Predictive performance of one-step-ahead volatility forecasts using realized observation driven models

J. de Deijn (2516200) A. Kahali (2542764) S. Schimmel (2559469) L. Scholte (2500114)

Vrije Universiteit Amsterdam (FEWEB)

Case Study - MSc Econometrics and Operations Research
Supervisors: Prof. Dr. S.J. Koopman, Dr. L.F. Hoogerheide
January 29, 2016

Abstract

This paper studies several methods with the aim of modeling and forecasting volatility in a more accurate way than the classical GARCH volatility model. We extend the GARCH and GAS observation driven models and some variations on these by incorporating the realized volatility measures into these models via a measurement equation. The realized measures we consider are the realized variance, bipower variation, and the realized kernel. Furthermore we use these realized volatilities calculated over the entire sample as a benchmark for the out-of-sample analysis, where we compare the forecasting performance of our implemented models. For these comparisons we use the mean absolute error, mean squared error and the logarithmic scoring rule as metrics, and use the Diebold-Mariano test to see whether differences in forecast performance between different models are significant. We use seven years of high frequency trade data of the PEPSICO (PEP) stock. Our results indicate that the realized models outperform the the non-realized models significantly. Generally, even the worst realized models outperform the best non-realized models. We also find evidence that asymmetric models perform generally better compared to the symmetric models that do not account for leverage effects. Also realized GAS models that include the leverage effect function τ outperform those that do not. Finally, we include a comparison between the realized GARCH and realized GAS model. The realized GARCH models generally outperform the realized GAS models. Although the realized GAS models seem to fit the data well in-sample, they do not necessarily perform well when it comes to out-of-sample forecasting.

Contents

1 Introduction

Modeling and forecasting the volatility in financial markets has been a popular subject in econometric and quantitative finance literature over the past decade. Engle (1982) was the first to introduce a new class of models, called Autoregressive Conditional Heteroskedasticity (ARCH) models, that could deal with time-varying volatility processes. This gained much interest as it was observed in financial literature that stock prices include a process known as volatility clustering. In other words, stock prices of tomorrow can be more volatile if stock prices have been unusually volatile recently.

Improvements in modeling and forecasting volatility processes have been made ever since. Many volatility models are variations on the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, first introduced by Bollerslev (1986). However, research has shown that standard GARCH models perform poorly in situations where volatility changes rapidly to new levels. As Andersen et al. (2005) point out, GARCH models adjust slowly to these new levels and therefore GARCH models are inappropriate for modeling financial returns where, with the introduction of high frequency data, volatility levels of financial returns are changing fast.

In the financial econometrics arena, researchers aim to find new and more informative measures for the current level of volatility. However, the existence of jumps in high frequency stock returns and the contamination of high frequency data by microstructure noise drives the realized variance disproportionally up, which may also be unfavorable and complicates matters more. Examples of such realized measures are the realized variance (RV), bipower variation (BV) and the realized kernel (RK). In this paper, we use the so-called realized GARCH framework of Hansen et al. (2012), that combines a GARCH structure with an integrated model for realized measures of volatility. For reasons mentioned at the beginning of this paragraph, the RK is the preferred measure to use.

Using high-frequency data of PEPSICO (PEP), we compare four different classes of models (including natural extensions) for forecasting daily volatility. First we start with GARCH models. Then we propose Generalized Autoregressive Score (GAS) models, introduced by Creal et al. (2013), that are based on a score updating scheme for the variance. Lastly, we use Realized GARCH and also Realized GAS models. We obtain forecasts on a fixed window basis, as well as on a rolling window basis. The general results that we found are in line with Hansen et al. (2012) and Barndorff-Nielsen et al. (2009), namely by use of the Diebold-Mariano test we find that realized models tend to outperform the non-realized models significantly. Generally, even the worst realized models outperform the best non-realized models. Moreover, we find that asymmetric models, such as (realized) EGARCH, perform typically better compared to symmetric models, which do not account for leverage effects. Also realized GAS models that include the leverage effect function τ outperform those that do not. Lastly, we include a comparison between the realized GARCH and realized GAS model. We conclude that the realized GAS models seem less able to adapt to data for forecasting compared to the more parsimonious realized GARCH models. Nevertheless, in-sample results show that the realized GAS models fit data better in likelihood comparisons, which could be the result of over-fitting.

The remainder of this paper is organized as follows. In Section 2, we elaborate on our methodology, the derivations of our used models, the forecasting procedures and the measures for comparing the forecast performance of the different models. In Section 3, we explain the simulations studies that we used. Then in Section 4, we apply our models on empirical

PEPSICO stock data and compare the performances of all models and draw results. Finally, Section 5 provides the conclusions of this paper and will point towards some suggestions for further research.

2 Method

In this section we discuss methods for modelling the volatility of financial stock price data. The main statistical framework for daily financial returns that we make use of is the Realized General Autoregressive Conditional Heteroskedasticity (Realized GARCH) model, developed by Hansen et al. (2012). This framework combines a GARCH structure with an integrated model for realized measures of volatility, such as the realized variance (RV), bipower variation (BV) and the realized kernel (RK).
The general structure of the Realized GARCH(p,q) model provided in the Hansen et al. (2012) paper is given by the following equations:

$$r_t = \mu + \sigma_t \varepsilon_t, \tag{1a}$$
$$\sigma_t^2 = v(\sigma_{t-1}, ..., \sigma_{t-p}, x_{t-1}, ..., x_{t-q}), \tag{1b}$$
$$x_t = m(\sigma_t, \varepsilon_t, u_t), \tag{1c}$$

where r_t is the return, x_t a realized measure of volatility and $\sigma_t^2 = \mathbb{V}\text{ar}[r_t|D_{t-1}]$ with $D_{t-1} = (r_{t-1}, x_{t-1}, r_{t-2}, x_{t-2}, ...)$. Moreover, $\varepsilon_t \sim i.i.d.(0,1)$ and $u_t \sim i.i.d.(0, \sigma_u^2)$, with ε_t and u_t being mutually independent. Here, Equation (1a) is the return equation, (1b) is the GARCH equation and (1c) is referred to as the measurement equation, because this equation relates the observed realized measure to the latent volatility.

2.1 GARCH models

The returns modelled by the observation-driven GARCH(p,q) model is given by:

$$r_t = \mu + \sigma_t \varepsilon_t, \qquad \text{where} \quad \{\varepsilon_t\} \sim i.i.d. f_\varepsilon, \tag{2a}$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i (r_{t-i} - \mu)^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-i}^2. \tag{2b}$$

In this study we start with the GARCH in its most simple form, the GARCH(1,1). Because of the specific characteristics in financial stock data, such as the leverage effect, we provide a variation on this model: the EGARCH and NGARCH model. Moreover, we vary with Gaussian and Student-t distributed disturbances. The literature is rich on documentations that the Student-t distribution outperforms the normal distribution in GARCH models, as it better accounts for the fatter tails in financial asset returns. Note that depending on the nature of the disturbances we need to standardize the return equation in order for it to have $\mathbb{V}\text{ar}[r_t] = \sigma_t^2$.

2.1.1 Gaussian GARCH and GARCH-t models

The parameter restrictions that we impose on the Gaussian GARCH(1,1) are $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and in order to ensure stationarity of the model we must also have $\alpha + \beta < 1$. For the starting value σ_0^2 the unconditional variance is used, which is equal to $\frac{\omega}{1-\alpha-\beta}$.

To model the returns r_t with a Gaussian GARCH(1,1) we have that $r_t \sim \mathcal{N}(\mu, \sigma_t^2)$ and that the conditional density of r_t on past data D_{t-1} is

$$f(r_t|D_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \exp\left\{ -\frac{(r_t - \mu)^2}{2\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)} \right\}. \tag{3}$$

To estimate the vector of parameters $\boldsymbol{\theta} = (\omega, \alpha, \beta)^\top$ in the Gaussian GARCH(1,1), we use that all observations are independent and therefore the joint density function of $\mathbf{r_T} = (r_1, ..., r_T)$ is given by

$$f(\mathbf{r_T}; \boldsymbol{\theta}) = f(r_1; \boldsymbol{\theta}) \times \prod_{t=2}^{T} f(r_t | D_{t-1}; \boldsymbol{\theta}). \tag{4}$$

As a result, the average log-likelihood functions (M-estimator for $\boldsymbol{\theta}$) of the Gaussian GARCH is

$$\begin{aligned} \ell(\boldsymbol{\theta}|r_t) &= \frac{1}{T}\sum_{t=2}^{T} \log f(r_t|D_{t-1}; \boldsymbol{\theta}) \\ &= -\frac{1}{2}\frac{1}{T}\sum_{t=2}^{T}\left[\log 2\pi + \log \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2) + \frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right]. \end{aligned} \tag{5}$$

The GARCH-t model has the same updating equation for the variance as the normal GARCH, but here $\{\varepsilon_t\} \sim TID(\lambda)$. To use this model we impose the same restrictions on the parameters as for the Gaussian GARCH and add that $\lambda > 2$. Again, we start with the unconditional variance. As a consequence of the fat tails of the standardized Student-t density, the GARCH-t is a more robust model.

Using the derivations for the density function of the standardized Student-t distribution (see Appendix A), we obtain the following average log-likelihood function for the GARCH(1,1)-t that we optimize in order to get estimates for $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda)^\top$

$$\begin{aligned} \ell(\boldsymbol{\theta}|r_t) &= \frac{1}{T}\sum_{t=2}^{T} \log f(r_t|D_{t-1}; \boldsymbol{\theta}) \\ &= \frac{1}{T}\sum_{t=2}^{T}\left[\log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda-2)\right. \\ &\qquad \left. - \log\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \frac{\lambda+1}{2}\log\left(1 + \frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)(\lambda-2)}\right)\right]. \end{aligned} \tag{6}$$

2.1.2 EGARCH

The EGARCH model allows for the asymmetric effects between positive and negative asset returns, where negative shocks typically have a larger impact on the variance than positive shocks of the same size. This is called the leverage effect. The EGARCH(m,s) assumes that

$$a_t = \sigma_t \varepsilon_t, \qquad \text{where} \quad \{\varepsilon_t\} \sim i.i.d. f_\varepsilon, \tag{7a}$$

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{s} \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m} \beta_j \log(\sigma_{t-j}^2). \tag{7b}$$

Note that here a positive a_{t-i} contributes $\alpha_i(1+\gamma_i)|\varepsilon_{t-i}|$ to the log volatility, whereas a negative a_{t-i} gives $\alpha_i(1-\gamma_i)|\varepsilon_{t-i}|$. The γ_i parameter thus signifies the leverage effect of a_{t-i} and we expext γ_i to be negative (larger effect of negative shocks on future volatility). The vector of parameters is $\boldsymbol{\theta} = (\omega, \gamma, \alpha, \beta)^\top$ in the Gaussian EGARCH(1,1). As the starting value for the

variance we use the sample variance of the data. Moreover, we have an EGARCH model where $\{\varepsilon_t\} \sim TID(\lambda)$ and $\boldsymbol{\theta} = (\omega, \gamma, \alpha, \beta, \lambda)^\top$.

The log-likelihood functions of both EGARCH(1,1) models are similar to the log-likelihood functions of their standard GARCH counterparts, which are given in Equations (5) and (6) respectively.

2.1.3 NGARCH

A last extension of the GARCH framework is a non-linear GARCH (NGARCH) model. The NGARCH model is represented as

$$a_t = \sigma_t \varepsilon_t, \qquad \text{where} \quad \{\varepsilon_t\} \sim i.i.d. f_\varepsilon, \tag{8a}$$

$$\sigma_t^2 = \omega + \alpha(r_{t-1} - \gamma\sigma_{t-1})^2 + \beta\sigma_{t-1}^2. \tag{8b}$$

When optimizing the likelihood we use the following variables as input $\boldsymbol{\theta} = (\omega, \alpha, \beta, \gamma)^\top$. The assumptions for this model are $\omega > 0$, $\beta \geq 0$, $\gamma \geq 0$, and in order to ensure stationarity we must have that $\gamma + \beta < 1$. Similar to the EGARCH model, the NGARCH model accounts for the leverage effect which is that negative returns have a larger effect on future volatility than positive returns of the same magnitude. Here this is captured in the parameter γ.

We can also vary the distribution of the disturbances. The average log-likelihood of a NGARCH model with Gaussian innovations is similar to the log-likelihood of the standard GARCH, as given in Equation (5). Moreover, the disturbances can also be assumed to have a Student-t distibution. In that case the log-likelihood becomes as in Equation (6).

2.2 Realized GARCH models

A new framework, introduced by Hansen et al. (2012), is the realized GARCH, for the joint modeling of returns and realized measures of volatility. In this framework the measurement equation ties the realized measure to the latent volatility. Below we will give several linear and log linear realized models.

2.2.1 Linear Realized GARCH(1,1) models

The linear realized GARCH(1,1) model is defined by the following equations

$$r_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0,1), \tag{9a}$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \gamma x_{t-1}, \tag{9b}$$

$$x_t = \xi + \phi\sigma_t^2 + \tau(\varepsilon_t) + u_t = \xi + \phi\sigma_t^2 + \tau\left(\frac{r_t}{\sigma_t}\right) + u_t, \qquad u_t \sim i.i.d.(0, \sigma_u^2). \tag{9c}$$

Here $\tau(\varepsilon_t) = \tau_1\varepsilon_t + \tau_2(\varepsilon_t^2 - 1)$ represents a leverage function that captures the asymmetric response in volatility to return shocks, as explained before. Note that $\mathbb{E}[\tau(\varepsilon_t)] = 0$ for all distributions that have a zero mean and unit variance. For the joint log density of r_t and x_t we obtain

$$\log f(\boldsymbol{r}_T, \boldsymbol{x}_T | \boldsymbol{r}_T; \boldsymbol{\theta}) = \log f(\boldsymbol{r}_T; \boldsymbol{\theta}) + \log f(\boldsymbol{x}_T; \boldsymbol{\theta}). \tag{10}$$

The log likelihood function of $\boldsymbol{\theta} = (\omega, \beta, \gamma, \xi, \phi, \tau_1, \tau_2, \sigma_u)^\top$ is then given by

$$\ell(\boldsymbol{r}_T, \boldsymbol{x}_T; \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} \log f_r(r_t|\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) + \frac{1}{T}\sum_{t=2}^{T} \log f_x(x_t|r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}). \tag{11}$$

In this model it is natural to assume that $\omega > 0$, $\beta \geq 0$ and $\gamma \geq 0$ to ensure that the variance σ_t^2 is positive, and that $\xi > 0$ and $\phi \geq 0$ to ensure that the conditional expectation of x_t given σ_t^2 is positive. Furthermore, it is assumed that $\beta + \phi\gamma < 1$, in order to ensure the stationarity of the processes. The initial value σ_1^2 is taken equal to the sample variance of the r_t.

The log-likelihood expression for the linear realized GARCH(1,1) models depends on the distribution of ε_t and u_t. If $\varepsilon_t \sim \mathcal{N}(0,1)$ then it follows that

$$\log f_r(r_t|\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = -\frac{1}{2}\left[\log 2\pi + \log \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2) + \frac{(r_t)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right]. \tag{12}$$

On the other hand, if $\varepsilon_t \sim TID(\lambda)$, then we have that

$$\begin{aligned}\log f_r(r_t|\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda-2) \qquad 2005/06/28ver:1.3\\ &\quad - \log\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \frac{\lambda+1}{2}\log\left(1 + \frac{(r_t)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)(\lambda-2)}\right). \end{aligned} \tag{13}$$

Similarly, for x_t it follows that if $u_t \sim \mathcal{N}(0, \sigma_u^2)$

$$\begin{aligned}\log f_x(x_t|r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= -\frac{1}{2}\left[\log 2\pi + \log\sigma_u^2 + \frac{(x_t - \xi - \phi\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau(\varepsilon_t))^2}{\sigma_u^2}\right]\\ &= -\frac{1}{2}\left[\log 2\pi + \log\sigma_u^2 + \frac{\left(x_t - \xi - \phi\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau\left(\frac{r_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right)\right)^2}{\sigma_u^2}\right], \end{aligned} \tag{14}$$

and if $u_t \sim TID(\lambda_u)$ it becomes

$$\begin{aligned}\log f_x(x_t|r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi \qquad 2005/06/28ver\\ &\quad - \frac{1}{2}\log(\lambda_u-2) - \log\sigma_u - \frac{\lambda_u+1}{2}\log\left(1 + \frac{(x_t - \xi - \phi\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau(\varepsilon_t))^2}{\sigma_u^2(\lambda_u-2)}\right)\\ &= \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u-2) \qquad 2005/06/28ver\\ &\quad - \log\sigma_u - \frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(x_t - \xi - \phi\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau\left(\frac{r_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right)\right)^2}{\sigma_u^2(\lambda_u-2)}\right). \end{aligned} \tag{15}$$

Therefore the log-likelihood of the linear realized GARCH(1,1) model is always a combination of Equations (12), (13), (14) and (15) depending on the distributions of ε_t and u_t.

2.2.2 Log Linear Realized GARCH(1,1)

Note that in the above linear realized GARCH(1,1) framework, it cannot be garuanteed that x_t (and also σ_t^2) is strictly positive. A negative value for u_t can make x_t negative, as a result of which σ_{t+1}^2 can become negative. Better would be to have an explicit restriction for $x_t > 0$, such as $u_t > -(\xi + \phi\sigma_t^2 + \tau(\varepsilon_t))$. Another natural solution to this is to take logarithms. Yet we arrive at the log linear realized GARCH(1,1) models, given by

$$r_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0,1), \tag{16a}$$

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \gamma \log x_{t-1}, \tag{16b}$$

$$\begin{aligned} \log x_t &= \xi + \phi \log \sigma_t^2 + \tau(\varepsilon_t) + u_t \\ &= \xi + \phi \log \sigma_t^2 + \tau\left(\frac{r_t}{\sigma_t}\right) + u_t, \qquad u_t \sim i.i.d.(0, \sigma_u^2), \end{aligned} \tag{16c}$$

where we may assume $\gamma > a$, $\phi > a$ and $\beta > a$ for some $a > 0$ and we assume that $(\beta + \phi\gamma) < 1$ to ensure a stationary process. Again, the initial value $\log(\sigma_1^2)$ is equal to the logarithm of the sample variance of the returns. Due to the logarithm of σ_t^2 and x_t in (16b) and (16c), ω and ξ can be negative, so that these are unrestricted (just like τ_1 and τ_2).

For the log likelihood of the joint density of r_t and x_t we need to find similar expressions as in the previous subsection. Note that the log densities for r_t stay the same as before in (12) and (13). For x_t we obtain for normal and Student-t distributed u_t the following expressions respectively

$$\begin{aligned} &\log f_x(x_t | r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = \\ &\quad -\frac{1}{2}\left[\log 2\pi + \log \sigma_u^2 + \frac{\left(\log x_t - \xi - \phi \log \sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau\left(\frac{r_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right)\right)^2}{\sigma_u^2}\right], \end{aligned} \tag{17}$$

$$\begin{aligned} &\log f_x(x_t | r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = \log \Gamma\left(\frac{\lambda_u + 1}{2}\right) - \log \Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log \pi \\ &\quad - \frac{1}{2}\log(\lambda_u - 2) - \log \sigma_u \\ &\quad - \frac{\lambda_u + 1}{2} \log\left(1 + \frac{\left(\log x_t - \xi - \phi \log \sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \tau\left(\frac{r_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right). \end{aligned} \tag{18}$$

2005/06/28*ver*

2.2.3 Log Linear Realized EGARCH(1,1)

We define the log linear realized EGARCH(1,1) model as

$$r_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0,1), \tag{19a}$$

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \alpha\tau(\varepsilon_{t-1}) + \gamma u_{t-1}, \tag{19b}$$

$$\log x_t = \xi + \phi \log \sigma_t^2 + \tau(\varepsilon_t) + u_t, \qquad u_t \sim i.i.d.(0, \sigma_u^2). \tag{19c}$$

Now β is the persistence parameter, $\tau(\varepsilon_t)$ is the leverage effect, and $\tau(\varepsilon_{t-1}) + \gamma u_{t-1}$ captures the volatility shock.

The log likelihood function of the realized EGARCH can be obtained as the sum

$$\ell(\boldsymbol{r}_T, \boldsymbol{x}_T; \boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T} \log f_r(r_t|\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) + \frac{1}{T}\sum_{t=2}^{T} \log f_x(x_t|r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}), \tag{20}$$

for which the log linear expressions for $f_r(r_t|\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta})$ and $f_x(x_t|r_t, \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta})$ remain the same as obtained in equations (12), (13) and (17), (18).

2.3 The GAS framework

"*GAS is only for smart people.*" S.J. Koopman (January 2016).

The Generalized Autoregressive Score (GAS) framework, developed by Creal et al. (2013), has a time-varying factor f_t that determines the variance in stock returns. The general framework is given by

$$a_t = g(f_t, \varepsilon_t), \quad \text{with} \quad \{\varepsilon_t\} \sim i.i.d. f_\varepsilon, \tag{21a}$$

$$f_{t+1} = \omega + \alpha s(a_t, f_t) + \beta f_t, \tag{21b}$$

where $s(a_t, f_t)$ is a GAS shock. This shock is a scaled score function and the scaling matrix here is the Fisher information matrix and therefore $s(a_t, f_t)$ is given by

$$s(a_t, f_t) = \nabla_t S(f_t), \quad \nabla_t = \frac{\partial \log f_a(a_t|f_t; \boldsymbol{\theta})}{\partial f_t}, \quad S(f_t) = \mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]^{-1}. \tag{22}$$

2.3.1 GAS-t model

Given that financial returns are generally not normally distributed and typically have an excess kurtosis and long tails, a Student-t disribution for the disturbances of the error is desirable (Franses and Van Dijk, 2000). Creal et al. (2013) show that the GAS model uses the non-normality of the data in order to determine that time varying update function. The GARCH updates the volatility with the squared returns (which is only a good algorithm in the case of thin tails because then large excess returns are attributed to excess returns) such that extreme realizations could be nothing more than a result of the large mass in the fat-tailed density rather than the possible increased volatility.

If we apply the GAS framework with a time-varying variance $\left(f_t = \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)\right)$ to the log-density of the standardized Student-t distribution we obtain the GAS(1,1)-t model, which is defined as

$$r_t = \mu + a_t = \mu + \sigma_t\sqrt{\frac{\lambda-2}{\lambda}}\varepsilon_t, \quad \varepsilon_t \sim TID(\lambda), \tag{23a}$$

$$f_{t+1} = \omega + \alpha\left[\left(\frac{\lambda+1}{\lambda-2}\right)\left(1 + \frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - f_t\right]\left(\frac{\lambda+3}{\lambda}\right) + \beta f_t. \tag{23b}$$

The vector of parameters is given by $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda)^\top$, where $\omega > 0$, $\alpha < \beta < 1$ and $\lambda > 2$. In Appendix B, one can find the derivations of the updating equation in (23b). Note that

in this case r_t is scaled in such a way that $\mathbb{V}\text{ar}[r_t] = \sigma_t^2 = f_t$. Now we can clearly observe that the GAS(1,1)-t model is different from the GARCH(1,1)-t, because it puts less weight on extreme observations of $(r_t - \mu)$, and now extreme observations are less likely to be induced by a substantial higher volatility but rather the result of the heavier tails of the standardized Student-t distribution.

The log-likelihood function that is optimized to find estimates is similar to one of the GARCH(1,1)-t in Equation (93).

2.3.2 Skewed GAS-t model

If, besides fat tails, we also want to take into account the asymmetry that is observed in financial asset returns, we apply the GAS framework to the asymmetric Student-t disturbances of Zhu and Galbraith (2010). In this case you acknowledge the fact that the distribution of the error terms is skewed towards negative or positive values. The standardized skewed Student-t (SST) GAS(1,1) model is given by

$$r_t = \mu + a_t = \mu + \sigma_t \frac{\varepsilon_t}{\sqrt{Q(\lambda,\alpha)}}, \quad \varepsilon_t \sim SST(\lambda), \tag{24a}$$

$$f_{t+1} = \omega + \alpha \left[\frac{(\lambda+1)Q(\lambda,\alpha)}{4\lambda(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu))^2} \left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu))^2}\right)^{-1} (r_t-\mu)^2 - f_t \right] \times \left(\frac{\lambda+3}{\lambda}\right) + \beta f_t. \tag{24b}$$

The derivations of the variance update equation can be found in Appendix D. The log-likelihood function (see Appendix C) is derived as

$$\begin{aligned} \ell(\boldsymbol{\theta}|r_t) &= \frac{1}{T}\sum_{t=2}^{T} \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) \\ &= \frac{1}{T}\sum_{t=2}^{T} \left[\log K(\lambda) + \frac{1}{2}\log Q(\lambda,\alpha) - \frac{1}{2}\log f_t \right. \\ &\quad \left. - \frac{\lambda+1}{2}\log\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)\right]. \end{aligned} \tag{25}$$

2.3.3 GAS(1,1)-Laplace

Another possible more fat-tailed distribution relative to the Gaussian distribution is that the disturbances follow a standardized Laplace distribution. In this case, the model captures the leptokurtic behaviour of stock returns; moderate changes happen less frequently because historical values are clustered around the mean. It also means that large fluctuations are more likely within the fat tails. The GAS(1,1)-Laplace model is defined as

$$a_t = \sigma_t \frac{\varepsilon_t}{\sqrt{2}b}, \qquad \text{where} \qquad \varepsilon_t \sim \mathcal{L}(b), \tag{26a}$$

$$f_{t+1} = \omega + \alpha\left[-2f_t + 2\sqrt{2}\sqrt{f_t}|a_t|\right] + \beta f_t. \tag{26b}$$

These derivations can be found in Appendix F. Moreover, the log-likelihood of the GAS(1,1)-Laplace, as in Appendix E, is derived as

$$\ell(\boldsymbol{\theta}|r_t) = \frac{1}{T}\sum_{t=2}^{T}\log f_r(r_t|f_t, D_{t-1};\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=2}^{T}\left[-\frac{1}{2}\log 2 - \frac{1}{2}\log f_t - \frac{|r_t|\sqrt{2}}{\sqrt{f_t}}\right]. \tag{27}$$

2.3.4 GAS(1,1)-GED

The final GAS model that we consider is the GAS(1,1)-GED which is a member of the exponential family and incorporates a normal and Laplace distribution. The density function of a general GED distribution is given by:

$$f(x_t;\nu,\lambda) = \frac{\nu}{2\lambda\Gamma(\nu^{-1})}\exp\left(-\left|\frac{x_t}{\lambda}\right|\right). \tag{28}$$

The distribution is defined by three parameters: $\mu = 0$ is the mode of the distribution; $\lambda \in (0,\infty)$ which defines the dispersion of the distribution and $\nu \in (0,\infty)$ which controls the skewness. Then it follows that for $\nu = 2$ we obtain a normal distribution with mean zero and variance $\frac{\lambda^2}{2}$, for $\nu = 1$ we obtain the Laplace distribution and for $\nu \to \infty$ we have a uniform distribution on the interval $(-\lambda, \lambda)$.

The GAS(1,1)-GED model is defined as (see also Appendix H)

$$r_t = \sigma_t\varepsilon_t\sqrt{\frac{\Gamma(\nu^{-1})}{\lambda^2\Gamma(3\nu^{-1})}}, \qquad \text{where} \qquad \varepsilon_t \sim GED(\lambda,\nu), \tag{29a}$$

$$f_{t+1} = \omega + \alpha\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}-1}} - \frac{2f_t}{\nu}\right] + \beta f_t. \tag{29b}$$

The log-likelihood (see Appendix G) of the GAS(1,1)-GED used for optimization in the estimation of the parameters is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}|r_t) &= \frac{1}{T}\sum_{t=2}^{T}\log f_r(r_t|f_t, D_{t-1};\boldsymbol{\theta}) \\ &= \frac{1}{T}\sum_{t=2}^{T}\left[\log K(\nu) - \frac{1}{2}\log\sigma_t^2(\boldsymbol{\theta},\sigma_1^2) - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\left(\frac{1}{\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)}\right)^{\frac{\nu}{2}}\right], \\ &\text{where } K(\nu) = \frac{\nu}{\psi(\nu)2^{1+\frac{1}{\nu}}\Gamma(\nu^{-1})}, \quad \psi(\nu) = \left(\frac{2^{-\frac{2}{\nu}}\Gamma(\nu^{-1})}{\Gamma\left(\frac{3}{\nu}\right)}\right)^{\frac{1}{2}}.\end{aligned} \tag{30}$$

2.4 Realized GAS models

As for the GARCH models, also implement the Realized versions of the GAS models discussed above. Note that for the GAS models we only make use of log linear Realized models. In the linear Realized GARCH models we assumed Student-t distributed disturbances most of the time, but it turns out that this is not a realistic assumption to make (see Subsection 4.2). Better would have been for example a one-sided Gamma distribution. This in combination

with the simulation methods that we use, log linear models are better suited for our purposes and therefore the linear realized models are disregarded here. Moreover, note that the log linear realized GAS(1,1)-Skewed-t is not further discussed here, since the estimated skewness parameter turned out to be 0.499. A value of 0.50 corresponds to the symmetric case of this specification of the skewed t distribution. In all the realized GAS models we have taken the scaling matrix to be $S(f_t) = 1$ for simplicity of the derivations, which is has less impact in the log linear setting. In the log linear case where we use $\log f_t$ instead of f_t we have that for the non-realized GAS models the score is a function of the standardized residuals, such that the data is already desirably scaled. In that case letting $S(f_t) = 1$ is less bothersome. Moving in the same direction we have chosen to use the log linear version for all the realized GAS models.

2.4.1 Log Linear Realized Gaussian GAS(1,1)

Then, the Gaussian log linear realized GAS(1,1) model is given by

$$r_t = \sqrt{f_t}\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \tag{31a}$$

$$\log x_t = \xi + \phi \log f_t + \tau(\varepsilon_t) + \sqrt{\frac{\lambda_u - 2}{\lambda_u}}\sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{31b}$$

$$\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) + \tau_2\left(\frac{r_t^2}{f_t} - 1\right).$$

The average conditional log likelihood function of this model is then given by (see Appendix I)

$$\begin{aligned}
\ell(\boldsymbol{r_T}, \boldsymbol{x_T}; \boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r_T}, \boldsymbol{x_T} | f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r_T} | f_t, D_{t-1}; \boldsymbol{\theta}) + \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{x_T} | \boldsymbol{r_T}, f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \Bigg[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log f_t - \frac{r_t^2}{2f_t} + \log\Gamma\left(\frac{\lambda_u + 1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) \\
&\quad -\frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2) - \log\sigma_u \\
&\quad -\frac{\lambda_u + 1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)\Bigg].
\end{aligned} \tag{32}$$

In Appendix I we have also derived the corresponding variance updating equation, which is given by

$$
\begin{aligned}
f_{t+1} = \omega + \alpha \Bigg\{ \Bigg[& -\frac{1}{2f_t} + \frac{r_t^2}{2f_t^2} - \frac{\lambda_u + 1}{2} \\
& \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)^2}{\sigma_u^2 (\lambda_u - 2)} \right)^{-1} \\
& \times \left(\frac{2 \left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)}{\sigma_u^2 (\lambda_u - 2)} \right) \\
& \times \left(-\frac{\phi}{f_t} - \frac{1}{2} \tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2 \left(\frac{r_t^2}{f_t^2} \right) \right) \Bigg] \times f_t \Bigg\} + \beta f_t.
\end{aligned} \tag{33}
$$

Note that for all realized GAS models (also the ones that follow), it holds that $\mathbb{E}\left[\tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right] = 0$. Therefore we could also consider to exclude $\tau\left(\frac{r_t}{\sqrt{f_t}}\right)$ from the model, which makes matters less complex. However, as we wanted to test the empirical effect of $\tau\left(\frac{r_t}{\sqrt{f_t}}\right)$ we considered GAS models both with and without $\tau\left(\frac{r_t}{\sqrt{f_t}}\right)$.

2.4.2 Log Linear Realized GAS(1,1)-t

A natural next step is to obtain the realized GAS(1,1) model where also $\varepsilon_t \sim TID(\lambda)$. In that case the realized GAS model is given by

$$
r_t = \sqrt{f_t} \sqrt{\frac{\lambda - 2}{\lambda}} \varepsilon_t, \qquad \varepsilon_t \sim TID(\lambda), \tag{34a}
$$

$$
\log x_t = \xi + \phi \log f_t + \tau(\varepsilon_t) + \sqrt{\frac{\lambda_u - 2}{\lambda_u}} \sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{34b}
$$

$$
\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) + \tau_2 \left(\frac{r_t^2}{f_t} - 1 \right).
$$

The average conditional log likelihood expression and the updating equation of this model are derived in Appendix J and are given by

$$
\begin{aligned}
\ell(\boldsymbol{r}_T, \boldsymbol{x}_T; \boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r}_T, \boldsymbol{x}_T | f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r_T} | f_t, D_{t-1}; \boldsymbol{\theta}) + \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{x_T} | \boldsymbol{r_T}, f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \Bigg[-\log \pi + \log \Gamma\left(\frac{\lambda+1}{2}\right) - \log \Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log(\lambda-2) - \frac{1}{2}\log f_t \\
&\quad -\frac{\lambda+1}{2}\log\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right) + \log \Gamma\left(\frac{\lambda_u+1}{2}\right) - \log \Gamma\left(\frac{\lambda_u}{2}\right) \\
&\quad -\frac{1}{2}\log(\lambda_u - 2) - \log \sigma_u \\
&\quad -\frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)\Bigg].
\end{aligned} \tag{35}
$$

$$
\begin{aligned}
\log f_{t+1} = \omega + \alpha \Bigg\{ \Bigg[&-\frac{1}{2f_t} - \frac{\lambda+1}{2}\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right)^{-1}\left(-\frac{r_t^2}{(\lambda-2)f_t^2}\right) \\
&-\frac{\lambda_u+1}{2} \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)^{-1} \\
&\times \left(\frac{2\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)}{\sigma_u^2(\lambda_u - 2)}\right) \\
&\times \left(-\frac{\phi}{f_t} - \frac{1}{2}\tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2\left(\frac{r_t^2}{f_t^2}\right)\right)\Bigg] \times f_t \Bigg\} + \beta f_t.
\end{aligned} \tag{36}
$$

2.4.3 Log Linear Realized GAS(1,1)-GED

The last realized GAS model that we turn to is the log linear realized GAS(1,1)-GED. This model is denoted by

$$
r_t = \sigma_t \varepsilon_t \sqrt{\frac{\Gamma(\nu^{-1})}{\lambda^2 \Gamma(3\nu^{-1})}}, \qquad \text{where} \qquad \varepsilon_t \sim GED(\lambda, \nu), \tag{37a}
$$

$$
\log x_t = \xi + \phi \log f_t + \sqrt{\frac{\lambda_u - 2}{\lambda_u}}\sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{37b}
$$

$$
\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) + \tau_2\left(\frac{r_t^2}{f_t} - 1\right).
$$

The average conditional log likelihood expression and the updating equation of this model are derived in Appendix K and are given by

$$
\begin{aligned}
\ell(\boldsymbol{r}_T, \boldsymbol{x}_T; \boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r}_T, \boldsymbol{x}_T | f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{r}_T | f_t, D_{t-1}; \boldsymbol{\theta}) + \frac{1}{T}\sum_{t=2}^{T} \log f(\boldsymbol{x}_T | \boldsymbol{r}_T, f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T} \Bigg[\log K(\nu) - \frac{1}{2}\log f_t - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu} \left(\frac{1}{f_t}\right)^{\frac{\nu}{2}} \\
&\quad + \log\Gamma\left(\frac{\lambda_u + 1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2) \\
&\quad - \log\sigma_u - \frac{\lambda_u + 1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)\Bigg], \qquad (38)
\end{aligned}
$$

$$
\begin{aligned}
\log f_{t+1} = \omega + \alpha \Bigg\{ \Bigg[& -\frac{1}{2f_t} + \frac{\nu}{4}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu} \frac{1}{f_t^{\frac{\nu}{2}+1}} - \frac{\lambda_u + 1}{2} \\
& \times \left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)^{-1} \\
& \times \left(\frac{2\left(\log x_t - \xi - \phi\log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)}{\sigma_u^2(\lambda_u - 2)}\right) \\
& \times \left(-\frac{\phi}{f_t} - \frac{1}{2}\tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2\left(\frac{r_t^2}{f_t^2}\right)\right)\Bigg] \times f_t \Bigg\} + \beta f_t. \qquad (39)
\end{aligned}
$$

2.5 Realized Volatility Measures

In this subsection we discuss realized variance measures, which are generally denoted by x_t in the realized models. These realized measures are calculated using high frequency stock price data and we forecast these realized returns using our models. We discuss three different realized measures: (i) realized variance (RV), (ii) bipower variation (BV) and (iii) realized kernel (RK). We will elaborate on each of these realized measures below.

2.5.1 Realized Variance

The most obvious realized measure for stock returns (variation in underlying stock prices) is probably the realized variance. The realized variance on day t is given by

$$
RV^{(t)} = \sum_{i=1}^{n} r_{i,t}^2, \qquad (40)
$$

where $r_{i,t} = 100 \times (\log p_{i,t} - \log p_{i-1,t})$ is the intraday return for time interval $i = 1, 2, ..., n$ on day $t = 1, 2, ..., T$. In order to do this, we used 5-minute frequency for the realized variance data.

2.5.2 Bipower Variation

A second realized volatility measure is the bipower variation. The bipower variation on day t is denoted as

$$BV^{(t)} = \frac{1}{\mu_1^2} \sum_{i=2}^{n} |r_{i,t}| \times |r_{i+1,t}|, \tag{41}$$

where $\mu \sim \mathcal{N}(0,1)$ and $\mu_1 = \mathbb{E}|\mu| = \sqrt{\frac{2}{\pi}}$. Again, $r_{i,t} = 100 \times (\log p_{i,t} - \log p_{i-1,t})$ is the intraday return for time interval $i = 1, 2, ..., n$ on day $t = 1, 2, ..., T$. Again, we used 5-minute intervals. As one can observe, the BV multiplies a certain return by the next/previous return (instead of a squared return), which leads to the fact that a large jump has less effect on the BV that day compared to the RV. Therefore the BV is regarded as a more robust realized measure than the RV.

2.5.3 Realized Kernel

A third measure of realized volatility that is discussed here is the RK. The RK is an even more robust realized volatility measure compared to RV and BV and was first introduced by Barndorff-Nielsen and Shephard (2006). They argued that high-frequency prices are measured with microstructure noise and jumps, which, as a result, actually overestimates the variance. Therefore the RV (and BV) may not be a good measure. The realized kernel is given by

$$RK^{(t)}(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \qquad \gamma_h = \sum_{j=|h|+1}^{n} r_{j,t} \times r_{j-|h|,t}. \tag{42}$$

Here $r_{j,t}$ are the intraday returns for $j = 1, 2, ..., n$ on day $t = 1, 2, ..., T$ and $k(x)$ is a kernel weight function. As Barndorff-Nielsen et al. (2009) propose in their article, we will use the Parzen kernel function, which is given by

$$k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & 0 \leq |x| \leq \frac{1}{2} \\ 2(1-|x|)^3 & \frac{1}{2} \leq |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{43}$$

Note that, in contrast to the RV and the BV, we do not need a frequency in order to calculate the RK. The optimal bandwith, as discussed in Barndorff-Nielsen et al. (2009), is defined by

$$H^* = c^* \xi^{\frac{4}{5}} n^{\frac{3}{5}}, \qquad c^* = \frac{k''(0)^2}{k(0,0)}, \tag{44}$$

$$\xi^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma_u^4 du}}, \tag{45}$$

where $c^* = \left(\frac{12^2}{0.269}\right)^{\frac{1}{5}} = 3.5134$. Note that the result for H^* is rounded to the nearest integer.

We proceed as follows:

$$\hat{\xi}^2 = \frac{\hat{\omega}}{\hat{IV}}, \tag{46}$$

$$\hat{\omega}^2_{(i)} = \frac{RV^{(i)}_{\text{dense}}}{2n_{(i)}}, \qquad i = 1, 2, ..., q, \tag{47}$$

$$\hat{\omega}^2 = \frac{1}{q}\sum_{i=1}^{q} \hat{\omega}^2_{(i)}, \tag{48}$$

$$\hat{IV} = RV_{\text{sparse}}, \tag{49}$$

$$RV_{\text{sparse}} = \frac{1}{1200}\sum_{i=1}^{1200} RV^{(i)}_{\text{sparse}}. \tag{50}$$

Here $\hat{\omega}^2$ is an estimate for ω^2, which is the variance in the log returns caused by microstructure noise. We calculate different RV_{dense} by using shifting intervals of, on average, 2 minutes. So, in other words, q is such that every q-th observation is, on average, 2 minutes apart ($q \approx 25$). Thus for example,

$$RV^{(1)}_{\text{dense}} = \sum_{i=1}^{N} r_i^2,$$

where r_i is the i-th return given by $i = 0, 25, 50, ...$, and

$$RV^{(2)}_{\text{dense}} = \sum_{i=1}^{N} r_i^2,$$

where $i = 1, 26, 51, ...$, and so forth. Similarly, to calculate RV_{sparse} we repeat the same steps as for the calculation of the different $RV^{(i)}_{\text{dense}}$, but instead of $q = 25$, we use $q = 1200$.

2.6 Forecasting

The realized volatility measures, which are denoted by x_{t+1} on day $t+1$, are used as benchmark to evaluate the performance of the volatility forecasts. That is, the realized measure is compared to the variance forecast by using the models described in Section 2. For these forecast computations we use both a rolling and fixed window procedure, with 880 in-sample observations and we construct 882 out-of-sample (883 for the open-to-close returns) one-step-ahead forecasts. Next, we will use the Mean Squared Error (MSE), Mean Absolute Error (MAE) and the Diebold-Mariano (DM) test statistics to compare the performance of these models.

2.6.1 Predictive accuracy forecasts

We can compare the forecast performance of different models discussed above by means of comparing the MAE or MSE, which are based on forecast errors. For the open-to-close data we used a correction factor c, which is calculated as

$$c = \frac{\mathbb{V}\text{ar}[r_t^{o-c}] + \mathbb{V}\text{ar}[r_t^{c-c}]}{\mathbb{V}\text{ar}[r_t^{o-c}]}. \tag{51}$$

Then the absolute errors for open-to-close and close-to-close on day t are denoted by

$$AE_t^{o-c} = |c\hat{\sigma}_t^{2(o-c)} - cx_t| \quad t = 1, ..., K, \tag{52a}$$

$$AE_t^{c-c} = |\hat{\sigma}_t^{2(c-c)} - cx_t| \quad t = 1, ..., K, \tag{52b}$$

where $\hat{\sigma}_t^2$ is the variance forecast using one of the models above and x_t is a benchmark realized volatility measure. Then the MAE is calculated as

$$MAE = \frac{1}{K}\sum_{t=1}^{K} AE_t. \tag{53}$$

Similarly, the MSE can be obtained as

$$MSE = \frac{1}{K}\sum_{t=1}^{K} SE_t, \tag{54}$$

where

$$SE_t^{o-c} = (c\hat{\sigma}_t^{2(o-c)} - cx_t)^2 \quad t = 1, ..., K, \tag{55a}$$

$$SE_t^{c-c} = (\hat{\sigma}_t^{2(c-c)} - cx_t)^2 \quad t = 1, ..., K. \tag{55b}$$

Note that in general the MAE is preferred over the MSE, as it is more robust to outliers. The MAE assigns equal weight to the data, whereas the MSE emphasized extremes (i.e. square of a number smaller than 1 becomes even smaller, and large numbers become even larger).

2.6.2 Diebold Mariano Test

The Diebold-Mariano test is used to test whether differences in forecast performances between different models are significant. Denote the one-step-ahead forecast error at time $t+1$ given time t as

$$\varepsilon_{t+1|t} = \hat{x}_{t+1|t} - x_{t+1}, \tag{56}$$

where $\hat{x}_{t+1|t}$ is the variance forecast (using a certain model from section 2.1-2.4) and x_{t+1} is a realized benchmark measure. Then we can measure the accuracy of the forecast by the absolute errors loss function and the squared errors loss function, which are respectively defined as

$$L(x_{t+1}, \hat{x}_{t+1|t}) = L(\varepsilon_{t+1|t}) = |\varepsilon_{t+1|t}|, \tag{57a}$$

$$L(x_{t+1}, \hat{x}_{t+1|t}) = L(\varepsilon_{t+1|t}) = (\varepsilon_{t+1|t})^2. \tag{57b}$$

In this article, we use the DM statistic by comparing absolute error outcomes and squared error outcomes of two models with the same realized volatility benchmark. This test is based on two loss differentials (absolute errors and squared errors)

$$d_t = L(\varepsilon_{t+1|t}^i) - L(\varepsilon_{t+1|t}^j) = |\varepsilon_{t+1|t}^i| - |\varepsilon_{t+1|t}^j|, \tag{58a}$$

$$d_t = L(\varepsilon_{t+1|t}^i) - L(\varepsilon_{t+1|t}^j) = (\varepsilon_{t+1|t}^i)^2 - (\varepsilon_{t+1|t}^j)^2, \tag{58b}$$

where $\varepsilon_{t+1|t}^i$ denotes the one-step-ahead forecast error at time t for model i.

The DM test statistic is given by

$$DM = \frac{\bar{d}}{s_d}, \tag{59}$$

under the null hypothesis of $\mathbb{E}[d_t] = 0$. Here s_d is the robust Newey-West standard error to account for the presence of autocorrelation in the time series of d_{t+1}. This autocorrelation is likely given the fact that if a model is superior at time t, it is more likely that this model will also produce superior forecasts in subsequent periods.

The Newey-West standard errors are obtained as

$$\frac{1}{n}(X^\top X)^{-1}\hat{V}(X^\top X)^{-1}, \tag{60}$$

where $\hat{V}$ is defined as

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i^\top + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{j-i} e_i e_j (x_i x_j^\top + x_j x_i^\top), \tag{61}$$

and where the Bartlett kernel with bandwidth $4(T/100)^{\frac{2}{9}}$ is used, i.e. $w_{j-i} = 1 - \frac{j-i}{B}$ for $j - i < B$ and $w_{j-i} = 0$ for $h \geq B$, where B equals the integer obtained as the rounded value $[4(T/100)^{\frac{2}{9}}]$.

2.6.3 Logarithmic Scoring Rule

A different measure for assessing the predictive accuracy of competing density forecasts is the logarithmic scoring rule (LSR), proposed by Diks et al. (2010). The score is calculated as the logarithm of the probability estimate for the actual outcome.

$$S_t(\hat{f}_t; r_{t+1}) = \log \hat{f}_t(r_{t+1}) \tag{62}$$

The LSR is calculated by taking r_{t+1} (close-to-close and open-to-close data) in combination with a forecasted $\hat{\sigma}^2_{t+1}$ and the estimated parameters $\hat{\boldsymbol{\theta}}$ and plug this into the log-likelihood function of a particular model. Then the average of all these log-likelihoods (one for every forecast) is the LSR. Next, we can apply a DM test again to assess which differences between competing models are significant. The test statistic that is being used here is

$$t = \frac{\bar{d}^*}{s_d}, \tag{63}$$

where $\bar{d}^*$ denotes the sample average of two log scoring rules differences. The advantage of the LSR compared to other measures is that it is unimportant whether the realized volatility measure x_t (RK_t, RV_t or BV_t) is an accurate measure for the true variance σ_t^2 or not. Only the quality of the forecast of the density of r_{t+1} plays a role. Even if x_t lies on a different level than σ_t^2, it could be worthwhile to obtain LSR estimated for Realized models as long as the realized volatility measure is correlated with σ_t^2.

3 Simulations

In this section we discuss the different simulation methods that have been used to ensure the correct implementation of the realized measures and the (realized) volatility models. By simulating high-frequency data using the Geometric Brownian Motion, we can validate the different volatility models. This is described in Subsection 3.1. The GARCH and GAS models, both realized and non-realized, are validated by the simulation of return data generated by the model itself and then estimating the model using this simulated data. Moreover, we have used Monte Carlo simulation methods to extract the asymptotic results of the models. This is described in Subsection 3.2.

3.1 Simulation of high frequency data

In order to simulate the stock price process $S(t)$, we make use of the Geometric Brownian Motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \tag{64}$$

where $W(t)$ is Brownian motion. The solution of this first-order ODE is

$$S(t) = S(0)\exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\right) \tag{65}$$

and can be discretized to obtain the update equation

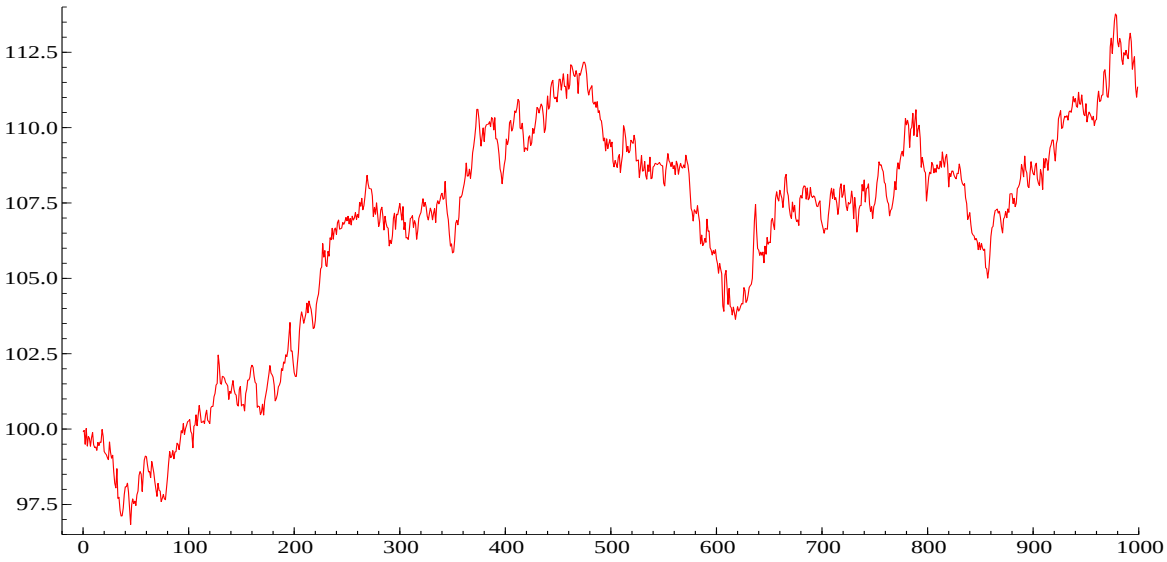
$$S_j = S_{j-1}\exp\left((\mu - \frac{1}{2}\sigma_d^2)h + \sigma_d\sqrt{h}Z_j\right). \tag{66}$$

This equation yields an exact method to simulate stock prices per second. Therefore, we need $h = 1/N$, with $N = 6.5 \cdot 60 \cdot 60$ the number of seconds per trading day, and $Z_j \sim \mathcal{N}(0,1)$. We can account for a small drift μ and a constant or varying volatility σ_d. In case of a time-varying σ_d, we assume that it only varies between days and therefore simulate a daily volatility $\tilde{\sigma}_d$ using the GARCH(1,1) update equation as in Equation (2b). Note that this daily volatility needs to be scaled to get a σ_d that is in proportion with the amount of volatility within one second: $\sigma_d^2 = \tilde{\sigma}_d^2/N$.

In order to incorporate micro-structure noise in the simulated data, we add a normally distributed disturbance term to the exact simulation. For this disturbance term we take the volatility equal to the volatility σ_d of the exact simulation. The reason for this is to have the noise such that it is in proportion to the volatility of the stock price process. An example of a simulation outcome can be found in Figure 1.

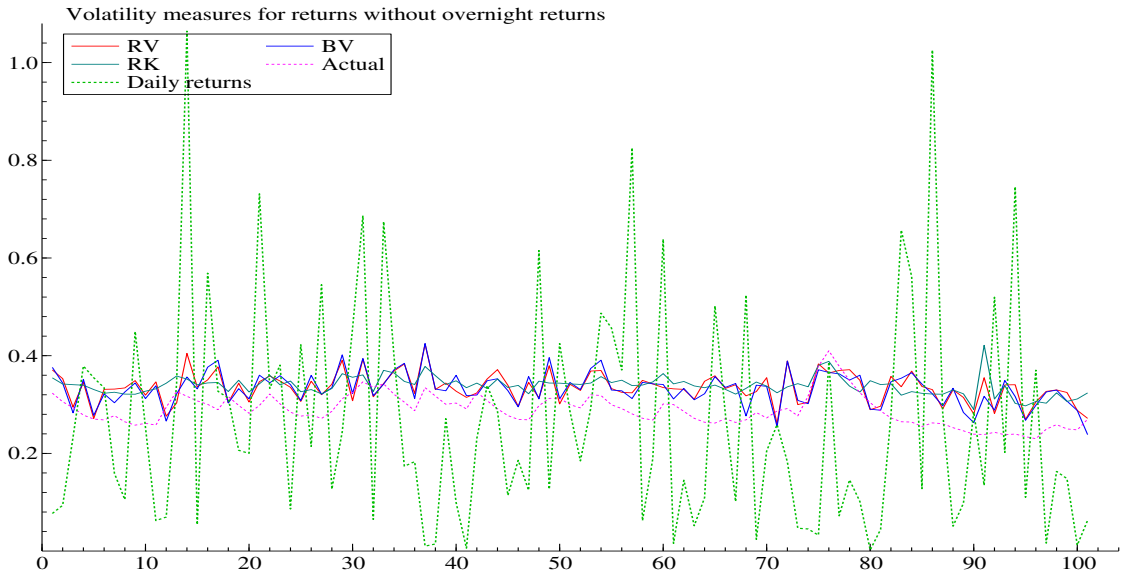
Using the simulated high frequency data we compute the realized measures. We take the five minute returns to calculate the RV and BV for the simulated data. The realized kernel is calculated using a taking a random subset of high frequency prices of the simulated stock prices per second. To create a realistic setting, the size of this random subset representing the ('cleaned') trades on a trading day is chosen at random as well. In Figure 2 we show the simulated realized measures and the actual volatility. We observe that the realized measures captures the changes in the true volatility well. This confirms the correct implementation of the realized measures.

Figure 1: Simulated stock prices



Note: this figure shows the simulated closing prices for 1000 days.

Figure 2: Simulation of realized measures



Note: this figure shows a simulation of the realized measures based on high frequency data generated by the Geometric Brownian Motion. The actual volatility is displayed in pink. The absolute returns are displayed in green.

3.2 Simulation of volatility models

We use simulation methods to ensure the correct implementation of the volatility models. By generating random disturbances of the underlying probability distribution of the volatility, we

generate a conditional variance series and use this to create an artificial returns series. The correct implementation of the volatility model and its corresponding conditional log likelihood function can be tested by estimating the model based on the return data. For a sufficiently large T, the number of simulated returns, the model should return a vector of estimated parameters $\hat{\boldsymbol{\theta}}_T$ that should not be significantly different from the true parameter vector $\boldsymbol{\theta_0}$ that have been used for the simulation. Here we use the following asymptotic result:

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta_0} \quad \text{as} \quad T \to \infty. \tag{67}$$

In addition, we validate the estimations by using a Monte Carlo simulation. By performing 10,000 estimation based on different simulations of 1,000 return observations (using a random seed), we can draw the distribution of the estimated parameters. Namely, it must hold that:

$$\hat{\boldsymbol{\theta}}_T \overset{approx}{\sim} \mathcal{N}(\boldsymbol{\theta}_0, \Omega\Sigma\Omega^\top/T) \quad \text{as} \quad T \to \infty, \tag{68}$$

where, in this case, Σ is the asymptotic variance of the standardized derivative of the (joint) log-likelihood $\ell(r_t, r_{t-1}, \boldsymbol{\theta}_0)$. When this derivative function in $\boldsymbol{\theta}_0$ is a *martingale difference sequence*, then Σ is just the variance of the terms in this sequence. Hence, under the null hypothesis $H_0 : \boldsymbol{\theta}_0 = \boldsymbol{\theta}_0^*$ and assuming *correct specification* for sake of simplicity, we know that we can estimate Σ by

$$\hat{\Sigma}_T = \frac{1}{T}\sum_{t=2}^{T} \nabla\ell(r_t, r_{t-1}, \boldsymbol{\theta}_0^*)^2 \tag{69}$$

and, in addition, that $\Omega = \Sigma^{-1}$. We note that this correct specification assumption may influence the estimated standard errors of our estimates and that more robust measures of Σ and Ω are necessary to get more reliable errors. However, for time reasons we chose to focus on other aspects in this study and to take the simplified error estimates.

In Appendix O, Figure 10, we show the asymptotic distribution of the parameters for the GAS-GED model and GARCH-$\mathcal{N}$ model as an example of our simulation. This figure tells us that the consistency and asymptotic normality are in most cases responsible assumptions, although the estimates of ω and β show a small bias.

4 Results

We obtain the high frequency trade data over the period January 1, 2008 to December 31, 2014 of the PEPSICO stock (ticker PEP) from the WRDS database. The raw data yields 61,248,192 observations. We continue by performing the data-cleaning procedure as devised by Barndorff-Nielsen et al. (2009). Based on the cleaned data we determine the opening and closing prices in order to compute the open-to-close returns and close-to-close returns, yielding 1762 and 1763 observations respectively. Moreover, based on the cleaned data we determine the Realized Kernel, Realized Variance, and Bipower Variation. Finally, we estimate the realized and non-realized GARCH and GAS models using the return series. The results are described in this last section.

4.1 Data cleaning procedure

A data cleaning procedure for HF data, as devised by Barndorff-Nielsen et al. (2009), was used to filter the data and contained the following steps:

P1: Delete entries with a time stamp outside the 9:30-16:00h, which is already done by the database.

P2: Delete entries with a price equal to zero.

P3: Retain entries originating from a single exchange. In this case, exchange "T" (NASDAQ) has been chosen because it contained the most trades.

T1: Delete entries with corrected trades.

T2: Delete entries with abnormal Sale Condition.

T3: Replace multiple prices at same time by median.

An overview of the data cleaning procedure is presented in Table 1. Based on the remaining 4,628,907 observations, we determine the open and close prices for each day and compute the close-to-close and open-to-close returns.

4.2 Descriptive statistics

In Figure 3 we show the log return series for the open-to-close prices and the close-to-close prices. Note that we have one return observation less for the close-to-close series. As expected the open-to-close series show somewhat more extreme returns. This can also be seen in Table 2, where the standard deviation of the close-to-close returns is slightly higher compared to that of the open-to-close returns. Also, the close-to-close returns show a higher degree of negative skewness. This can be explained by the fact that large negative overnight returns may be present. In Appendix N we have included the density and auto-correlation function of the close-to-close and open-to-close returns. We observe that for both return series there is little significant auto-correlation. However, for the squared return series the auto-correlation is significant at almost every lag. That justifies the use of volatility models. If we consider the return distribution, we observe that in both cases there is extreme peakiness around the mean (i.e. leptokurtosis) and fatter tails compared to the normal distribution. Moreover, for the close-to-close returns it is evident that the distribution is left-skewed. This justifies the use of

Table 1: Data cleaning procedure

Year	Trades before cleaning	Cleaning Step						Trades after cleaning
		P1	P2	P3	T1	T2	T3	
2008	10,562,687	13,589	-	7,230,945	-	843	2,177,377	1,139,933
2009	10,584,036	-	-	8,011,456	-	619	1,751,548	820,413
2010	8,381,664	-	-	6,403,100	-	876	1,392,469	585,219
2011	9,977,189	-	-	7,610,922	-	681	1,745,902	619,684
2012	7,223,842	11,993	-	5,486,593	-	596	1,289,896	434,764
2013	6,546,168	-	-	4,933,021	-	9,135	1,103,558	500,454
2014	7,972,606	-	-	6,097,612	-	179,697	1,166,857	528,440
Total	61,248,192	25,582	-	45,773,649	-	192,447	10,627,607	4,628,907

Note: this table shows the results of the data cleaning procedure as devised by Barndorff-Nielsen et al. (2009).

Table 2: Descriptive statistics log returns and realized measures

Log returns		
	Close-to-Close	Open-to-Close
N	1762	1763
Mean	0.013	0.034
Std. Dev.	1.189	1.021
Skewness	-0.618	-0.128
Kurtosis	18.475	11.188
Maximum	8.166	7.898
Minimum	-14.049	-8.343

Realized Measures			
	BV	RK	RV
N	1763	1763	1763
Mean	1.081	1.375	1.173
Std. Dev.	2.284	4.390	2.544
Skewness	9.626	21.729	9.872
Kurtosis	149.244	624.385	151.315
Maximum	49.766	140.466	54.829
Minimum	0.062	0.036	0.071

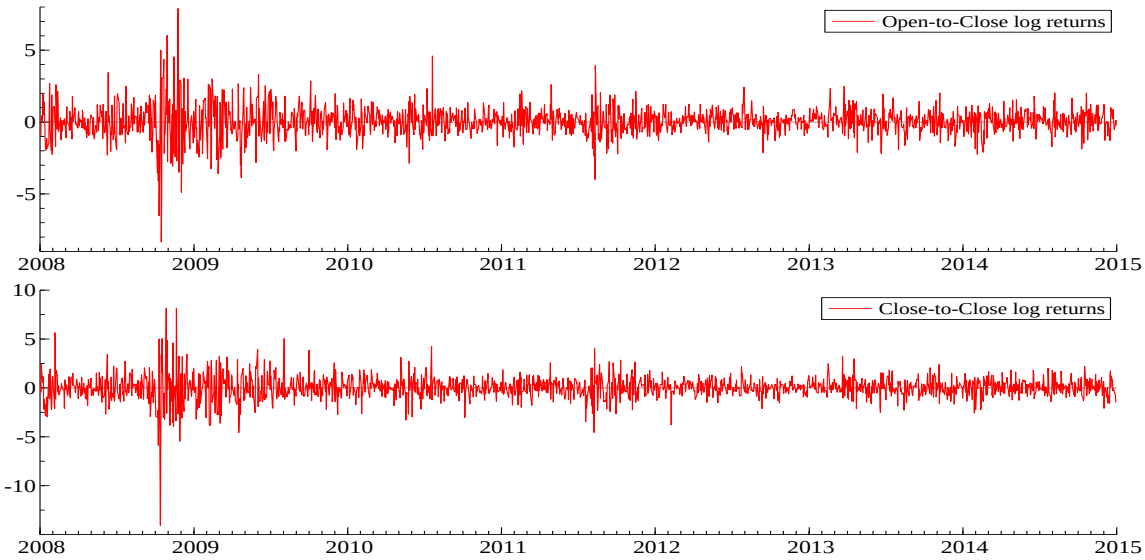
Note: this table shows the descriptive statistics of the close-to-close, open-to-close returns, the realized kernel (RK), realized variance (RV), and bipower variation (BV).

fat tailed and asymmetric distributions for the volatility models (e.g. GED,Laplace, Student t or skewed Student t).

The plots of the realized measures are presented in Figure 4. For the RV and BV have been calculated on the five minute frequency and the first non-zero return of every trading day is set to zero to correct for overnight returns.

Figure 9 in appendix N shows the distribution and QQ-plots of the realized measures and the log of the realized measures. From the log realized measures it is obvious that the distribution exhibits fatter tails compared to the normal distribution and is right-skewed. This calls for other distributions than the normal distribution for the error term in the realized measure equation of the log-linear realized models. To investigate whether or not this actually makes a differences in the predictive accuracy of the one-step-ahead forecasts, we have included both t and normal disturbances for the error term in the realized measure equation of the realized GARCH and the log-linear realized GARCH models. It should be noted that for the

Figure 3: Log returns



Note: this figure shows the time series of the log returns (close-to-close and open-to-close).

Figure 4: Realized measures

Note: this figure shows the time series of the realized measures RV, BV, RK. For the RV and BV are computed based on 5-minute price intervals.

linear models we should in fact use a distribution for the error such that the x_t does not become negative, in which case we would suggest a fat tailed Gamma distribution for instance. For this reason, with the exception of the Realized GARCH(1,1) model we have chosen to implement the log-linear version of the realized models and we consider the linear models with fat-tailed one sided distributions for further research. Considering the skewness we observe in the histogram and the QQ-plots of the logarithm of the realized measures (especially looking at the QQ-plot against the Student-t distribution we see that one side is captured by the distribution while the other side exhibits more mass in the tail) a fat tailed (possibly leptokurtic) skewed distribution such as the skewed-t distribution or the asymmetric Laplace distribution are appealing options. However due to time constraints we were unable to implement these in the realized models. As a default then we have chosen to implement a Student-t distributed error for the measurement equation in all the models except the Realized GARCH(1,1) which we use as a comparison tool.

4.3 In-Sample Results

4.3.1 Parameter estimates

Non-realized models. In Table 12, all parameter estimates for the GARCH and GAS models are provided for the open-to-close and the close-to-close data. Note that all parameter estimates are significant, according to the standard errors we have calculated. For all GARCH models, the parameter β is quite close to 1, which indicates a strong persistence of the past volatility on future variance. As explained before in Section 2, we choose to use EGARCH models, as well as NGARCH models, because of the presence of the leverage effect in these models. As was explained before, in the EGARCH model (see Equations (7a), (7b)) the parameter γ signifies the leverage effect and we expect it to be negative in order to have a larger effect of negative shocks in the volatility on future volatility levels. Table 12 shows that this is indeed the case. Note that for our data it is the case the leverage effect is larger for the close-to-close data. Whether this is a general result is unknown, yet a possible explanation based on our data may be that close-to-close returns generally shows more extreme peaks compared to open-to-close returns. In this sense, higher negative returns can lead to a strong effect on the volatility and hence the leverage effect should be more present.

In the NGARCH model (see Equations (8a), (8b)), the leverage effect is obtained through the cross-term $r_{t-1}\sigma_{t-1}$ in the volatility driver $\alpha(r_{t-1}-\gamma\sigma_{t-1}) = \alpha r_{t-1}^2+\alpha\gamma^2\sigma_{t-1}^2-2\alpha\gamma r_{t-1}\sigma_{t-1}$. Note that for our data, both α and γ are positive. This means that positive values of r_t produce less volatility because then it holds that $-2\alpha\gamma r_{t-1}\sigma_{t-1} < 0$. Similarly it holds for negative r_t that $-2\alpha\gamma r_{t-1}\sigma_{t-1} > 0$, which indeed increases future volatility. Therefore the leverage effect depends on the parameter values of α and γ, and corresponds in this case with what we would expect for the leverage effect.

Similarly for the GAS models, almost all parameters turn out to be statistically significant and have realistic values. Again, the persistence of past returns is quite high, indicated by a β close to 1 in all GAS models. The first extension of the GAS-t models that we described is the skewed GAS-t, because of the presence of a leverage effect in this model. Here the parameter γ indicates the asymmetry. In our definition of the model, that follows Zhu and Galbraith (2010), a value of $\gamma = 0.5$ indicates no skewness. As the table shows, there is very little, if not negligible, skewness since we have $\gamma = 0.496$. In the GAS-GED model, ν is a shape parameter. As was explained in Section 2.3.4, this family allows for tails that are either

heavier than normal (when $\nu < 2$) or lighter than normal (when $\nu > 2$) and nests a continuum of symmetric, leptokurtic densities spanning from the Laplace ($\nu = 1$) to the normal density ($\nu = 2$). A value of $\nu = 1.491$ indicates a density halfway between a normal and a Laplace.

Realized GARCH models. We will now have a closer look at the parameter estimates of the realized EGARCH models as given in Table 13 and those of the realized GARCH-$\mathcal{N}$ models as given in Tables 14 and 15. General observations here are for example that the parameters ω and ξ, representing intercepts in respectively the volatility update equation and the measurement equation, are often around zero. Secondly, we have that τ_1 and τ_2 are always small and of opposite sign. In this way, if the noise term z_t is above expectation, this has both a negative effect (through τ_1) and a positive effect (through τ_2) on x_t. Together with the scale of these parameters, this suggests that they have a minor impact on the model. Furthermore, the parameter ϕ is always about 1 so that the realized measure has approximately the same scale as the variance of the returns. The last general observation we can make is that in the log-linear models with $u_t \sim \mathcal{T}(\lambda_u)$, the degrees of freedom λ_u is very high, suggesting a normal distribution for u_t (the error term of $\log(x_t)$) may be fine as well. In that case, the error term of x_t has a log-normal distribution, which has fatter tails than the normal distribution. This explains why λ_u is small (around 4) in the *linear* realized GARCH-$\mathcal{N}$ model with $u_t \sim \mathcal{T}(\lambda_u)$.

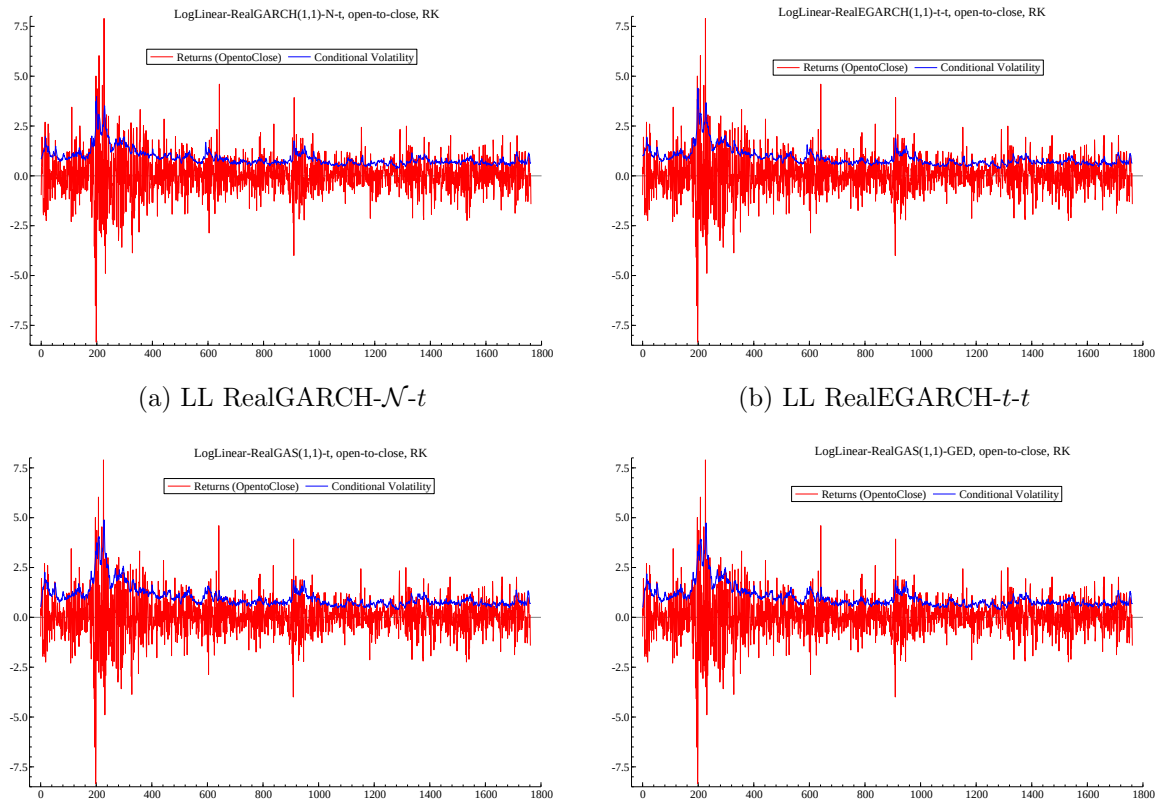
Next, we look at the differences between the models for open-to-close and close-to-close returns. Obvious differences are expected since the close-to-close returns are more volatile since they contain the overnight information as well. This extra volatility is not directly covered by the realized measures, since these are all based on open-to-close returns. Therefore, the impact of the overnight activity should be captured by the parameters. In the linear realized GARCH models, this is done by reducing ϕ so that f_t in the measurement equation is scaled down to equal the scale of x_t again. In the realized log-linear GARCH and EGARCH models, this rescaling in the measurement equation is done by reducing the intercept ξ. Besides this rescaling, we also expect more mass in the volatility of the returns. In the log-linear realized GARCH models, ω is increased to compensate that the influence of $\log(x_t)$ is not high enough. In the linear realized GARCH models, we see a slight increase in both ω and γ and, finally, in the realized EGARCH models we have a higher α to induce this effect.

An additional observation for the realized EGARCH-t model is that λ_z is slightly lower for the close-to-close returns, again showing the fatter tails due to the overnight returns. Besides, the realized EGARCH models show much more dependence in the volatility ($\beta \approx 0.97$) than the realized GARCH models ($\beta \approx 0.65$), which give more weight to the realized measure.

Realized GAS models. In Tables 16 and 17 we have tabulated the in-sample estimates of the RealGAS(1,1)-t. We see the estimates of both the model which includes the leverage function $\tau(z_t)$ and one that doesn't, inputting all calculated realized measures (RV, BV, RK) as x_t. One thing we immediately notice is that in all cases the parameter β is very close to one and the parameter α is between 0.05 and 0.1 which is quite small. These parameter estimates are in contrast with the estimates of the RealGARCH modesl (with the exception of the RealEGARCH) where the α corresponds with the γ in the RealGARCH framework. The two update equations in the different frameworks however are vastly different. In the RealGAS case the high β parameter in fact confirms the high persistence of conditional volatlity as expected. We observe a ϕ which is close to a one which indicates a linear relationship between the realized measure and the latent volatility in the measurement equation. Furthermore as expected $\tau_1 < 0$

and $\tau_2 > 0$, where the estimated are found to be significant. In Figure 5 below we have plotted the estimated conditional volatility for a small selection of the estimated models.

Figure 5: Realized measures



(a) LL RealGARCH-$\mathcal{N}$-t

(b) LL RealEGARCH-t-t

(c) LL RealGAS-t

(d) LL RealGAS-GED

Note: The in-sample (full sample) estimated conditional variance (blue) plotted for the respective models against the open-to-close returns (red).

4.3.2 Likelihood fit

Before continuing to the out-of-sample analysis we observe a couple of results in terms of the likelihood values found for the realized models. Keeping in mind that the MLE ony minimizes the distance between the DGP and the model in KL-distance asymptotically we cannot immediately conclude that a higher likelihood is better. Therefore in Table 18 we have also tabulated the number of parameters to be able to connect the likelihood value found with the goodness of the fit. We can quite quickly notice that the realized GAS models have a higher likelihood compared to realized GARCH models with the same number of parameters. For instance looking at the realized GAS-GED without the τ we notice that it has a much higher likelihood than the highest realized GARCH models for all realized measures. We note that in general the realized GAS models offer a better in sample fit to the data, and later juxtapose this to the predictive ability of the realized GAS models.

4.4 Out-of-Sample Results

4.4.1 Fixed window

In the fixed window estimation, a fixed sample of 880 observations is used to estimate the parameters of the different volatility models. Next, we obtain the one-step-ahead forecasts based on those parameter estimates. That implies that one-step-ahead forecasts are made based on the estimates $\hat{\theta}$. It is important to emphasize the fact that the financial crisis of 2008 is part of this estimation sample. This may have a significant effect on the parameter estimates, due to the large realized measures in that period. It is interesting to see whether or not this has an effect on the predictive accuracy of the one-step-ahead forecasts, compared to the case where we use a rolling window estimation procedure. Namely, for the latter the financial crisis is only included in the first 200 forecasts, thereafter the financial crisis is not part of the estimation sample.

Open-to-close In Table 19 (see Appendix P), we have presented the MAE, MSE and LSR statistics of the one-step-ahead forecasts based on the Open-to-Close returns. The first column shows the different realized measures used in the realized models, while the realized measures in the top of the table denote the benchmark used to calculate the MAE and MSE statistic.

As expected, the MAE and MSE statistic show that both the realized GARCH and GAS models strongly outperform the regular GARCH and GAS models. By the use of Diebold-Mariano tests, we can conclude that in most cases even the worst realized GAS and GARCH models significantly outperform the best non-realized models for most benchmarks. This result is shown in Table 3.

If we dig further into the realized models, we observe that all models perform generally the best with the BV as benchmark. This is also a bit counter intuitive, as we would expect that at least the realized measure that serves as input for the model should best predict itself. However, it should be pointed out that this differences are rather small and in most cases not significantly different from each other.

The Diebold-Mariano test shows that the best realized GARCH model significantly outperforms the best realized GAS model, based on the MAE statistic. It is even more interesting to note that even the worst realized GARCH even outperforms the best realized GAS (see Table 4). To the contrary, these results are not found for the MSE statistic. That is, in only one case the best realized GARCH outperforms the best realized GAS. More importantly, we cannot show that the best realized GAS is outperformed by the worst realized GARCH. The difference in the results of the MAE and MSE statistic can be explained by the fact that GAS models perform worse in terms of being close to the "true" volatility (MAE), i.e. the used benchmark, but perform generally well on the 'peaks'. Namely, the MSE measure is smaller if the model peaks at the same time as the used benchmark.

We also find that for the MAE in almost all cases the realized EGARCH model outperforms the linear and log-linear realized GARCH specifications. Also for the non-realized models the EGARCH performs particularly well. This confirms the suggestions in the literature that in financial data there exists a leverage effect. A similar result is found when looking at the realized GAS models. Here the GAS models that include a leverage effect (τ function) generally outperform those without. However, this is not necessarily the case when looking at the MSE measure. Here the realized EGARCH models are generally outperformed by the regular models that do not have asymmetric effects. Also the leverage function in the realized

Table 3: Comparison of best non-realized and worst realized models (fixed window open-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH t	EGARCH t	EGARCH t
Comparison with			
Worst Realized GARCH	Real GARCH $\mathcal{N}$ (t) (RK)	Real GARCH $\mathcal{N}$ (t) (RK)	Real GARCH $\mathcal{N}$ (t) (RK)
DM	5.1577***	5.0157***	5.4334***
Worst Realized GAS	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)
DM	4.1765***	2.8712***	3.4460***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH $\mathcal{N}$	EGARCH t	EGARCH $\mathcal{N}$
Comparison with			
Worst Realized GARCH	Real EGARCH $\mathcal{N}$ (t) (RK)	Real EGARCH t (t) (RK)	Real EGARCH $\mathcal{N}$ (t) (RK)
DM	0.59206	3.2370***	2.2761**
Worst Realized GAS	LL Real GAS GED (τ) (RK)	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)
DM	1.1083	2.9163***	2.9805***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold Mariano statistic for the best non-realized model based on the MAE and MSE statistic compared to (1) the worst realized GAS and (2) the worst realized GARCH model. If the DM is positive, it indicates that the test is in favor of the worst realized models.

GAS models leads to lower MSE measures for each benchmark. This result can be explained by the fixed window estimation. As the financial crisis of 2008 is in the estimation sample, the leverage parameters are significantly present in the models. However, as these parameters are not updated after each observation, these models will overestimate volatility in calmer periods. This explains that these models perform worse based on the MSE measure. Because in this case the effect of the overestimated volatility on the MSE is enlarged even further, due to the square in its calculations.

Apart from the comparison of the different classes of models, it is important to check whether the specification of the error term distribution of the realized measure actually matters. As explained in Section 2, we would expect that in the log-linear specification of the realized models the Student-t distribution better fits the data compared to the Gaussian distribution.

Table 4: Comparison of best realized GAS and worst realized models GARCH (fixed window open-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
Comparison with			
Best Realized GAS	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)
DM	-1.8838*	-2.9419***	-2.4579***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
Comparison with			
Best Realized GAS	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)
DM	-0.14353	-1.3240	-0.77648
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold-Mariano (DM) statistic for the best realized GAS based on the MAE and MSE statistic compared to the worst realized GARCH. If the DM statistic is positive, it indicates that the test is in favor of the worst realized GARCH.

This is confirmed by the MAE and MSE statistic. In almost all cases, log linear realized models with Student-t disturbances in the measurement equation outperform the models with normally distributed errors.

The log-scoring rule shows different results compared to the MAE and MSE. Where the latter two measures only emphasize the predictive ability of the models, the log-scoring rule considers the fit of the out-of-sample return observation in the predicted density. It is interesting that the realized EGARCH with Student-t disturbances is the model with the best fit, for the RV, BV and the RK as realized measures. For the non-realized measures the Laplace distribution has the highest log-score. From this we can derive the fact that even though the LSR does not say so much about the quality of the forecasts with realized EGARCH models, these models offer a good fit for the underlying density of the returns. Again, the importance of the EGARCH models and the presence of leverage effect should be stressed here.

Close-to-close For the close-to-close returns, we find similar results. See Table 20 in appendix N for the MAE, MSE and LSR statistics of the one-step-ahead forecasts. Again, the realized GARCH models outperform the non-realized models on both the MAE and MSE measure. However, opposed to the results for the open-to-close returns, we do not find confirming

evidence that the worst realized GAS and GARCH models outperform the best non-realized models (see Table 7 in Appendix L). While for the MAE statistic all worst realized models outperform the best non-realized, we find that for the MSE statistic only two out of six times the best non-realized model is outperformed by the worst realized model (at the 10% significance level). This result is not surprising, as we have calculated all the realized measures based on open-to-close returns. We do not find remarking significant differences between the best realized GAS and best realized GARCH model based on the MAE statistic. However, the best realized GAS outperforms the best realized GARCH significantly on the and MSE statistic for the close-to-close fixed window (see Table 8 in Appendix L). This can be explained by the fact that the realized GAS model incorporates the properties of a Student-t distribution in its updating equation. Therefore it may be a better fit to the close-to-close data because this data is characterized by more volatile data (outliers). Moreover, not only is the realized GAS model a better fit, it is also the case that these parameters are estimated once and then repeatedly used in the fixed window estimation procedure.

As pointed out for the open-to-close returns, it is interesting to point out that the RV as input for the realized models shows lower MAE and MSE statistics with RK as benchmark, compared to the models where RK itself serves as an input. Again, by the Diebold-Mariano test we find that these differences, however, are in most cases not significantly different from one another.

For the MAE statistic we find that the EGARCH models perform particularly well for the realized and non-realized models. An explanation might be that due to the inclusion of the overnight returns (more volatile returns), the asymmetric effect is more present and hence the EGARCH model that incorporates this effect outperforms models that do not. For the MSE statistic we find that the asymmetric models are generally outperformed by the symmetric ones. Again, as explained before for the open-to-close case, one of the reasons could be that the parameters are fit against a highly volatile period, causing relatively strong leverage effects. Then volatility may be overestimated in calmer periods.

The log-scoring rule shows similar results and prefers the EGARCH models in most cases. Moreover, the importance of the leverage effect can also be shown by looking at the realized GAS models. Here the realized models that include the leverage effect, i.e. the τ function, perform significantly better than the GAS models that do not incorporate this asymmetry.

Finally, we find somewhat different results for the distribution of the kernel. Whereas for the open-to-close returns the Student-t distribution is preferred over the normal distribution, we have the opposite holding true for the close-to-close returns. Note however that these differences are relatively small for all benchmarks and are not significantly different from each other.

In the paragraphs that follow we will outline the results from a rolling estimation procedure and discuss how the outcomes are different from the results described above in the fixed window. By comparing the two methods we can draw conclusions on whether or not the estimation method has an effect on the predictive accuracy of the forecasts.

4.4.2 Rolling window estimation

The rolling window estimation takes a similar approach as that of the fixed window estimation. However, after each one-step-ahead forecast of the volatility, we move the estimation sample by one observation and then estimate the parameters again. In this way, all the recent dynamics of the time series are included in the estimation process, which by intuition should result

in higher quality forecasts compared to the fixed sample estimation. It is also important to emphasize the fact that for the fixed window estimation the financial crisis of 2008 is in the estimation sample. This could drastically affect the estimated parameter vector. For the rolling estimation, the recent financial crisis is only included in the first 200 observations.

Open-to-close The MAE and MSE statistics for open-to-close returns are presented in table 21 in appendix N. Firstly, similar to the fixed window case, we observe that the RV as input for the realized models shows lower MAE and MSE statistics with RK as benchmark, compared to the models where RK itself serves as an input.

Moreover, we directly observe that for both the MAE and MSE statistics, the non-realized models are typically outperformed by the realized GARCH specifications, as is suggested in the literature (specifically the article of Hansen et al. (2012)). Also the worst realized GARCH models perform better than the best non-realized specifications (see Table 5). This becomes especially clear from the MAE statistic. From the MSE statistic this is less evident since this is only the case where the RV or BV is used as benchmark at the 10% significance level.

The same does not necessarily hold for the realized GAS models. That is, some of the GAS models perform substantially worse compared to the non-realized models. We even find that, based on the MAE statistic, the best non-realized model significantly outperforms the worst realized GAS model on the 1% significance level. For the MSE statistic we find somewhat similar results, but these are not always significantly different from one another as tested by the Diebold-Mariano test. We do, however, find that the best realized GAS model outperforms the best non-realized model for both the MSE and MAE, so a suitable GAS specification can still yield a realized model that outperforms non-realized models.

If we compare the realized GAS models to the realized GARCH models, we find that for the MAE statistic that the best realized GAS model is outperformed by the best realized GARCH model. In addition, even the worst realized GARCH model outperforms the best realized GAS model (see Table 6). In both cases, however, this conclusion cannot be drawn by looking at the MSE. This implies that while the realized GARCH models are in general closer to the 'true' benchmark, they are not necessarily better in predicting the peaks of the model. As was mentioned before, the GAS model incorporates Student-t properties as a result of which is fits more volatile data ('peaks') better. Consequently, in terms of MSE the models do not differ significantly from one another.

It is obvious that in terms of MAE the asymmetric GARCH models perform especially well. For the non-realized models, the EGARCH with Gaussian disturbances is the best model with the RK or RV as benchmark, while the GAS with skewed Student-t disturbances is the best for the BV. Also for the realized models we see that the EGARCH models are generally preferred in terms of MAE over the other models. This gives evidence for the importance of accounting for the leverage effect in financial time series. This is also obvious by looking at the realize GAS models. Namely, the GAS models that include the leverage effect (i.e. τ function) show better MAE statistics compared to those of without the leverage function. Similar results are found when looking at the MSE statistic. Based on the MSE, the EGARCH scores best of all non-realized models. Also the realized GAS functions that include the leverage effect generally show lower MSE statistics compared to those without the τ function. In terms of the log-scoring the EGARCH models are preferred. As mentioned before this means that not necessarily the forecasts are most accurate based on a realized EGARCH model, but this model best captures the underlying density process.

Table 5: Comparison of best non-realized and worst realized models (rolling window open-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH $\mathcal{N}$	GAS Skewed t	EGARCH $\mathcal{N}$
Comparison with			
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
DM	3.2990***	2.9419***	3.4536***
Worst Realized GAS	LL Real GAS t (no τ) (RK)	LL Real GAS t (no τ) (RK)	LL Real GAS t (no τ) (RK)
DM	-3.1079***	-5.5166***	-4.3757***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH $\mathcal{N}$	EGARCH $\mathcal{N}$	EGARCH $\mathcal{N}$
Comparison with			
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
DM	1.2525	1.8829*	1.6575*
Worst Realized GAS	LL Real GAS $\mathcal{N}$ (no τ) (RV)	LL Real GAS t (no τ) (RK)	LL Real GAS t (no τ) (RK)
DM	1.7455*	-2.1623**	-1.5159
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold Mariano statistic for the best non-realized model based on the MAE and MSE statistic compared to (1) the worst realized GAS and (2) the worst realized GARCH model. If the DM is positive, it indicates that the test is in favor of the worst realized models.

By further looking at the models, we compare the Student-t and Gaussian disturbances for the realized measure equation in the (log)-linear realized models. We find that there is a tendency for the Student-t distribution to produce better forecasts in terms of MAE and MSE. Hence these differences turn out to be rather small. It may be the case that this benefits the close-to-close data more, as this data is typically more volatile which is then better fitted to a Student-t distribution.

Close-to-close The results for the close-to-close returns are presented in Table 22 in Appendix N. We immediately observe that the MAE and MSE statistics are higher compared to the case where we have open-to-close returns. This is not surprising given the fact that all realized measures have been computed using open-to-close returns. As expected and in

Table 6: Comparison of best realized GAS and worst realized models GARCH (rolling window open-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
Comparison with			
Best Realized GAS	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)
DM	-1.8838*	-2.9419***	-2.4579***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Worst Realized GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (RK)
Comparison with			
Best Realized GAS	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)
DM	-0.14353	-1.3240	-0.77648
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold-Mariano (DM) statistic for the best realized GAS based on the MAE and MSE statistic compared to the worst realized GARCH. If the DM statistic is positive, it indicates that the test is in favor of the worst realized GARCH.

line with the results for open-to-close and the fixed window, we find that in general there is evidence that suggests that realized models outperform the non-realized models. For example, we find that the worst realized GARCH model still outperforms the best non-realized model for the MAE. For the realized GAS model, we do not find a significant difference between the worst realized GAS model and the best non-realized model, based on the MAE (see Table 9 in Appendix L). By comparing the best realized GAS and best realized GARCH model, we find no significant differences on the 5% confidence level. Consequently, similar to the fixed window case we find that the realized GAS models perform generally better on the close-to-close returns (see Table 10 in Appendix L).

Moreover, in contrast to the case of the close-to-close returns for the fixed window, the realized EGARCH model does not consistently outperform the regular realized models. However, based on the log-scoring rule these models are still preferred. Additionally, we notice that for the realized GAS models the specifications that include the leverage effect (i.e. τ function) show better MAE and MSE statistics compared to the models that do not incorporate this.

4.4.3 Comparisons

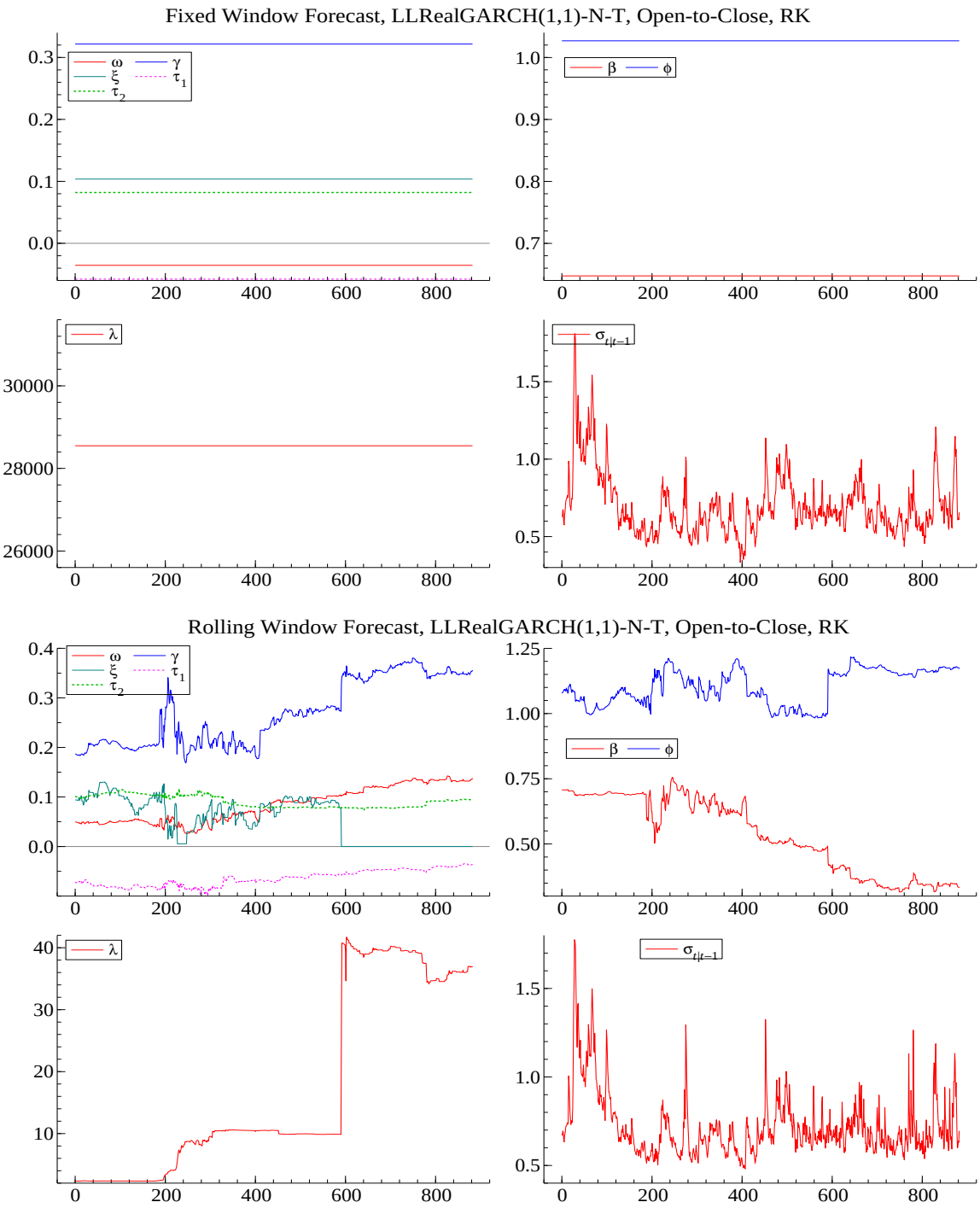
Given the previous results a couple of important observations can be made. First we find that in general (except a few uncertain cases) the worst realized GARCH models performed better than the best non-realized models when forecasting both with a fixed window as well as forecasting with a rolling window. However a peculiar observation is that the worst realized GAS model is better than all the non-realized models when forecasting with a fixed window, but when forecasting with a rolling window the worst realized GAS model is outperformed by the best non-realized model. Consequenly, we test whether the realized GAS models perform better in a fixed window setting than in a rolling window setting. We perform a similar test for realized GARCH models, which we expect to perform better in a rolling window setting. The results can be found in Appendix L. We find that the best performing GAS model the log-linear realized GAS-GED in the rolling window is outperformed by the same model in the fixed window, whereas for the best realized GARCH model this distinction is much less clear.

This result could indicate weaker adaptive ability of the realized GAS models (those that we have implemented) in comparison to the realized GARCH models. To better try to understand why it could be that our realized GAS models generally perform better in the fixed window forecast setting than in the rolling window forecast we observe the two graphs below in Figures 6 and 7. Note the elements that we analyse in these graphs and the conclusions we make about them are representative for all other realized GARCH and realized GAS models we have analysed.

First looking at Figure 6, in the upper part the fixed window is given and the bottom section of the figure shows the rolling window setting. In the rolling window we see that the β estimate starts of rather high and decreases with time (the fixed window keeps the β high) thus the dependence on the latent volatility decreases over time. At the same time we see that the γ parameter increases with time thus giving more weight to the realized volatility measure. When looking at these graphs we can observe a change in structure over time, which can be due the fact we have included the crisis period in our analysis. Now if we consider the rolling window setting for the realized GAS model in Figure 7 we notice that β, which like the γ in the realized GARCH setting is a coefficient of the dependence on the previous latent volatility, starts off very high and remains high. If we look at the α parameter it starts off low and remain the same much like the fixed window setting. Now the α parameter is much harder to interpret because it is the coefficient of the score of the log density function, however it is peculiar that the similar dynamics are estimated as a result of these parameters in the crisis as well as after the crisis. This indicates that the realized GAS model is less capable of adapting to the data when it comes to forecasting.

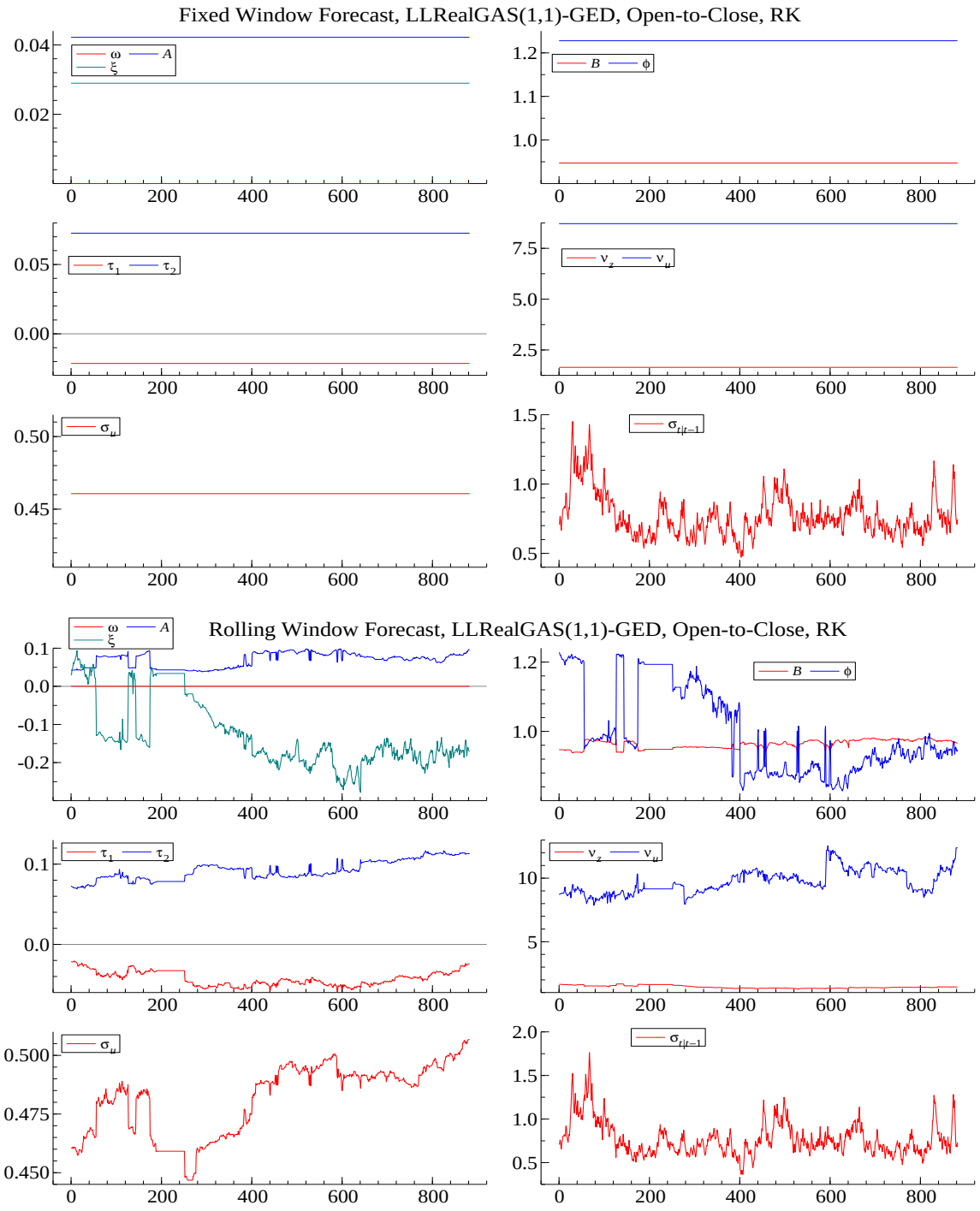
The reason why in the fixed window setting the realized GAS models perform better is that in that setting all models are estimated over a fixed data sample, for instance we see that the realized GAS-t performs well in fixed window forecasting, because the window was taken to be the first 880 days, which are the dates of the crisis. We know that the realized GAS-t incorporates the fat tail properties of the Student-t distribution quite well in its update equation thus has a better ability to cope with excessive returns as witnessed in during the first 880 days. The parameter estimated during that period are thus more reliable in comparison to other models when using the same estimates to forecast in period of more stability in the returns.

Figure 6: Fixed and rolling window estimates and forecasts



Note: In this figure we observe that for the rolling window γ estimates increase in value over time while the β decreases. The ϕ remains around one which again shows a linear relationship between the realised measure and the latent volatility (in logs). Clearly in the fixed case this remains the same.

Figure 7: Fixed and rolling window estimates and forecasts



Note: In this figure we observe that for the rolling window A and B do not change over time. We find this result in all the Realized GAS models This could indicate that the Realized GAS is less adaptive to changes.The ϕ remains around one which again shows a linear relationship between the realised measure and the latent volatility (in logs).

5 Conclusion

This study compares the standard volatility modelling framework to that of the realized measures. In total, ten regular volatility models and ten realized model specification have been estimated. Three realized measures are used as input for the realized models, i.e. the Realized Kernel, the Realize Variance, and the Bipower Variation. By the use of simulation methods the correct implementation of the realized measures and the the maximum likelihood estimation of the models have been tested extensively.

Using the return data from January 2008 to December 2014 on the PEPSICO stock, both open-to-close and close-to-close, we create one-step-ahead volatility forecasts based on the different models. This is performed on a fixed window basis and a rolling window basis. Our results are in line with those found in Hansen et al. (2012) and Barndorff-Nielsen et al. (2009). That is, by the use of Diebold-Mariano tests we find that the realized models outperform the the non-realized models significantly. Using the realized measures as benchmark for the volatility forecasts, we find that based on the MAE and MSE statistics even the worst realized models outperform the best non-realized models.

We also find evidence that asymmetric models perform generally better compared to the symmetric models that do not account for leverage effects. Also realized GAS models that include the leverage effect function τ outperform those that do not.

This research also includes a comparison between realized GAS and realized GARCH models. We find that the implemented realized GAS in general seems to be less able to adapt to data for forecasting than the more parsimonious realized GARCH models. Especially for the open-to-close returns, the realized GARCH models generally outperform the realized GAS models. However, in-sample results indicate that the realized GAS models fit the data better in likelihood comparisons. This in-sample fit of the realized GAS model could be a case of over fitting. For example, Hansen (2009) shows that in-sample over fitting results in out-of-sample under fitting (even more so when the considered models nest the true DGP).

Chatfield (2006) reports the phenomenon where adding complexity typically results in a better fit to the data in-sample, while out-of-sample the opposite is the case. An interesting case would be when one starts the rolling window forecasting with a realized GAS model that performs well out-of-sample there and then moves toward a different model that performs better out-of-sample as the window moves forward. This could for instance be achieved by dynamic model averaging.

Possible over-fitting due to model complexity is but one possible reason for our results, there are many more. For instance looking at the graphs of the rolling window series of estimates we find evidence indicating a structural break (between the crisis and the time of less financial instability) and it could be that the realized GAS models are less able to cope with this. Blasques et al. (2015) show that parameter updates based on the score will always reduce the local Kullback-Leibler divergence between the true conditional density and the model-implied conditional density. However, this is only the case if the underlying process is continuous (hence containing no structure breaks). The same has not yet been proven for discontinuous responses, and investigation into these cases could be a topic for future research. Furthermore due to time constraints we were unable to compare the performance of the realized GAS updating algorithm between the case where we use the Fisher information matrix as a scaling factor and when use no scaling (i.e. $S(f_t) = 1$). We were unable to check these possible hypotheses, thereby exhaustively finding the most probable underlying reason for our results.

Hence, this study opens doors for further research. For example, the realized GAS models can be extended by adding a skewed disturbance term rather than a symmetric in the realized measure equation. The histograms of the log realized measures show that the data is somewhat skewed and has fat tailed. A possible extension could therefore be to implement a skewed t distribution. For the linear model, one could implement a fat-tailed gamma distribution. Another possibility is to add even more different asymmetric GARCH (e.g. the GJR-GARCH or HEAVY model) or GAS models and extensively (thereby also investigating the effect of an information matrix based scaling) compare the forecasting performance of these models. Finally, in our research we limit ourselves to the performance of one-step-ahead forecasts over a fixed period based on three realized measures. One could check the robustness of the results by creating multiple forecasting periods, by performing multiple-step-ahead forecasts, and using different realized measures. For example, Vortelinos (2014) provides an overview of range-based volatility estimators for high frequency data.

References

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using Bipower Variation. *Journal of financial Econometrics*, 4(1):1–30.

Blasques, F., Koopman, S., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Chatfield, C. (2006). Model uncertainty. *Encyclopedia of Environmetrics*.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized Autoregressive Score Models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Diks, C., Panchenko, V., and Van Dijk, D. (2010). Out-of-sample comparison of copula specifications in multivariate density forecasts. *Journal of Economic Dynamics and Control*, 34(9):1596–1609.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Franses, P. H. and Van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press.

Hansen, P. R. (2009). In-sample fit and out-of-sample fit: Their joint distribution and its implications for model selection. *preliminary version April*, 23:2009.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Vortelinos, D. I. (2014). Optimally sampled realized range-based volatility estimators. *Research in International Business and Finance*, 30:34–50.

Zhu, D. and Galbraith, J. W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics*, 157(2):297–305.

A Log-likelihood GARCH-t

Recall that the GARCH-t model is given by

$$a_t = \sigma_t \varepsilon_t, \quad \text{where} \quad \varepsilon_t \sim TID(\lambda), \tag{70a}$$
$$\sigma_t^2 = \phi(\sigma_{t-1}^2, a_{t-1}; \boldsymbol{\theta}). \tag{70b}$$

We only know the distribution of the innovations ε_t, which is given by

$$f_\varepsilon(\varepsilon_t, \lambda) = \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi\lambda}} \left(1 + \frac{\varepsilon_t^2}{\lambda}\right)^{-\frac{\lambda+1}{2}}. \tag{71}$$

To find the density function of a_t we have to use the following transformation:

$$f_a(a_t; \boldsymbol{\theta}) = f_\varepsilon\left(h^{-1}(a_t; \boldsymbol{\theta}), \boldsymbol{\theta}\right) \left|\frac{\partial h^{-1}(a_t, \boldsymbol{\theta})}{\partial a_t}\right|, \tag{72a}$$
$$\text{where } h^{-1}(a_t, \boldsymbol{\theta}) = \frac{a_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} = \varepsilon_t. \tag{72b}$$

$$f_a(a_t | D_{t-1}; \boldsymbol{\theta}) = \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi\lambda}} \left(1 + \frac{a_t^2}{\lambda\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right)^{-\frac{\lambda+1}{2}} \left|\frac{1}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right| \tag{73}$$
$$= \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi\lambda}\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \left(1 + \frac{a_t^2}{\lambda\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right)^{-\frac{\lambda+1}{2}}. \tag{74}$$

Note that in (73) we do not need the absolute signs because h is strictly increasing. From the model in (70a) and (70b) we can derive that $\mathbb{V}\text{ar}[a_t] = \sigma_t^2 \mathbb{V}\text{ar}[\varepsilon_t] = \sigma_t^2 \left(\frac{\lambda}{\lambda-2}\right)$ since $\varepsilon_t \sim TID(\lambda)$. Therefore we use the following scaling function to obtain a standardized Student-t density

$$\mathbb{V}\text{ar}[a_t] = \sigma_t^2 \left(\frac{\lambda-2}{\lambda}\right) \underbrace{\mathbb{V}\text{ar}[\varepsilon_t]}_{=\frac{\lambda}{\lambda-2}} = \sigma_t^2, \tag{75}$$

which can be obtained if $a_t = \sigma_t \sqrt{\frac{\lambda-2}{\lambda}} \varepsilon_t$. Then $h^{-1}(a_t; \boldsymbol{\theta})$ becomes

$$h^{-1}(a_t; \boldsymbol{\theta}) = \varepsilon_t = \frac{a_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)\sqrt{\frac{\lambda-2}{\lambda}}}. \tag{76}$$

$$f_a(a_t | D_{t-1}; \boldsymbol{\theta}) = \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi(\lambda-2)}\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \left(1 + \frac{a_t^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)(\lambda-2)}\right)^{-\frac{\lambda+1}{2}}. \tag{77}$$

And using that $r_t = \mu + a_t$ we finally obtain the density function of r_t, given by

$$f_r(r_t | D_{t-1}; \boldsymbol{\theta}) = \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi(\lambda-2)}\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \left(1 + \frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)(\lambda-2)}\right)^{-\frac{\lambda+1}{2}}. \tag{78}$$

Now the log-likelihood of the of the returns with Student-t innovations can be derived as well as a M-estimator $\hat{\theta}$ for $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda)^T$.

$$\begin{aligned}
\ell(\boldsymbol{\theta}|r_T) &= \frac{1}{T}\sum_{t=2}^{T} \log f_r(r_T|D_{t-1};\boldsymbol{\theta}) \\
&= \frac{1}{T}\sum_{t=2}^{T}\left[\frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\pi(\lambda-2)}\sigma_t(\boldsymbol{\theta},\sigma_1^2)}\right. \\
&\qquad \left. -\frac{\lambda+1}{2}\log\left(1+\frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)(\lambda-2)}\right)\right] \\
&= \frac{1}{T}\sum_{t=2}^{T}\left[\log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda-2)\right. \\
&\qquad \left. -\log\sigma_t(\boldsymbol{\theta},\sigma_1^2) - \frac{\lambda+1}{2}\log\left(1+\frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)(\lambda-2)}\right)\right]. \qquad (79)
\end{aligned}$$

B Updating equation GAS(1,1)-t

From the result in Appendix A, we know that the log-density of r_t with standardized Student-t distributed innovations is given by

$$\begin{aligned}\log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log\pi \\ &\quad - \frac{1}{2}\log(\lambda-2) - \log\sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \frac{\lambda+1}{2}\log\left(1 + \frac{(r_t-\mu)^2}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)(\lambda-2)}\right). \end{aligned} \tag{80}$$

To obtain the GARCH updating equation we need the score function that indicates the sensitivity of the log-likelihood function w.r.t. the time-varying parameter f_t and is given by

$$\begin{aligned}\nabla_t &= \frac{\partial \log f_r(r_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} \\ &= -\frac{1}{2}\frac{1}{f_t} - \frac{\lambda+1}{2}\left[\left(\frac{1}{1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}}\right)\left(-\frac{(r_t-\mu)^2}{\lambda-2}\frac{1}{f_t^2}\right)\right] \\ &= -\frac{1}{2f_t} + \frac{(\lambda+1)(r_t-\mu)^2}{2(\lambda-2)f_t^2}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1} \\ &= \frac{\lambda+1}{2(\lambda-2)f_t^2}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}. \end{aligned} \tag{81}$$

Moreover by using the inverse of the Fisher information matrix, we use the local curvature of the log density function of $f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})$ in the scaling.

$$S_t = \mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right]^{-1}, \tag{82}$$

$$\begin{aligned}\nabla_t\nabla_t^\top &= \left[\frac{\lambda+1}{2(\lambda-2)f_t^2}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}\right]^2 \\ &= \frac{(\lambda+1)^2}{4(\lambda-2)^2f_t^4}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}(r_t-\mu)^4 \\ &\quad - \frac{\lambda+1}{2(\lambda-2)f_t^3}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 + \frac{1}{4f_t^2}, \end{aligned} \tag{83}$$

$$\begin{aligned}\mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right] &= \frac{1}{4f_t^2} + \frac{(\lambda+1)^2}{4(\lambda-2)^2f_t^4}\underbrace{\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}(r_t-\mu)^4\right]}_{(*)} \\ &\quad - \frac{\lambda+1}{2(\lambda-2)f_t^3}\underbrace{\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2\right]}_{(**)}. \end{aligned} \tag{84}$$

To find an expression for (*) and (**) we use the property for Z-estimators that maximize

by setting the score to zero, hence it holds that $\mathbb{E}_{t-1}[\nabla_t] = 0$.

$$\begin{aligned}
&\mathbb{E}_{t-1}[\nabla_t] = \mathbb{E}_{t-1}\left[\frac{\lambda+1}{2(\lambda-2)f_t^2}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}\right] = 0 \\
&\Leftrightarrow \frac{\lambda+1}{2(\lambda-2)f_t^2}\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2\right] - \frac{1}{2f_t} = 0 \qquad 2005/06/28ver \\
&\Leftrightarrow \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2\right] = \frac{2(\lambda-2)f_t^2}{2f_t(\lambda+1)} = \frac{(\lambda-2)f_t}{(\lambda+1)} \qquad (85)
\end{aligned}$$

Equation (85) provides an expression for (**). Then by the Information Matrix Equality (IME), which states that the expected value of the hessian of the log-likelihood function equals the negative of the expected value of the outer product of its gradient, $\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t}\right] =$ -$\mathbb{E}_{t-1}\left[\frac{\partial \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t}\frac{\partial \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t}\right] = -\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$ provides an alternative expression for $\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$. To obtain an expression for (*) we derive this alternative expression for $\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$ and compare it to the one in Equation (84).

$$\begin{aligned}
\frac{\partial^2 \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t} &= -\frac{2(\lambda+1)}{2(\lambda-2)}(r_t-\mu)^2\frac{1}{f_t^3}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1} \\
&\quad - \frac{\lambda+1}{2(\lambda-2)f_t^2}(r_t-\mu)^2\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\left(-\frac{(r_t-\mu)^2}{(\lambda-2)}\frac{1}{f_t^2}\right) + \frac{1}{2f_t^2} \\
&= -\frac{(\lambda+1)}{(\lambda-2)f_t^3}(r_t-\mu)^2\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1} \\
&\quad + \frac{(\lambda+1)}{2(\lambda-2)f_t^4}(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2} + \frac{1}{2f_t^2} \qquad (86)
\end{aligned}$$

Then by taking expectations and applying the result of (85) we obtain

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t}\right] &= -\frac{(\lambda+1)}{(\lambda-2)f_t^3}\underbrace{\mathbb{E}_{t-1}\left[(r_t-\mu)^2\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}\right]}_{=\frac{(\lambda-2)f_t}{\lambda+1}} \\
&\quad + \frac{(\lambda+1)}{2(\lambda-2)^2 f_t^4}\mathbb{E}_{t-1}\left[(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\right] + \frac{1}{2f_t^2} \\
&= \frac{(\lambda+1)}{2(\lambda-2)^2 f_t^4}\mathbb{E}_{t-1}\left[(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\right] - \frac{1}{2f_t^2}. \qquad (87)
\end{aligned}$$

Then by the IME, the negative of (84) should equal (87).

$$\frac{1}{2f_t^2} - \frac{(\lambda+1)}{2(\lambda-2)^2 f_t^4}\mathbb{E}_{t-1}\left[(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\right]$$

$$= \frac{1}{4f_t^2} + \frac{(\lambda+1)^2}{4(\lambda-2)^2 f_t^4}\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}(r_t-\mu)^4\right] - \frac{1}{2f_t^2} \tag{88}$$

$$\Leftrightarrow \frac{(\lambda+1)^2+2(\lambda+1)}{4(\lambda-2)^2 f_t^4}\mathbb{E}_{t-1}\left[(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\right] = \frac{3}{4f_t^2} \tag{89}$$

$$\Leftrightarrow \mathbb{E}_{t-1}\left[(r_t-\mu)^4\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}\right] = \frac{3(\lambda-2)^2 f_t^2}{(\lambda+1)(\lambda+3)} \tag{90}$$

Now that we know the expressions for (*) and (**), $\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$ can be reduced to

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right] &= \frac{1}{4f_t^2} + \frac{(\lambda+1)^2}{4(\lambda-2)^2 f_t^4}\underbrace{\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-2}(r_t-\mu)^4\right]}_{=\frac{3(\lambda-2)^2 f_t^2}{(\lambda+1)(\lambda+3)}} \\
&\quad - \frac{\lambda+1}{2(\lambda-2)f_t^3}\underbrace{\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2\right]}_{=\frac{(\lambda-2)f_t}{\lambda+1}} \\
&= \frac{1}{4f_t^2} + \frac{3(\lambda+1)^2(\lambda-2)^2 f_t^2}{4(\lambda+1)(\lambda-2)^2(\lambda+3)f_t^4} - \frac{(\lambda+1)(\lambda-2)f_t}{2(\lambda+1)(\lambda-2)f_t^3} \\
&= \frac{3(\lambda+1)}{4(\lambda+3)f_t^2} - \frac{1}{4f_t^2} = \frac{1}{4f_t^2}\left(\frac{3(\lambda+1)}{\lambda+3}-1\right) \\
&= \frac{\lambda}{2f_t^2(\lambda+3)}.
\end{aligned} \tag{91}$$

Therefore,

$$S_t = \mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right]^{-1} = \frac{2f_t^2(\lambda+3)}{\lambda}. \tag{92}$$

Now to get the update equation

$$\begin{aligned}
f_{t+1} &= \omega + \alpha s_t + \beta f_t = \omega + \alpha \nabla_t S_t + \beta f_t \\
&= \omega + \alpha\left[\frac{\lambda+1}{2(\lambda-2)f_t^2}\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}\right]\left[\frac{2f_t^2(\lambda+3)}{\lambda}\right] + \beta f_t \\
&= \omega + \alpha\left[\left(\frac{\lambda+1}{\lambda-2}\right)\left(1+\frac{(r_t-\mu)^2}{f_t(\lambda-2)}\right)^{-1}(r_t-\mu)^2 - f_t\right]\left(\frac{\lambda+3}{\lambda}\right) + \beta f_t.
\end{aligned} \tag{93}$$

C Log-likelihood Skewed Student-t

In the derivations of the skew Student-t density we follow the parameterization of Zhu and Galbraith (2010). Here the skew Student-t is a special case of an asymmetric Student-t. The general form of the density function of the error terms is

$$f_\varepsilon(\varepsilon_t;\lambda,\alpha) = \begin{cases} K(\lambda)\left[1+\frac{1}{\lambda}\left(\frac{\varepsilon_t}{2\alpha}\right)^2\right]^{-\frac{\lambda+1}{2}} & \text{if } \varepsilon_t > 0 \\ K(\lambda)\left[1+\frac{1}{\lambda}\left(\frac{\varepsilon_t}{2\alpha}\right)^2\right]^{-\frac{\lambda+1}{2}} & \text{if } \varepsilon_t \leq 0 \end{cases} \tag{94}$$

$$= K(\lambda)\left(1+\frac{\varepsilon_t^2}{4\lambda(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(\varepsilon_t))^2}\right)^{-\frac{\lambda+1}{2}}, \tag{95}$$

where $K(\lambda) = \frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\sqrt{\pi\lambda}}$ and $\alpha \in (0,1)$ is the skewness parameter. This reparameterization of the Skew Student-t is ultimately equivalent to those of Hansen (1994), who first introduced an extension for skew Student-t, and Fernandez Steel (1998). Then the mean and variance of a skew Student-t ($\lambda = \lambda_1 = \lambda_2$, in asymmetric Student-t) random variable are

$$\mathbb{E}[\varepsilon_t] = 4K(\lambda)\left[-\alpha^2\frac{\lambda}{\lambda-1} + (1-\alpha)^2\frac{\lambda}{\lambda-1}\right] = 4K(\lambda)\left[(1-2\alpha)\frac{\lambda}{\lambda-1}\right], \tag{96}$$

$$\begin{aligned}\mathbb{V}\text{ar}[\varepsilon_t] &= 4\left[\alpha^3\frac{\lambda}{\lambda-1} + (1-\alpha)^3\frac{\lambda}{\lambda-1}\right] - 16K(\lambda)^2\left[-\alpha^2\frac{\lambda}{\lambda-1} + (1-\alpha)^2\frac{\lambda}{\lambda-1}\right] \\ &= 4\left[\alpha^3\frac{\lambda}{\lambda-1} + (1-\alpha)^3\frac{\lambda}{\lambda-1}\right] - 16K(\lambda)^2\left[(1-2\alpha)\frac{\lambda}{\lambda-1}\right] = Q(\lambda,\alpha).\end{aligned} \tag{97}$$

Using a similar standardization procedure as before where $r_t = \mu + a_t = \mu + \sigma_t\varepsilon_t$ and $\varepsilon_t \sim SST(\lambda,\alpha)$. Then $\mathbb{V}\text{ar}[a_t] = \sigma_t^2\mathbb{V}\text{ar}[\varepsilon_t] = \sigma_t^2 Q(\lambda,\alpha)$ and $\mathbb{V}\text{ar}[a_t] = \sigma_t^2$ can be obtained if $a_t = \sigma_t\frac{\varepsilon_t}{\sqrt{Q(\lambda,\alpha)}}$ because then it holds that $\mathbb{V}\text{ar}[a_t] = \sigma_t^2\frac{1}{Q(\lambda,\alpha)}\mathbb{V}\text{ar}[\varepsilon_t] = \sigma_t^2\frac{1}{Q(\lambda,\alpha)}Q(\lambda,\alpha) = \sigma_t^2$. To find the density function of a_t we need to use a transformation procedure again in order to get it for

$$a_t = \sigma_t\frac{\varepsilon_t}{\sqrt{Q(\lambda,\alpha)}}, \qquad \varepsilon_t \sim SST(\lambda,\alpha), \tag{98a}$$

$$\sigma_t^2 = \phi(\sigma_{t-1}^2, a_t;\boldsymbol{\theta}). \tag{98b}$$

Because $\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)$ is known conditional on past data, the same transformation theorem as before is applied

$$f_a(a_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2), D_{t-1};\boldsymbol{\theta}) = f_\varepsilon\left(h^{-1}(a_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2), D_{t-1};\boldsymbol{\theta})\right)\left|\frac{\partial h^{-1}(a_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2), D_{t-1};\boldsymbol{\theta})}{\partial a_t}\right|, \tag{99a}$$

$$\text{where } h^{-1}(a_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2), D_{t-1};\boldsymbol{\theta}) = \frac{a_t\sqrt{Q(\lambda,\alpha)}}{\sigma_t(\boldsymbol{\theta},\sigma_1^2)} = \varepsilon_t, \tag{99b}$$

and we get the following expression

$$f_a(a_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2),D_{t-1};\boldsymbol{\theta}) = K(\lambda)\left(\frac{\sqrt{Q(\lambda,\alpha)}}{\sigma_t(\boldsymbol{\theta},\sigma_1^2)}\right)\left[1+\frac{\left(\frac{a_t\sqrt{Q(\lambda,\alpha)}}{\sigma_t(\boldsymbol{\theta},\sigma_1^2)}\right)^2}{4\lambda\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}\left(\frac{a_t\sqrt{Q(\lambda,\alpha)}}{\sigma_t(\boldsymbol{\theta},\sigma_1^2)}\right)\right)^2}\right]^{-\frac{\lambda+1}{2}}$$
$$= K(\lambda)\sqrt{\frac{Q(\lambda,\alpha)}{\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)}}\left[1+\frac{a_t^2Q(\lambda,\alpha)}{4\lambda\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(a_t)\right)^2}\right]^{\frac{\lambda+1}{2}}. \tag{100}$$

Now that we have the density function of a_t, also the log density of of $u_t = r_t - \mu$ can easily be determined.

$$\log f_r(r_t|\sigma_t^2(\boldsymbol{\theta},\sigma_1^2),D_{t-1};\boldsymbol{\theta}) = \log K(\lambda) + \frac{1}{2}\log Q(\lambda,\alpha) - \frac{1}{2}\log\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)$$
$$-\frac{\lambda+1}{2}\log\left[1+\frac{(r_t-\mu)^2Q(\lambda,\alpha)}{4\lambda\sigma_t^2(\boldsymbol{\theta},\sigma_1^2)\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right]. \tag{101}$$

D Updating equation Skewed Student-t GAS(1,1)

Now that we have the log-density of the skew Student-t in Appendix C, we can use this to derive the update equation for the time-varying variance in the skew Student-t GAS(1,1) model.

The score w.r.t. the time-varying parameter $f_t = \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)$ is derived as

$$\begin{aligned}
\nabla_t &= \frac{\partial \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} \\
&= -\frac{1}{2}\frac{1}{f_t} - \frac{\lambda+1}{2}\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}\left(-\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\frac{1}{f_t^2}\right) \\
&= \frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2}\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}. \quad (102)
\end{aligned}$$

Again, for the scaling matrix S_t we use the inverse Fisher information matrix $S_t = \mathbb{E}[\nabla_t \nabla_t^\top]^{-1}$.

$$\begin{aligned}
\nabla_t \nabla_t^\top &= \left[\frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2}\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}\right]^2 \\
&= \frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2 \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4}\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4 \\
&\quad - \frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3}\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2 \qquad 2005/06/2 \\
&\quad + \frac{1}{4f_t^2}. \quad (103)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{t-1}[\nabla_t \nabla_t^\top] &= \frac{1}{4f_t^2} + \frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2 \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad \times \underbrace{\mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right]}_{(*)} \\
&\quad - \frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3} \\
&\quad \times \underbrace{\mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2\right]}_{(**)}. \quad (104)
\end{aligned}$$

Again, as before we will derive expressions for (*) and (**) using that $\mathbb{E}_{t-1}[\nabla_t] = 0$ for Z-estimators and the IME.

$$\mathbb{E}_{t-1}[\nabla_t] = \mathbb{E}_{t-1}\left[\frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2}\right.$$
$$\left.\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2 - \frac{1}{2f_t}\right] = 0 \tag{105}$$

$$\Leftrightarrow -\frac{1}{2f_t} + \frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2}$$
$$\mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2\right] = 0 \tag{106}$$

$$\Leftrightarrow \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2\right]$$
$$= \frac{8\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2}{(\lambda+1)Q(\lambda,\alpha)} = \frac{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}{(\lambda+1)Q(\lambda,\alpha)} \tag{107}$$

Yet, if we use this expression for (**), equation (104) reduces to

$$\mathbb{E}_{t-1}[\nabla_t \nabla_t^\top] = \frac{1}{4f_t^2} + \frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4}$$
$$\times \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right]$$
$$- \frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3}\frac{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}{(\lambda+1)Q(\lambda,\alpha)}$$
$$= \frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4}$$
$$\times \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right] - \frac{1}{4f_t^2}. \tag{108}$$

To apply the IME, we also need the alternative expression

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t}\right] &= \frac{1}{2f_t^2} - \frac{(\lambda+1)Q(\lambda,\alpha)}{4\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3} \\
&\quad \times \mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^2\right] \\
&\quad + \frac{(\lambda+1)(Q(\lambda,\alpha))^2}{32(\lambda)^2\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad \times \mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-1}(r_t-\mu)^4\right] \\
&= \frac{1}{2f_t^2} - \frac{(\lambda+1)Q(\lambda,\alpha)}{4\lambda\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3} \underbrace{\frac{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}{(\lambda+1)Q(\lambda,\alpha)}}_{\text{By equation (50)}} \\
&\quad + \frac{(\lambda+1)(Q(\lambda,\alpha))^2}{32(\lambda)^2\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad \times \mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right] \\
&= \frac{(\lambda+1)(Q(\lambda,\alpha))^2}{32(\lambda)^2\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad \times \mathbb{E}_{t-1}\left[\left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right] - \frac{1}{2f_t^2}. \qquad (109)
\end{aligned}$$

Then by the IME, $\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t}\right] = -\mathbb{E}_{t-1}\left[\frac{\partial \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t}\frac{\partial \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t}\right] = -\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$ and using this with equation (108) $\left(\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]\right)$ and the negative of equation (109) $\left(-\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t,D_{t-1};\boldsymbol{\theta})}{\partial f_t \partial f_t}\right]\right)$, we obtain:

$$
\begin{aligned}
&\frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad\times \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right]-\frac{1}{4f_t^2} \\
&= -\frac{(\lambda+1)(Q(\lambda,\alpha))^2}{32(\lambda)^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad\times \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right]+\frac{1}{2f_t^2} \qquad (110)\\
&\Leftrightarrow \frac{3}{4f_t^2}=\frac{2(\lambda+1)Q(\lambda,\alpha)^2+(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4} \\
&\quad\times \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right] \qquad (111)\\
&\Leftrightarrow \mathbb{E}_{t-1}\left[\left(1+\frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}\right)^{-2}(r_t-\mu)^4\right]= \\
&\frac{48(\lambda)^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^2}{(\lambda+1)(\lambda+3)Q(\lambda,\alpha)^2}. \qquad (112)
\end{aligned}
$$

Now that we have expressions for (*) and (**), $\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]$ can be reduced to

$$
\begin{aligned}
\mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right] &= \frac{1}{4f_t^2}+\left(\frac{(\lambda+1)^2 Q(\lambda,\alpha)^2}{64\lambda^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^4}\right) \\
&\quad\times\left(\frac{48(\lambda)^2\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^4 f_t^2}{(\lambda+1)(\lambda+3)Q(\lambda,\alpha)^2}\right) \\
&\quad-\left(\frac{(\lambda+1)Q(\lambda,\alpha)}{8\lambda\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^3}\right)\left(\frac{4\lambda f_t\left(\alpha-\mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2}{(\lambda+1)Q(\lambda,\alpha)}\right) \qquad (113)\\
&= \frac{3(\lambda+1)}{4f_t^2(\lambda+3)}-\frac{1}{4f_t^2}=\frac{\lambda}{2(\lambda+3)f_t^2}. \qquad (114)
\end{aligned}
$$

As a result,

$$
S_t=\mathbb{E}_{t-1}\left[\nabla_t\nabla_t^\top\right]^{-1}=\frac{2(\lambda+3)f_t^2}{\lambda}. \qquad (115)
$$

Now to get the update equation

$$
\begin{aligned}
f_{t+1} &= \omega + \alpha s_t + \beta f_t = \omega + \alpha \nabla_t S_t + \beta f_t \\
&= \omega + \alpha \left[\frac{2(\lambda+3) f_t^2}{\lambda} \right] \left[\frac{(\lambda+1) Q(\lambda,\alpha)}{8\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2 f_t^2} \right. \\
&\quad \left. \times \left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2} \right)^{-1} (r_t-\mu)^2 - \frac{1}{2f_t} \right] + \beta f_t \\
&= \omega + \alpha \left[\frac{(\lambda+1)(\lambda+3) Q(\lambda,\alpha)}{4\lambda^2 \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2} \left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2} \right)^{-1} \right. \\
&\quad \left. \times (r_t-\mu)^2 - \frac{(\lambda-3) f_t}{\lambda} \right] + \beta f_t \\
&= \omega + \alpha \left(\frac{\lambda+3}{\lambda} \right) \left[\frac{(\lambda+1) Q(\lambda,\alpha)}{4\lambda \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2} \right. \\
&\quad \left. \times \left(1 + \frac{(r_t-\mu)^2 Q(\lambda,\alpha)}{4\lambda f_t \left(\alpha - \mathbb{1}_{\{\mathbb{R}_+\}}(r_t-\mu)\right)^2} \right)^{-1} (r_t-\mu)^2 - f_t \right] + \beta f_t. \qquad (116)
\end{aligned}
$$

E Log density GAS(1,1)-Laplace

The standardized GAS(1,1)-Laplace model is given by

$$r_t = \sigma_t \frac{\varepsilon_t}{\sqrt{2}b}, \qquad \text{where} \qquad \varepsilon_t \sim \mathcal{L}(b), \tag{117a}$$

$$\sigma_t^2 = \phi(\sigma_{t-1}^2, a_{t-1}; \boldsymbol{\theta}), \tag{117b}$$

where ϕ is unknown and $b > 0$ is the scale parameter of the Laplace distribution. The density function of the disturbances is given by the following expression

$$f_\varepsilon(\varepsilon_t; b) = \frac{1}{2b} \exp\left(-\frac{|\varepsilon_t|}{b}\right). \tag{118}$$

Note that in this case $\mathbb{V}\text{ar}[r_t] = \sigma_t^2 \frac{1}{2b^2} \mathbb{V}\text{ar}[\varepsilon_t] = \sigma_t^2 \frac{1}{2b^2} 2b^2 = \sigma_t^2$. To obtain the density function we have to use the transformation theorem (see Appendix A) again, with

$$h^{-1}(r_t | \sigma_t(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = \frac{r_t \sqrt{2} b}{\sigma_t} = \varepsilon_t, \tag{119}$$

and we obtain the following expression

$$\begin{aligned} f_r(r_t | \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= \frac{1}{2b} \exp\left(-\frac{\left|\frac{r_t\sqrt{2}b}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right|}{b}\right) \left(\frac{\sqrt{2}b}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right) \\ &= \frac{1}{\sqrt{2}\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \exp\left(-\frac{|r_t|\sqrt{2}}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right). \end{aligned} \tag{120}$$

Then the log density is given by

$$\log f_r(r_t | \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = -\frac{1}{2}\log 2 - \log \sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \frac{|r_t|\sqrt{2}}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}. \tag{121}$$

F Updating equation GAS(1,1)-Laplace

Top obtain the variance updating equation for the standardized GAS(1,1)-Laplace we follow similar steps as before. From Appendix E we know that the standardized Laplace log density, where $f_t = \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)$ is given conditional on past data D_{t-1}, is

$$\log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) = -\frac{1}{2}\log 2 - \frac{1}{2}\log f_t - \frac{|r_t|\sqrt{2}}{\sqrt{f_t}}.$$

The score w.r.t. the time-varying parameter is derived as

$$\nabla_t = \frac{\partial \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} = -\frac{1}{2f_t} + \frac{|r_t|}{\sqrt{2}f_t^{3/2}}. \tag{122}$$

Again, for the scaling matrix S_t we use the inverse Fisher information matrix $S_t = \mathbb{E}[\nabla_t \nabla_t^\top]^{-1}$. To obtain this we use the property for Z-estimators that $\mathbb{E}[\nabla_t] = 0$ and the IME.

$$\begin{aligned}
\mathbb{E}_{t-1}\left[\nabla_t\right] = \mathbb{E}_{t-1}\left[-\frac{1}{2f_t} + \frac{|r_t|}{\sqrt{2}f_t^{3/2}}\right] &= 0 \\
\Leftrightarrow -\frac{1}{2f_t} + \frac{1}{\sqrt{2}f_t^{3/2}}\mathbb{E}_{t-1}|r_t| &= 0 \\
\Leftrightarrow \mathbb{E}_{t-1}|r_t| = \frac{1}{2f_t}\frac{\sqrt{2}f_t^{3/2}}{1} &= \frac{\sqrt{f_t}}{\sqrt{2}}
\end{aligned} \tag{123}$$

$$\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t} = \frac{1}{2f_t^2} - \frac{3|r_t|}{2\sqrt{2}f_t^{5/2}} \tag{124}$$

$$\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t}\right] = \frac{1}{2f_t^2} - \frac{3}{2\sqrt{2}f_t^{5/2}}\underbrace{\mathbb{E}_{t-1}|r_t|}_{\frac{\sqrt{f_t}}{\sqrt{2}}} = \frac{1}{2f_t^2} - \frac{3}{4f_t^2} = -\frac{1}{4f_t^2} \tag{125}$$

$$\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right] \overset{\text{IME}}{=} -\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t}\right] = \frac{1}{4f_t^2} \tag{126}$$

Hence the scaling matrix becomes $S_t = \mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]^{-1} = 4f_t^2$. Then it follows that the variance updating equation is

$$\begin{aligned}
f_{t+1} &= \omega + \alpha s_t + \beta f_t = \omega + \alpha \nabla_t S_t + \beta f_t \\
&= \omega + \alpha\left[-\frac{1}{2f_t} + \frac{|r_t|}{\sqrt{2}f_t^{3/2}}\right] 4f_t^2 + \beta f_t \\
&= \omega + \alpha\left[-2f_t + 2\sqrt{2}\sqrt{f_t}|r_t|\right] + \beta f_t.
\end{aligned} \tag{127}$$

G Log density GAS(1,1)-GED

The standardized GAS(1,1)-GED model is defined as

$$r_t = \sigma_t \varepsilon_t \sqrt{\frac{\Gamma(\nu^{-1})}{\lambda^2 \Gamma(3\nu^{-1})}}, \qquad \text{where} \qquad \varepsilon_t \sim GED(\lambda, \nu), \tag{128a}$$

$$\sigma_t^2 = \phi(\sigma_{t-1}^2, a_{t-1}; \boldsymbol{\theta}). \tag{128b}$$

Here λ is the scaling parameter and ν is a shape parameter. Note that

$$\mathbb{V}\text{ar}[r_t] = \sigma_t^2 \frac{\Gamma(\nu^{-1})}{\lambda^2 \Gamma(3\nu^{-1})} \mathbb{V}\text{ar}[\varepsilon_t] = \sigma_t^2 \frac{\Gamma(\nu^{-1})}{\lambda^2 \Gamma(3\nu^{-1})} \frac{\lambda^2 \Gamma(3\nu^{-1})}{\Gamma(\nu^{-1})} = \sigma_t^2. \tag{129}$$

We only know the density function of the innovations, which is given by

$$f_\varepsilon(\varepsilon_t; \lambda\nu) = \frac{\nu}{2\lambda\Gamma(\nu^{-1})} \exp\left(-\left|\frac{\varepsilon_t}{\lambda}\right|^\nu\right). \tag{130}$$

Then by the transformation theorem

$$h^{-1}(r_t | \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) = \frac{r_t}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2) \sqrt{\frac{\Gamma(\nu^{-1})}{\lambda^2 \Gamma(3\nu^{-1})}}}. \tag{131}$$

$$\begin{aligned}
f_r(r_t | \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= \frac{\nu}{2\lambda\Gamma(\nu^{-1})} \exp\left(-\left|\frac{r_t \sqrt{\lambda^2 \Gamma(3\nu^{-1})}}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2) \sqrt{\Gamma(\nu^{-1})} \lambda}\right|^\nu\right) \frac{\sqrt{\lambda^2 \Gamma(3\nu^{-1})}}{\sqrt{\Gamma(\nu^{-1})} \sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \\
&= \frac{\nu \Gamma(3\nu^{-1})}{2\Gamma(\nu^{-1}) \sqrt{\Gamma(\nu^{-1})} \sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \exp\left(-\left|\frac{r_t \sqrt{\Gamma(3\nu^{-1})}}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right|^\nu\right) \\
&= \frac{\nu}{\psi(\nu) 2^{1+\frac{1}{\nu}} \Gamma(\nu^{-1}) \sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \exp\left(-\frac{1}{2}\left|\frac{r_t \sqrt{\Gamma(3\nu^{-1})}}{2^{-\frac{1}{\nu}} \sqrt{\Gamma(\nu^{-1})} \sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right|^\nu\right) \\
&= \frac{K(\nu)}{\sigma_t(\boldsymbol{\theta}, \sigma_1^2)} \exp\left(-\frac{1}{2}\left|\frac{r_t}{\psi(\nu) \sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right|^\nu\right),
\end{aligned} \tag{132}$$

$$\text{where} \quad K(\nu) = \frac{\nu}{\psi(\nu) 2^{1+\frac{1}{\nu}} \Gamma(\nu^{-1})}, \quad \psi(\nu) = \left(\frac{2^{-\frac{2}{\nu}} \Gamma(\nu^{-1})}{\Gamma\left(\frac{3}{\nu}\right)}\right)^{\frac{1}{2}}.$$

Then it follows that the log density is given by

$$\begin{aligned}
\log f_r(r_t | \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2), D_{t-1}; \boldsymbol{\theta}) &= \log K(\nu) - \log \sigma_t(\boldsymbol{\theta}, \sigma_1^2) - \frac{1}{2}\left|\frac{r_t}{\psi(\nu) \sigma_t(\boldsymbol{\theta}, \sigma_1^2)}\right|^\nu \\
&= \log K(\nu) - \frac{1}{2} \log \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2) - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^\nu \left(\frac{1}{\sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)}\right)^{\frac{\nu}{2}}.
\end{aligned} \tag{133}$$

H Updating equation GAS(1,1)-GED

To obtain the variance updating equation for the standardized GAS(1,1)-GED we follow similar steps as before. From Appendix G we know that the standardized GED log density, where $f_t = \sigma_t^2(\boldsymbol{\theta}, \sigma_1^2)$ is given conditional on past data D_{t-1}, is

$$\log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) = \log K(\nu) - \frac{1}{2}\log f_t - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\left(\frac{1}{f_t}\right)^{\frac{\nu}{2}}.$$

The score w.r.t. the time-varying parameter is derived as

$$\begin{aligned}\nabla_t = \frac{\partial \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} &= -\frac{1}{2f_t} - \frac{1}{2}\left(-\frac{\nu}{2}\right)\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}+1}} \\ &= -\frac{1}{2f_t} + \frac{\nu}{4}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}+1}}. \end{aligned} \tag{134}$$

Again, for the scaling matrix S_t we use the inverse Fisher information matrix $S_t = \mathbb{E}[\nabla_t \nabla_t^\top]^{-1}$. To obtain this we use the property for Z-estimators that $\mathbb{E}[\nabla_t] = 0$ and the IME.

$$\mathbb{E}_{t-1}\left[\nabla_t\right] = \frac{\nu}{4}\frac{1}{f_t^{\frac{\nu}{2}+1}}\mathbb{E}_{t-1}\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\right] - \frac{1}{2f_t} = 0 \tag{135}$$

$$\Leftrightarrow \mathbb{E}_{t-1}\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\right] = \frac{1}{2f_t}f_t^{\frac{\nu}{2}+1}\frac{4}{\nu} = \frac{2f_t^{\frac{\nu}{2}}}{\nu} \tag{136}$$

$$\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t} = \frac{1}{2f_t^2} - \frac{\nu}{4}\left(\frac{\nu}{2}+1\right)\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}+2}} \tag{137}$$

$$\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t}\right] = \frac{1}{2f_t^2} - \frac{\nu}{4}\left(\frac{\nu}{2}+1\right)\mathbb{E}_{t-1}\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\right]\frac{1}{f_t^{\frac{\nu}{2}+2}} \tag{138}$$

$$\begin{aligned}\mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right] &\overset{\text{IME}}{=} -\mathbb{E}_{t-1}\left[\frac{\partial^2 \log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t \partial f_t}\right] \\ &= \frac{\nu}{4}\left(\frac{\nu}{2}+1\right)\underbrace{\mathbb{E}_{t-1}\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\right]}_{\frac{2f_t^{\frac{\nu}{2}}}{\nu}}\frac{1}{f_t^{\frac{\nu}{2}+2}} - \frac{1}{2f_t^2} \\ &= \frac{\nu}{4}\left(\frac{\nu}{2}+1\right)\left(\frac{2}{\nu f_t^2}\right) - \frac{1}{2f_t^2} = \frac{\nu}{4f_t^2} \end{aligned} \tag{139}$$

Hence the scaling matrix becomes $S_t = \mathbb{E}_{t-1}\left[\nabla_t \nabla_t^\top\right]^{-1} = \frac{4f_t^2}{\nu}$. Then it follows that the variance updating equation is

$$\begin{aligned} f_{t+1} &= \omega + \alpha s_t + \beta f_t = \omega + \alpha \nabla_t S_t + \beta f_t \\ &= \omega + \alpha\left(-\frac{1}{2f_t} + \frac{\nu}{4}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}+1}}\right)\frac{4f_t^2}{\nu} + \beta f_t \\ &= \omega + \alpha\left[\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\frac{1}{f_t^{\frac{\nu}{2}-1}} - \frac{2f_t}{\nu}\right] + \beta f_t. \end{aligned} \tag{140}$$

I Log density and update equation LL Real Gaussian GAS models

Since the Gaussian log linear realized GAS(1,1) model is given by

$$r_t = \sqrt{f_t}\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \tag{141a}$$

$$\log x_t = \xi + \phi \log f_t + \tau(\varepsilon_t) + \sqrt{\frac{\lambda_u - 2}{\lambda_u}}\sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{141b}$$

$$\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) + \tau_2\left(\frac{r_t^2}{f_t} - 1\right). \tag{141c}$$

We want to obtain the log density functions of r_t and x_t and from before we know that

$$\log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) = -\frac{1}{2}\left[\log 2\pi + \log f_t + \frac{r_t^2}{f_t}\right]$$

$$\begin{aligned}\log f_x(x_t|r_t, f_t, D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda_u + 1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2) \\ &\quad - \log\sigma_u - \frac{\lambda_u + 1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right).\end{aligned}$$

$$\begin{aligned}\log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta}) &= \log f(\boldsymbol{r_T}|f_t, D_{t-1}; \boldsymbol{\theta}) + \log f(\boldsymbol{x_T}|\boldsymbol{r_T}, f_t, D_{t-1}; \boldsymbol{\theta}) \\ &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log f_t - \frac{r_t^2}{2f_t} + \log\Gamma\left(\frac{\lambda_u + 1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) \\ &\quad - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2) - \log\sigma_u \\ &\quad - \frac{\lambda_u + 1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right).\end{aligned} \tag{142}$$

Then it follows that

$$\begin{aligned}\nabla_t &= \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial \log f_t} = \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} \times f_t \\ &= \Bigg[-\frac{1}{2f_t} + \frac{r_t^2}{2f_t^2} - \frac{\lambda_u + 1}{2} \\ &\quad \times \left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)^{-1} \\ &\quad \times \left(\frac{2\left(\log x_t - \xi - \phi\log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)}{\sigma_u^2(\lambda_u - 2)}\right) \\ &\quad \times \left(-\frac{\phi}{f_t} - \frac{1}{2}\tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2\left(\frac{r_t^2}{f_t^2}\right)\right)\Bigg] \times f_t,\end{aligned} \tag{143}$$

and the updating equation is given by

$$\begin{aligned}\log f_{t+1} &= \omega + \alpha \nabla_t S_t + \beta \log f_t = \omega + \alpha \nabla_t \cdot 1 + \beta \log f_t \\ &= \omega + \alpha \left\{ \left[-\frac{1}{2f_t} + \frac{r_t^2}{2f_t^2} - \frac{\lambda_u + 1}{2} \right.\right. \\ &\quad \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)^2}{\sigma_u^2 (\lambda_u - 2)} \right)^{-1} \\ &\quad \times \left(\frac{2 \left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)}{\sigma_u^2 (\lambda_u - 2)} \right) \\ &\quad \left. \left. \times \left(-\frac{\phi}{f_t} - \frac{1}{2} \tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2 \left(\frac{r_t^2}{f_t^2} \right) \right) \right] \times f_t \right\} + \beta \log f_t. \end{aligned} \tag{144}$$

J Log density and update equation LL Realized GAS(1,1)-t

Since the Gaussian log linear realized GAS(1,1) model is given by

$$r_t = \sqrt{f_t}\sqrt{\frac{\lambda_u - 2}{\lambda_u}}\varepsilon_t, \qquad \varepsilon_t \sim TID(\lambda), \tag{145a}$$

$$\log x_t = \xi + \phi \log f_t + \tau(\varepsilon_t) + \sqrt{\frac{\lambda_u - 2}{\lambda_u}}\sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{145b}$$

$$\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) + \tau_2\left(\frac{r_t^2}{f_t} - 1\right). \tag{145c}$$

We want to obtain the log density functions of r_t and x_t and from before we know that

$$\begin{aligned}
\log f_r(r_t|f_t, D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda-2) \\
&\quad - \log f_t - \frac{\lambda+1}{2}\log\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right) \\
\log f_x(x_t|r_t, f_t, D_{t-1}; \boldsymbol{\theta}) &= \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u-2) \\
&\quad - \log\sigma_u - \frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u-2)}\right).
\end{aligned}$$

$$\begin{aligned}
\log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta}) &= \log f(\boldsymbol{r_T}|f_t, D_{t-1}; \boldsymbol{\theta}) + \log f(\boldsymbol{x_T}|\boldsymbol{r_T}, f_t, D_{t-1}; \boldsymbol{\theta}) \\
&= -\log\pi + \log\Gamma\left(\frac{\lambda+1}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log(\lambda-2) \\
&\quad - \log f_t - \frac{\lambda+1}{2}\log\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right) \\
&\quad + \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log(\lambda_u-2) - \log\sigma_u \\
&\quad - \frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u-2)}\right).
\end{aligned} \tag{146}$$

Then it follows that

$$\begin{aligned}
\nabla_t &= \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial \log f_t} = \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} \times f_t \\
&= \Bigg[-\frac{1}{2f_t} - \frac{\lambda+1}{2}\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right)^{-1}\left(-\frac{r_t^2}{(\lambda-2)f_t^2}\right) \\
&\quad -\frac{\lambda_u+1}{2} \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)^{-1} \\
&\quad \times \left(\frac{2\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)}{\sigma_u^2(\lambda_u - 2)}\right) \\
&\quad \times \left(-\frac{\phi}{f_t} - \frac{1}{2}\tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2\left(\frac{r_t^2}{f_t^2}\right)\right)\Bigg] \times f_t, \qquad (147)
\end{aligned}$$

and the updating equation is given by

$$\begin{aligned}
\log f_{t+1} &= \omega + \alpha \nabla_t S_t + \beta \log f_t = \omega + \alpha \nabla_t \cdot 1 + \beta \log f_t \\
&= \omega + \alpha \Bigg\{ \Bigg[-\frac{1}{2f_t} - \frac{\lambda+1}{2}\left(1 + \frac{r_t^2}{(\lambda-2)f_t}\right)^{-1}\left(-\frac{r_t^2}{(\lambda-2)f_t^2}\right) \\
&\quad -\frac{\lambda_u+1}{2} \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)^2}{\sigma_u^2(\lambda_u - 2)}\right)^{-1} \\
&\quad \times \left(\frac{2\left(\log x_t - \xi - \phi \log f_t - \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) - \tau_2\left(\frac{r_t^2}{f_t}\right) + \tau_2\right)}{\sigma_u^2(\lambda_u - 2)}\right) \\
&\quad \times \left(-\frac{\phi}{f_t} - \frac{1}{2}\tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2\left(\frac{r_t^2}{f_t^2}\right)\right)\Bigg] \times f_t \Bigg\} + \beta \log f_t. \qquad (148)
\end{aligned}$$

K Log density and update equation LL Realized GAS(1,1)-GED

Since the Gaussian log linear realized GAS(1,1) model is given by

$$r_t = \sqrt{f_t}\varepsilon_t \sqrt{\frac{\Gamma(\nu^{-1})}{\lambda^2\Gamma(3\nu^{-1})}}, \qquad \text{where} \qquad \varepsilon_t \sim GED(\lambda,\nu), \tag{149a}$$

$$\log x_t = \xi + \phi \log f_t + \tau(\varepsilon_t) + \sqrt{\frac{\lambda_u - 2}{\lambda_u}}\sigma_u u_t, \qquad u_t \sim TID(\lambda_u), \tag{149b}$$

$$\text{where} \quad \tau(\varepsilon_t) = \tau_1(\varepsilon_t) + \tau_2(\varepsilon_t^2 - 1) = \tau_1\left(\frac{r_t}{\sqrt{f_t}}\right) + \tau_2\left(\frac{r_t^2}{f_t} - 1\right). \tag{149c}$$

We want to obtain the log density functions of r_t and x_t. From Appendix G we know that

$$\log f_r(r_t|f_t, D_{t-1};\boldsymbol{\theta}) = \log K(\nu) - \frac{1}{2}\log f_t - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\left(\frac{1}{f_t}\right)^{\frac{\nu}{2}},$$

$$\text{where} \quad K(\nu) = \frac{\nu}{\psi(\nu)2^{1+\frac{1}{\nu}}\Gamma(\nu^{-1})}, \quad \psi(\nu) = \left(\frac{2^{-\frac{2}{\nu}}\Gamma(\nu^{-1})}{\Gamma\left(\frac{3}{\nu}\right)}\right)^{\frac{1}{2}},$$

and from before we also have that

$$\log f_x(x_t|r_t, f_t, D_{t-1};\boldsymbol{\theta}) = \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2)$$
$$- \log\sigma_u - \frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right).$$

Then it follows that

$$\begin{aligned}
\log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1};\boldsymbol{\theta}) &= \log f(\boldsymbol{r_T}|f_t, D_{t-1};\boldsymbol{\theta}) + \log f(\boldsymbol{x_T}|\boldsymbol{r_T}, f_t, D_{t-1};\boldsymbol{\theta}) \\
&= \log K(\nu) - \frac{1}{2}\log f_t - \frac{1}{2}\left|\frac{r_t}{\psi(\nu)}\right|^{\nu}\left(\frac{1}{f_t}\right)^{\frac{\nu}{2}} \\
&\quad + \log\Gamma\left(\frac{\lambda_u+1}{2}\right) - \log\Gamma\left(\frac{\lambda_u}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\lambda_u - 2) \\
&\quad - \log\sigma_u - \frac{\lambda_u+1}{2}\log\left(1 + \frac{\left(\log x_t - \xi - \phi\log f_t - \tau\left(\frac{r_t}{\sqrt{f_t}}\right)\right)^2}{\sigma_u^2(\lambda_u - 2)}\right).
\end{aligned} \tag{150}$$

As before, the next step is to obtain the score. In order to do this we refer to the appendices before. In particular, Appendix H for the derivative of the log density of the "GED-part".

$$\begin{aligned}
\nabla_t &= \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial \log f_t} = \frac{\partial \log f(\boldsymbol{r_T}, \boldsymbol{x_T}|f_t, D_{t-1}; \boldsymbol{\theta})}{\partial f_t} \times f_t \\
&= \Bigg[-\frac{1}{2f_t} + \frac{\nu}{4} \left| \frac{r_t}{\psi(\nu)} \right|^{\nu} \frac{1}{f_t^{\frac{\nu}{2}+1}} - \frac{\lambda_u + 1}{2} \\
&\quad \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)^2}{\sigma_u^2 (\lambda_u - 2)} \right)^{-1} \\
&\quad \times \left(\frac{2 \left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)}{\sigma_u^2 (\lambda_u - 2)} \right) \\
&\quad \times \left(-\frac{\phi}{f_t} - \frac{1}{2} \tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2 \left(\frac{r_t^2}{f_t^2} \right) \right) \Bigg] \times f_t, \qquad (151)
\end{aligned}$$

and the updating equation is given by

$$\begin{aligned}
\log f_{t+1} &= \omega + \alpha \nabla_t S_t + \beta f_t = \omega + \alpha \nabla_t \cdot 1 + \beta \log f_t \\
&= \omega + \alpha \Bigg\{ \Bigg[-\frac{1}{2f_t} + \frac{\nu}{4} \left| \frac{r_t}{\psi(\nu)} \right|^{\nu} \frac{1}{f_t^{\frac{\nu}{2}+1}} - \frac{\lambda_u + 1}{2} \\
&\quad \times \left(1 + \frac{\left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)^2}{\sigma_u^2 (\lambda_u - 2)} \right)^{-1} \\
&\quad \times \left(\frac{2 \left(\log x_t - \xi - \phi \log f_t - \tau_1 \left(\frac{r_t}{\sqrt{f_t}} \right) - \tau_2 \left(\frac{r_t^2}{f_t} \right) + \tau_2 \right)}{\sigma_u^2 (\lambda_u - 2)} \right) \\
&\quad \times \left(-\frac{\phi}{f_t} - \frac{1}{2} \tau_1 r_t f_t^{-\frac{3}{2}} + \tau_2 \left(\frac{r_t^2}{f_t^2} \right) \right) \Bigg] \times f_t \Bigg\} + \beta \log f_t. \qquad (152)
\end{aligned}$$

L Diebold-Mariano tests

Table 7: Comparison of best non-realized and worst realized models (fixed window close-to-close)

		MAE	
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH t	EGARCH t	EGARCH t
		Comparison with	
Worst Realized GARCH	Real GARCH $\mathcal{N}$ (t) (RK)	Real GARCH $\mathcal{N}$ (t) (RK)	Real GARCH $\mathcal{N}$ (t) (RK)
DM	2.7775***	4.3364***	4.1028***
Worst Realized GAS	LL Real GAS t (no τ) (RK)	LL Real GAS $\mathcal{N}$ (no τ) (RK)	LL Real GAS $\mathcal{N}$ (no τ) (RK)
DM	2.7096***	4.9158***	4.4993***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

		MSE	
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	EGARCH t	EGARCH t	EGARCH t
		Comparison with	
Worst Realized GARCH	Real EGARCH $\mathcal{N}$ (t) (RK)	Real EGARCH $\mathcal{N}$ (t) (RK)	Real EGARCH $\mathcal{N}$ (t) (RK)
DM	-0.97434	0.95719	0.09812
Worst Realized GAS	LL Real GAS t (τ) (BV)	LL Real GAS t (no τ) (RK)	LL Real GAS t (no τ) (BV)
DM	-0.47807	1.2318	0.45537
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold Mariano statistic for the best non-realized model based on the MAE and MSE statistic compared to (1) the worst realized GAS and (2) the worst realized GARCH model. If the DM is positive, it indicates that the test is in favor of the worst realized models.

Table 8: Comparison of best realized GAS and worst realized models GARCH (fixed window close-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Best Realized-GAS	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)
Comparison with			
Best Realized-GARCH	Real EGARCH t (t) (BV)	Real EGARCH t (t) (BV)	Real EGARCH t (t) (BV)
DM	-1.2349	2.6861***	0.88018
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Best Realized-GAS	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)	LL Real GAS t (τ) (RK)
Comparison with			
Best Realized-GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)
DM	-2.5925***	-1.7058**	-2.1346**
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold-Mariano (DM) statistic for the best realized GAS based on the MAE and MSE statistic compared to the best realized GARCH. If the DM statistic is positive, it indicates that the test is in favor of the best realized GARCH.

Table 9: Comparison of best non-realized and worst realized models (rolling window close-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	GAS-GED	GAS-GED	GAS-GED
Comparison with			
Worst Realized GARCH	Real GARCH $\mathcal{N}(\mathcal{N})$ (RV)	Real GARCH $\mathcal{N}(\mathcal{N})$ (RV)	Real GARCH $\mathcal{N}(\mathcal{N})$ (RV)
DM	1.8147*	2.4235***	2.4167***
Worst Realized GAS	LL Real GAS $t(no\tau)$ (RK)	LL Real GAS $t(no\tau)$ (BV)	LL Real GAS $t(no\tau)$ (RK)
DM	-0.79832	1.2189	-0.066101
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Best non-Realized	GAS-Skewed t	GAS-GED	GAS-GED
Comparison with			
Worst Realized GARCH	Real GARCH $\mathcal{N}(\mathcal{N})$ (RK)	Real GARCH $\mathcal{N}(\mathcal{N})$ (RK)	Real GARCH $\mathcal{N}(\mathcal{N})$ (RK)
DM	-1.2480	0.23436	-0.41411
Worst Realized GAS	LL Real GAS $t(no\tau)$ (RK)	LL Real GAS $t(no\tau)$ (RK)	LL Real GAS $t(no\tau)$ (RK)
DM	-1.8605*	-1.3238	-1.6865*
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold Mariano statistic for the best non-realized model based on the MAE and MSE statistic compared to (1) the worst realized GAS and (2) the worst realized GARCH model. If the DM is positive, it indicates that the test is in favor of the worst realized models.

Table 10: Comparison of best realized GAS and worst realized models GARCH (rolling window close-to-close)

MAE			
Benchmark	**RK**	**BV**	**RV**
Best Realized-GAS	LL Real GAS $t(\tau)$ (RK)	LL Real GAS $t(\tau)$ (RK)	LL Real GAS $t(\tau)$ (RK)
Comparison with			
Best Realized-GARCH	LL Real GARCH $\mathcal{N}(\mathcal{N})$ (BV)	LL Real GARCH $\mathcal{N}(\mathcal{N})$ (BV)	LL Real GARCH $\mathcal{N}(\mathcal{N})$ (BV)
DM	0.18847	1.9116*	1.2036
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

MSE			
Benchmark	**RK**	**BV**	**RV**
Best Realized-GAS	LL Real GAS $t(\tau)(RK)$	LL Real GAS $t(\tau)(RK)$	LL Real GAS $t(\tau)(RK)$
Comparison with			
Best Realized-GARCH	Real EGARCH $\mathcal{N}(t)$ (RK)	Real EGARCH $\mathcal{N}(t)$ (RK)	Real GARCH $\mathcal{N}(t)$ (BV)
DM	-1.8350*	-0.83974	-1.2671
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

Note: this table shows the Diebold-Mariano (DM) statistic for the best realized GAS based on the MAE and MSE statistic compared to the best realized GARCH. If the DM statistic is positive, it indicates that the test is in favor of the best realized GARCH.

Table 11: DM-tests for comparison between rolling and fixed window

	MAE Open-to-Close		
Benchmark	**RK**	**BV**	**RV**
Best Rolling Realized-GAS	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)	LL Real GAS GED (τ) (BV)
Best Fixed Realized-GAS	**LL Real GAS t (τ) (BV)**	**LL Real GAS t (τ) (BV)**	**LL Real GAS t (τ) (BV)**
DM	4.5912***	6.4129***	5.3827***
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

	MSE Open-to-close		
Benchmark	**RK**	**BV**	**RV**
Best Rolling Realized-GAS	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)	LL Real GAS $\mathcal{N}$ (no τ) (BV)
Best Fixed Realized-GAS	**LL Real GAS GED (no τ) (BV)**	LL Real GAS t (no τ) (RK)	LL Real GAS t (no τ) (RK)
DM	1.4202*	1.1847	1.0011
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

	MAE Open-to-Close		
Benchmark	**RK**	**BV**	**RV**
Best Rolling Realized-GARCH	Real EGARCH t (t) (BV)	Real EGARCH t (t) (BV)	Real EGARCH t (t) (BV)
Best Fixed Realized-GARCH	Real EGARCH t (t) (RV)	**Real EGARCH $\mathcal{N}$ t (BV)**	**Real EGARCH $\mathcal{N}$ t (BV)**
DM	-0.25212	3.297***	1.4223*
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

	MSE Open-to-close		
Benchmark	**RK**	**BV**	**RV**
Best Rolling Realized-GARCH	Real GARCH $\mathcal{N}$ (t) (BV)	**Real GARCH $\mathcal{N}$ (t) (BV)**	Real GARCH $\mathcal{N}$ (t) (BV)
Best Fixed Realized-GARCH	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)	Real GARCH $\mathcal{N}$ ($\mathcal{N}$) (BV)
DM	0.34286	-1.3555*	-0.54521
*Note: *** significance at the 1% confidence level, ** at 5%, * at 10%.*			

M In-Sample Results

Table 12: Parameter estimates non-realized models

GARCH(1,1)-$\mathcal{N}$		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.056 (0.018)	0.05 (0.022)
ω	0.031 (0.01)	0.031 (0.01)
α	0.082 (0.014)	0.082 (0.014)
β	0.893 (0.021)	0.893 (0.021)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2212.489	-2502.097

GARCH(1,1)-t		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.046 (0.017)	0.035 (0.02)
ω	0.026 (0.01)	0.021 (0.016)
α	0.119 (0.025)	0.086 (0.021)
β	0.855 (0.032)	0.903 (0.032)
ν	7.421 (1.29)	5.228 (0.853)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2186.016	-2441.390

EGARCH(1,1)-$\mathcal{N}$		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.044 (0.018)	0.02 (0.021)
ω	-0.168 (0.023)	-0.112 (0.016)
α	0.204 (0.027)	0.147 (0.02)
β	0.967 (0.008)	0.973 (0.006)
γ	-0.344 (0.085)	-0.659 (0.131)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2210.297	-2479.056

EGARCH(1,1)-t		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.039 (0.017)	0.015 (0.016)
ω	-0.17 (0.027)	-0.121 (0.022)
α	0.209 (0.034)	0.159 (0.029)
β	0.972 (0.009)	0.976 (0.007)
γ	-0.301 (0.095)	-0.564 (0.143)
ν	7.802 (1.351)	6.16 (0.836)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2184.889	-2426.607

NGARCH(1,1)-$\mathcal{N}$		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.039 (0.019)	0.017 (0.022)
ω	0.031 (0.008)	0.035 (0.009)
α	0.109 (0.019)	0.07 (0.013)
β	0.839 (0.025)	0.853 (0.023)
γ	0.364 (0.102)	0.813 (0.166)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2204.925	-2484.136

NGARCH(1,1)-t		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.035 (0.018)	0.015 (0.02)
ω	0.026 (0.009)	0.029 (0.01)
α	0.109 (0.023)	0.073 (0.018)
β	0.85 (0.029)	0.862 (0.027)
γ	0.326 (0.116)	0.748 (0.193)
ν	8.024 (1.43)	6.007 (0.799)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2181.429	-2430.083

GAS(1,1)-t		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.046 (0.017)	0.035 (0.02)
ω	0.026 (0.01)	0.021 (0.016)
A	0.119 (0.025)	0.086 (0.021)
B	0.855 (0.032)	0.903 (0.032)
ν	7.421 (1.29)	5.228 (0.853)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2186.016	-2441.390

GAS(1,1)-GED		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.049 (0.018)	0.033 (0.02)
ν	1.491 (0.067)	1.328 (0.055)
ω	0.029 (0.009)	0.031 (0.012)
A	0.108 (0.02)	0.087 (0.018)
B	0.965 (0.013)	0.972 (0.013)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2190.281	-2447.595

GAS(1,1)-Laplace		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.042 (0.008)	0.028 (0.011)
ω	0.031 (0.013)	0.036 (0.017)
A	0.103 (0.021)	0.091 (0.02)
B	0.97 (0.016)	0.974 (0.016)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-1616.447	-1859.108

GAS(1,1)-Skewed t		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
μ	0.04 (0.034)	0.021 (0.038)
ω	0.026 (0.009)	0.033 (0.014)
A	0.104 (0.019)	0.095 (0.019)
B	0.971 (0.013)	0.973 (0.015)
ν	8.358 (1.525)	5.739 (0.725)
γ	0.496 (0.014)	0.493 (0.015)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2187.977	-2435.523

Table 13: Parameter estimates Log-linear Realized EGARCH models

Log Linear Real EGARCH-$\mathcal{N}$ RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.009 (0.008)	-0.001 (0.008)
α	0.63 (0.124)	2.61 (1.277)
β	0.971 (0.012)	0.964 (0.013)
γ	0.29 (0.039)	0.293 (0.04)
ξ	-0.022 (0.046)	-0.355 (0.045)
ϕ	1.096 (0.055)	1.154 (0.061)
τ_1	-0.068 (0.021)	-0.027 (0.015)
τ_2	0.079 (0.014)	0.018 (0.01)
λ_u	262979 (4.105)	169998 (7.488)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3984	-4250

Log Linear Real EGARCH-t RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.007 (0.006)	-0.001 (0.005)
α	0.533 (0.084)	-15.511 (0.229)
β	0.978 (0.007)	0.968 (0.008)
γ	0.292 (0.032)	0.283 (0.031)
ξ	-0.038 (0.047)	-0.356 (0.047)
ϕ	1.052 (0.051)	1.134 (0.069)
τ_1	-0.06 (0.015)	0.003 (0.001)
τ_2	0.081 (0.009)	-0.003 (0)
λ_u	96075 (4.246)	30625355 (59.739)
λ_z	9.281 (1.732)	6.958 (0.196)
σ_u	0.799 (0.01)	0.803 (0.01)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3829	-4085

Log Linear Real EGARCH-$\mathcal{N}$ BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.009 (0.008)	-0.001 (0.007)
α	0.67 (0.133)	2.374 (1.133)
β	0.971 (0.012)	0.964 (0.013)
γ	0.283 (0.038)	0.289 (0.039)
ξ	-0.101 (0.046)	-0.435 (0.045)
ϕ	1.099 (0.056)	1.153 (0.061)
τ_1	-0.069 (0.021)	-0.03 (0.016)
τ_2	0.075 (0.014)	0.019 (0.011)
λ_u	82038 (5.127)	81160 (5.063)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3992	-4258

Log Linear Real EGARCH-t BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.007 (0.006)	-0.001 (0.006)
α	0.573 (0.092)	2.271 (0.7)
β	0.978 (0.007)	0.968 (0.008)
γ	0.28 (0.031)	0.272 (0.031)
ξ	-0.115 (0.048)	-0.436 (0.054)
ϕ	1.061 (0.052)	1.14 (0.062)
τ_1	-0.063 (0.015)	-0.027 (0.01)
τ_2	0.076 (0.01)	0.021 (0.007)
λ_u	71178 (3.147)	57231 (151.997)
λ_z	9.449 (1.906)	6.969 (1.089)
σ_u	0.81 (0.01)	0.81 (0.01)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3853	-4101

Log Linear Real EGARCH-$\mathcal{N}$ RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.009 (0.008)	-0.001 (0.007)
α	0.542 (0.098)	2.506 (1.186)
β	0.971 (0.011)	0.961 (0.013)
γ	0.267 (0.037)	0.27 (0.035)
ξ	0.077 (0.046)	-0.261 (0.046)
ϕ	1.114 (0.055)	1.188 (0.062)
τ_1	-0.073 (0.022)	-0.028 (0.015)
τ_2	0.096 (0.014)	0.02 (0.011)
λ_u	52403 (3.27)	84384 (3.726)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-4015	-4288

Log Linear Real EGARCH-t RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.007 (0.006)	-0.001 (0.006)
α	0.476 (0.074)	2.383 (0.741)
β	0.979 (0.007)	0.965 (0.009)
γ	0.268 (0.031)	0.263 (0.03)
ξ	0.059 (0.048)	-0.262 (0.056)
ϕ	1.063 (0.05)	1.16 (0.064)
τ_1	-0.064 (0.016)	-0.025 (0.009)
τ_2	0.098 (0.01)	0.022 (0.008)
λ_u	56961 (3.552)	40942 (3.617)
λ_z	8.939 (1.754)	6.822 (1.009)
σ_u	0.831 (0.01)	0.837 (0.01)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3898	-4160

Table 14: Parameter estimates Log-linear Realized GARCH-$\mathcal{N}$ models

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.003 (0.013)	0.109 (0.015)
β	0.642 (0.024)	0.637 (0.024)
γ	0.306 (0.023)	0.299 (0.023)
ξ	-0.028 (0.039)	-0.362 (0.039)
ϕ	1.099 (0.044)	1.141 (0.052)
τ_1	-0.048 (0.011)	-0.013 (0.011)
τ_2	0.085 (0.007)	0.03 (0.006)
σ_u	0.452 (0.008)	0.473 (0.008)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3290	-3640

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.001 (0.015)	0.119 (0.019)
β	0.604 (0.045)	0.607 (0.045)
γ	0.337 (0.039)	0.326 (0.037)
ξ	-0.033 (0.046)	-0.37 (0.044)
ϕ	1.093 (0.055)	1.122 (0.059)
τ_1	-0.048 (0.024)	-0.015 (0.024)
τ_2	0.084 (0.014)	0.029 (0.012)
λ_u	38008 (3.349)	81965 (5.117)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3989	-4275

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.026 (0.013)	0.131 (0.016)
β	0.649 (0.024)	0.642 (0.024)
γ	0.296 (0.023)	0.294 (0.023)
ξ	-0.104 (0.04)	-0.442 (0.039)
ϕ	1.11 (0.045)	1.14 (0.052)
τ_1	-0.051 (0.011)	-0.014 (0.012)
τ_2	0.081 (0.007)	0.033 (0.006)
σ_u	0.466 (0.008)	0.483 (0.008)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3341	-3675

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.027 (0.015)	0.144 (0.021)
β	0.61 (0.044)	0.608 (0.045)
γ	0.33 (0.038)	0.324 (0.037)
ξ	-0.112 (0.046)	-0.45 (0.044)
ϕ	1.098 (0.055)	1.123 (0.059)
τ_1	-0.051 (0.024)	-0.015 (0.024)
τ_2	0.08 (0.014)	0.032 (0.013)
λ_u	36028 (133.882)	68807 (2.479)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3999	-4282

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.024 (0.012)	0.076 (0.013)
β	0.667 (0.024)	0.655 (0.024)
γ	0.281 (0.022)	0.277 (0.022)
ξ	0.069 (0.039)	-0.271 (0.04)
ϕ	1.109 (0.042)	1.165 (0.053)
τ_1	-0.05 (0.012)	-0.01 (0.012)
τ_2	0.102 (0.007)	0.035 (0.007)
σ_u	0.489 (0.008)	0.518 (0.009)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3431	-3803

Log Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	-0.031 (0.014)	0.082 (0.016)
β	0.624 (0.044)	0.626 (0.042)
γ	0.314 (0.037)	0.3 (0.034)
ξ	0.066 (0.046)	-0.279 (0.045)
ϕ	1.108 (0.054)	1.152 (0.06)
τ_1	-0.05 (0.024)	-0.011 (0.024)
τ_2	0.102 (0.014)	0.034 (0.012)
λ_u	43479 (2.217)	48695 (4.292)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-4021	-4318

Table 15: Parameter estimates Linear Realized GARCH-$\mathcal{N}$ models

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.057 (0.012)	0.078 (0.017)
β	0.614 (0.026)	0.618 (0.026)
γ	0.283 (0.028)	0.375 (0.038)
ξ	0 (0.066)	0 (0.067)
ϕ	1.183 (0.091)	0.882 (0.069)
τ_1	-0.042 (0.042)	-0.011 (0.041)
τ_2	0.093 (0.024)	0.045 (0.022)
σ_u	1.719 (0.029)	1.725 (0.029)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-5640	-5913

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.054 (0.013)	0.07 (0.016)
β	0.628 (0.037)	0.64 (0.035)
γ	0.273 (0.033)	0.358 (0.043)
ξ	0 (0.051)	0 (0.051)
ϕ	1.058 (0.083)	0.778 (0.061)
τ_1	-0.036 (0.024)	-0.019 (0.023)
τ_2	0.074 (0.014)	0.023 (0.013)
λ_u	4.099 (0.302)	4.075 (0.3)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-4362	-4642

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.055 (0.012)	0.075 (0.016)
β	0.618 (0.025)	0.62 (0.025)
γ	0.303 (0.03)	0.407 (0.04)
ξ	0 (0.059)	0 (0.059)
ϕ	1.097 (0.083)	0.812 (0.062)
τ_1	-0.044 (0.037)	-0.001 (0.036)
τ_2	0.087 (0.021)	0.04 (0.02)
σ_u	1.515 (0.026)	1.521 (0.026)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-5416	-5689

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.058 (0.014)	0.076 (0.018)
β	0.602 (0.041)	0.612 (0.04)
γ	0.315 (0.039)	0.418 (0.052)
ξ	0 (0.05)	0 (0.05)
ϕ	0.99 (0.079)	0.729 (0.059)
τ_1	-0.037 (0.024)	-0.021 (0.023)
τ_2	0.065 (0.014)	0.024 (0.013)
λ_u	4.466 (0.35)	4.458 (0.348)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-4323	-4599

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim \mathcal{N}$ RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.046 (0.01)	0.07 (0.014)
β	0.716 (0.029)	0.728 (0.028)
γ	0.177 (0.022)	0.216 (0.028)
ξ	0.138 (0.114)	0.081 (0.118)
ϕ	1.215 (0.108)	0.964 (0.089)
τ_1	-0.059 (0.091)	0.077 (0.09)
τ_2	0.137 (0.052)	0.087 (0.048)
σ_u	3.773 (0.064)	3.776 (0.064)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-7030	-7299

Linear Real GARCH-$\mathcal{N}$ with $u_t \sim t$ RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0.057 (0.012)	0.074 (0.014)
β	0.682 (0.034)	0.689 (0.029)
γ	0.192 (0.025)	0.255 (0.031)
ξ	0 (0.056)	0 (0.056)
ϕ	1.167 (0.093)	0.853 (0.068)
τ_1	-0.052 (0.024)	-0.01 (0.024)
τ_2	0.095 (0.015)	0.029 (0.013)
λ_u	3.36 (0.22)	3.338 (0.217)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-4476	-4766

Table 16: Parameter estimates Log-linear Realized GAS-t models

Log Linear REAL GAS-t (no τ) RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.005)
A	0.073 (0.01)	0.084 (0.01)
B	0.972 (0.006)	0.969 (0.006)
ξ	-0.218 (0.061)	-0.362 (0.045)
ϕ	0.977 (0.058)	0.97 (0.048)
σ_u	0.472 (0.009)	0.47 (0.009)
λ_u	6.656 (1.368)	7.725 (1.313)
λ_z	13.308 (2.842)	14.699 (3.464)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3354	-3551

Log Linear REAL GAS-t (with τ) RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.005)
A	0.08 (0.011)	0.084 (0.01)
B	0.974 (0.006)	0.97 (0.006)
ξ	-0.195 (0.053)	-0.357 (0.044)
ϕ	0.959 (0.049)	0.965 (0.047)
σ_u	0.455 (0.009)	0.471 (0.009)
τ_1	-0.026 (0.011)	0.019 (0.01)
τ_2	0.081 (0.008)	0.01 (0.005)
λ_u	7.238 (1.587)	8.029 (1.416)
λ_z	13.808 (3.034)	14.008 (3.196)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3277	-3547

Log Linear REAL GAS-t (no τ) BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.005)
A	0.073 (0.01)	0.083 (0.01)
B	0.973 (0.005)	0.97 (0.006)
ξ	-0.255 (0.054)	-0.432 (0.045)
ϕ	0.989 (0.054)	0.986 (0.049)
σ_u	0.483 (0.009)	0.481 (0.009)
λ_z	7.376 (1.451)	7.79 (1.314)
λ_u	14.809 (3.434)	15.857 (3.867)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3390	-3593

Log Linear REAL GAS-t (with τ) BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.005)
A	0.079 (0.01)	0.084 (0.01)
B	0.975 (0.005)	0.971 (0.006)
ξ	-0.229 (0.047)	-0.426 (0.044)
ϕ	0.981 (0.048)	0.98 (0.048)
σ_u	0.469 (0.009)	0.481 (0.009)
τ_1	-0.026 (0.011)	0.02 (0.01)
τ_2	0.073 (0.007)	0.013 (0.006)
λ_z	8.233 (1.748)	8.158 (1.435)
λ_u	15.099 (3.485)	15.238 (3.65)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3325	-3588

Log Linear REAL GAS-t (no τ) RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.007)	0 (0.005)
A	0.082 (0.013)	0.096 (0.012)
B	0.969 (0.006)	0.965 (0.007)
ξ	-0.195 (0.079)	-0.293 (0.045)
ϕ	0.969 (0.06)	0.963 (0.048)
σ_u	0.517 (0.011)	0.514 (0.01)
λ_z	5.607 (1.247)	7.528 (1.286)
λ_u	10.417 (1.886)	11.884 (2.357)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3512	-3703

Log Linear REAL GAS-t (with τ) RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.007)	0 (0.004)
A	0.09 (0.014)	0.068 (0.008)
B	0.971 (0.006)	0.97 (0.006)
ξ	-0.173 (0.073)	-0.331 (0.05)
ϕ	0.947 (0.049)	1.059 (0.043)
σ_u	0.493 (0.01)	0.485 (0.01)
τ_1	-0.028 (0.012)	-0.025 (0.011)
τ_2	0.104 (0.011)	0.1 (0.008)
λ_z	5.887 (1.469)	6.246 (0.958)
λ_u	10.054 (1.858)	10.203 (1.902)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-3414	-3618

Table 17: Parameter estimates Log-linear Realized GAS-GED models

Log Linear Real GAS-GED (no τ) RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.004)
A	0.071 (0.011)	0.08 (0.01)
B	0.971 (0.006)	0.968 (0.006)
ξ	-0.201 (0.05)	-0.359 (0.041)
ϕ	0.991 (0.063)	0.997 (0.05)
σ_u	0.472 (0.009)	0.47 (0.009)
λ_z	1.456 (0.073)	1.475 (0.063)
λ_u	13.099 (2.75)	16.662 (4.362)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2352	-2551

Log Linear Real GAS-GED (with τ) RV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.005)	0 (0.004)
A	0.078 (0.012)	0.073 (0.009)
B	0.973 (0.006)	0.973 (0.005)
ξ	-0.183 (0.045)	-0.374 (0.04)
ϕ	0.972 (0.053)	0.988 (0.05)
σ_u	0.455 (0.009)	0.55 (0.018)
τ_1	-0.025 (0.011)	0.018 (0.01)
τ_2	0.077 (0.007)	0.015 (0.004)
λ_z	1.49 (0.075)	1.461 (0.061)
λ_u	14.099 (3.13)	4 (0.271)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2275	-2585

Log Linear Real GAS-GED (no τ) BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.004)	0 (0.004)
A	0.071 (0.01)	0.079 (0.009)
B	0.973 (0.005)	0.968 (0.006)
ξ	-0.243 (0.047)	-0.43 (0.042)
ϕ	1.006 (0.058)	1.013 (0.051)
σ_u	0.483 (0.009)	0.48 (0.009)
λ_z	1.478 (0.071)	1.477 (0.063)
λ_u	14.643 (3.358)	17.259 (4.492)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2387	-2593

Log Linear Real GAS-GED (with τ) BV		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.004)	0 (0.004)
A	0.076 (0.011)	0.081 (0.01)
B	0.974 (0.005)	0.97 (0.006)
ξ	-0.222 (0.043)	-0.425 (0.041)
ϕ	0.996 (0.051)	1.004 (0.049)
σ_u	0.469 (0.009)	0.48 (0.009)
τ_1	-0.026 (0.011)	0.02 (0.01)
τ_2	0.071 (0.007)	0.014 (0.005)
λ_z	1.518 (0.073)	1.495 (0.064)
λ_u	15.225 (3.541)	16.482 (4.209)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2323	-2587

Log Linear Real GAS-GED (no τ) RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.006)	0 (0.005)
A	0.081 (0.014)	0.092 (0.011)
B	0.968 (0.006)	0.963 (0.007)
ξ	-0.166 (0.054)	-0.29 (0.041)
ϕ	0.974 (0.066)	0.989 (0.05)
σ_u	0.517 (0.011)	0.513 (0.01)
λ_z	1.419 (0.076)	1.467 (0.062)
λ_u	10.385 (1.859)	12.749 (2.662)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2512	-2705

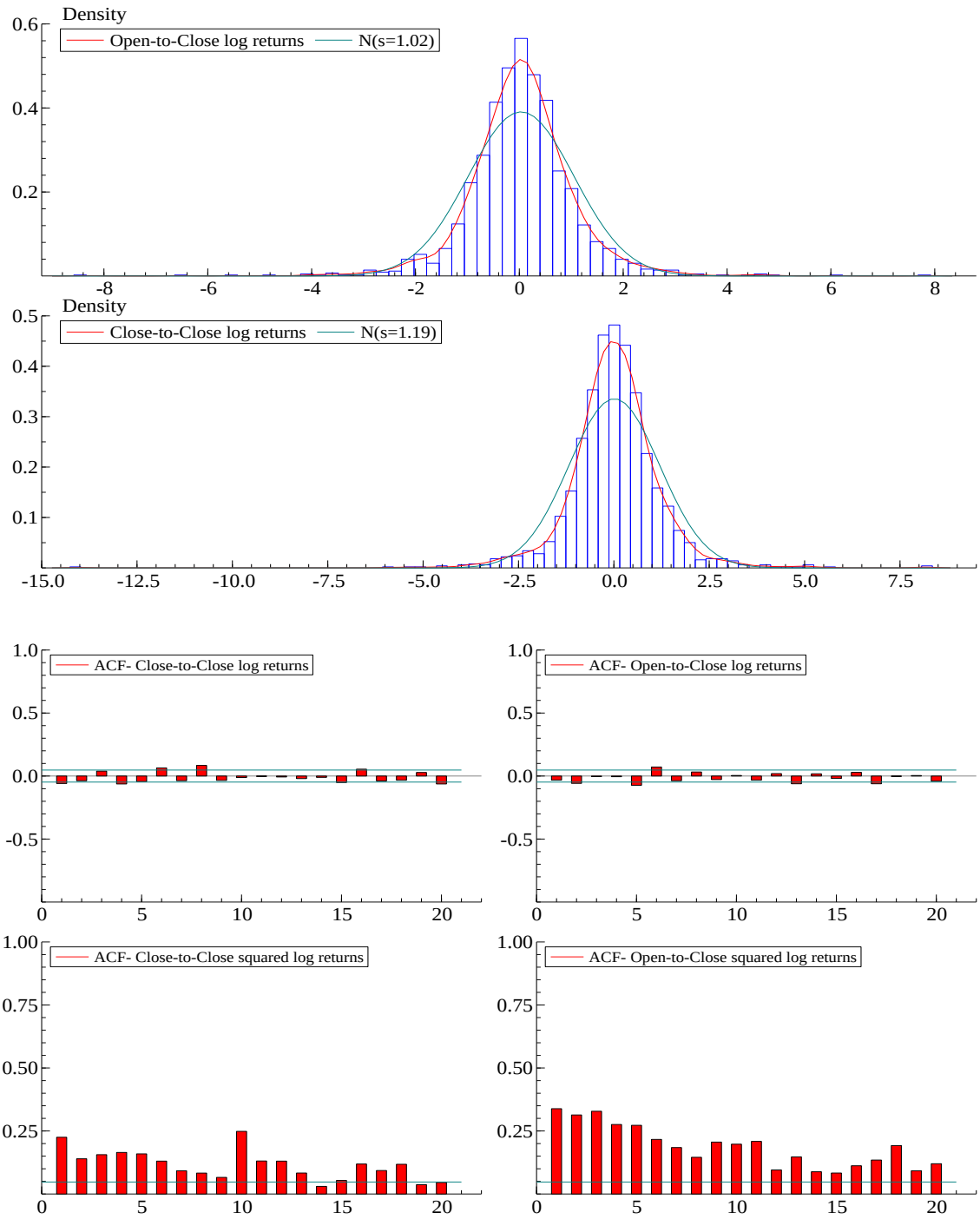
Log Linear Real GAS-GED (with τ) RK		
$\boldsymbol{\theta}$	Open-to-Close	Close-to-Close
ω	0 (0.006)	0 (0.005)
A	0.089 (0.016)	0.094 (0.012)
B	0.97 (0.006)	0.965 (0.007)
ξ	-0.15 (0.05)	-0.285 (0.04)
ϕ	0.955 (0.056)	0.979 (0.048)
σ_u	0.493 (0.011)	0.513 (0.01)
τ_1	-0.027 (0.012)	0.023 (0.011)
τ_2	0.098 (0.008)	0.013 (0.006)
λ_z	1.446 (0.079)	1.484 (0.063)
λ_u	10.215 (1.919)	12.228 (2.5)
$\mathcal{L}_T(\boldsymbol{r};\boldsymbol{\theta})$	-2414	-2700

Table 18: In-sample likelihoods realized models

RM	**Models**	**Likelihood**		
		Open-to-Close	**Close-to-Close**	# Parameters
BV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-5416	-5689	8
BV	Real-GARCH(1,1) $\mathcal{N}$ (t)	-4323	-4599	8
BV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-3341	-3675	8
BV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	-3999	-4282	8
BV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	-3992	-4258	9
BV	Real-EGARCH(1,1) t (t)	-3853	-4101	11
BV	LL Real-GAS(1,1) t (no τ)	-3390	-3593	8
BV	LL Real-GAS(1,1) t (τ)	-3325	-3588	10
BV	LL Real-GAS(1,1) GED (no τ)	-2387	-2593	8
BV	LL Real-GAS(1,1) GED (τ)	-2323	-2587	10
RV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-5640	-5913	8
RV	Real-GARCH(1,1) $\mathcal{N}$ (t)	-4362	-4642	8
RV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-3290	-3640	8
RV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	-3989	-4275	8
RV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	-3984	-4250	9
RV	Real-EGARCH(1,1) t (t)	-3829	-4085	11
RV	LL Real-GAS(1,1) t (no τ)	-3354	-3551	8
RV	LL Real-GAS(1,1) t (τ)	-3277	-3547	10
RV	LL Real-GAS(1,1) GED (no τ)	-2352	-2551	8
RV	LL Real-GAS(1,1) GED (τ)	-2275	-2585	10
RK	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-7030	-7299	8
RK	Real-GARCH(1,1) $\mathcal{N}$ (t)	-4476	-4766	8
RK	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	-3431	-3803	8
RK	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	-4021	-4318	8
RK	Real-EGARCH(1,1) $\mathcal{N}$ (t)	-4015	-4288	9
RK	Real-EGARCH(1,1) t (t)	-3898	-4160	11
RK	LL Real-GAS(1,1) t (no τ)	-3512	-3703	8
RK	LL Real-GAS(1,1) t (τ)	-3414	-3618	10
RK	LL Real-GAS(1,1) GED (no τ)	-2512	-2705	8
RK	LL Real-GAS(1,1) GED (τ)	-2414	-2700	10

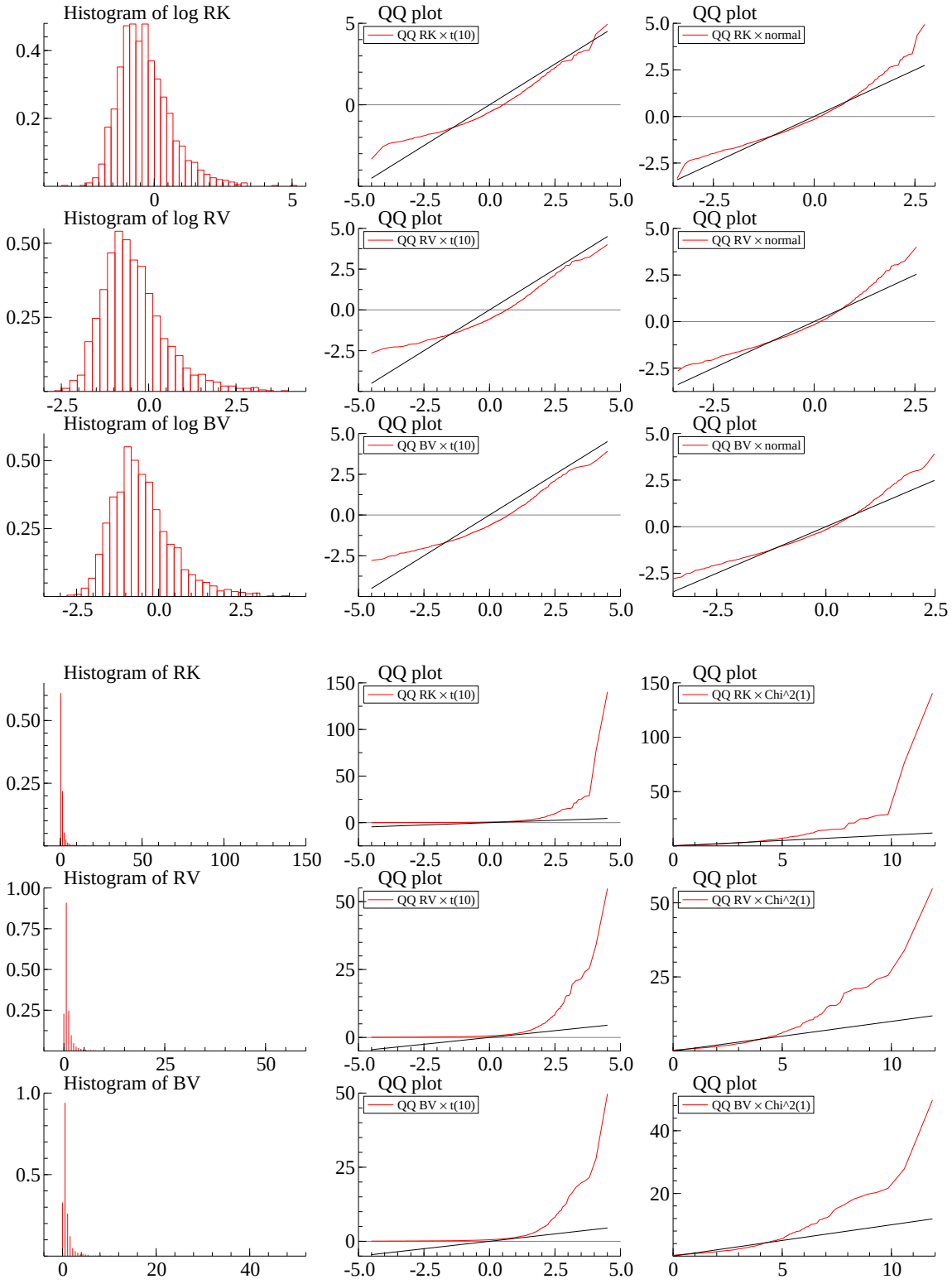
N Log returns and realized measures

Figure 8: Density and auto-correlation function log returns



Note: this figure shows the density of the return series (open-to-close and close-to-close) and the ACF plots of the return series and the squared return series.

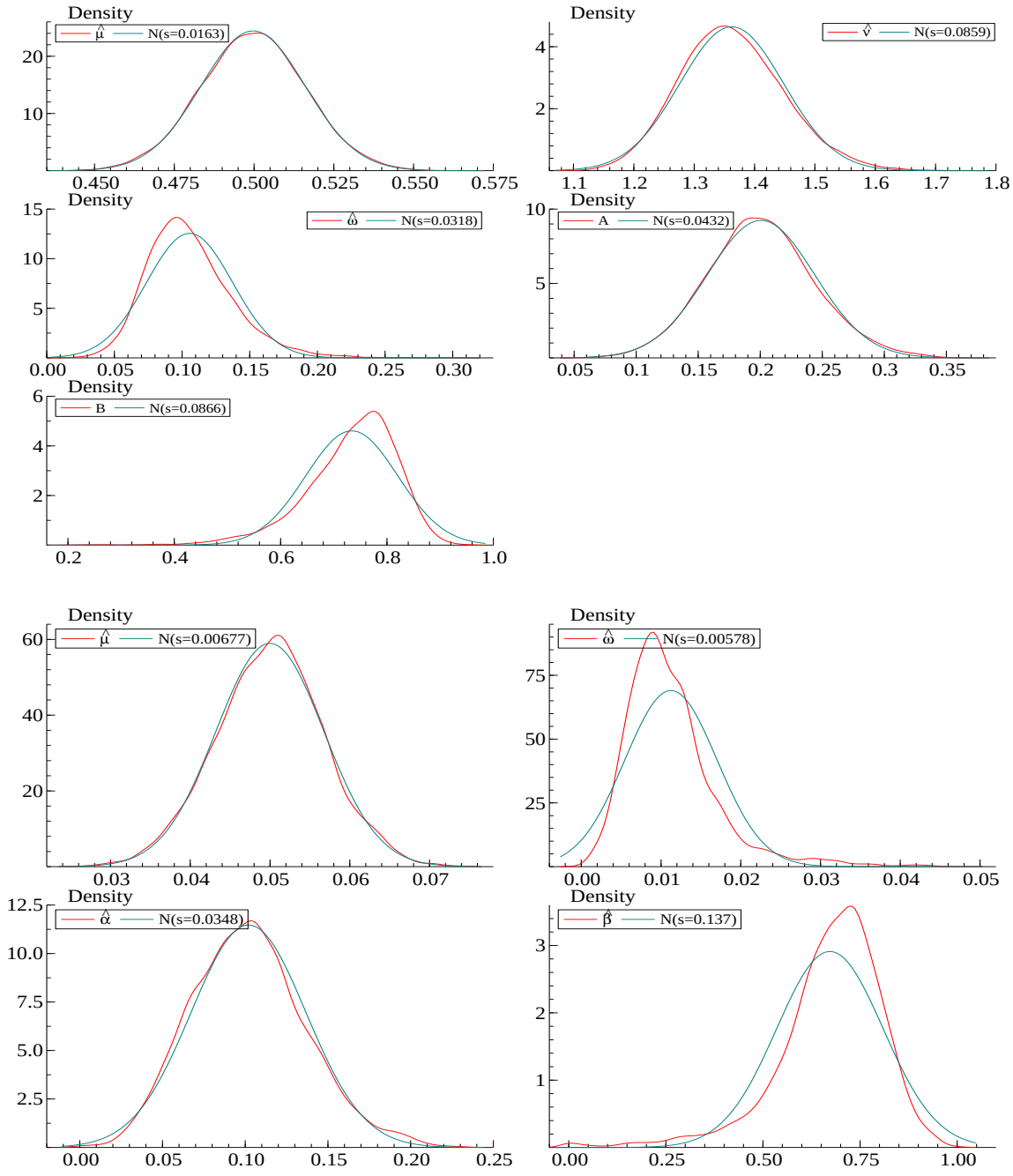
Figure 9: Realized measures



Note: this figure shows the distribution and QQ-plots of the (log) realized measures.

O Monte Carlo Simulation

Figure 10: Monte Carlo simulation of volatility models



Note: this figure shows the distribution of the estimated parameters of the GAS-GED model (above) with $\boldsymbol{\theta_0} = (\mu, \omega, A, B, \nu) = (0.50, 0.10, 0.20, 0.75, 1.35)$ and the GARCH-$\mathcal{N}$ model (below) with $\boldsymbol{\theta_0} = (\mu, \omega, \alpha, \beta) = (0.05, 0.01, 0.10, 0.70)$ obtained by Monte Carlo simulation. In total 10,000 estimations have been performed on 1,000 simulated return series.

P MAE and MSE statistics

Table 19: Fixed window: Open-to-Close

MODEL		MAE			MSE			LSR
		Realized measure			*Realized measure*			
RM		RK	BV	RV	RK	BV	RV	
	GARCH(1,1)-$\mathcal{N}$	0.457	0.419	0.421	0.647	0.453	0.498	-1.086
	GARCH(1,1)-t	0.458	0.420	0.423	0.653	0.461	0.505	-1.069
	EGARCH(1,1)-$\mathcal{N}$	0.438	0.399	0.400	**0.605**	0.404	**0.449**	-1.086
	EGARCH(1,1)-t	**0.435**	**0.393**	**0.396**	0.607	**0.403**	0.449	-1.069
	NGARCH(1,1)-$\mathcal{N}$	0.443	0.405	0.406	0.619	0.421	0.466	-1.082
	NGARCH(1,1)-t	0.439	0.399	0.401	0.620	0.419	0.465	-1.066
	GAS(1,1)-t	0.441	0.396	0.400	0.620	0.414	0.463	-1.069
	GAS(1,1)-GED	0.446	0.404	0.407	0.628	0.425	0.474	-1.069
	GAS(1,1)-Laplace	0.527	0.516	0.504	0.685	0.547	0.569	**-0.737**
	GAS(1,1)-Skewed t	0.446	0.403	0.406	0.624	0.423	0.470	-1.071
BV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.362	0.315	0.318	0.521	0.286	0.344	-2.745
BV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.362	0.315	0.318	**0.520**	**0.286**	**0.343**	-1.063
BV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.358	0.295	0.305	0.543	0.295	0.358	-6.446
BV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.357	0.294	0.304	0.538	0.292	0.354	-1.069
BV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.357	**0.281**	**0.298**	0.577	0.305	0.377	-1.072
BV	Real-EGARCH(1,1) t (t)	**0.356**	0.283	0.299	0.563	0.299	0.368	**-1.058**
BV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.389	0.356	0.353	0.546	0.330	0.382	-1.071
BV	LL Real-GAS(1,1) t (no τ)	0.379	0.333	0.335	0.539	0.330	0.377	-1.627
BV	LL Real-GAS(1,1) t (τ)	0.375	0.323	0.329	0.546	0.327	0.378	-1.290
BV	LL Real-GAS(1,1) GED (no τ)	0.382	0.344	0.343	0.537	0.326	0.375	-1.057
BV	LL Real-GAS(1,1) GED (τ)	0.376	0.331	0.333	0.543	0.325	0.376	-1.058
RV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.365	0.318	0.320	**0.523**	**0.289**	**0.348**	-2.873
RV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.366	0.319	0.321	0.528	0.294	0.352	-1.065
RV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.358	0.294	0.305	0.543	0.296	0.359	-6.497
RV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.357	0.294	0.306	0.536	0.292	0.354	-1.069
RV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	**0.356**	**0.283**	**0.299**	0.572	0.303	0.376	-1.072
RV	Real-EGARCH(1,1) t (t)	0.356	0.284	0.300	0.559	0.298	0.367	**-1.058**
RV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.394	0.363	0.360	0.555	0.337	0.390	-1.073
RV	LL Real-GAS(1,1) t (no τ)	0.390	0.358	0.356	0.545	0.331	0.382	-1.631
RV	LL Real-GAS(1,1) t (τ)	0.386	0.341	0.343	0.551	0.344	0.390	-1.311
RV	LL Real-GAS(1,1) GED (no τ)	0.393	0.362	0.359	0.551	0.335	0.387	-1.061
RV	LL Real-GAS(1,1) GED (τ)	0.387	0.353	0.351	0.561	0.337	0.392	-1.062
RK	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.371	0.315	0.321	0.549	0.306	0.367	-3.664
RK	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.374	0.323	0.327	0.547	0.307	0.367	-1.069
RK	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.360	0.294	0.307	0.554	0.300	0.365	-5.858
RK	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	**0.360**	0.293	0.306	0.545	**0.294**	**0.358**	-1.076
RK	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.361	**0.282**	**0.301**	0.589	0.311	0.386	-1.080
RK	Real-EGARCH(1,1) t (t)	0.361	0.284	0.302	0.582	0.308	0.381	**-1.063**
RK	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.394	0.361	0.359	0.562	0.339	0.394	-1.078
RK	LL Real-GAS(1,1) t (no τ)	0.389	0.356	0.355	**0.540**	0.322	0.375	-1.634
RK	LL Real-GAS(1,1) t (τ)	0.410	0.380	0.375	0.566	0.384	0.418	-1.349
RK	LL Real-GAS(1,1) GED (no τ)	0.393	0.361	0.360	0.552	0.331	0.385	-1.063
RK	LL Real-GAS(1,1) GED (τ)	0.391	0.356	0.355	0.575	0.345	0.402	-1.066

Note: this table shows MAE and MSE statistics for the (non-) realized models for different benchmarks.

Table 20: Fixed window: Close-to-Close

MODEL		MAE			MSE			LSR
		Realized measure			*Realized measure*			
RM		RK	BV	RV	RK	BV	RV	
	GARCH(1,1)-$\mathcal{N}$	0.483	0.461	0.458	0.649	0.471	0.508	-1.257
	GARCH(1,1)-t	0.483	0.462	0.459	0.655	0.482	0.517	-1.220
	EGARCH(1,1)-$\mathcal{N}$	0.451	0.428	0.426	0.622	0.432	0.471	-1.247
	EGARCH(1,1)-t	**0.442**	**0.415**	**0.415**	**0.617**	**0.422**	**0.462**	-1.215
	NGARCH(1,1)-$\mathcal{N}$	0.467	0.447	0.443	0.632	0.441	0.481	-1.247
	NGARCH(1,1)-t	0.450	0.427	0.425	0.624	0.425	0.468	-1.215
	GAS(1,1)-t	0.461	0.436	0.435	0.618	0.441	0.479	-1.219
	GAS(1,1)-GED	0.462	0.435	0.434	0.623	0.437	0.479	-1.225
	GAS(1,1)-Laplace	0.598	0.583	0.575	0.835	0.617	0.673	**-0.911**
	GAS(1,1)-Skewed t	0.470	0.448	0.447	0.629	0.462	0.496	-1.220
BV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.386	0.327	0.335	**0.609**	**0.352**	**0.418**	-2.921
BV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.386	0.327	0.335	0.609	0.352	0.419	-1.235
BV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.386	0.319	0.331	0.628	0.363	0.433	-6.276
BV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.389	0.321	0.333	0.629	0.364	0.434	-1.245
BV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.382	0.295	0.317	0.663	0.363	0.447	-1.256
BV	Real-EGARCH(1,1) t (t)	**0.381**	**0.294**	**0.316**	0.667	0.365	0.450	**-1.223**
BV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.392	0.325	0.338	0.626	0.369	0.437	-1.243
BV	LL Real-GAS(1,1) t (no τ)	0.390	0.316	0.334	0.636	0.374	0.444	-1.790
BV	LL Real-GAS(1,1) t (τ)	0.390	0.315	0.333	0.637	0.374	0.443	-1.512
BV	LL Real-GAS(1,1) GED (no τ)	0.390	0.318	0.334	0.635	0.372	0.442	-1.223
BV	LL Real-GAS(1,1) GED (τ)	0.390	0.317	0.333	0.636	0.372	0.442	-1.224
RV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.390	0.333	0.341	**0.614**	0.356	**0.424**	-3.048
RV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.389	0.331	0.339	0.616	0.357	0.425	-1.236
RV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.386	0.320	0.332	0.627	0.362	0.433	-6.270
RV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.389	0.322	0.334	0.628	0.363	0.434	-1.245
RV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.382	0.304	0.323	0.642	**0.354**	0.434	-1.247
RV	Real-EGARCH(1,1) t (t)	**0.382**	**0.296**	**0.318**	0.665	0.364	0.450	**-1.222**
RV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.393	0.327	0.341	0.627	0.369	0.438	-1.244
RV	LL Real-GAS(1,1) t (no τ)	0.391	0.319	0.337	0.633	0.373	0.443	-1.789
RV	LL Real-GAS(1,1) t (τ)	0.390	0.318	0.336	0.633	0.373	0.442	-1.520
RV	LL Real-GAS(1,1) GED (no τ)	0.391	0.320	0.337	0.634	0.372	0.443	-1.223
RV	LL Real-GAS(1,1) GED (τ)	0.395	0.327	0.341	0.624	0.369	0.436	-1.224
RK	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.396	0.338	0.346	0.626	0.367	0.435	-3.836
RK	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.397	0.340	0.347	0.624	0.366	0.434	-1.239
RK	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.391	0.321	0.334	0.629	0.362	0.434	-5.751
RK	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.393	0.323	0.336	0.629	0.362	0.434	-1.249
RK	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.389	**0.298**	**0.322**	0.674	0.370	0.456	-1.260
RK	Real-EGARCH(1,1) t (t)	0.387	0.301	0.322	0.662	0.364	0.448	**-1.222**
RK	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.397	0.336	0.346	0.616	0.366	0.432	-1.245
RK	LL Real-GAS(1,1) t (no τ)	0.398	0.331	0.345	0.625	0.377	0.442	-1.790
RK	LL Real-GAS(1,1) t (τ)	**0.372**	0.314	0.323	**0.565**	**0.327**	**0.385**	-1.557
RK	LL Real-GAS(1,1) GED (no τ)	0.396	0.329	0.343	0.624	0.372	0.438	-1.223
RK	LL Real-GAS(1,1) GED (τ)	0.395	0.327	0.341	0.624	0.369	0.436	-1.224

Note: this table shows MAE and MSE statistics for the non-realized and realized models. The first column represents the realized measure used in the realized measure. The realized measures in column three to eight represent the benchmark used for the calculation of the MAE and MSE.

Table 21: Rolling window: Open-to-Close

MODEL		MAE			MSE			LSR
		Realized measure			*Realized measure*			
RM		RK	BV	RV	RK	BV	RV	
	GARCH(1,1)-$\mathcal{N}$	0.416	0.363	0.374	0.634	0.412	0.468	-1.081
	GARCH(1,1)-t	0.417	0.363	0.374	0.633	0.410	0.466	-1.061
	EGARCH(1,1)-$\mathcal{N}$	**0.408**	0.358	**0.364**	**0.596**	**0.380**	**0.431**	-1.081
	EGARCH(1,1)-t	0.410	0.359	0.365	0.601	0.386	0.436	-1.064
	NGARCH(1,1)-$\mathcal{N}$	0.410	0.357	0.366	0.621	0.399	0.452	-1.078
	NGARCH(1,1)-t	0.410	0.355	0.364	0.622	0.401	0.453	-1.062
	GAS(1,1)-t	0.413	0.358	0.368	0.616	0.397	0.450	-1.063
	GAS(1,1)-GED	0.412	0.359	0.369	0.623	0.403	0.457	-1.063
	GAS(1,1)-Laplace	0.455	0.424	0.420	0.619	0.445	0.481	**-0.733**
	GAS(1,1)-Skewed t	0.411	**0.354**	0.365	0.612	0.385	0.442	-1.065
BV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.359	0.305	0.310	0.523	0.285	0.344	-1.063
BV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.355	0.300	0.306	**0.522**	**0.282**	**0.342**	-1.064
BV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.356	0.296	0.304	0.544	0.295	0.358	-1.067
BV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.356	0.296	0.305	0.533	0.290	0.350	-1.066
BV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.357	0.296	0.304	0.539	0.294	0.356	-1.066
BV	Real-EGARCH(1,1) t (t)	**0.355**	**0.293**	**0.303**	0.540	0.295	0.357	-1.053
BV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.389	0.351	0.349	0.554	0.341	0.391	-1.074
BV	LL Real-GAS(1,1) t (no τ)	0.417	0.387	0.381	0.572	0.381	0.421	-1.632
BV	LL Real-GAS(1,1) t (τ)	0.415	0.384	0.378	0.571	0.370	0.414	-1.369
BV	LL Real-GAS(1,1) GED (no τ)	0.389	0.347	0.347	0.562	0.351	0.401	-1.059
BV	LL Real-GAS(1,1) GED (τ)	0.387	0.343	0.344	0.562	0.342	0.393	-1.061
RV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.361	0.307	0.313	**0.527**	**0.289**	0.349	-1.065
RV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.358	0.304	0.310	0.530	0.290	0.351	-1.066
RV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.357	0.296	0.306	0.543	0.296	0.359	-1.068
RV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.356	0.297	0.306	0.532	0.290	0.351	-1.067
RV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	**0.355**	0.295	**0.304**	0.535	0.291	0.353	-1.066
RV	Real-EGARCH(1,1) t (t)	0.355	**0.294**	0.305	0.539	0.295	0.358	**-1.054**
RV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.407	0.357	0.363	0.626	0.385	0.448	-1.244
RV	LL Real-GAS(1,1) t (no τ)	0.433	0.410	0.404	0.594	0.415	0.452	-1.633
RV	LL Real-GAS(1,1) t (τ)	0.431	0.406	0.399	0.595	0.404	0.445	-1.391
RV	LL Real-GAS(1,1) GED (no τ)	0.398	0.358	0.358	0.569	0.364	0.411	-1.061
RV	LL Real-GAS(1,1) GED (τ)	0.400	0.358	0.358	0.583	0.371	0.420	-1.064
RK	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.371	0.316	0.323	0.550	0.311	0.372	-1.068
RK	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.368	0.312	0.319	0.547	0.307	0.368	-1.071
RK	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.360	0.299	0.308	0.548	0.299	0.363	-1.071
RK	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.359	0.298	**0.307**	**0.533**	**0.290**	**0.351**	-1.071
RK	Real-EGARCH(1,1) $\mathcal{N}$ (t)	**0.359**	**0.298**	0.308	0.536	0.291	0.354	-1.070
RK	Real-EGARCH(1,1) t (t)	0.359	0.298	0.309	0.542	0.297	0.359	**-1.055**
RK	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.403	0.372	0.369	0.559	0.355	0.401	-1.082
RK	LL Real-GAS(1,1) t (no τ)	0.447	0.432	0.422	0.601	0.437	0.466	-1.636
RK	LL Real-GAS(1,1) t (τ)	0.440	0.419	0.410	0.598	0.413	0.449	-1.402
RK	LL Real-GAS(1,1) GED (no τ)	0.411	0.380	0.376	0.571	0.380	0.420	-1.066
RK	LL Real-GAS(1,1) GED (τ)	0.405	0.371	0.367	0.581	0.372	0.418	-1.069

Note: this table shows MAE and MSE statistics for the non-realized and realized models. The first column represents the realized measure used in the realized measure. The realized measures in column three to eight represent the benchmark used for the calculation of the MAE and MSE.

Table 22: Rolling window: Close-to-Close

MODEL		MAE			MSE			LSR
		Realized measure			*Realized measure*			
RM		RK	BV	RV	RK	BV	RV	
	GARCH(1,1)-$\mathcal{N}$	0.428	0.382	0.388	0.623	0.402	0.453	-1.245
	GARCH(1,1)-t	0.428	0.385	0.389	0.619	0.402	0.452	-1.217
	EGARCH(1,1)-$\mathcal{N}$	0.421	0.384	0.386	0.610	0.398	0.443	-1.243
	EGARCH(1,1)-t	0.422	0.385	0.387	0.609	0.402	0.446	-1.216
	NGARCH(1,1)-$\mathcal{N}$	0.423	0.381	0.385	0.626	0.401	0.452	-1.243
	NGARCH(1,1)-t	**0.420**	0.380	0.383	0.618	0.399	0.448	-1.216
	GAS(1,1)-t	0.426	0.387	0.391	0.604	0.409	0.451	-1.216
	GAS(1,1)-GED	0.420	**0.377**	**0.381**	0.606	**0.395**	**0.443**	-1.220
	GAS(1,1)-Laplace	0.456	0.432	0.429	0.617	0.441	0.474	**-0.886**
	GAS(1,1)-Skewed t	0.423	0.382	0.387	**0.601**	0.399	0.443	-1.217
BV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.400	0.347	0.352	0.614	0.366	0.429	-1.238
BV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.395	0.342	0.347	0.612	**0.361**	0.425	-1.237
BV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	**0.390**	**0.337**	**0.342**	0.620	0.367	0.433	-1.239
BV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.394	0.340	0.346	0.616	0.365	0.429	-1.239
BV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.396	0.342	0.350	0.612	0.362	0.426	-1.235
BV	Real-EGARCH(1,1) t (t)	0.399	0.345	0.353	0.626	0.379	0.441	**-1.208**
BV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.405	0.355	0.360	0.624	0.383	0.445	-1.242
BV	LL Real-GAS(1,1) t (no τ)	0.417	0.387	0.381	**0.572**	0.381	**0.421**	-1.632
BV	LL Real-GAS(1,1) t (τ)	0.413	0.354	0.363	0.646	0.407	0.468	-1.584
BV	LL Real-GAS(1,1) GED (no τ)	0.412	0.356	0.363	0.640	0.401	0.462	-1.220
BV	LL Real-GAS(1,1) GED (τ)	0.409	0.351	0.359	0.640	0.397	0.459	-1.221
RV	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.401	0.350	0.355	0.619	0.368	0.434	-1.239
RV	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.396	0.344	0.349	0.617	0.365	0.431	-1.238
RV	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	**0.390**	**0.338**	**0.343**	0.620	**0.366**	**0.433**	-1.240
RV	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.394	0.341	0.347	0.617	0.365	0.431	-1.241
RV	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.402	0.351	0.358	**0.613**	0.368	0.432	-1.234
RV	Real-EGARCH(1,1) t (t)	0.401	0.349	0.356	0.635	0.387	0.451	**-1.208**
RV	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.407	0.357	0.363	0.626	0.385	0.448	-1.244
RV	LL Real-GAS(1,1) t (no τ)	0.417	0.361	0.370	0.651	0.416	0.476	-1.788
RV	LL Real-GAS(1,1) t (τ)	0.414	0.356	0.366	0.649	0.410	0.472	-1.587
RV	LL Real-GAS(1,1) GED (no τ)	0.413	0.358	0.366	0.643	0.404	0.466	-1.222
RV	LL Real-GAS(1,1) GED (τ)	0.410	0.354	0.362	0.641	0.400	0.462	-1.222
RK	Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.412	0.360	0.365	0.638	0.389	0.454	-1.242
RK	Real-GARCH(1,1) $\mathcal{N}$ (t)	0.407	0.355	0.360	0.630	0.380	0.445	-1.241
RK	LL Real-GARCH(1,1) $\mathcal{N}$ ($\mathcal{N}$)	0.395	**0.342**	**0.347**	0.619	0.366	0.433	-1.242
RK	LL Real-GARCH(1,1) $\mathcal{N}$ (t)	0.398	0.345	0.350	0.614	0.364	0.429	-1.242
RK	Real-EGARCH(1,1) $\mathcal{N}$ (t)	0.400	0.346	0.353	0.611	0.361	0.426	-1.237
RK	Real-EGARCH(1,1) t (t)	0.409	0.358	0.365	0.628	0.387	0.450	**-1.209**
RK	LL Real-GAS(1,1) $\mathcal{N}$ (no τ)	0.414	0.366	0.370	0.624	0.387	0.448	-1.245
RK	LL Real-GAS(1,1) t (no τ)	0.430	0.375	0.382	0.657	0.427	0.485	-1.789
RK	LL Real-GAS(1,1) t (τ)	**0.391**	0.350	0.350	**0.569**	**0.342**	**0.396**	-1.628
RK	LL Real-GAS(1,1) GED (no τ)	0.425	0.372	0.378	0.649	0.415	0.474	-1.223
RK	LL Real-GAS(1,1) GED (τ)	0.422	0.367	0.373	0.645	0.408	0.468	-1.224

Note: this table shows MAE and MSE statistics for the non-realized and realized models. The first column represents the realized measure used in the realized measure. The realized measures in column three to eight represent the benchmark used for the calculation of the MAE and MSE.