Bachelor Thesis Quantitative Finance

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Two Models for Returns, Volatility and Volatility of Volatility

Supervisor:
dr. R. Lange

Author[1]*:*
Jasper Brakel

Second Assessor:
dr. O. Kleen

July 3, 2022

Abstract

We compare two different models that allow for a time-varying volatility of volatility against each other and three benchmark models. The models, ART-GARCH from Ding (2021b) and (A)SHARV from Ding (2021a), are simple to implement and able to now- and forecast volatility and volatility of volatility simultaneously. We show that both models have a better fit as well as better now- and forecasting performance than other GARCH-type models. We find no evidence that the ART-GARCH type models have a better goodness-of-fit than ASHARV. We also find no empirical evidence that (A)SHARV is outperformed by the ART-GARCH type models in nowcasting nor in forecasting.

Keywords: ART-GARCH, (A)SHARV, Volatility of volatility, forecasting

[1]The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

1 Introduction

In this paper we compare two models that both allow for conditional heteroskedasticity in the volatility. The first one, from Ding (2021a), is called the stochastic heteroskedastic autoregressive volatility (SHARV) model and the asymmetric extension of SHARV is called ASHARV. We use (A)SHARV to refer to both SHARV and ASHARV. Second, we use the model proposed in Ding (2021b) called augmented real-time generalized autoregressive conditional heteroskedasticity (ART-GARCH). We also compare these models to other GARCH-type models. We are interested to see if the ART-GARCH model is able to outperform (A)SHARV.

Volatility is widely used in finance. It is used as a proxy for risk and therefore useful risk management, derivative pricing and hedging, market making, market timing, portfolio selection and many other financial activities. Obtaining reliable and estimations and forecasts is crucial. Almost all the financial uses of volatility models entail forecasting aspects of future returns. The two main methods for modelling volatility are GARCH-type models and stochastic volatility (SV) models. The main difference lies in their information set. Namely, whether the volatility is a random process itself. Breitung and Hafner (2016) and Politis (2007) state that most GARCH-type and SV models ignore crucial information from the current observation and therefore inefficiently now- and forecast the volatility. Breitung and Hafner (2016) propose a model where they do use the current return to estimate the volatility, but for this reason the model is only able to nowcast.

Smetanina (2017) proposes a new model where she attempts to link the GARCH-type and SV models into a hybrid: Real-Time GARCH. This model includes the lagged squared return and current squared return innovation in the volatility process, i.e,

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 +_{t-1}^2 +\psi\varepsilon_{t-1}^2, \tag{1}$$

where $\varepsilon_t \equiv r_t/\sigma_t$ are i.i.d(0, 1) with $\mathrm{E}\varepsilon_t^4 < \infty$. The RT-GARCH model aims to use all available information efficiently. She states that this model is able to obtain better volatility forecasts than the standard GARCH-type models.

All these models assume that the volatility of volatility is constant over time. However, Corsi et al. (2008) find empirical evidence of volatility clustering in realized variance and that accounting for this clustering improves the accuracy of point forecasts. True volatility should also exhibit time-variation and clustering as realised measures are consistent estimators of the implied volatility. In the GARCH and SV literature, Ding (2021a) is the first to propose a model with conditional heteroskedasticity in the volatility. The model is based upon the idea of RT-GARCH, but allowing for time-varying volatility of volatility. This model is called SHARV. He find empirical evidence that his model has a better fit, a more efficient parameter estimator as well as more accurate volatility and VaR forecasts than other common GARCH-type models.

Nowadays there is an abundance of high frequency data, due to the rise of automated trading. Because of that a new approach to volatility modelling has arisen, which focuses on modelling daily realised measures (RM). Volatility is no longer treated as a latent variable, but can be consistently estimated using intraday data. Nevertheless, GARCH and SV models are still widely used by researchers and practitioners for several reasons. First, Hansen and Lunde (2006) state that high frequency data are contaminated by microstructure noise. Also, these RM-based models rely on the availability of high frequency data, which is not also available for illiquid assets. Third, they have no direct continuous time analogues, which makes them not ideal for purposes as derivative pricing.

Ding (2021a) proposes a simple model for returns, volatility and volatility of volatility, called (A)SHARV. This model is a simplification of his earlier proposed model the Augmented RT-GARCH (Ding (2021b)). The ART-GARCH model includes a constant term and the squared lagged return. Ding (2021a) argued that these terms were unnecessary. The ART-GARCH model is more extensive and potentially superior to the simplified (A)SHARV. Our main research question is: "Can ART-GARCH beat (A)SHARV?".

We compare (A)SHARV against the ART-GARCH models and three benchmark models: GARCH, GJR-GARCH and RT-GARCH. We compare the different models and perform an empirical analysis. We perform in-sample analysis, where we look at QQ-plots of the standardised residuals and also nowcast the volatility and compare the performance through MSE and the model confidence set as proposed by Hansen et al. (2011). Second, we analyse the models out-of-sample by comparing forecasting performance. Again, we compare the models by means of MSE and the 95% MCS. We use the realized variance as a proxy for volatility. We use daily open-to-close (log) returns of the SP500, DJIA and AEX from 3 January 2000 to 31 December 2019.

We find no empirical evidence that ART-GARCH can beat (A)SHARV. The ART-GARCH models have similar forecasting performance as (A)SHARV. In nowcasting it performs slightly worse. In terms of BIC, ASHARV outperforms the other models. The QQ-plots show that ASHARV has the best goodness-of-fit. We do find that both ART-GARCH and (A)SHARV outperform the benchmark models. Comparing QQ-plots, we see that both models have a better goodness-of-fit than the benchmark models. Also in terms of BIC and log likelihood the models are favoured over the benchmark models. Lastly, we find that both models have a better now- and forecasting performance than the other GARCH-type models.

2 Augmented RT-GARCH

2.1 Importance of the volatility of volatility

Ding (2021a) showed why the volatility of volatility should be incorporated in a volatility model as following. Suppose that the true data generating process is defined as in, Andersen (1994), by

$$r_t = f(K_t)_t, \tag{2}$$
$$K_t = \omega + \beta K_{t-1} + (\alpha + \gamma K_{t-1})u_t, \tag{3}$$

where $u_t - 1$ are i.i.d.(0,1) white noise and $f(.)$ is a positive measurable function of K_t. Also we assume that $\mathrm{E}[\varepsilon_t u_t] = 0$. Here, K_t has conditional heteroskedasticity in the volatility. Most SV models assume $\gamma = 0$ which means constant volatility of volatility and most GARCH models use γr_{t-1}^2 instead of the stochastic term. This results in a measurement error in K_t. Nelson (1992) and Nelson and Foster (1995) showed that as the time between observations goes to zero, misspecified GARCH models can still consistently now- and forecast the true volatility. However, the measurement errors are not negligible for fixed time intervals. To show this, suppose we estimate a GARCH model on the sample, which follows

$$\tilde{K}_t = (\omega + \alpha) + \beta \tilde{K}_{t-1} + \gamma r_{t-1}^2. \tag{4}$$

If $\tilde{K}_t = K_t$, then $\tilde{K}_t$ has a measurement error $v_t = \alpha(u_t - 1) + \gamma(K_{t-1}u_t - r_{t-1}^2)$. The measurement error can be large as, $\mathrm{E}v_t = 0$, but the conditional mean, $\mathrm{E}[v_t|F_{t-1}] = \gamma(K_{t-1} - r_{t-1}^2)$, is generally not zero, where F_{t-1} is σ generated by all known returns at time $t-1$.

Ding (2021a) stated that the main reason to incorporate time-varying volatility of volatility into the model is that even when the DGP in (3) is misspecified, the average measurement error should still be smaller than those of models with constant volatility of volatility. To clarify this, suppose τ_t is the true volatility of volatility and $K_{t-1} = \tilde{K}_{t-1}$ such that the measurement error is given by $v_t = (\tau_t - \alpha - \gamma K_{t-1})(u_t - 1)$ with conditional mean zero. Usually when volatility is small the volatility of volatility is also small and vice versa. Therefore, $|\tau_t - \alpha - \gamma K_{t-1}|$ should be on average smaller than $|\tau_t - c|$. This still holds even if the innovation term of K_t is misspecified as long as it is positively correlated with u_t and shares similar first two moments. So, for GARCH, we need that $\rho(u_t, \varepsilon_{t-1}^2) > 0$ to make $|\tau_t u_t - \gamma r_{t-1}^2|$ small. This is hard to proof empirically, but it is reasonable to assume that $\rho(u_t, \varepsilon_t^2) > 0$. Therefore, we use ε_t^2 to approximate u_t.

2.2 Augmented RT-GARCH models

We use the ART-GARCH models as in Ding (2021b). Let the joint process $(r_t,\, \sigma_t^2)$ satisfy

$$r_t = \sigma_t \varepsilon_t, \tag{5}$$
$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \phi(r_{t-1}^-)^2 + (\psi_1 + \psi_2\sigma_{t-1}^2)\varepsilon_t^2 + \eta(\varepsilon_t^-)^2, \tag{6}$$

where $\varepsilon_t \equiv r_t/\sigma_t$ are i.i.d. random variables symmetric around zero with the first two moments equal to 0 and 1, respectively and $\mathrm{E}\varepsilon^4 < \infty$. Also $x^- = \min(0, x)$. To ensure $\sigma_t^2 > 0$ with probability one, we require the parameter vector $(\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)' \geq 0$ with at least one of the inequalities being strict. We specify three different ART-GARCH models. First, the full specification with no additional restrictions is called the ART-GJR-GARCH-F model. Leverage effects come both from current and lagged negative returns. Second the ART-GJR-GARCH with leverage effect only from current negative return, i.e., $\phi = 0$. And last, the symmetric model, ART-GARCH with $\phi = \eta = 0$. The model nests RT-GARCH from Smetanina (2017) and GARCH. Setting $\psi_2 = \phi = \eta = 0$ makes the joint process a RT-GARCH model. By additionally setting $\psi_1 = 0$, RT-GARCH becomes GARCH. Because all models are nested we can perform the specification tests for model selection.

This model specification allows us to model the volatility and volatility of volatility simultaneously. To see this we express (6) as an AR(1) process with stochastic coefficient,

$$\sigma_t^2 = \Phi_0 + \Phi_{1,t-1}\sigma_{t-1}^2 + z_t, \tag{7}$$

where

$$\Phi_0 = \alpha + \psi_1 + 0.5\eta, \tag{8}$$
$$\Phi_{1,t-1} = \beta + \psi_2 + \gamma\varepsilon_{t-1}^2 + \phi(\varepsilon_{t-1}^-)^2, \tag{9}$$

and

$$z_t = (\psi_1 + \psi_2\sigma_{t-1}^2)^2(\varepsilon_t^2 - 1) + \eta((\varepsilon_t^-)^2 - 0.5), \tag{10}$$

is a martingale difference sequence (MDS) with conditional variance

$$\mathrm{E}[z_t^2|F_{t-1}] = \kappa(\psi_1 + \psi_2\sigma_{t-1}^2)^2 + \kappa\eta(\psi_1 + \psi_2\sigma_{t-1}^2) + (0.5\kappa + 0.25)\eta^2, \tag{11}$$

where $\kappa = \mathrm{E}\varepsilon_t^4 - 1$. By definition, $\mathrm{E}[z_t^2|F_{t-1}]$ is the conditional variance of σ_t^2 at time $t-1$. It is a quadratic function of the volatility. Therefore, the volatility and volatility of volatility can be estimated through one filter even though $\mathrm{E}[z_t^2|F_{t-1}]$ is stochastic.

Also the model can capture multiple lags. Specifically,

$$\sigma_t^2 = \alpha + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 + \sum_{j=1}^{q}(\gamma_j + \phi_j I_{t-j})r_{t-j}^2 + \varepsilon_t^2\sum_{j=1}^{l}(\phi_1 + \phi_{j+1}\sigma_{t-j}^2 + \eta I_t), \tag{12}$$

where I_t is an indicator function that equals 1 if $r_t < 0$. However, in this paper we focus on the class of ART-GARCH(1,1,1) models.

3 (Asymmetric) SHARV

Ding (2021a) argues that the ART-GARCH models can be simplified. First, he showed empirical evidence that the quasi-maximum likelihood estimates of the constant term are close to zero for all models which include the current squared return innovation. Also he stated that you do not need to include the squared lagged return in the volatility process. This because of two reasons. First, he showed that σ_t^2 in (12) can be approximated by,

$$\sigma_t^2 \approx \sum_{j=0}^{\infty} \beta^j \frac{a_{t-1-j}}{b_{t-1-j}} r_{t-j}^2, \tag{13}$$

where $b_{t-1} = \beta\sigma_{t-1}^2$ and $a_{t-1} = \psi_1 + \psi_2\sigma_{t-1}^2$. This shows that even though the SHARV specification does not include the squared lagged return, it implicitly assigns time-varying weights to all past squared returns. Second, $\sigma_{t-1}^2\varepsilon_t^2$ is a more accurate measure of σ_t^2 than r_{t-1}^2. Which corresponds with Hansen et al. (2012) where they find that r_{t-1}^2 becomes insignificant after including realized measures of the volatility in the process. Removing the constant term and squared lagged return from the ART-GARCH specification results in a new simplified model, called the (A)SHARV model.

3.1 SHARV model

We use the SHARV model as in Ding (2021a). Let the joint process (r_t, σ_t^2) satisfy

$$r_t = \sigma_t\varepsilon_t, \tag{14}$$

$$\sigma_t^2 = \beta\sigma_{t-1}^2 + (\psi_1 + \psi_2\sigma_{t-1}^2)\varepsilon_t^2, \tag{15}$$

where $\varepsilon_t, \psi_1, \psi_2$ and $\beta > 0$ are the same as in ART-GARCH. SHARV is fully nested in the ART-GARCH model. The main innovation of the SHARV model, as well as the ART-GARCH models, is the term $\psi_2\sigma_{t-1}^2\varepsilon_t^2$ which accounts for the conditional heteroskedasticity in σ_t^2. The main drawback here is that the model cannot allow for skewness in ε_t and $\mathrm{E}[r_t|F_{t-1}] = 0$ simultaneously. However, in 2.3 we show that SHARV can be extended to an asymmetric SHARV that allows for skewness in the error term while r_t is really close to an MDS.

SHARV can also include multiple lags. SHARV(p,q,l) is given by

$$\sigma_t^2 = \sum_{j=1}^{p} \beta_j\sigma_{t-j}^2 + \sum_{j=1}^{q}(\psi_{1j} + \sum_{k=1}^{l} \psi_{2j,k}\sigma_{t+1-j-k}^2)\varepsilon_{t+1-j}^2. \tag{16}$$

We focus on SHARV(1,1,1) for the remainder of this paper.

3.2 Asymmetric SHARV model

We use the ASHARV as in Ding (2021a). This because SHARV does not capture the well documented leverage effect between asset returns and their volatility. The ASHARV is defined as follows

$$r_t = \mu\sigma_{t-1} + \sigma_t\varepsilon_t, \tag{17}$$

$$\sigma_t^2 = \beta\sigma_{t-1}^2 + (\psi_1 + \psi_2\sigma_{t-1}^2)\varepsilon_t^2 + (\eta + \omega\sigma_{t-1}^2)(\varepsilon_t^-)^2, \tag{18}$$

where $x^- = \min(x, 0)$ and ε_t satisfy the same conditions as in SHARV. ASHARV is not nested in the ART-GARCH model as μ and ω are added. This model includes a drift term $\mu\sigma_{t-1}$. This because $\sigma_t\varepsilon_t$ is no longer an odd function of ε_t and thus not an MDS. Ding (2021a) showed that without the drift term, $\mathrm{E}[r_t|F_{t-1}]$ is always negative, which is not a desired feature.

Different as in other asymmetric GARCH models, the leverage effect is not lagged. The conditional variance of σ_t^2 is given by

$$\mathrm{Var}(\sigma_t^2|F_{t-1}) = \kappa(\psi_1+\psi_2\sigma_{t-1}^2)^2 + (0.5\kappa+0.25)(\eta+\omega\sigma_{t-1}^2)^2 + \kappa(\psi_1+\psi_2\sigma_{t-1}^2)(\eta+\omega\sigma_{t-1}^2),$$

where $\kappa = \mathrm{E}\varepsilon_t^4 - 1$. We can easily see that σ_t is weakly stationary when $\beta+\psi_2+\omega/2 < 1$ and the unconditional volatility is defined as follows, $\mathrm{E}\sigma_t^2 = (\psi_1+\eta/2)/(1-\beta-\psi_2-\omega/2)$.

4 Properties of ART-GARCH and (A)SHARV

We state some assumptions and theorems needed for model estimation and forecasting. These assumptions and theorems are obtained from Ding (2021b) and Ding (2021a).

Assumption 2.1. Let ε_t be i.i.d. random variables symmetric around zero with $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}\varepsilon_t^2 = 1$ and $\mathrm{E}\varepsilon_t^4 < \infty$.

Theorem 2.1 Ding If (r_t, σ_t^2) are generated by either (5) and (6) (ART-GARCH models), (14) and (15) (SHARV) or (17) and (18) (ASHARV), ε_t satisfies Assumption 2.1 and $\theta = (\alpha,\beta,\gamma,\phi,\psi_1,\psi_2,\eta,\omega,\mu)' \geq 0$, then the conditional density of the returns process is given by

$$f_r(y|F_{t-1}) = \begin{cases} \frac{y}{d_1(y,\sigma_{t-1}^2;\theta)d_2(y,\sigma_{t-1}^2;\theta)} f_\varepsilon(d_2(y,\sigma_{t-1}^2;\theta)), & \text{for } y \neq 0, \\ \frac{1}{\sqrt{b_{t-1}}} f_\varepsilon(0), & \text{for } y = 0, \end{cases} \tag{19}$$

where $f_\varepsilon(.)$ is the pdf of ε_t,

$$d_1(y,\sigma_{t-1}^2;\theta) = \sqrt{b_{t-1}^2 + 4a_{t-1}y^2 + 4c_{t-1}(y^-)^2}, \tag{20}$$

and

$$d_2(y,\sigma_{t-1}^2;\theta) = \begin{cases} \mathrm{sign}(y)\sqrt{\frac{d_1(y,\sigma_{t-1}^2;\theta)-b_{t-1}}{2a_{t-1}+2c_{t-1}I_{(y<0)}}}, & \text{if } (\psi_1,\psi_2,\eta,\omega)' \neq \mathbf{0}, \\ y/\sqrt{b_{t-1}}, & \text{if } (\psi_1,\psi_2,\eta,\omega)' = \mathbf{0}, \end{cases} \tag{21}$$

where

$$a_{t-1} = \psi_1 + \psi_2\sigma_{t-1}^2, \tag{22}$$
$$b_{t-1} = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \phi(r_{t-1})^2, \tag{23}$$
$$c_{t-1} = \eta + \omega\sigma_t^2. \tag{24}$$

Theorem 2.2 Ding Let (r_t, σ_t^2) be generated by either (5) and (6) (ART-GARCH models), (14) and (15) (SHARV) or (17) and (18) (ASHARV) and let ε_t satisfy Assumption 2.1 and $\mathrm{E}\varepsilon_t^6 < \infty$. Then, the quasi maximum likelihood estimator (QMLE) $\hat{\theta}$ of the true parameter θ_0 is given by $\hat{\theta} = \arg\max_{\theta\in\Theta} L_T(\theta)$, where $L_T(\theta)$ is the quasi-log-likelihood function of $\tilde{r}_t = r_t - \mu\sigma_{t-1}$, given by

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \tag{25}$$

where

$$l_t(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}d_2(\tilde{r}_t,\sigma_{t-1}^2;\theta)^2 + \log\frac{\tilde{r}_t}{d_1(\tilde{r}_t,\sigma_{t-1}^2;\theta)d_2(\tilde{r}_t,\sigma_{t-1}^2;\theta)}, \tag{26}$$

where d_1 and d_2 are given in (20) and (21).

4.1 Now- and forecasting ART-GARCH

Next, we show the filters for volatility nowcasting and forecasting for the ART-GARCH models as proposed by Ding (2021b). Note there are two different concepts of volatility. Namely, instantaneous volatility σ_t^2 and conditional variance $\mathrm{Var}[r_t|F_{t-1}]$. Since $\mathrm{E}[r_t|F_{t-1}]$ is close to zero, we approximate $\mathrm{Var}[r_t|F_{t-1}]$ by $\mathrm{E}[r_t^2|F_{t-1}]$ which we regard as the conditional variance and call σ_t^2 just volatility. Note that the true conditional variance is $\mathrm{E}[(r_{t+n}-\mathrm{E}[r_{t+n}|F_t]^2|F_t] \approx \mathrm{E}[r_{t+n}^2|F_t]$ for all n ≥ 1. We define the volatility nowcast (27), one-step ahead volatility forecast (28), one-step ahead conditional variance forecast (29), two-step ahead volatility forecast (30), two-step ahead conditional variance forecast (31) and finally the multi-period ahead forecasts in Theorem 2.3.

$$\sigma_t^2 = \frac{1}{2}b_{t-1} + \frac{1}{2}\sqrt{b_{t-1}^2 + 4a_{t-1}r_t^2 + 4\eta(r_t^-)^2}, \tag{27}$$

$$\mathrm{E}[\sigma_{t+1}^2|F_t] = \alpha + \psi_1 + \frac{1}{2}\eta + (\beta+\psi_2)\sigma_t^2 + \gamma r_t^2 + \phi(r_t^-)^2, \tag{28}$$

$$\mathrm{E}[r_{t+1}^2|F_t] = \alpha + (\psi_1 + \frac{1}{2}\eta)\mathrm{E}\varepsilon_t^4 + (\beta + \psi_2\mathrm{E}\varepsilon_t^4)\sigma_t^2 + \gamma r_t^2 + \phi(r_t^-)^2, \tag{29}$$

$$\mathrm{E}[\sigma_{t+2}^2|F_t] = \alpha + \psi_1 + \frac{1}{2}\eta + \frac{1}{4}\phi\eta\mathrm{E}\varepsilon_t^4 + (\beta+\psi_2)\mathrm{E}[\sigma_{t+1}^2|F_t] + (\gamma + \frac{1}{2}\phi)\mathrm{E}[r_{t+1}^2|F_t], \tag{30}$$

$$\mathrm{E}[r_{t+2}^2|F_t] = \alpha + (\psi_1 + \frac{1}{2}\eta + \frac{1}{4}\phi\eta)\mathrm{E}\varepsilon_t^4 + (\beta + \psi_2\mathrm{E}\varepsilon_t^4)\mathrm{E}[\sigma_{t+1}^2|F_t] + (\gamma + \frac{1}{2}\phi)\mathrm{E}[r_{t+1}^2|F_t]. \tag{31}$$

Theorem 2.3 Ding Let (r_t, σ_t^2) be generated by (5) and (6) and let ε_t satisfy Assumption 2.1. Then for n $\geq 3, n \in \mathrm{Z}^+$, the n-step ahead volatility forecast and conditional variance are respectively given by

$$\mathrm{E}[\sigma_{t+n}^2|F_t] = \mathrm{E}\sigma_t^2 + \Phi_1(\mathrm{E}[\sigma_{t+n-1}^2|F_t] - \mathrm{E}\sigma_t^2) + \Phi_2(\mathrm{E}[\sigma_{t+n-2}^2|F_t] - \mathrm{E}\sigma_t^2), \tag{32}$$

and

$$\mathrm{E}[r_{t+n}^2|F_t] = \mathrm{E}r_t^2 + \Phi_1(\mathrm{E}[r_{t+n-1}^2|F_t] - \mathrm{E}r_t^2) + \Phi_2(\mathrm{E}[r_{t+n-2}^2|F_t] - \mathrm{E}r_t^2), \tag{33}$$

where $\Phi_1 = \beta + \gamma + \psi_2 + \frac{1}{2}\phi$ and $\Phi_2 = \kappa\psi_2(\gamma + \frac{1}{2}\phi)$ with $\kappa = \mathrm{E}\varepsilon_t^4 - 1$.

4.2 Now- and forecasting (A)SHARV

Last, we show the filters for volatility nowcasting and forecasting for (A)SHARV as in Ding (2021a). We assume that $\mathrm{E}\varepsilon_t^3 = 0$ for the rest of the paper. The volatility nowcast is defined as

$$\sigma_t^2 = \frac{1}{2}b_{t-1} + \frac{1}{2}\sqrt{b_{t-1}^2 + 4a_{t-1}\tilde{r}_t^2 + 4c_{t-1}(\tilde{r}_t^-)^2}. \tag{34}$$

The n-step ahead volatility forecast, for $n \in Z^+$, is given by

$$\mathrm{E}[\sigma_{t+n}^2|F_t] = \psi_1 + \frac{1}{2}\eta + (\beta + \psi_2 + \frac{1}{2}\omega)\mathrm{E}[\sigma_{t+n-1}^2|F_t]. \tag{35}$$

The n-step ahead conditional variance forecast, for $n \in Z^+$, is given by

$$\mathrm{E}[\tilde{r}_{t+n}^2|F_t] = (\psi_1 + \frac{1}{2}\eta)\mathrm{E}\varepsilon_t^4 + (\beta + (\psi_2 + \frac{1}{2}\omega)\mathrm{E}\varepsilon_t^4)\mathrm{E}[\sigma_{t+n-1}^2|F_t]. \tag{36}$$

5 Comparison to other volatility models

We compare the ART-GARCH models and (A)SHARV against GARCH, GJR-GARCH and RT-GARCH. We consider the news impact curve (NIC) from Engle and Ng (1993) as in Ding (2021b). For the ART-GARCH models, the NIC is given by

$$\mathrm{E}[r_{t+1}^2|F_t] = \overline{\alpha} + 0.5\overline{\beta}\left(\overline{b} + \sqrt{\overline{b^2} + 4\overline{a}r_t^2 + 4\eta(r_t^-)^2}\right) + \gamma r_t^2 + \phi(r_t^-)^2, \tag{37}$$

where we have set $\varepsilon \sim N(0,1)$, $\overline{\alpha} = \alpha + 3(\psi_1 + 0.5\eta_1), \overline{\beta} = \beta + 3\psi_2, \overline{a} = \psi_1 + \psi_2\overline{\sigma}_2$ and $\overline{b} = \alpha + \beta\overline{\sigma}_2 + \gamma\overline{r}_2 + \phi\overline{r_2^-}$ with $\overline{\sigma}_2, \overline{r}_2$ and $\overline{r_2^-}$ being the unconditional levels of σ_t^2, r_t^2 and $(r_t^-)^2$ given by

$$\mathrm{E}\sigma_t^2 = \frac{\alpha + \psi_1 + 0.5\eta + 0.25\phi\eta\mathrm{E}\varepsilon_t^4 + (\gamma + 0.5\phi)(\psi_1 + 0.5\eta)\kappa}{1 - (\beta + \psi_2 + \gamma + 0.5\phi + \kappa\psi_2(\gamma + 0.5\phi))}, \tag{38}$$

$$\mathrm{E}r_t^2 = \frac{\alpha + (\psi_1 + 0.5\eta + 0.25\phi\eta)\mathrm{E}\varepsilon_t^4 + \kappa(\alpha\psi_2 - \beta(\psi_1 + 0.5\eta) + 0.25\phi\eta\psi_2\mathrm{E}\varepsilon_t^4)}{1 - (\beta + \psi_2 + \gamma + 0.5\phi + \kappa\psi_2(\gamma + 0.5\phi))}, \tag{39}$$

$$\mathrm{E}(r_t^-)^2 = 0.5\mathrm{E}r_t^2 + 0.5\eta\mathrm{E}\varepsilon_t^4. \tag{40}$$

For (A)SHARV, the NIC as in Ding (2021a) is given by

$$\begin{aligned}\mathrm{Var}(r_{t+1}|\tilde{F}_t) =& 3\psi_1 + 1.5\eta + 0.5(\beta + 3\psi_2 + 1.5\omega)\left(\overline{b} + \sqrt{\overline{b}^2 + 4\overline{a}r_t^2 + 4\overline{c}(r_t^-)^2}\right)\\ &- \left(\mathrm{E}[\sigma_{t+1}\varepsilon t + 1|\tilde{F}_t]\right)^2,\end{aligned} \tag{41}$$

where we have set $\varepsilon \sim N(0,1)$, $\overline{a} = \psi_1 + \psi_2\mathrm{E}\sigma_t^2$, $\overline{b} = \beta\mathrm{E}\sigma_t^2$ and $\overline{c} = \eta + \omega\mathrm{E}\sigma_t^2$. $\mathrm{E}\sigma_t^2$ is defined in section 3.2. The second order approximation of $\mathrm{E}[\sigma_{t+1}\varepsilon_{t+1}|\tilde{F}_t]$ is given by

$$\mathrm{E}[\sigma_{t+1}\varepsilon_{t+1}|\tilde{F}_t] \approx \mu\sigma_{t-1} + \frac{c_{t-1}}{2\sqrt{b_{t-1}}}\mathrm{E}(\varepsilon_t^-)^3 - \frac{2a_{t-1}c_{t-1} + c_{t-1}^2}{8b_{t-1}^1.5}\mathrm{E}(\varepsilon_t^-)^5, \tag{42}$$

where $a_t = \psi_1 + \psi_2\sigma_t^2$, $b_t = \beta\sigma_t^2$ and $c_t = \eta + \omega\sigma_t^2$.

We plot the news impact curves of the ART-GARCH models and (A)SHARV against the benchmark models, to see how the conditional variance responds to different values of r_t. The NIC's for DJIA returns are shown in Figure 1, for the other indices the NIC's can be found in Appendix A. Figure 1 shows that for small values of r_t, ART-GARCH responds faster than RT-GARCH and GARCH, but SHARV responds even slightly faster. While for large values of r_t, we see the opposite. Responses of SHARV are lowest and also ART-GARCH's response is smaller than RT-GARCH and GARCH. This feature is exactly what we want. Namely, we like the volatility to respond rapidly to standard shocks but temper these abnormal shocks. Also for the asymmetric models ASHARV seems to perform best. It reacts quickly at the small negative returns but has a smaller respond to the large negative shocks. The volatility of volatility seems to act like a scaling factor to temper the effect of abnormal shocks. These results are similar to what Ding (2021a) and Ding (2021b) found.

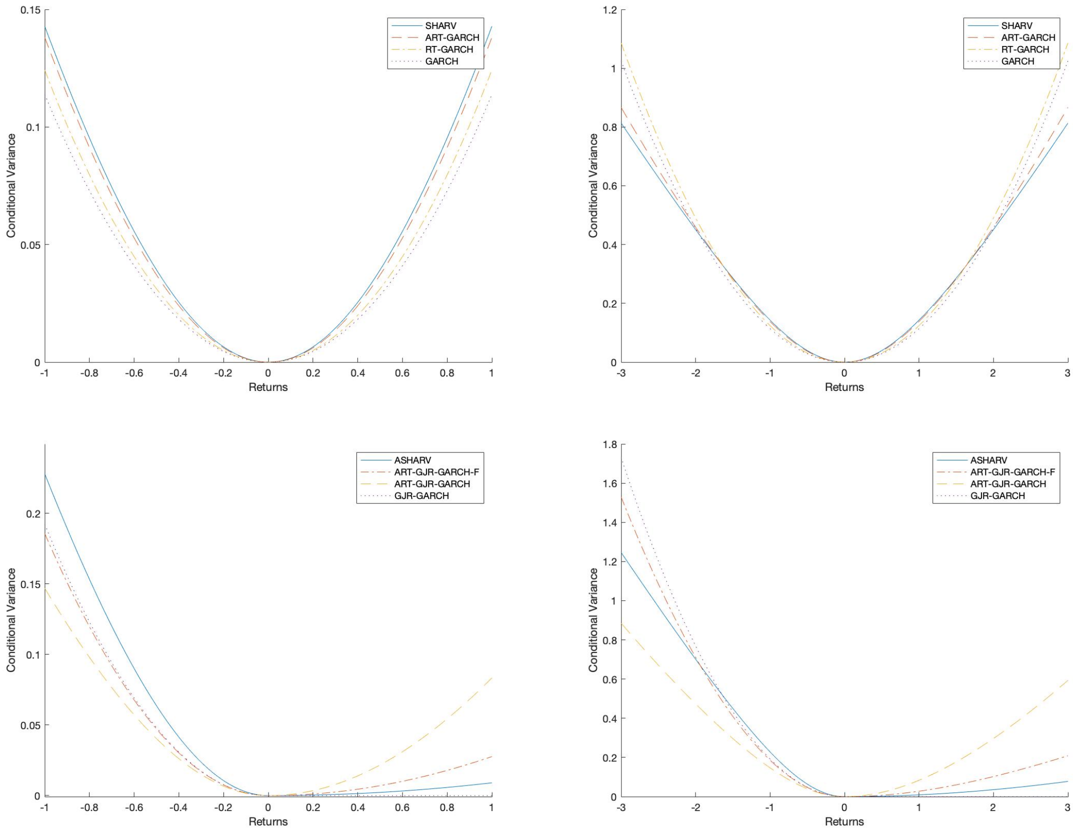


Figure 1: News impact curves for small and large values of r_t. All models' parameters are estimated from DJIA index daily returns.

Finally, as discussed in Ding (2021a), ART-GARCH models and (A)SHARV have some advantages over SV models. The main difference is that in our models, we include the current return innovation directly in the volatility process, whereas the SV model as in Breitung and Hafner (2016) do so in the log volatility specification. In general, for SV models, estimation is more difficult. Also the conditional variance is not available in closed from, which makes comparative statistics such as NIC's complicated. Computation would need numerical methods.

6 Empirical analysis

The aim of our empirical analysis is to compare the two different model specifications of Ding (2021b) and Ding (2021a) to each other and the three benchmarks models: GARCH, GJR-GARCH and RT-GARCH. We compare the goodness-of-fit using QQ-plots. We compare now- and forecasting ability of the volatility through MSE comparison and the MCS of Hansen et al. (2011). Realized variance (RV) is used as a proxy for the volatility. We compare 1-, 2-, 5,- and 10-step ahead volatility forecasts. Where we make use of an expanding window and update estimation every 50 observations.

6.1 Data description

We use daily open-to-close (log) returns of the S&P 500, Dow Jones Industrial Average (DJIA) and the Amsterdam Exchange Index (AEX) from 3 January 2000 to 31 December 2019. Realized variance is computed using 5-min intraday returns. All data is obtained from Oxford-Man Institute of Quantitative Finance.

Table 1: Parameter estimates of all models.

	α	β	γ	ψ_1	ψ_2	η	ϕ	ω	μ	BIC	LogL
S&P 500											
ASHARV		0.8893		0.0062	0.0001	0.0120		0.2037	0.0911	12310	-6129
SHARV		0.8833		0.0088	0.1093					12578	-6276
ART-GJR-GARCH-F	0.0000	0.8785	0.0000	0.0027	0.0188	0.0426	0.1337			12396	-6168
ART-GJR-GARCH	0.0000	0.8757	0.0330	0.0000	0.0472	0.0578				12489	-6219
ART-GARCH	0.0000	0.8760	0.0207	0.0100	0.0895					12592	-6275
RT-GARCH	0.0000	0.8680	0.1027	0.0208						12650	-6309
GJR-GARCH	0.0167	0.8864	0.0000				0.1923			12615	-6290
GARCH	0.0134	0.8775	0.1122							12813	-6394
DJIA											
ASHARV		0.8942		0.0068	0.0028	0.0105		0.1872	0.0957	12243	-6096
SHARV		0.8812		0.0090	0.1076					12475	-6225
ART-GJR-GARCH-F	0.0000	0.8817	0.0000	0.0036	0.0248	0.0366	0.1248			12319	-6130
ART-GJR-GARCH	0.0000	0.8816	0.0235	0.0000	0.0562	0.0524				12398	-6174
ART-GARCH	0.0000	0.8756	0.0158	0.0100	0.0954					12490	-6224
RT-GARCH	0.0000	0.8670	0.1023	0.0213						12555	-6260
GJR-GARCH	0.0160	0.8891	0.0000				0.1928			12513	-6239
GARCH	0.0131	0.8763	0.1139							12709	-6342
AEX											
ASHARV		0.9207		0.0050	0.0047	0.0063		0.1359	0.0221	12751	-6350
SHARV		0.8987		0.0082	0.0925					12896	-6435
ART-GJR-GARCH-F	0.0002	0.9013	0.0000	0.0000	0.0356	0.0313	0.0721			12800	-6370
ART-GJR-GARCH	0.0000	0.8980	0.0110	0.0000	0.0630	0.0363				12846	-6397
ART-GARCH	0.0051	0.8880	0.0028	0.0029	0.1006					12912	-6434
RT-GARCH	0.0000	0.8774	0.0907	0.0226						12991	-6478
GJR-GARCH	0.0104	0.9188	0.0000				0.1326			12936	-6451
GARCH	0.0116	0.8941	0.0965							13110	-6542

6.2 In-sample analysis

Table 1 shows the parameter estimates, BIC and log likelihood for all models and benchmark models. It does not include standard errors as there are multiple parameters at the boundary of the parameter space. The significance of the parameters can be checked through the difference in log likelihood. We clearly see an improved log likelihood after adding parameters and therefore these parameters are significant. For all ART-GARCH models the constant term α equals zero. Also for GJR-GARCH γ equals zero for all three indices. As expected, lagged volatility β is most influential in volatility estimation. In terms of BIC, we see that all ART-GARCH models as well as (A)SHARV are preferred over the benchmark models. ASHARV seems favorite with ART-GJR-GARCH-F and ART-GJR-GARCH right behind it.

Figure 2 shows the QQ-plots of the standardised residuals of the 8 different models for DJIA returns. The QQ-plots for the other indices can be found in Appendix A. It is clearly visible that GARCH has the worst goodness-of-fit. GJR-GARCH, RT-GARCH are also quite far off, especially in the left bottom corner. SHARV and ART-GARCH have significantly better goodness-of-fit then the benchmark models. However, ASHARV, ART-GJR-GARCH and ART-GJR-GARCH-F perform even better, especially in the extremes. These three models have almost standard normal quantiles. The plots indicate that the efficiency of our QMLE approaches that of MLE since $\varepsilon_t \sim N(0, 1)$ seems a reasonable assumption.

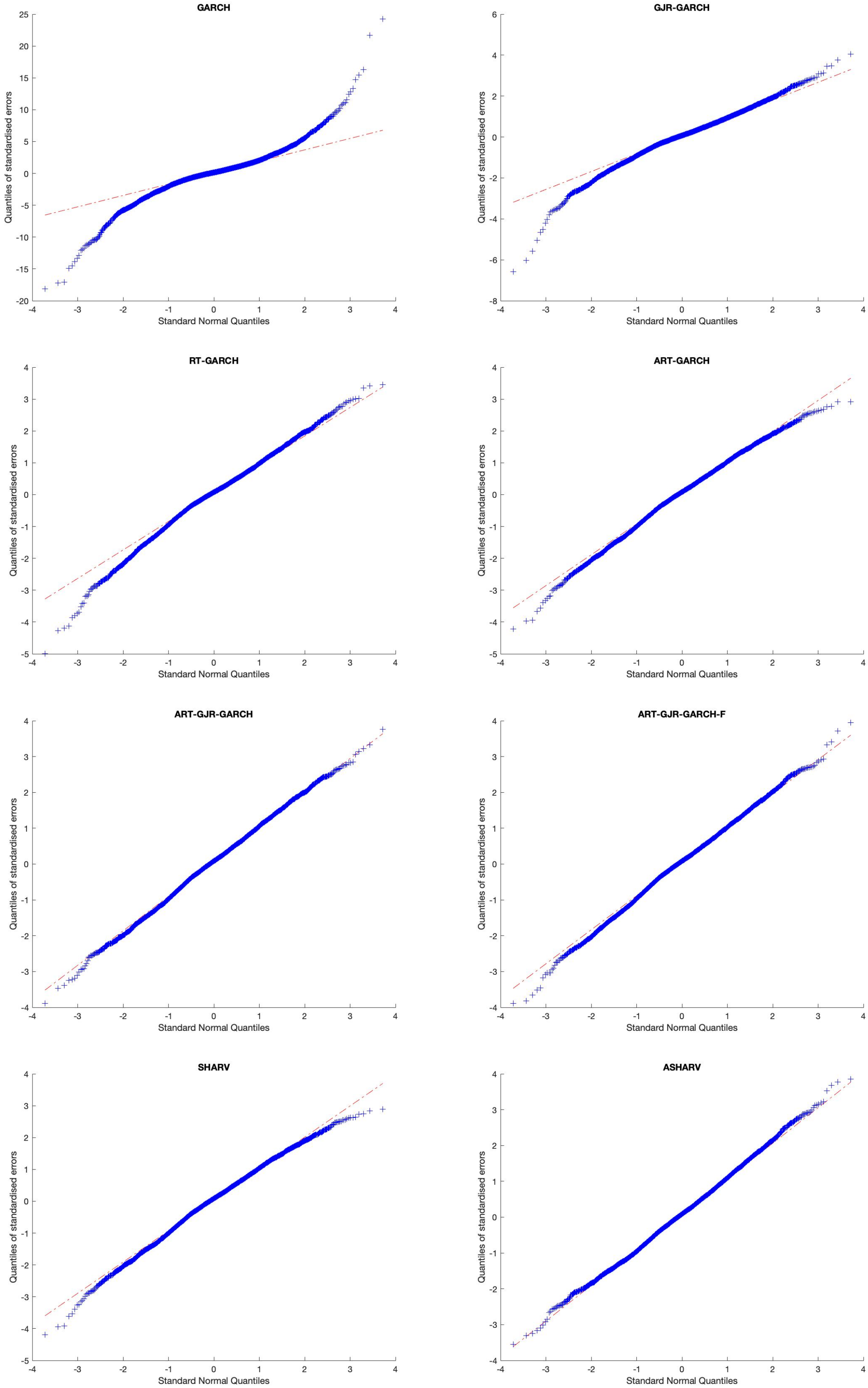


Figure 2: QQ plots of the standardised residuals of DJIA index returns.

Table 2 shows the MSE and 95% MCS. Evaluating the models we use the sample from 04 January 2010 to 31 December 2019. We see that for all indices, the ART-GJR-GARCH-F and ASHARV are in the 95% MCS with probability one. In the benchmark models we see that GJR-GARCH is in the MCS for the AEX index with a p-value of 1. GARCH is never included in the MCS. In terms of MSE, we see that, espcially for the S&P500 and DJIA, the three benchmark models are outperformed by the other models. The MSE is lowest for ASHARV and ART-GJR-GARCH-F. Therefore, we draw the same conclusion as from the MCS, and ART-GJR-GARCH-F and ASHARV seem superior to the other models in nowcasting the volatility, with a slight advantage for ASHARV in terms of MSE.

Table 2: Volatility nowcast comparison using RV as proxy for volatility.

	S&P 500		**DJIA**		**AEX**	
	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
GARCH	1.3487	0.0000	2.1941	0.0000	0.7323	0.0280
GJR-GARCH	1.2752	0.3296	2.0980	0.1102	0.6780	1.0000
RT-GARCH	1.2138	0.0000	2.1141	0.0000	0.7060	0.2102
ART-GARCH	1.1909	0.0698	2.0489	0.1644	0.7228	0.5814
ART-GJR-GARCH	1.1524	0.0792	2.0168	0.9950	0.7127	0.8686
ART-GJR-GARCH-F	1.1198	1.0000	1.9665	1.0000	0.6525	1.0000
SHARV	1.1926	0.0918	2.0505	0.1754	0.7292	0.2102
ASHARV	1.0372	1.0000	1.8978	1.0000	0.6707	1.0000

6.3 Out-of-sample analysis

Finally, we compare the forecasting performance of the models. Table 3 shows that for the 1-step ahead volatility forecasts, ART-GJR-GARCH-F and ASHARV are always in the MCS with a probability equal to 1. Looking at the MSE of the 1-step ahead forecasts, we see that ART-GJR-GARCH-F outperforms the other models for all three indices. For 2-step ahead forecasts the results are more close and also ART-GJR-GARCH is in the 95% MCS with probability one for all indices. Also in terms of MSE, ASHARV, ART-GJR-GARCH and ART-GJR-GARCH-F perform best. In the 5-step ahead forecasts we see a shift towards SHARV and ART-GJR-GARCH. This change continues in the 10-step ahead forecasts. Here we see that SHARV is the only model that is always in the MCS with a p-value of one. The MSE confirms that SHARV outperforms the other models for the 10-step ahead volatility forecasts. But, differences to ART-GJR-GARCH and ASHARV are very small. Where ASHARV continues to perform well in the 5- and 10-step ahead forecasts, ART-GJR-GARCH-F is almost the worst performing model in terms of MSE and MCS p-values.

ART-GARCH models and (A)SHARV are always together in the MCS and also differ not much in terms of MSE. To see if ART-GARCH has better forecasting performance than (A)SHARV we perform the superior predictive ability (SPA) test from Hansen (2005). We compare the best performing ART-GARCH model, in terms of MSE, against (A)SHARV for 1-,2-,5-,10-step ahead forecasts. We see that the ART-GARCH models are not able to reject the null hypothesis that the benchmark models, (A)SHARV, are not inferior to the new models. These results suggest that the ART-GARCH model has no better forecasting performance than (A)SHARV.

Table 3: Out-of-sample volatility forecasts comparison using RV as proxy for volatility.

	S&P 500		**DJIA**		**AEX**	
	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
1-step						
GARCH	1.2241	0.0000	2.7425	0.0000	0.6767	0.3798
GJR-GARCH	1.1390	0.0316	2.6260	1.0000	0.6588	1.0000
RT-GARCH	1.1907	0.0320	2.7045	0.2192	0.6744	0.8754
ART-GARCH	1.2212	0.1908	2.7581	0.2594	0.6839	0.1014
ART-GJR-GARCH	1.1659	1.0000	2.6879	0.7428	0.6705	1.0000
ART-GJR-GARCH-F	1.1093	1.0000	2.6009	1.0000	0.6524	1.0000
SHARV	1.2306	0.1908	2.7694	0.2594	0.6876	0.1348
ASHARV	1.1361	1.0000	2.6618	1.0000	0.6582	1.0000
2-step						
GARCH	1.3190	0.0000	2.8794	0.0000	0.7154	0.3746
GJR-GARCH	1.2884	0.0404	2.8518	0.0400	0.7032	1.0000
RT-GARCH	1.2839	0.0368	2.8332	0.3056	0.7147	0.4642
ART-GARCH	1.2851	0.2594	2.8459	0.5756	0.7117	0.1960
ART-GJR-GARCH	1.2482	1.0000	2.7960	1.0000	0.7003	1.0000
ART-GJR-GARCH-F	1.2455	1.0000	2.7879	1.0000	0.6937	1.0000
SHARV	1.2928	0.2594	2.8547	0.5756	0.7133	0.3122
ASHARV	1.2330	1.0000	2.7931	1.0000	0.6903	1.0000
5-step						
GARCH	1.4498	0.0368	3.0431	0.1144	0.7650	0.6842
GJR-GARCH	1.4755	0.1252	3.0970	0.1984	0.7638	0.4702
RT-GARCH	1.4353	0.0287	3.0110	0.0608	0.7771	0.0702
ART-GARCH	1.3994	0.9882	2.9767	1.0000	0.7555	0.9982
ART-GJR-GARCH	1.3810	1.0000	2.9483	1.0000	0.7481	1.0000
ART-GJR-GARCH-F	1.4622	0.2020	3.0772	0.2292	0.7756	0.3998
SHARV	1.3871	0.9882	2.9682	1.0000	0.7501	0.9982
ASHARV	1.3782	1.0000	2.9629	1.0000	0.7427	1.0000
10-step						
GARCH	1.4990	0.0560	3.0961	0.1832	0.8167	0.4112
GJR-GARCH	1.5217	0.8528	3.1592	0.1634	0.8169	0.2342
RT-GARCH	1.5030	0.0478	3.0619	0.1524	0.8544	0.3708
ART-GARCH	1.5772	0.5976	3.0621	0.8512	0.8305	0.9936
ART-GJR-GARCH	1.4402	1.0000	3.0000	1.0000	0.8755	0.9936
ART-GJR-GARCH-F	1.6001	0.1524	3.3547	0.1606	1.2323	0.3592
SHARV	1.4238	1.0000	3.0058	1.0000	0.7873	1.0000
ASHARV	1.4386	0.8528	3.0265	0.8512	0.7956	0.9936

Table 4: SPA test p-values of ART-GJR-GARCH-F against (A)SHARV

	S&P 500		**DJIA**		**AEX**	
	1-step	2-step	1-step	2-step	1-step	2-step
SHARV	0.1140	0.2140	0.1100	0.2130	0.0880	0.1910
ASHARV	0.2650	1.0000	0.1860	0.4130	0.2190	1.0000

Table 5: SPA test p-values of ART-GJR-GARCH against (A)SHARV

	S&P 500		**DJIA**		**AEX**	
	5-step	10-step	5-step	10-step	5-step	10-step
SHARV	0.2930	1.0000	0.1000	0.4090	0.4260	1.0000
ASHARV	1.0000	1.0000	0.1850	0.1620	1.0000	1.0000

7 Conclusion

In this paper, we aimed to answer the following research question: "Can ART-GARCH beat (A)SHARV?". We find no evidence that the ART-GARCH models outperform (A)SHARV. In terms of forecasting performance, (A)SHARV is always inside the 95% MCS of Hansen et al. (2011) with probability one. We find similar results for nowcasting performance. Also the QQ-plots showed that ASHARV has at least the same goodness-of-fit as the ART-GARCH models. Looking at the BIC, ASHARV is again preferred over the ART-GARCH models. The news impact curves also suggest that (A)SHARV has the best features.

We did find empirical evidence that the ART-GARCH models and (A)SHARV have a better fit as well as a better now- and forecasting performance than the benchmark models. These findings are in line with Ding (2021a) and Ding (2021b).

It has to be noted that we only tested our model on indices. We did not use our models on singular stocks or an exchange rate. Also, due to the absence of high-frequency data, we were unable to compare the realized quarticity to the filtered volatility of volatility.

We end with some suggestions for further research. It is interesting to investigate if the performance of both models could be improved by incorporating some realized measures as in Hansen et al. (2012). Also ASHARV not only differs from the ART-GARCH models in volatility dynamics but also in the mean dynamics. It is of interest to see what part of performance difference is caused by the difference in volatility dynamics and what part due to the different mean dynamics of the models.

References

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

Breitung, J. and Hafner, C. M. (2016). A simple model for now-casting volatility series. *International Journal of Forecasting*, 32(4):1247–1255.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.

Ding, Y. D. (2021a). A simple joint model for returns, volatility and volatility of volatility. *Journal of Econometrics*, forthcoming.

Ding, Y. D. (2021b). Augmented real-time garch: A joint model for returns, volatility and volatility of volatility. *Working Paper. University of Cambridge.* Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3779497`.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

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.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Nelson, D. B. (1992). Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics*, 52(1-2):61–90.

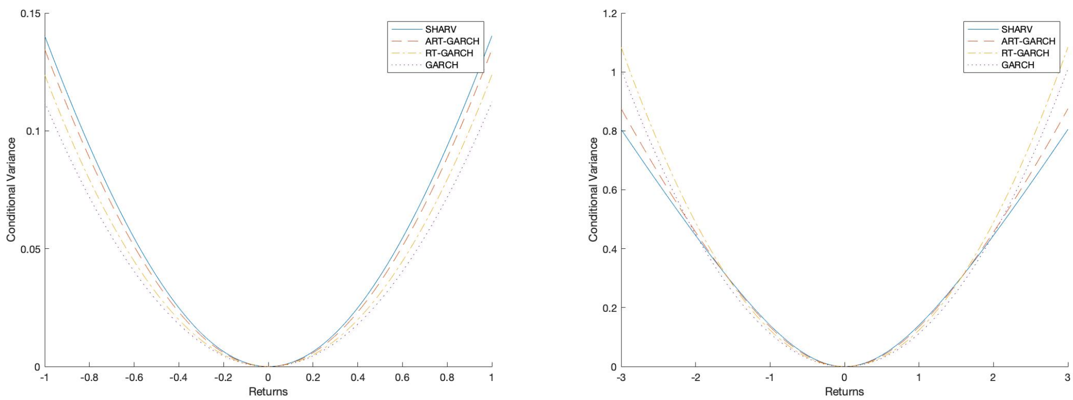
Nelson, D. B. and Foster, D. P. (1995). Filtering and forecasting with misspecified ARCH models II: making the right forecast with the wrong model. *Journal of Econometrics*, 67(2):303–335.

Politis, D. N. (2007). Model-free versus model-based volatility prediction. *Journal of Financial Econometrics*, 5(3):358–359.

Smetanina, E. (2017). Real-time GARCH. *Journal of Financial Econometrics*, 15(4):561–601.

Appendix A. Additional figures

Figure 4 - Figure 7 show the news impact curves for S&P 500 and AEX, as well as the QQ-plots for these indices.



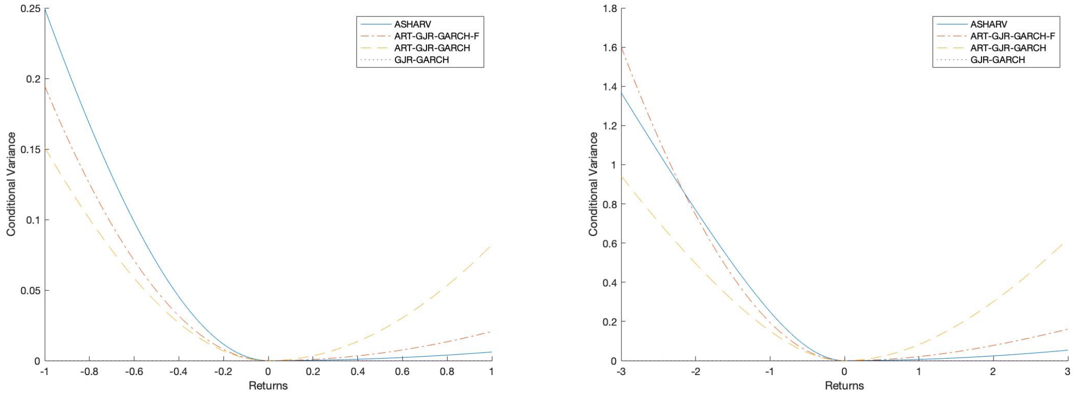


Figure 4: News impact curves for small and large values of r_t. All models' parameters are estimated from S&P 500 index daily returns.

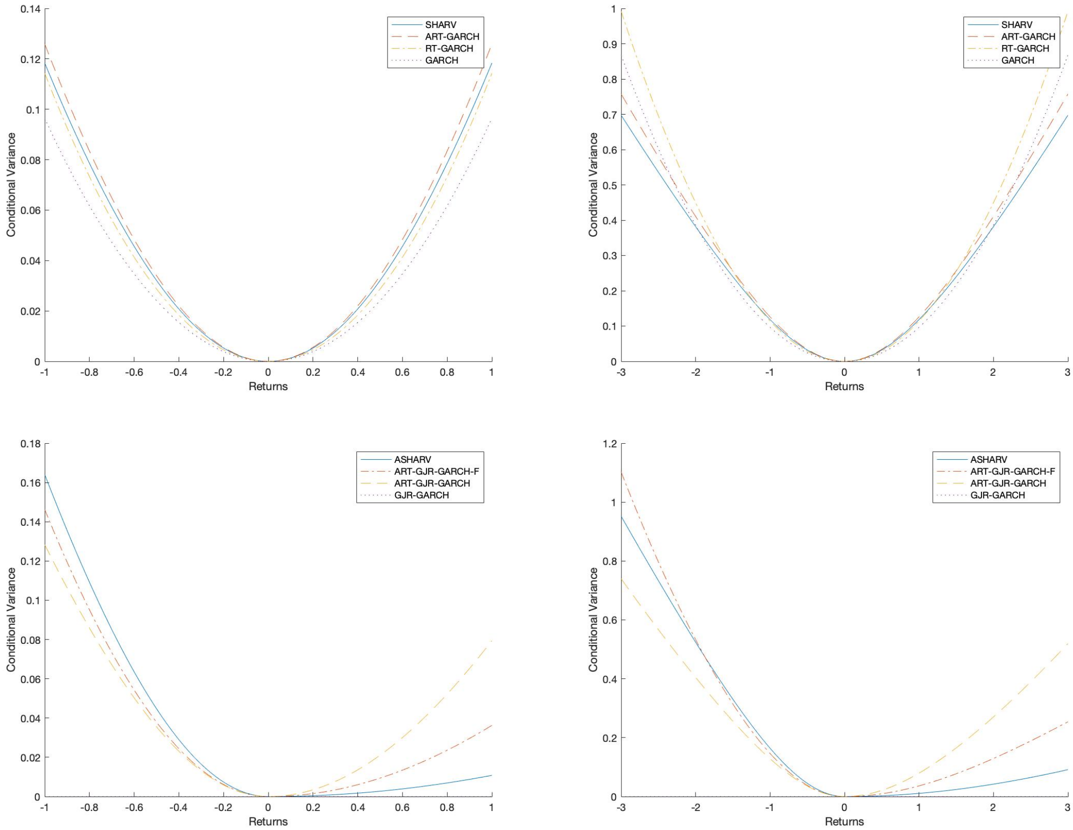


Figure 5: News impact curves for small and large values of r_t. All models' parameters are estimated from AEX index daily returns.

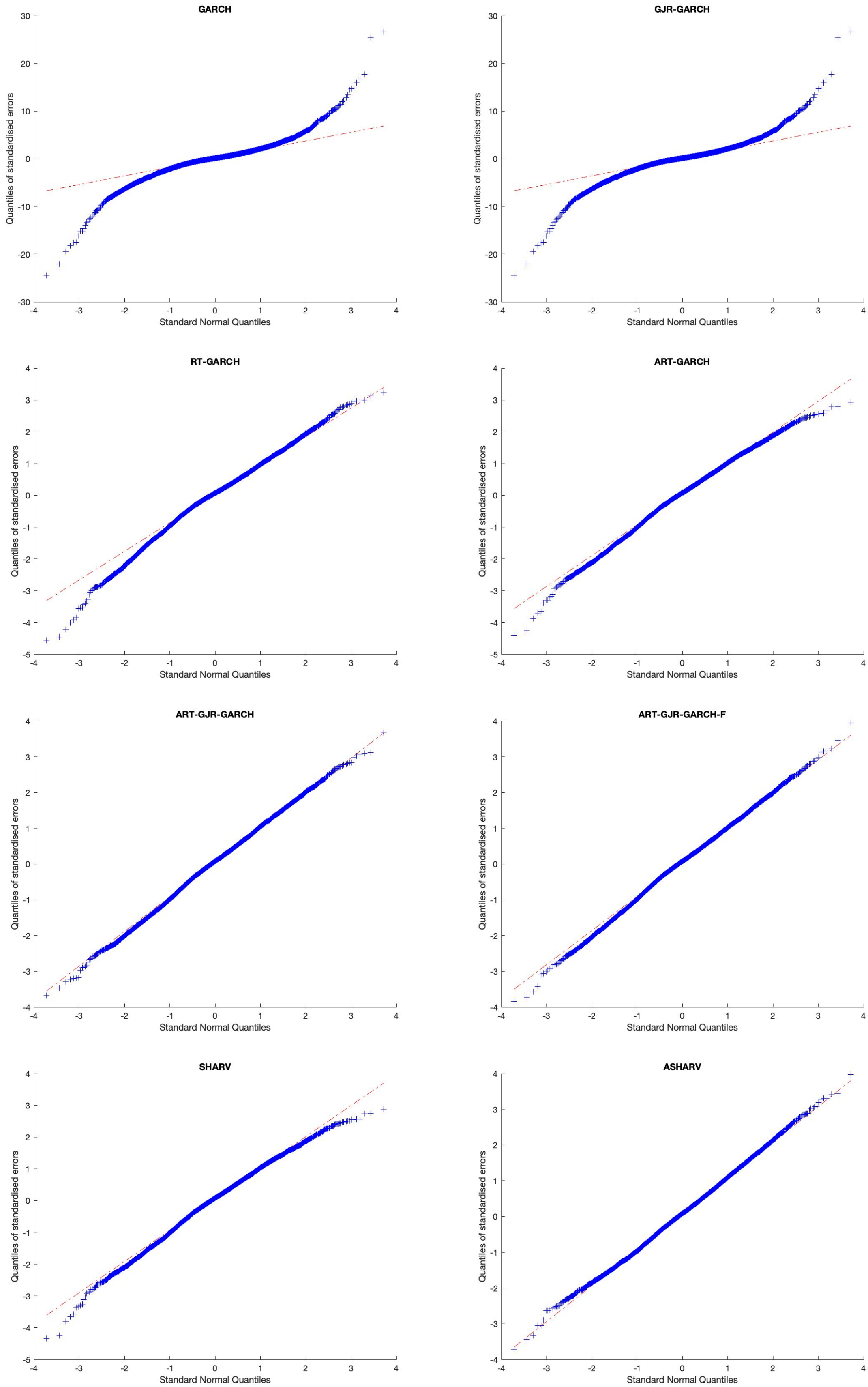


Figure 6: QQ plots of the standardised residuals of S&P 500 index returns.

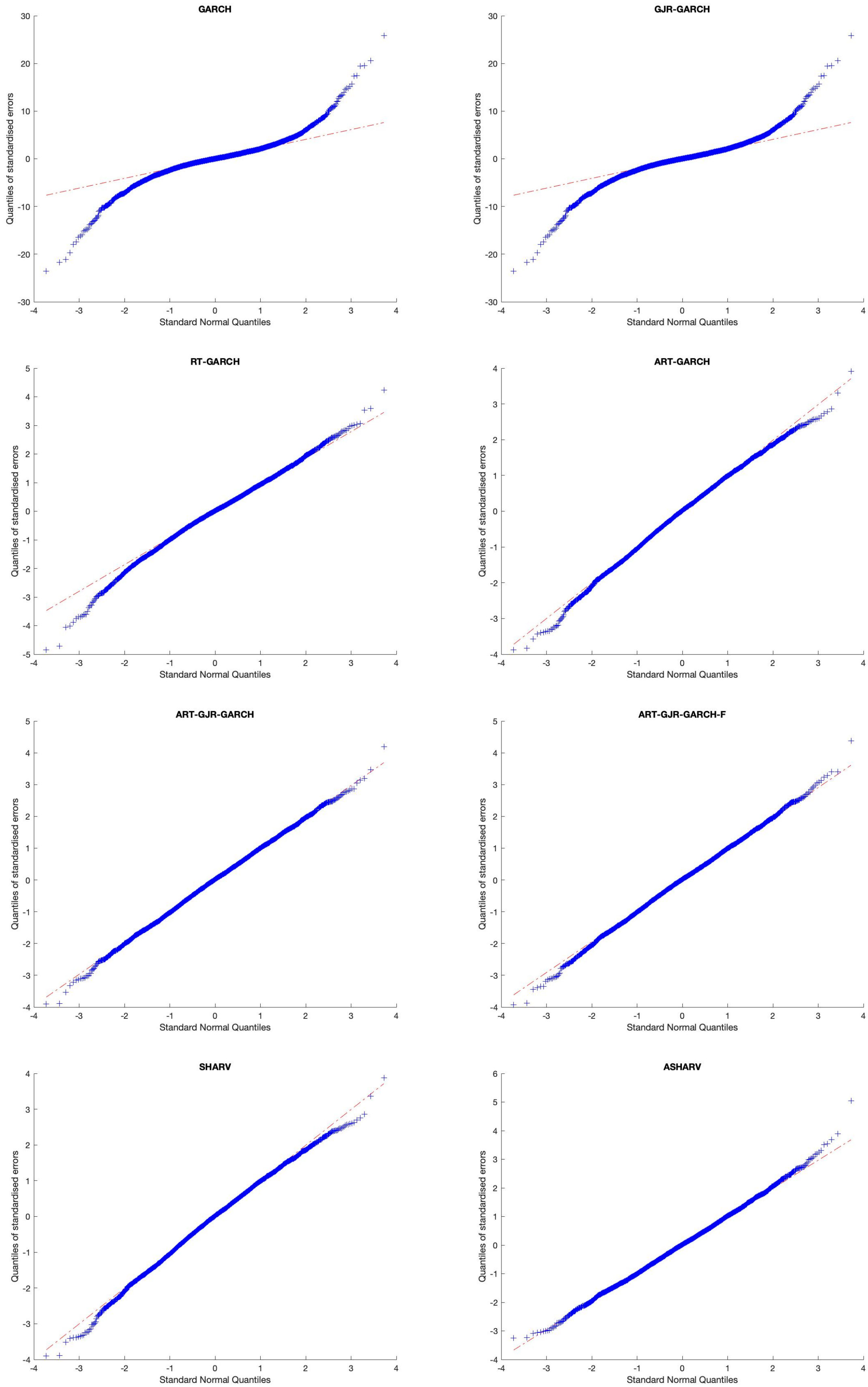


Figure 7: QQ plots of the standardised residuals of AEX index returns.

Appendix B. Code

Attached to this paper is a zip-file that contains all data and code used to construct estimates, nowcasts, forecasts, tests and figures. It contains four folders. The first, Data Excel, contains all data used in excel format.

The second folder, MATLAB, contains all code used and the data in MATLAB format. The MATLAB code consists of 7 chapters. A1 is a script to load the data. All A2 scripts are for model estimation. The scripts that start with A3 are to construct NIC's. The A4 script is for the QQ-plot generation. A5 scripts estimate the MSE and losses that are later used for the 95% MCS from Hansen et al. (2011). N-step ahead forecasts can be computed using the eight A6 scripts, corresponding to the eight different models. A7 scripts are for the MCS and SPA test. Then there are the functions which are used in the scripts.

The third folder, R, contains the data and code used in R to compute the p-values for the 95% model confidence set. We use the MCSprocedure from package 'MCS'. The last folder contains all constructed figures for this paper.