000	500
MLS (Gaussian)	-0.0716	-0.1426	-0.8002	-3.5185
MLS (student-t)	-0.8285	-0.5776	-3.5121	-5.1267

Table 8: MLS for varying numbers of observations in case Gaussian and student-t distributed jumps.

expression

$$PI(G_{t_{i+1}}) := \left(G_{t_i} - 1.96 \cdot \sqrt{\widehat{\sigma}^2_{t_{i+1}-}(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)/\widehat{\lambda}_i}, G_{t_i} + 1.96 \cdot \sqrt{\widehat{\sigma}^2_{t_{i+1}-}(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)/\widehat{\lambda}_i} \right) .$$

For a simulated log-price process G, with observation times $t_1, \ldots, t_{n+1}$ and normally distributed jumps, the empirical quality of the prediction interval was tested. As parameter values we chose $\boldsymbol{\vartheta} = (\alpha, \theta, \gamma, \mu) = (0.1, -0.3, 0.4, -3)$ and $\lambda = 2$. Starting with 100 observations, we reestimated the model for each new observation to get $\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i$ and computed in each step the volatility $\sigma^2_{t_{i+1}-}(\widehat{\boldsymbol{\vartheta}}_i, \widehat{\lambda}_i)$, $i = 100, \ldots, n$. Then we calculated the prediction interval $PI(G_{t_{i+1}})$ and counted over three time intervals $(t_{100}, t_{k_1}]$, $(t_{k_1}, t_{k_2}]$ and $(t_{k_2}, t_{n+1}]$, for $100 < k_1 < k_2 < n+1$, containing the same number of observations G_{t_i} how many of them are observed within $PI(G_{t_i})$.
We repeated the whole procedure 100 times. The average percentages of observations lying in the prediction intervals over the three time intervals were 94.04%, 94.47% and 94.55%, respectively. The results for all three intervals are close to the expected value of 95%. An increase in the prediction accuracy with the numbers of observations available to estimate the parameters can be observed. On the left hand side of Figure 2

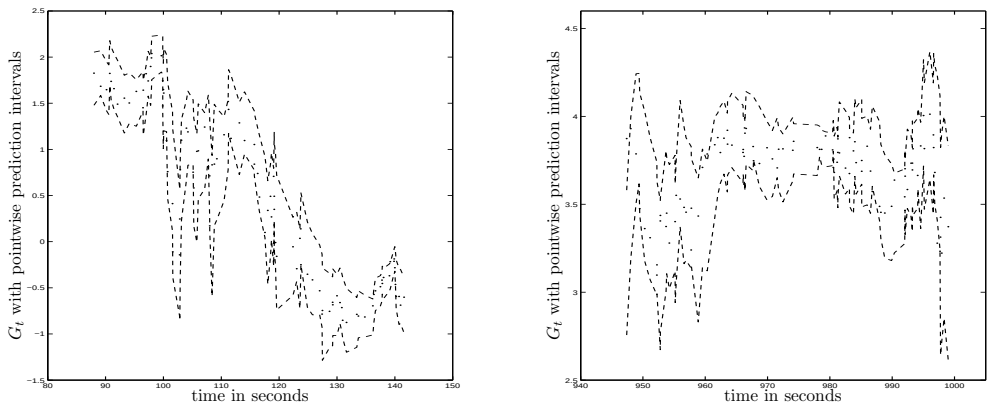

Figure 2: Simulated compound Poisson ECOGARCH$(1,1)$ observations with prediction intervals (*dotted line*) for $t_i \in [87.9, 141.6]$, $i = 201, \ldots, 300$ (*left*) and $t_i \in [947.9, 999]$, $i = n - 99, \ldots, n$ (*right*) .

we plotted the observations $G_{t_{201}}, \ldots, G_{t_{300}}$ over the time interval $[87.9, 141.6]$ and on the right hand side the last 100 observations $G_{t_{n-99}}, \ldots, G_{t_n}$ over the interval $[947.4, 999]$ together with the corresponding prediction interval for one simulation. More sophisticated evaluation methods of the quality of interval forecasts can be found in Clements and Taylor (2003).

5 Analysis of General Motors stock prices

In this section we will fit the compound Poisson ECOGARCH(1, 1) model to tick-by-tick data for the General Motors (GM) stock. The data spans four weeks starting form 6th of May 2002 until the 31st of May. Due to the Memorial Day there was no trading on the 27th at the NYSE. The data will be analysed on a daily basis. Therefore no strong seasonality effect will be present. On the other hand we have to take into account a market microstructure noise on this fine level. This will be done by considering mid quotes, which are the average of the last bid and ask quote just before the trade, as our price data. In particular this means, if we have observation points $t_1, \ldots, t_n$, then the log-price G_{t_i} is given by

$$G_{t_i} = \frac{1}{2}\left(\log(b_{t_i-}) + \log(a_{t_i-})\right) \cdot 1000\,, \qquad i = 1, \ldots, n,$$

where b_{t_i-} (a_{t_i-}) denotes the last bid (ask) quote just before time t_i. This will reduce the effect of bid-ask bounces. We also multiplied the log-price by 1 000 not to run into numerical difficulties due very small values of the volatility process. If equal transaction times occurred, the corresponding trades are combined to a single trade. Further we have omitted the first and last 5 minutes on each trading data due to possible irregularities during that time. Hence we only consider the trading between 9.35 and 15.55. The resulting log-price series together with the corresponding returns are given in Figure 3.

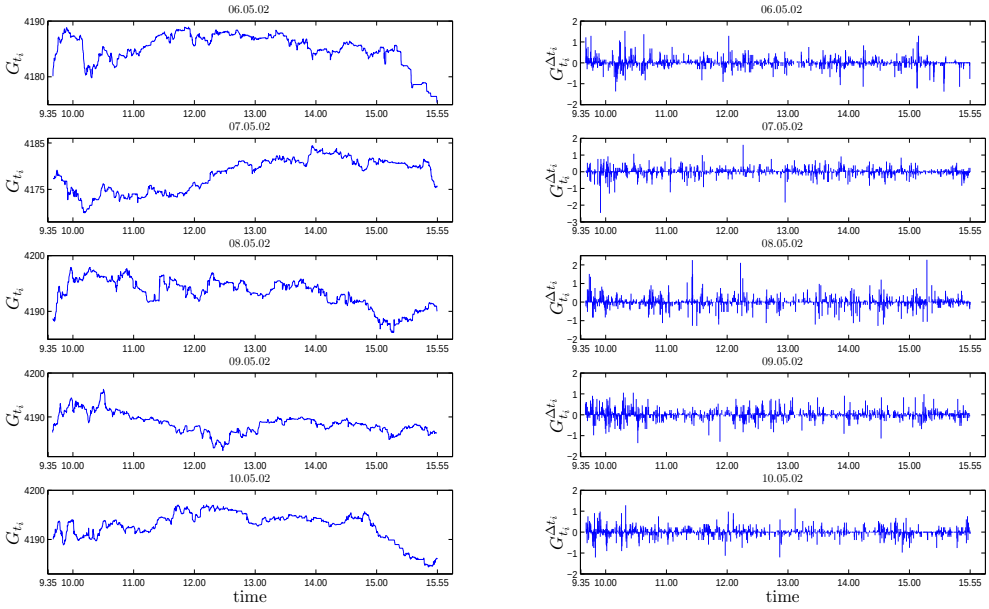

Figure 3: Tick-by-tick log-prices G_{t_i} (*left*) together with the corresponding returns $G_{t_i}^{\Delta t_i}$ (*right*) of GM over the time span 06.05.02-10.05.02.

Observe that we have transformed the observation time such that one unit corre-

sponds to 30 seconds in calendar time.

The parameters are then estimated as explained in Section 2. The results are reported in Table 9. The parameter values suggest that we have a leverage effect, which is the case if $\widehat{\theta} < 0$, on eleven of the nineteen days. For these dates we have that a positive jump to the log-price will increase the log-volatility less than a negative one ($-\widehat{\gamma} < \widehat{\theta} < 0$) as well as a positive positive jump decreasing the log-volatility while a negative one increases it ($\widehat{\theta} < -|\widehat{\gamma}|$). The estimated a_1 and γ for May 7, 2002 are significantly smaller than on the other days. Therefore we will expect rather strong dependence and no large jumps in the estimated volatility on this day.

Date	$\widehat{a}_1$	$\widehat{\theta}$	$\widehat{\gamma}$	$\widehat{\mu}$	$\widehat{\lambda}$
06.05.02	0.1720	-0.1435	0.2000	-1.8020	2.4757
07.05.02	0.0009	0.0273	0.0064	-1.4852	2.3579
08.05.02	0.2450	0.0837	0.1029	-1.6034	2.7004
09.05.02	0.1650	-0.0581	0.1754	-1.8155	2.4484
10.05.02	0.1369	-0.0160	0.1195	-2.0677	2.5888
13.05.02	0.0545	0.0327	0.0326	-2.3718	2.7663
14.05.02	0.1184	-0.0700	0.0866	-3.4486	2.5053
15.05.02	0.0645	0.0503	0.1129	-1.8520	2.4314
16.05.02	0.1329	0.0366	0.1817	-1.6567	2.5540
17.05.02	0.0406	-0.0496	0.0401	-2.0270	2.3457
20.05.02	0.0190	-0.0180	0.0549	-1.1827	2.1675
21.05.02	0.0353	-0.0474	0.0337	-2.0020	2.4313
22.05.02	0.0045	-0.0293	0.0142	-1.8056	2.0969
23.05.02	0.2314	-0.0257	0.2379	-1.7436	2.3559
24.05.02	0.1367	-0.0539	0.1666	-2.2541	1.9666
28.05.02	0.0053	0.0153	0.0107	-1.5602	2.4600
29.05.02	0.6487	-0.1167	0.0412	-2.5678	2.4108
30.05.02	0.3248	0.0794	0.2521	-1.6922	2.9888
31.05.02	0.0360	0.0451	0.1189	-0.9330	2.5922

Table 9: Estimated parameters for the GM data over the time span 06.05.02-31.05.02.

From now on we will just concentrate on the first week of observations, since the analysis on the other days is rather similar. Given the parameter estimates $(\widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n)$ we are able, due to equation (10), to estimate the volatility for the five days. The results are plotted on the left hand side in Figure 4. There one can observe the above mentioned behaviour on May 7, 2002.

Analogously to the simulation the remaining correlation in the squared fitted innovations was checked by performing a Ljung-Box test. The hypothesis of no correlation could not be rejected for the first four days, indicating a suitable fit of the data. On May 10th we rejected the hypothesis with a p-value of 0.02. Therefore we will exclude May 10th from the further analysis. These results are also confirmed by the empirical autocorrelation function of the squared fitted innovations $\widehat{Z}_i = G_{t_i}^{\Delta t_i}/\widehat{\sigma}_{t_i-}$, which can be seen on the right hand side of Figure 4.

As explained in Section 4.1 we are able to compute a prediction density

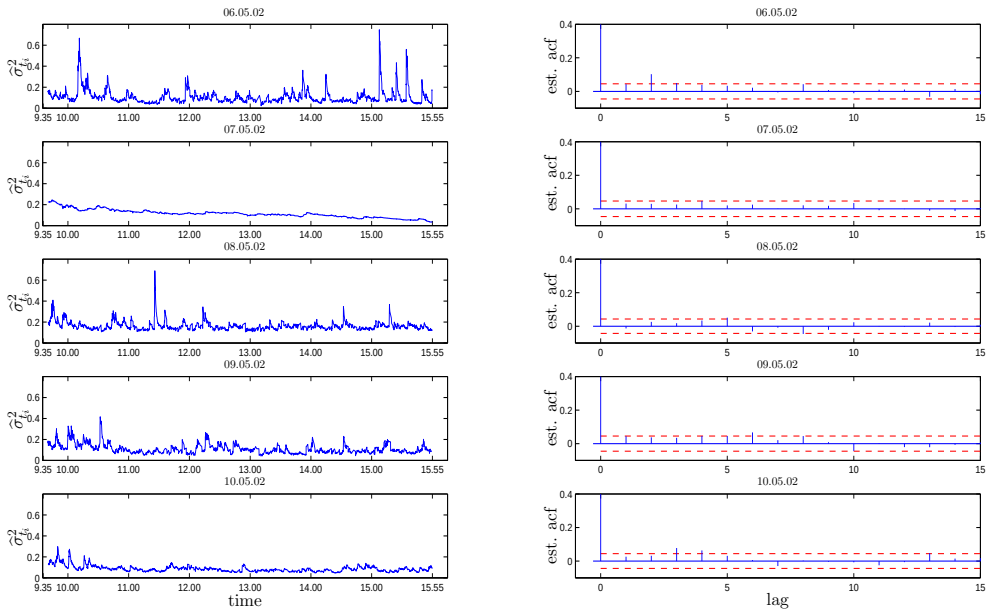

Figure 4: Estimated volatility $\widehat{\sigma}^2_{t_i}$ of GM tick-by-tick data (*left*) and estimated autocorrelation function of the squared fitted innovations (*right*) over the time span 06.05.02-10.05.02.

$p(s|\widehat{X}_{t_n}, \widehat{\boldsymbol{\vartheta}}_n, \widehat{\lambda}_n, \Delta t_{n+1})$ for the one-step ahead volatility $\sigma^2_{t_{n+1}}$. This is also done for the volatility of the last observation on each of the remaining four days. The densities are shown in Figure 5. One observes that the support of the density for May 7th is smaller than for the other days, due to the fact that log-price process on that day was not that volatile. Notice also that up to now we have seen in Figure 1 only the shape of the prediction density in the leverage case where a negative jump in the price increases the log-volatility process, while a positive one decreases it. Here we have for the May 8, 2002 an examples for the non-leverage case and on the other days the leverage case where a positive shock increases the log-volatility process less than a negative one. In the bottom panel we plotted the corresponding logarithmic prediction densities.

In Section 4.2 we introduced a prediction interval $PI(G_{t_{i+1}})$ for the one-step ahead log-price process $G_{t_{i+1}}$. For each of the four days we started with two-thirds of the observations to estimate the model. Then we calculated for each of the following time points t_i the prediction interval $PI(G_{t_{i+1}})$ for $i = \frac{2}{3}n, \ldots, n$. Afterwards we counted how many observations $G_{t_{i+1}}$ were actually observed within $PI(G_{t_{i+1}})$.

In Table 10 we report $J := \#\{G_{t_i} \in PI(G_{t_i}); i > \frac{2}{3}n\}$ the number of observations lying in the prediction interval, the percentage $J/\frac{n}{3}$ and the average prediction interval length $\widehat{PI} = \frac{3}{n}\sum_{j>\frac{2}{3}n} PI(G_{t_j})$.

The average length of the prediction intervals varies between five and six times the tick-size. Accepting this kind of uncertainty one has a coverage rate of roughly 95%. Taking significantly less than two-thirds of the observations leads to bias estimates and

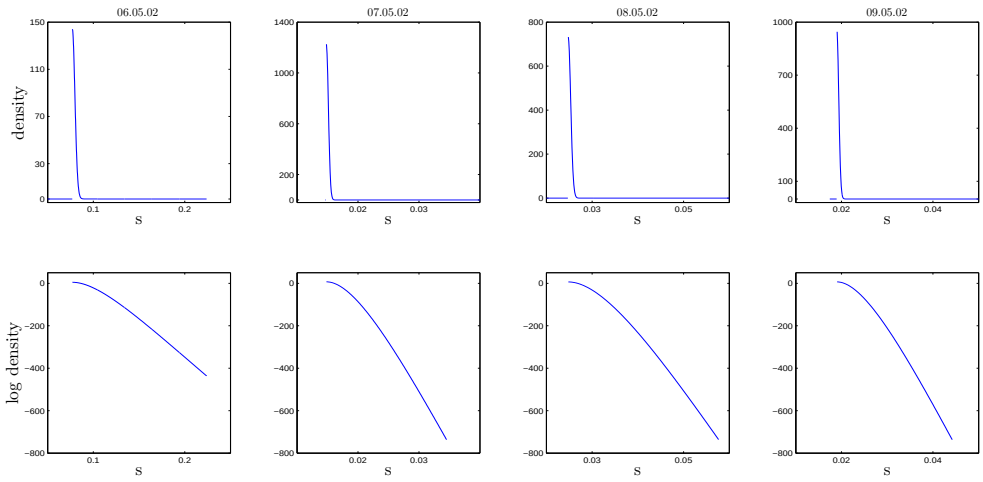

Figure 5: Estimated prediction density (*top*) and prediction log density (*bottom*) of the volatility for the last trade on the days 06.05.02 to 09.05.02.

Date	06.05.02	07.05.02	08.05.02	09.05.02
J	606	586	662	598
$J/\frac{n}{3}$	0.9543	0.9686	0.9553	0.9522
$\widehat{PI}$	0.0496	0.0602	0.0642	0.0474

Table 10: The number (percentage) of observations G_{t_i} lying in the prediction interval $PI(G_{t_i})$ on the days 06.05.02 to 09.05.02 together with the average prediction interval length.

imprecise prediction intervals. In Figure 6 we plotted the last third of the price process with the first observation centered at zero on each of the four days together with the corresponding prediction intervals.

If we decrease the prediction accuracy to 80%, the average length of the prediction interval decrease to about 65% of the former length. The percentage of observations lying in the prediction interval varies in this case between 89% and 92%.

6 Conclusions

Continuous time models are especially useful for the analysis of time series which are observed at unequally spaced time points, such as tick-by-tick trading data. Many papers dealing with continuous time models focus on model construction and theoretical properties. In contrast this paper deals with estimation and prediction in a continuous time model allowing for leverage effects. In particular the estimability and prediction capability of a compound Poisson ECOGARCH$(1,1)$ model included in the model class suggested by Haug and Czado (2007) are investigated.
A quasi maximum likelihood approach is developed for parameter estimation. It is

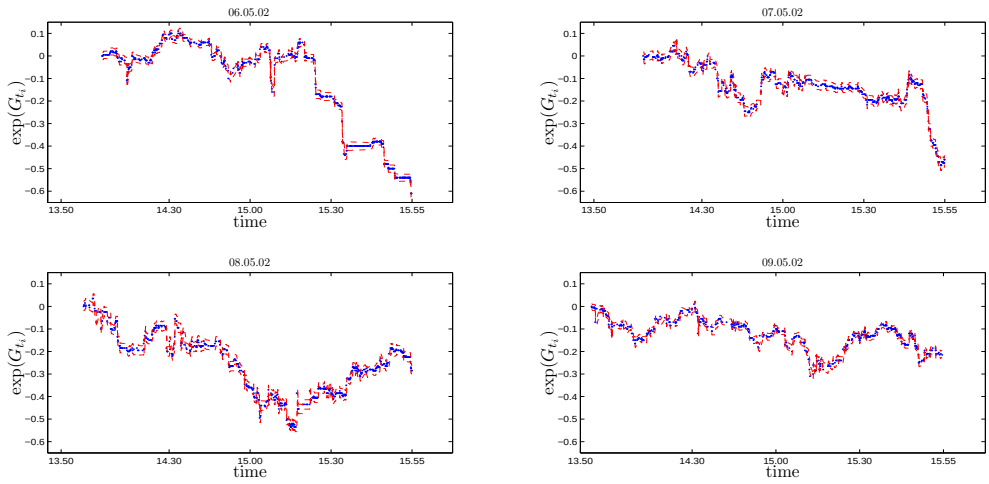

Figure 6: Price process centered at zero for the first observation on May 6th (*top left*), 7th (*top right*), 8th (*bottom left*) and 9th (*bottom left*) with corresponding prediction intervals.

based on an approximate recursion of the underlying state process of the log-volatility, thus allowing for the estimation of the log-volatility. Using a conditional representation and known volatilities quasi maximum likelihood estimates (QMLE) are constructed. The performance of these QMLE are investigated in a simulation study. This study showed that the performance is very satisfactory under normal innovations (absolute relative bias at most 7% for all investigated scenarios). In a second part of the simulations study the performance of the QMLE's is explored under t-distributed innovations. This misspecification results in a doubling of the relative bias, which is still acceptable. As a side product of the simulation study we can also quantify the leverage effect of the compound Poisson ECOGARCH$(1,1)$ model. The size of the leverage effect as characterised by the negative correlation between the current return and the future volatility increases as θ decreases and decreases as γ increases. We observe moderate correlation values of -0.25 thus showing that the model is capable of modelling such effects.
Besides estimation the second focus of the paper is prediction. For this one step ahead predictions of the volatility and the log-price are considered. For the volatility two point predictions, one based on recursion and the other on the mode of the prediction density, are constructed and compared in a simulation study. It shows a typical bias-variance trade off situation, the recursive point predictions have lower bias than the mode predictions, while the reverse is true for the mean square error. The increase in bias and mean square error is however tolerable if data with t-distributed innovations are usde instead of normal innovations thus indicating robustness.
Prediction densities of the volatility are constructed and evaluated using logarithmic scores. These are more sensitive to the misspecification of the innovation distribution. Logarithmic scores are also used to determine the number of observations necessary to achieve a reasonable prediction quality, indicating that 1500 observations are necessary. Finally predictions for the log-price are studied. Since sensible point predictions for

the log-price are not available, prediction intervals are constructed. The performance of these interval forecasts was evaluated using successively more observations for the prediction intervals showing only a slight improvement in the empirical coverage probability as n increases.
The application to General Motors stock prices demonstrates the usefulness of these models. They are able to identify and quantify leverage effect present in the data as well as quantify prediction uncertainties.

We like to note that this model assumes that all jumps are observable and that the driving Lévy process is of compound Poisson type. Although we successfully fitted the model to tick-by-tick data these assumptions certainly will not hold for all kinds of data. Therefore the development of estimation and prediction methods for the general ECOGARCH model is necessary and subject of current research.

References

Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge: Cambridge University Press.

Barndorff-Nielsen, O.E. and N.Shephard(2001). "Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion)." *Journal of the Royal Statistical Society, Series B 63*, 167–241.

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

Brockwell, P.J. (2001). "Lévy driven CARMA processes." *Annals of the Institute of Statistical Mathematics 53*, 113–124.

Brockwell, P.J. and R.A. Davis (1987). *Time Series: Theory and Methods*. New York: Springer.

Clements, M.P. and N. Taylor (2003). "Evaluating interval forecasts of high-frequency financial data." *Journal of Applied Econometrics 18*, 445–456.

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

Gneiting, T. and A.E. Raftery (2007). "Strictly proper scoring rules, prediction and estimation." *Journal of the American Statistical Association 102*, 359–378.

Haug, S. and C. Czado (2007). "An exponential continuous time GARCH process." *Journal of Applied Probability 44*, 960–976.

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

Ljung, G.M. and G.E.P. Box (1978). "On a measure of lack of fit in time series models." *Biometrika 65*, 297–303.

Nelson, D.B. (1991). "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica 59*, 347–370.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics 181. Berlin: Springer.