UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Augmented Real-Time GARCH: A Joint Model for Returns, Volatility and Volatility of Volatility

Dexter
Ding

Abstract

We propose a model that extends Smetanina's (2017) original RT-GARCH model by allowing conditional heteroskedasticity in the variance of volatility process. We show we are able to filter and forecast both volatility and volatility of volatility simultaneously in this simple setting. The volatility forecast function follows a second-order difference equation as opposed to first-order under GARCH(1,1) and RT-GARCH(1,1). Empirical studies confirm the presence of conditional heteroskedasticity in the volatility process and the standardised residuals of return are close to Gaussian under this model. We show we are able to obtain better in-sample nowcast and out-of-sample forecast of volatility.

Reference Details

CWPE	2112
Published	29 January 2021
Updated	19 February 2021
Key Words	GARCH, diffusion limit, forecasting, volatility of volatility
JEL Codes	C22, C32, C53, C58
Website	www.econ.cam.ac.uk/cwpe

Augmented Real-Time GARCH: A Joint Model for Returns, Volatility and Volatility of Volatility

Yashuang (Dexter) Ding*

University of Cambridge

Abstract

We propose a model that extends Smetanina's (2017) original RT-GARCH model by allowing conditional heteroskedasticity in the variance of volatility process. We show we are able to filter and forecast both volatility and volatility of volatility simultaneously in this simple setting. The volatility forecast function follows a second-order difference equation as opposed to first-order under GARCH(1,1) and RT-GARCH(1,1). Empirical studies confirm the presence of conditional heteroskedasticity in the volatility process and the standardised residuals of return are close to Gaussian under this model. We show we are able to obtain better in-sample nowcast and out-of-sample forecast of volatility.

Keywords: GARCH, SV, forecasting, volatility of volatility
JEL classification: C22, C32, C53, C58

*I am grateful to my supervisor, Oliver Linton, for his valuable comments and suggestions, as well as his constant supports and encouragement. I would also like to thank Alexey Onatskiy, Ekaterina Smetanina, Tristan Hennig and Christian Tien for useful comments. Correspondence to: Yashuang (Dexter) Ding, Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DD, UK. Email: yd274@cam.ac.uk.

1 Introduction

Volatility modelling is important in many areas of finance and economics from risk management to derivative pricing and asset allocation. There are two main approaches in volatility modelling: GARCH and its various extensions and hybrid models (Engle (1982), Bollerslev (1986), Nelson (1991), Glosten et al. (1986), Hansen et al. (2012), among others) regard volatility as determined solely by past information and share the same source of uncertainty with return process (see Francq and Zakoïan (2010) for an overview of GARCH models). On the other hand, stochastic volatility (SV) models (Heston (1993), Fong and Vasicek (1991), Longstaff and Schwartz (1992), among others) regard volatility as a latent variable driven by a different innovation term (see Shepherd (2005) for an overview of discrete and continuous time SV models). The main difference between GARCH and SV models lies in their information set, that is, whether the current information is incorporated in the volatility process. Nelson's (1990) diffusion approximation theorem links these two approaches when the sampling interval is increasing finer. Duan (1997) extends the theorem to include a wide class of popular GARCH-type models.

In discrete time, Smetanina (2017) attempts to link these two approaches by proposing a hybrid model called the Real-time GARCH (RT-GARCH), which incorporate the current return innovation in the volatility process. Specifically,

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \psi\epsilon_t^2, \tag{1.1}$$

where $\epsilon_t \equiv r_t/\sigma_t$ are i.i.d. random variables symmetric around zero with the first two moments equal to 0 and 1, respectively and $\mathbb{E}\epsilon^4 < \infty$.[1] The process σ_t^2 is no longer the conditional variance of return process since it is not $\mathcal{F}_{t-1}-$measurable, where $\mathcal{F}_{t-1}$ is the $\sigma-$ algebra generated by $r_0, ..., r_{t-1}$. RT-GARCH is closely related to Breitung and Hafner (2016), who include the current return innovation in the log volatility process. Ding (2020) derives the diffusion limit of RT-GARCH and show the enlarged RT-GARCH converges weakly to a bivariate Ornstein-Uhlenbeck process with an auxiliary process.

The aim of RT-GARCH model is to make efficient use of all internal information (Smetanina, 2017). However, the volatility process under RT-GARCH has a constant conditional variance which casts doubt on the efficiency. Time-varying volatility of volatility has long been considered as an important risk factor. Corsi and Mittnik (2008) and Bollerslev et al. (2009) have noted the volatility clustering of realised volatility (RV) and incorporate a GARCH type specification in the conditional variance of RV. Moreover, the Chicago Board Options Exchange (CBOE) has been publishing the implied volatility of VIX index (VVIX) since 2012. VVIX index is essentially the risk neutral expectation of the volatility of volatility of S&P 500 index options. Park (2015) argues that VVIX is

[1]The zero third moment assumption is to ensure the return process is a martingale difference sequence and the existence of fourth moment is needed for covariance stationarity as well as valid forecast, see Smetanina (2017).

a better 'tail risk indicator' than other existing measures. It is therefore, important to incorporate the volatility of volatility in existing volatility models.

In this paper, we propose an RT-GARCH-type model that jointly models the volatility and volatility of volatility while retaining the simple QML estimation framework. We call this model augmented RT-GARCH model. In this model, the volatility process, σ_t^2, has a time-varying conditional variance which is a quadratic function of lagged volatility. This can be related to the asymptotic variance of RV which is proportional to the integrated quarticity (IQ), $\int_{t-1}^{t} \sigma_s^4 ds$. Since RV is a noisy estimate of σ_t^2, this specification of volatility of volatility can be viewed as an approximation of IQ.

The volatility forecast function under the augmented RT-GARCH follows a second-order difference equation in contrast to first-order difference equation under RT-GARCH and GARCH models. This comes from the feedback of volatility of volatility on the squared return. In the empirical studies, we show that the new model produces not only a better fit of standardised return residuals, but also more accurate out-of-sample volatility forecasts over GARCH and RT-GARCH models.

The remainder of the paper is structured as follows. In section 2 we introduce the augmented RT-GARCH model. In section 3 we provide some statistical properties of augmented RT-GARCH. Section 4 presents the empirical analysis including in-sample goodness-of-fit and out-of-sample forecasts. Section 5 concludes. All proofs are included in appendix A and additional figures are in appendix B.

2 Augmented RT-GARCH

2.1 Main model

We present the general model with leverage effects in the fashion of GJR-GARCH. Specifically, the joint process (r_t, σ_t^2) satisfies

$$r_t = \sigma_t \epsilon_t, \tag{2.1}$$

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \phi(r_{t-1}^-)^2 + (\psi_1 + \psi_2\sigma_{t-1}^2)\epsilon_t^2 + \eta(\epsilon_t^-)^2, \tag{2.2}$$

where ϵ_t satisfy the same conditions as in (1.1) and $x^- = \min(0, x)$. 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 strictly larger. Since the leverage effects come from both current and lagged negative returns. We call this full specification the augmented RT-GJR-GARCH with feedback and write as ART-GJR-GARCH-F(1,1,1). We call the model with leverage effect only from the current negative return, i.e., $\phi = 0$, ART-GJR-GARCH(1,1,1) and the symmetric model, ART-GARCH(1,1,1) with $\phi = \eta = 0$. The numbers inside the bracket correspond to the numbers of lags of σ_t^2 and r_t^2 included in the autoregressive and variance terms. (2.1) - (2.2) nest Smetanina's (2017) RT-GARCH by setting $\psi_2 = \phi = \eta = 0$, which nests GARCH model with $\psi_1 = 0$. In what follows we will call the three variants of (2.1) - (2.2) the class of ART-GARCH models.

The reasons for choosing this particular specification are as follows: First, it allows us to model the volatility and volatility of volatility simultaneously. To see this, we can express (2.2) as an AR(1) process with stochastic coefficient,

$$\sigma_t^2 = \Phi_0 + \Phi_{1,t-1}\sigma_{t-1}^2 + z_t, \tag{2.3}$$

where

$$\Phi_0 = \alpha + \psi_1 + \tfrac{1}{2}\eta, \tag{2.4}$$

$$\Phi_{1,t-1} = \beta + \psi_2 + \gamma\epsilon_{t-1}^2 + \phi(\epsilon_{t-1}^-)^2, \tag{2.5}$$

and

$$z_t = (\psi_1 + \psi_2\sigma_{t-1}^2)(\epsilon_t^2 - 1) + \eta\big((\epsilon_t^-)^2 - \tfrac{1}{2}\big) \tag{2.6}$$

is a martingale difference sequence (MDS) with conditional variance

$$\mathbb{E}[z_t^2|\mathcal{F}_{t-1}] = \kappa(\psi_1 + \psi_2\sigma_{t-1}^2)^2 + \kappa\eta(\psi_1 + \psi_2\sigma_{t-1}^2) + (\tfrac{1}{2}\kappa + \tfrac{1}{4})\eta^2, \tag{2.7}$$

where $\kappa = \mathbb{E}\epsilon_t^4 - 1$. By definition, the RHS of (2.7) is the conditional variance of σ_t^2 at time $t-1$. We call $\mathbb{E}[z_t^2|\mathcal{F}_{t-1}]$ the pseudo-latent variable in ART-GARCH models. This is because although stochastic, it is a quadratic function of the volatility and only one filter is needed to estimate both the volatility and volatility of volatility from the observed return process. Note the specification (2.3) is not final since $\Phi_{1,t}$ and σ_t^2 are not independent and thus, we can not forecast σ_{t+n}^2 for $n > 1$ directly. We present the final expression of σ_t^2 and r_t^2 as an ARMA process in section 3. Finally, Nelson (1992) argues the GARCH filter works in a similar way as RV for high frequency data. Since the asymptotic variance of RV is proportional to IQ (see for example Barndorff-Nielsen and Shephard (2003)), that is,

$$\frac{\sum_{k=1}^{[t/h]} r_{kh}^2 - \int_0^t \sigma_s^2 ds}{\sqrt{2h\int_0^t \sigma_s^4 ds}} \xrightarrow{d} N(0,1), \tag{2.8}$$

as $h \downarrow 0$ for $kh \le t < (k+1)h$, where $[x]$ denotes the largest integral part less than or equal to x. Since $\mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}]$ is a quadratic function of σ_{t-1}^2, (2.7) can be viewed as a polynomial approximation of IQ. We can formally test the hypothesis against the presence of conditional heteroskedasticity in the variance of volatility process, i.e. $H_0 : \psi_2 = 0$ against $H_1 : \psi_2 > 0$. This can be done once we derive the quasi-likelihood function of ART-GARCH from which we can construct the quasi-likelihood ratio (QLR) statistics.

To include multiple lags, we can consider the ART-GJR-GARCH-F(p,q,l):

$$\sigma_t^2 = \alpha + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 + \sum_{j=1}^{q}(\gamma_j + \phi_j I_{t-j})r_{t-j}^2 + \epsilon_t^2\sum_{j=1}^{l}\big(\psi_1 + \psi_{j+1}\sigma_{t-j}^2 + \eta I_t\big), \tag{2.9}$$

where $I_t = \mathbb{1}_{(r_t<0)}$ and $\mathbb{1}_{(\cdot)}$ is the indicator function. For the rest of the paper, we only consider the class of ART-GARCH(1,1,1) models.

2.2 Comparison to other volatility models

ART-GARCH models, similar to RT-GARCH, assign time-varying albeit different weights to past squared returns. Specifically, it can be shown that (2.2) can be approximately expressed as

$$\sigma_t^2 \approx \frac{\alpha}{1-\beta} + \frac{(a_{t-1}+\eta I_t)r_t^2}{b_{t-1}} + \sum_{j=1}^{\infty} \big(\beta^j \frac{a_{t-1-j}+\eta I_{t-j}}{b_{t-1-j}} + \beta^{j-1}(\gamma+\phi I_{t-j})\big) r_{t-j}^2, \tag{2.10}$$

using a first order Taylor expansion, where

$$a_{t-1} = \psi_1 + \psi_2 \sigma_{t-1}^2, \tag{2.11}$$

$$b_{t-1} = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \phi(r_{t-1}^-)^2, \tag{2.12}$$

The weights on past squared returns, $(a_{t-1-j}+\eta I_{t-j})/b_{t-1-j}$, are more flexible then those of RT-GARCH, $(\psi_1+\eta I_{t-j})/b_{t-1-j}$. To see how this flexibility affects the volatility process, we can regard b_{t-1-j} as the predictable part of σ_t^2 since it is $\mathcal{F}_{t-1-j}-$measurable while $a_{t-1-j}+\eta I_{t-j}$ can be seen as the uncertainty part since they are the coefficients of ϵ_{t-j}. Both parts are time-varying and depend on lagged volatility whereas in RT-GARCH the uncertainty part is a constant. The ratio of these terms can be interpreted as how surprising the new observation is relative to the predictable part. The weights in (2.10) are then the standard GARCH weights adjusted by these surprising factors.

We next consider the news impact curve as defined in Engle and Ng (1993). For ART-GARCH (2.1) - (2.2), the news impact curve is given by

$$\mathbb{E}[r_{t+1}^2|\mathcal{F}_t] = \bar{\alpha} + \tfrac{1}{2}\bar{\beta}\Big(\bar{b} + \sqrt{\bar{b}^2 + 4\bar{a}r_t^2 + 4\eta(r_t^-)^2}\Big) + \gamma r_t^2 + \phi(r_t^-)^2, \tag{2.13}$$

where we have taken $\epsilon_t \sim N(0,1)$, $\bar{\alpha} = \alpha + 3(\psi_1 + 1/2\eta_1)$, $\bar{\beta} = \beta + 3\psi_2$, $\bar{b} = \alpha + \beta\bar{\sigma}_2 + \gamma\bar{r}_2 + \phi\bar{r}_2^-$ and $\bar{a} = \psi_1 + \psi_2\bar{\sigma}_2$ with $\bar{\sigma}_2$, $\bar{r}_2$ and $\bar{r}_2^-$ being the unconditional levels of σ_t^2, r_t^2 and $(r_t^-)^2$ whose exact expressions are given in (3.5), (3.6) and (3.7) in section 3, respectively. To see how the conditional variance responds to different values of r_t in our model, we plot the news impact curves in Figure 1 for all variants of ART-GARCH models against the benchmark GARCH, GJR-GARCH and RT-GARCH models. In the upper panel, for small values of r_t, ART-GARCH models respond faster than RT-GARCH and GARCH models. While for large values of r_t, the responses are smaller for ART-GARCH models as seen in the lower panel of Figure 1. This is a desirable feature since we would like the volatility to respond quickly to 'normal' shocks but downweigh large abnormal shocks. As Harvey (2013) points out quadratic response does not fit heavy tail distributions since large shocks are fed substantially into the volatility update and can lead to a lack of

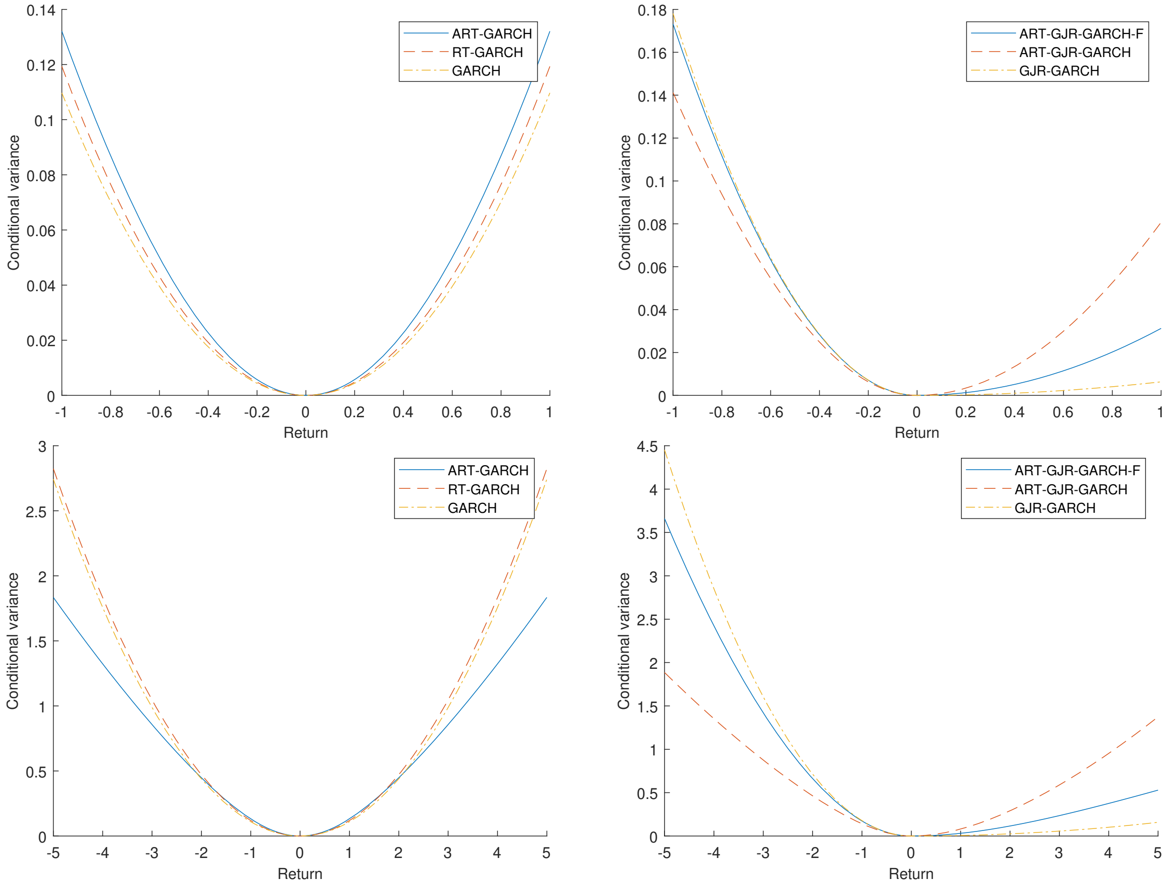


Figure 1: News impact curves for small and large values of r_t. All models' parameters are estimated from DJIA index daily returns which can be found in Table 1.

robustness. While still quadratic, the response in our model is substantially smaller for large values of r_t than RT-GARCH and GARCH models. The term $a_{t-1}\epsilon_t^2 + \eta(\epsilon_t^-)^2$ acts like a scaling factor to downweigh large shocks similar to using the score of conditional distribution in Harvey's (2013) DCS model. Note ART-GJR-GARCH model is less prone to negative shocks than other asymmetric models.

Finally, we compare our model to discrete time SV models. The ART-GARCH models, like RT-GARCH, are similar to the contemporaneous SV model of Taylor (1994). Both RT-GARCH and ART-GARCH share the same idea with Breitung and Hafner (2016). The main difference between them is ART-GARCH includes the current return innovation directly in the volatility process whereas Breitung and Hafner (2016) do so in the log volatility specification. SV models are generally more difficult to estimate especially when the volatility of volatility is also stochastic. ART-GARCH models, on the other hand, admit analytical form of quasi-likelihood function. Moreover, the conditional variance in SV models are not typically available in closed form. This makes comparative statistics, for example, news impact curve, complicated and the computation of the conditional variance requires numerical methods (Linton, 2019). ART-GARCH models, on the other hand, do not have these issues.

3 Properties of ART-GARCH

In this section, we present some statistical properties of ART-GARCH(1,1,1). We first state the assumption on return innovations ϵ_t.

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

Let $\mathcal{F}_t$ be the σ-algebra generated by $r_0, ..., r_t$. We now present the stationarity conditions for r_t and σ_t^2.

Theorem 3.1. *Let ϵ_t satisfy Assumption 1 and (r_t, σ_t^2) be generated by (2.1) and (2.2). Let $\theta = (\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)' \geq 0$ satisfy*

$$\mathbb{E} \log |\beta + (\gamma + \psi_2)\epsilon_0^2 + \phi(\epsilon_0^-)^2| < 0, \tag{3.1}$$

$$\mathbb{E}\big(\log |\alpha + \psi_1 \epsilon_0^2 + \eta(\epsilon_0^-)^2|\big)^+ < \infty, \tag{3.2}$$

where $(x)^+ = \max(x, 0)$ and ϵ_0 is the starting point of ϵ_t endowed with probability measure

$$\mathbb{P}((\sigma_0^2, \epsilon_0) \in \Gamma) = v_0(\Gamma) \text{ for any } \Gamma \in B(R^2), \tag{3.3}$$

where $B(R^2)$ is the Borel sets on R^2 and $v_0\big((\sigma_0^2, \epsilon_0) : 0 < \sigma_0^2 < \infty\big) = 1$. Then the joint process (r_t, σ_t^2) is strictly stationary.

Theorem 3.2. *Let ϵ_t satisfy Assumption 1 and (r_t, σ_t^2) be generated by (2.1) and (2.2). If $\theta = (\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)' \geq 0$ satisfy*

$$\beta + \psi_2 + \gamma + \tfrac{1}{2}\phi + \kappa\psi_2(\gamma + \tfrac{1}{2}\phi) < 1, \tag{3.4}$$

where $\kappa = \mathbb{E}\epsilon_t^4 - 1$, then the process (r_t, σ_t) is weakly stationary with unconditional second moment given by

$$\mathbb{E}\sigma_t^2 = \frac{\alpha + \psi_1 + \frac{1}{2}\eta + \frac{1}{4}\phi\eta\mathbb{E}\epsilon_t^4 + (\gamma + \frac{1}{2}\phi)(\psi_1 + \frac{1}{2}\eta)\kappa}{1 - \big(\beta + \psi_2 + \gamma + \frac{1}{2}\phi + \kappa\psi_2(\gamma + \frac{1}{2}\phi)\big)}, \tag{3.5}$$

and

$$\mathbb{E}r_t^2 = \frac{\alpha + (\psi_1 + \frac{1}{2}\eta + \frac{1}{4}\phi\eta)\mathbb{E}\epsilon_t^4 + \kappa\big(\alpha\psi_2 - \beta(\psi_1 + \frac{1}{2}\eta) + \frac{1}{4}\phi\eta\psi_2\mathbb{E}\epsilon_t^4\big)}{1 - \big(\beta + \psi_2 + \gamma + \frac{1}{2}\phi + \kappa\psi_2(\gamma + \frac{1}{2}\phi)\big)}. \tag{3.6}$$

Moreover,

$$\mathbb{E}(r_t^-)^2 = \tfrac{1}{2}\mathbb{E}r_t^2 + \tfrac{1}{4}\eta\mathbb{E}\epsilon_t^4. \tag{3.7}$$

Remark 1. *Since $\mathbb{P}(r_t < 0) = \mathbb{P}(r_t > 0) = 0.5$ because ϵ_t is symmetric around zero and $\sigma_t > 0$ a.s.. r_t has negative unconditional mean and skewness from (3.7).*

In order to analyse the kurtosis and fourth moment stationarity condition, we need a stronger assumption on ϵ_t.

Assumption 2. *In addition to Assumption* 1, *let* $\mathbb{E}\epsilon_t^8 < \infty$.

Theorem 3.3. *Let* ϵ_t *satisfy Assumption* 2 *and* (r_t, σ_t^2) *be generated by* (2.1) *and* (2.2). *Let* $\theta = (\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)' \geq 0$ *satisfy*

$$\xi_4 < 1, \tag{3.8}$$

$$\xi_1 + \xi_4 - \xi_1\xi_4 + \xi_2\xi_3 < 1, \tag{3.9}$$

where

$$\xi_1 = \Big(\beta^2 + \psi_2^2 u_4 + 2\beta\psi_2 + (\gamma^2 + \tfrac{1}{2}\phi^2 + \gamma\phi)\big(u_4 + \psi_2^2(u_8 - u_4^2) + 2\beta\psi_2(u_6 - u_4)\big)\Big), \tag{3.10}$$

$$\xi_2 = (2\gamma + \phi)\Big(\beta + \psi_2\big(1 + (\gamma^2 + \tfrac{1}{2}\phi^2 + \gamma\phi)(u_6 - u_4)\big)\Big), \tag{3.11}$$

$$\xi_3 = \Big(1 + \psi_2\big((u_6 - u_4)\psi_2 + 2\beta\kappa\big)\Big), \tag{3.12}$$

$$\xi_4 = \psi_2(2\gamma + \phi)\kappa, \tag{3.13}$$

with $u_n = \mathbb{E}\epsilon_t^n$ *for* $n \geq 1$ *and* $\kappa = u_4 - 1$. *Then the process* (r_t, σ_t) *is fourth moment stationary.*

Remark 2. *The exact expressions for* $\mathbb{E}\sigma_t^4$ *and* $\mathbb{E}r_t^4$ *are lengthy and can be found in the proof of Theorem* 3.3 *in appendix A. Since all parameters are restricted to be non-negative and ART-GARCH models nest RT-GARCH and GARCH models, it can be shown that* r_t *has excess unconditional kurtosis* > 0 *and* $\geq$ *those of GARCH and RT-GARCH.*

We next turn to the conditional properties of ART-GARCH models.

Theorem 3.4. *Let* ϵ_t *satisfy Assumption* 1 *and* (r_t, σ_t^2) *be generated by* (2.1) *and* (2.2). *Let* $\theta = (\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)' \geq 0$. *Then the transition density of the return process is given by*

$$f_r(y|\mathcal{F}_{t-1}) = \frac{y}{d_1(y, \sigma_{t-1}^2; \theta)d_2(y, \sigma_{t-1}^2; \theta)} f_\epsilon\big(d_2(y, \sigma_{t-1}^2; \theta)\big), \tag{3.14}$$

for $y \neq 0$, *where* $f_\epsilon(\cdot)$ *is the probability density function of* ϵ_t,

$$d_1(y, \sigma_{t-1}^2; \theta) = \sqrt{b_{t-1}^2 + 4a_{t-1}y^2 + 4\eta(y^-)^2}, \tag{3.15}$$

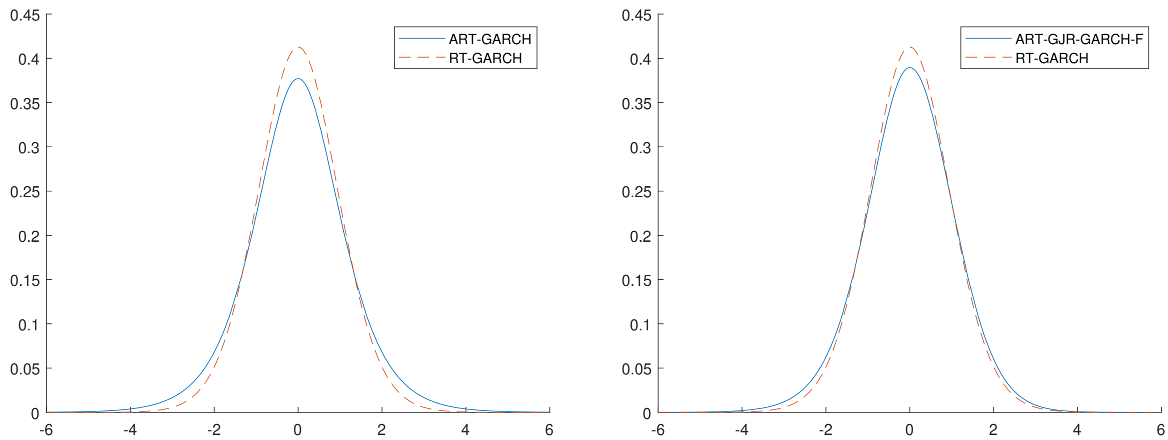
and

$$d_2(y, \sigma_{t-1}^2; \theta) = \begin{cases} \mathrm{sign}(y)\sqrt{\dfrac{d_1(y, \sigma_{t-1}^2; \theta) - b_{t-1}}{2a_{t-1} + 2\eta\mathbb{1}_{(y<0)}}}, & \text{for } (\psi_1, \psi_2, \eta)' \neq 0 \\ y/\sqrt{b_{t-1}}, & \text{for } (\psi_1, \psi_2, \eta)' = 0 \end{cases} \tag{3.16}$$

where a_{t-1} *and* b_{t-1} *are defined in* (2.11) *and* (2.12). *For* $y = 0$,

$$\lim_{y \to 0} f_r(y|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{b_{t-1}}} f_\epsilon(0). \tag{3.17}$$

Note at the true parameter vector θ_0, $\epsilon_t = d_2(r_t, \sigma_{t-1}^2; \theta_0)$. *The conditional cumulative distribution function of return process is given by* $F_r(y|\mathcal{F}_{t-1}) = F_\epsilon\big(d_2(y, \sigma_{t-1}^2; \theta)\big)$, *where* $F_\epsilon(\cdot)$ *is the cdf of* ϵ_t.



(a) Without leverages. (b) With leverages.

Figure 2: Transition densities of ART-GARCH and RT-GARCH models from unconditional levels. The parameters are estimated from S&P500 index daily returns and reported in Table 1.

Finally, the first-order approximation of the n-th conditional moment of y, for $n \in \mathbb{Z}^+$, *is given by*

$$\mathbb{E}[y^n|\mathcal{F}_{t-1}] \approx b_{t-1}^{n/2}\big(\mathbb{E}[d_2(y)^n] + \frac{na_{t-1}}{2b_{t-1}}\mathbb{E}[d_2(y)^{n+2}] + \frac{n\eta}{2b_{t-1}}\mathbb{E}[d_2(y)^{n+2}\mathbb{1}_{(y<0)}]\big), \tag{3.18}$$

where $\mathbb{E}[\cdot]$ *on the right hand side is taken w.r.t.* ϵ_t.

From (3.18), it is clear that r_t is not an MDS unless $\eta = 0$. Since

$$\mathbb{E}[r_t|\mathcal{F}_{t-1}] \approx \frac{\eta}{2\sqrt{b_{t-1}}}\mathbb{E}(\epsilon_t^-)^3, \tag{3.19}$$

which is clearly not zero. When ϵ_t are i.i.d. Gaussian, (3.19) becomes $-\eta/\sqrt{2\pi b_{t-1}}$. If both MDS and leverage effects are required, we can subtract (3.19) from (2.1) resulting in a variant of GARCH-in Mean or include only lagged leverage effect. The magnitude of (3.19) is assumed to be of smaller order, the same reason why GARCH models often ignore the mean of return series and it is the case from empirical estimates in section 5.[2] We plot the transition densities of ART-GARCH models against RT-GARCH in Figure 2.It is clear ART-GARCH is able to produce heavier-tails than RT-GARCH which already has heavier-tails compared to GARCH (Smetanina, 2017). Similar to stochastic volatility inducing heavy tails, stochastic volatility of volatility also contribute to the tails. The ART-GJR-GARCH-F is able to produce heavier left tail than RT-GARCH. It is also clear from Figure 2b that the magnitude of the conditional mean of ART-GJR-GARCH-F is close to zero.

If Assumption 2 is satisfied, then under ART-GJR-GARCH-F and ART-GJR-GARCH models, r_t has time-varying negative conditional skewness. Under all ART-GARCH models, r_t has time-varying excess conditional kurtosis. Since (3.19) is close to zero, we will

[2](3.19) is on average -0.034 compared to 1.312 for the unconditional second moment from S&P 500 estimates.

ignore the nonzero unconditional mean in order to keep the following expressions neat. The conditional kurtosis of ART-GARCH is given by

$$\frac{\mathbb{E}[r_t^4|\mathcal{F}_{t-1}]}{\mathbb{E}[r_t^2|\mathcal{F}_{t-1}]^2} = \frac{b_{t-1}^2\mathbb{E}\epsilon_t^4 + (2a_{t-1}+\eta)b_{t-1}\mathbb{E}\epsilon_t^6 + (a_{t-1}^2 + \frac{1}{2}\eta^2 + \eta a_{t-1})\mathbb{E}\epsilon_t^8}{\big(b_{t-1} + (a_{t-1} + \frac{1}{2}\eta)\mathbb{E}\epsilon_t^4\big)^2}, \tag{3.20}$$

and the first-order approximation of the conditional skewness is given by

$$\frac{\mathbb{E}[r_t^3|\mathcal{F}_{t-1}]}{\mathbb{E}[r_t^2|\mathcal{F}_{t-1}]^{3/2}} \approx \frac{\frac{3}{2}\eta\sqrt{b_{t-1}}\mathbb{E}(\epsilon_t^-)^5}{\big(b_{t-1} + (a_{t-1} + \frac{1}{2}\eta)\mathbb{E}\epsilon_t^4\big)^{3/2}}, \tag{3.21}$$

where a_{t-1} and b_{t-1} are defined in (2.11) and (2.12).

From (3.14) it is clear that all parameters of ART-GARCH models can be uniquely identified from the likelihood function. We next consider the asymptotic properties of the QML estimator based on Gaussian specification.

Theorem 3.5. *Let ϵ_t satisfy Assumption 1 and in addition, $\mathbb{E}\epsilon_t^6 < \infty$. Let (r_t, σ_t^2) be generated by (2.1) and (2.2). The QML estimator $\hat{\theta}$ of the true parameter θ_0 is given by $\hat{\theta} = \operatorname{argmax}_{\theta\in\Theta} L_T(\theta)$, where $L_t(\theta)$ is the quasi-log-likelihood function of r_t, that is,*

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \tag{3.22}$$

where

$$l_t(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}d_2(r_t, \sigma_{t-1};\theta)^2 + \log\frac{r_t}{d_1(r_t,\sigma_{t-1};\theta)d_2(r_t,\sigma_{t-1};\theta)}, \tag{3.23}$$

and $d_1(r_t,\sigma_{t-1};\theta)$, $d_2(r_t,\sigma_{t-1};\theta)$ are given in (3.15) and (3.16). Moreover, if $\theta_0 \in \Theta^{\mathrm{o}}$, where Θ^{o} is the interior of the parameter space Θ. Then,

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_{\theta_0}), \tag{3.24}$$

where $V_{\theta_0} = \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}$ with

$$\Sigma_1 = \mathbb{E}_{\theta_0}\Big[\frac{\partial l_t(\theta_0)}{\partial\theta_0}\frac{\partial l_t(\theta_0)}{\partial\theta_0'}\Big] \quad \textit{and} \quad \Sigma_2 = -\mathbb{E}_{\theta_0}\Big[\frac{\partial^2 l_t(\theta_0)}{\partial\theta_0\partial\theta_0'}\Big]. \tag{3.25}$$

Finally, provided a consistent estimator: $\hat{V}_{\hat{\theta}} \xrightarrow{p} V_{\theta_0}$,

$$\hat{V}_{\hat{\theta}}^{-1/2}\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0,1). \tag{3.26}$$

Theorem 3.5 also enables us to test the presence of conditional heteroskedasticity in the variance of volatility, i.e. $H_0 : \psi_2 = 0$ based on the QLR statistics. Given the non-negativity constraints on θ_0, the QLR statistics is on the boundary of the parameter space. The asymptotic distribution of the QLR statistics is therefore, nonstandard and requires corrections of the usual critical values as pointed out by Francq and Zakoïan (2009). Let $\hat{\theta}_{-1}$ be the restricted (by H_0) estimator of θ_0, the modified QLR statistics is given by $-2\big(L_T(\hat{\theta}_{-1}) - L_T(\hat{\theta})\big)/\hat{\kappa}$, where $\hat{\kappa}$ is a consistent estimator of κ. The modified asymptotic level is $c/2$ for a nominal asymptotic level of c under one restriction. That is, we reject the null at an asymptotic level c if the QLR statistics is larger than $\chi^2_{1,1-2c}$. See Francq and Zakoïan (2009) for detailed discussions.

We next discuss volatility forecasts under ART-GARCH models. Unlike GARCH models, in both RT-GARCH and ART-GARCH models there are two concepts of volatility: instantaneous volatility σ_t^2 and conditional variance $\mathrm{Var}[r_t|\mathcal{F}_{t-1}]$. Note in the case when $\eta \neq 0$, $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}]$ is no longer the conditional variance because r_t is no longer an MDS. However, since $\mathbb{E}[r_t|\mathcal{F}_{t-1}]$ is close to zero by (3.19), $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}]$ is approximately equal to $\mathrm{Var}[r_t|\mathcal{F}_{t-1}]$. Moreover, $\mathbb{E}[r_{t+n}|\mathcal{F}_t]$ do not have a closed form for $n > 1$. Therefore, with a little abuse of terminologies, we will regard $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}]$ as the conditional variance and call σ_t^2 simply the volatility. Readers should bear in mind that the true conditional variance is $\mathbb{E}[\big(r_{t+n} - \mathbb{E}[r_{t+n}|\mathcal{F}_t]\big)^2|\mathcal{F}_t] \approx \mathbb{E}[r_{t+n}^2|\mathcal{F}_t]$ for all $n \geq 1$. The volatility filtering equation is given by

$$\sigma_t^2 = \tfrac{1}{2}b_{t-1} + \tfrac{1}{2}\sqrt{b_{t-1}^2 + 4a_{t-1}r_t^2 + 4\eta(r_t^-)^2}. \tag{3.27}$$

The one-step volatility forecast is given by

$$\mathbb{E}[\sigma_{t+1}^2|\mathcal{F}_t] = \alpha + \psi_1 + \tfrac{1}{2}\eta + (\beta + \psi_2)\sigma_t^2 + \gamma r_t^2 + \phi(r_t^-)^2, \tag{3.28}$$

and the one-step conditional variance forecast is given by

$$\mathbb{E}[r_{t+1}^2|\mathcal{F}_t] = \alpha + (\psi_1 + \tfrac{1}{2}\eta)\mathbb{E}\epsilon_t^4 + \big(\beta + \psi_2\mathbb{E}\epsilon_t^4\big)\sigma_t^2 + \gamma r_t^2 + \phi(r_t^-)^2. \tag{3.29}$$

The one-step ahead forecast equations are similar to those of RT-GARCH except for the autoregressive parameter which takes into account the feedback from the volatility of volatility. The forecast equations start to differ towards multi-step ahead predictions. Specifically, the two-step ahead forecast function for volatility is given by

$$\mathbb{E}[\sigma_{t+2}^2|\mathcal{F}_t] = \alpha + \psi_1 + \tfrac{1}{2}\eta + \tfrac{1}{4}\phi\eta\mathbb{E}\epsilon_t^4 + (\beta + \psi_2)\mathbb{E}[\sigma_{t+1}^2|\mathcal{F}_t] + (\gamma + \tfrac{1}{2}\phi)\mathbb{E}[r_{t+1}^2|\mathcal{F}_t], \tag{3.30}$$

and for conditional variance,

$$\mathbb{E}[r_{t+2}^2|\mathcal{F}_t] = \alpha + (\psi_1 + \tfrac{1}{2}\eta + \tfrac{1}{4}\phi\eta)\mathbb{E}\epsilon_t^4 + \big(\beta + \psi_2\mathbb{E}\epsilon_t^4\big)\mathbb{E}[\sigma_{t+1}^2|\mathcal{F}_t] + (\gamma + \tfrac{1}{2}\phi)\mathbb{E}[r_{t+1}^2|\mathcal{F}_t]. \tag{3.31}$$

Finally, the multi-period forecasts are given in the following theorem:

Theorem 3.6. *Let ϵ_t satisfy Assumption 1 and (r_t, σ_t^2) be generated by (2.1) and (2.2). Then for any $n \geq 3, n \in \mathbb{Z}^+$, the n-step volatility forecast is given by*

$$\mathbb{E}[\sigma_{t+n}^2|\mathcal{F}_t] = \mathbb{E}\sigma_t^2 + \Phi_1(\mathbb{E}[\sigma_{t+n-1}^2|\mathcal{F}_t] - \mathbb{E}\sigma_t^2) + \Phi_2(\mathbb{E}[\sigma_{t+n-2}^2|\mathcal{F}_t] - \mathbb{E}\sigma_t^2), \tag{3.32}$$

where $\mathbb{E}\sigma_t^2$ is given in (3.5), $\Phi_1 = \beta + \gamma + \psi_2 + \frac{1}{2}\phi$ and $\Phi_2 = \kappa\psi_2(\gamma + \frac{1}{2}\phi)$ with $\kappa = \mathbb{E}\epsilon_t^4 - 1$. The initial conditions for (3.32), $\mathbb{E}[\sigma_{t+2}^2|\mathcal{F}_t]$ and $\mathbb{E}[\sigma_{t+1}^2|\mathcal{F}_t]$, are given in (3.28) and (3.30).

The n-step conditional variance forecast is given by

$$\mathbb{E}[r_{t+n}^2|\mathcal{F}_t] = \mathbb{E}r_t^2 + \Phi_1(\mathbb{E}[r_{t+n-1}^2|\mathcal{F}_t] - \mathbb{E}r_t^2) + \Phi_2(\mathbb{E}[r_{t+n-2}^2|\mathcal{F}_t] - \mathbb{E}r_t^2), \tag{3.33}$$

where $\mathbb{E}r_t^2$ is given in (3.6). The initial conditions for (3.33), $\mathbb{E}[r_{t+2}^2|\mathcal{F}_t]$ and $\mathbb{E}[r_{t+1}^2|\mathcal{F}_t]$, are given in (3.29) and (3.31).

Remark 3. *Which volatility to use depends on the purposes and what volatility proxy is used for evaluation. For example, realised measures (RM) are frequently used as volatility proxies and they are consistent estimators of the integrated volatility (IV), $\int_{t-1}^{t} \sigma_s^2 ds$. In this case, instantaneous volatility should be used. On the other hand, if we are interested in the overall fluctuations of future returns, the conditional variance should provide more insights.*

Theorem 3.6 states the forecast function of ART-GARCH models follows a second-order difference equation with two persistence parameters Φ_1 and Φ_2. In contrast, under both RT-GARCH and standard GARCH models the forecast functions are first-order difference equations. This is because in ART-GARCH models, there is a feedback from volatility of volatility on the squared return. This is the reason why we say the AR(1) with stochastic coefficient representation in (2.3) is not final. Indeed, for the symmetric case, both (r_t^2, σ_t^2) can be expressed as a VARMA(2,1) process. Specifically,

$$\begin{aligned}\begin{bmatrix} r_t^2 \\ \sigma_t^2 \end{bmatrix} &= \begin{bmatrix} \alpha + \psi_1 \mathbb{E}\epsilon_t^4 + \kappa(\alpha\psi_2 - \beta\psi_1) \\ \alpha + \psi_1 + \kappa\gamma\psi_1 \end{bmatrix} + (\beta + \gamma + \psi_2) \begin{bmatrix} r_{t-1}^2 \\ \sigma_{t-1}^2 \end{bmatrix} \\ &+ \kappa\gamma\psi_2 \begin{bmatrix} r_{t-2}^2 \\ \sigma_{t-2}^2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ e_t \end{bmatrix} + \begin{bmatrix} 0 & \gamma \\ \kappa\psi_2 & -\beta - \psi_2 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ e_{t-1} \end{bmatrix},\end{aligned} \tag{3.34}$$

where z_t is defined in (2.6) and

$$e_t = r_t^2 - \sigma_t^2 - \kappa(\psi_1 + \psi_2\sigma_{t-1}^2) \tag{3.35}$$

is an MDS. See appendix A for the derivation of (3.34).

4 Empirical Analysis

In this section we present some empirical results using daily open-to-close returns of S&P 500 and Dow Jones Industrial Average (DJIA) indices, JPMorgan Chase (JPM) and Apple Inc. (AAPL) stock prices and EUR/USD exchange rate. Our purpose is not to select the best model for volatility modelling but to compared the new model with three benchmark models: GARCH, GJR-GARCH and RT-GARCH. We use QQ plot compare the goodness-of-fit. Subsequently we compare the filtered volatility. We also report the filtered volatility of volatility. Finally, we compare the out-of-sample 1-, 2-, 5-, 10- and 15-step volatility forecasts. We use mean squared error (MSE) which is a robust criterion in the sense of Patton (2011) for forecast comparison.[3] We use both RV and bipower variation (BPV) as proxies for the true volatility and RQ for the volatility of volatility.

[3] A robust criterion consistently ranks forecast performance using a variety of volatility proxies as long as they are consistent estimators of volatility itself.

Table 1: Parameter estimates of the ART-GARCH models

	α	β	γ	ψ_1	ψ_2	η	ϕ	BIC	QLR
S&P 500	0.0000	0.8856	0.0000	0.0022	0.0212	0.0423	0.1257	14214	54.54
	(0.0014)	(0.0086)	(0.0016)	(0.0017)	(0.0076)	(0.0063)	(0.0155)		
	0.0000	0.8807	0.0332	0.0000	0.0454	0.0593	-	14315	58.80
	(0.0019)	(0.0100)	(0.0108)	(0.0018)	(0.0108)	(0.0082)			
	0.0000	0.8810	0.0201	0.0097	0.0872	-	-	14424	37.33
	(0.0024)	(0.0100)	(0.0112)	(0.0033)	(0.0136)				
DJIA	0.0000	0.8886	0.0000	0.0027	0.0289	0.0358	0.1112	13956	44.59
	(0.0009)	(0.0082)	(0.0009)	(0.0013)	(0.0068)	(0.0057)	(0.0137)		
	0.0000	0.8861	0.0245	0.0000	0.0528	0.0526	-	14036	53.54
	(0.0009)	(0.0085)	(0.0094)	(0.0010)	(0.0104)	(0.0073)			
	0.0000	0.8807	0.0166	0.0098	0.0905	-	-	14135	38.51
	(0.0015)	(0.0098)	(0.0110)	(0.0025)	(0.0121)				
JPM	0.0117	0.8885	0.0000	0.0000	0.0762	0.0212	0.0516	19936	11.72
	(0.0093)	(0.0127)	(0.0100)	(0.0097)	(0.0113)	(0.0125)	(0.0125)		
	0.0000	0.9051	0.0000	0.0000	0.0872	0.0434	-	19951	7.80
	(0.0090)	(0.0109)	(0.0087)	(0.0101)	(0.0110)	(0.0121)			
	0.0145	0.8919	0.0000	0.0000	0.1052	-	-	19958	78.30
	(0.0089)	(0.0113)	(0.0087)	(0.0098)	(0.0119)				
AAPL	0.0000	0.9118	0.0083	0.0054	0.0633	0.0420	0.0188	22241	1.56
	(0.0069)	(0.0092)	(0.0068)	(0.0108)	(0.0094)	(0.0181)	(0.0110)		
	0.0000	0.9148	0.0109	0.0000	0.0670	0.0542	-	22235	6.03
	(0.0061)	(0.0091)	(0.0073)	(0.0062)	(0.0091)	(0.0158)			
	0.0000	0.9143	0.0114	0.0203	0.0691	-	-	22239	41.94
	(0.0093)	(0.0096)	(0.0075)	(0.0112)	(0.0101)				
EUR/USD	0.0000	0.9348	0.0000	0.0005	0.0619	0.0015	0.0000	4684.4	0.00
	(0.0023)	(0.0144)	(0.0090)	(0.0028)	(0.0187)	(0.0017)	(0.0073)		
	0.0000	0.9348	0.0000	0.0005	0.0619	0.0015	-	4676.3	0.44
	(0.0019)	(0.0162)	(0.0066)	(0.0023)	(0.0155)	(0.0016)			
	0.0000	0.9340	0.0000	0.0013	0.0626	-	-	4669.0	44.79
	(0.0018)	(0.0170)	(0.0092)	(0.0021)	(0.0152)				

Note: The first line refers to ART-GJR-GARCH-F, followed by ART-GJR-GARCH and ART-GARCH. The standard errors in parentheses are calculated numerically. The last two columns are the BIC and likelihood ratio statistics which compare each model to the subsequent model and ART-GARCH to RT-GARCH. The modified critical value for QLR test is 2.706 at 5% asymptotic level for one restriction. The BIC for RT-GARCH is 14490 for S&P 500, 14203 for DJIA, 20106 for JPM, 22314 for AAPL and 4750.5 for EUR/USD.

4.1 Data description

Our sample spans from 2 January 1998 to 31 December 2019 and for EUR/USD exchange rate from 2 January 2009 to 31 December 2019. The daily data are obtained from Yahoo! Finance and their 1-min intraday high frequency data are from FirstRate Data LLC. For out-of-sample forecast evaluation we divide the sample into estimation and forecast periods. The out-of-sample period contains the last 1500 observations. We use an expansion window for estimation and update for every 50 observations. For the calculations of RV and BPV, we use 5-min intraday returns since ultra high frequency data are typically contaminated by market microstructure noise (see for example, Hansen and Lunde (2006)). Other methods to consistently estimate IV using high frequency data include Zhang et al. (2005), Barndorff-Nielsen et al. (2008), Kalnina and Linton (2008) and Podolskij and Vetter (2009) among others.

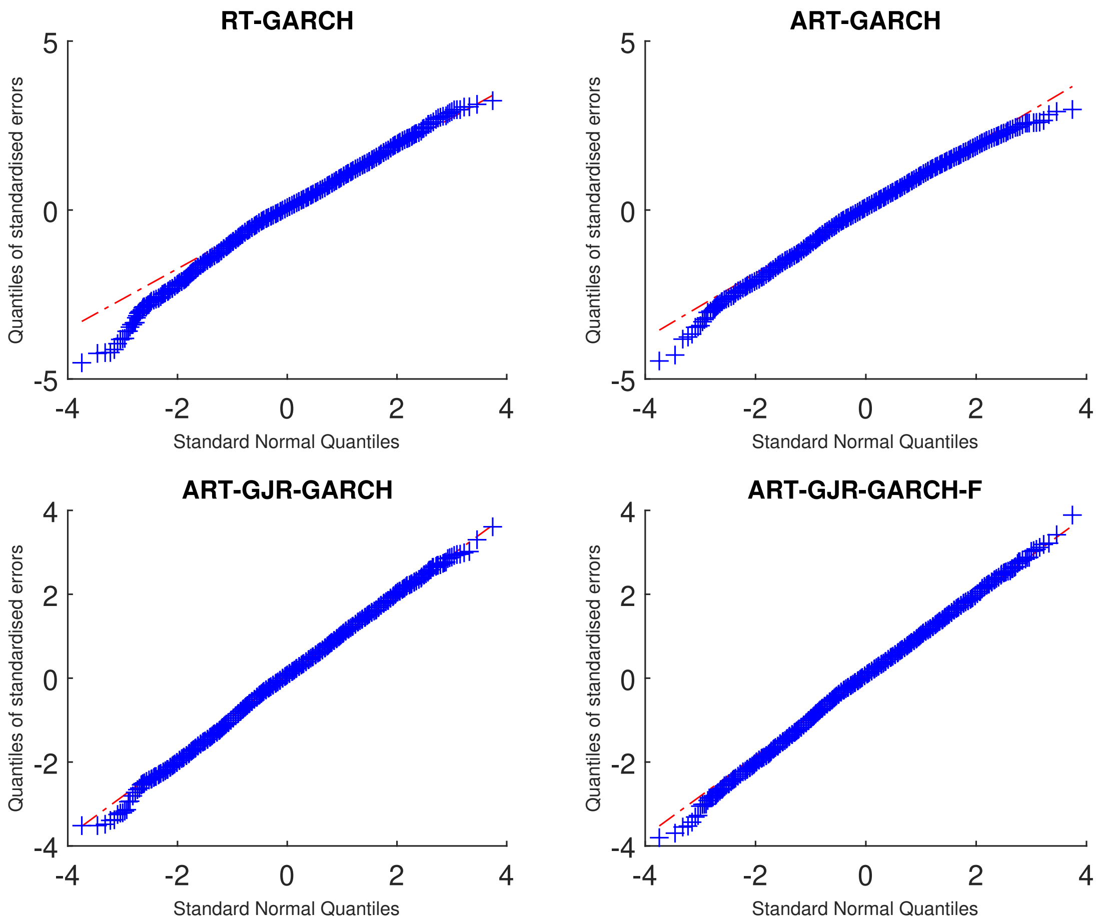


Figure 3: QQ plots of the standardised returns for S&P 500 index

4.2 Full sample analysis

The parameter estimates for full sample are reported in Table 1. We have imposed the covariance stationarity condition (3.4) to penalise overfitting. α is generally insignificant and close to zero for all ART-GARCH models. In terms of lowest BIC, ART-GARCH models are always selected over the benchmark models. Moreover, ART-GJR-GARCH-F is preferred for S&P 500, DJIA and JPM while ART-GJR-GARCH is preferred for AAPL and for EUR/USD, ART-GARCH. The QLR statistics suggests the hypothesis, $H_0 : \psi_2 = 0$, is rejected in all cases. Thus, there is strong evidence suggesting the presence of conditional heteroskedasticity in the variance of volatility. Moreover, all three ART-GARCH models are distinguishable from each other except for EUR/USD in which case we fail to reject the hypothesis $H_0 : \phi = \eta = 0$ and for AAPL where we fail to reject $H_0 : \eta = 0$. Note the standard errors are only indicative since clearly one or more coefficients are located on the boundary of parameter space. We refer to Francq and Zakoïan (2009) for the asymptotic variance of QMLE for the boundary case.

We show the QQ plots of standardised residuals under ART-GARCH and RT-GARCH for S&P 500 index in Figure 3 (for other assets see appendix B). All three ART-GARCH models have significant better goodness-of-fit over RT-GARCH, especially in the left tail.

Table 2: Volatility filtering comparison using RV as volatility proxy

	S&P 500		DJIA		JPM		AAPL		EURUSD	
	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
GARCH	1.1825	0.001	1.2002	0.001	4.2540	0.000	4.4926	0.002	0.0646	0.001
GJR-GARCH	1.2448	0.006	1.2399	0.004	4.1033	0.000	4.4658	0.002	0.0641	0.001
RT-GARCH	1.0434	0.006	1.0706	0.004	3.9953	0.002	4.3046	0.003	0.0651	0.000
ART-GARCH	0.9371	0.006	0.9853	0.004	3.9159	0.004	3.9891	0.003	0.0589	0.115*
ART-GJR-GARCH	0.8598	1.000*	0.9248	1.000*	3.8449	0.237*	3.9530	0.906*	0.0588	1.000*
ART-GJR-GARCH-F	0.9794	0.006	1.0047	0.004	3.7512	1.000*	3.9518	1.000*	0.0588	0.115*

Note: p_{MCS} are the p-values of Model Confidence Set of Hansen et al. (2011). The models marked with $*$ fall in the model confidence set $\hat{\mathcal{M}}^*_{95\%}$.

This is particularly true for ART-GJR-GARCH and ART-GJR-GARCH-F whose quantiles are almost identical to those of standard normal. This provides justification for better volatility filter and forecasts of ART-GARCH models. The improvement holds for all the other assets except for EUR/USD. As seen in Figure 8, all models have almost the same goodness-of-fit for EUR/USD and none of them results in a better fit of the tails.

The MSE of filtered volatility and the 95 percentile model confidence set (MSC), $\hat{\mathcal{M}}^*_{95\%}$, as defined in Hansen et al. (2011) are reported in Table 2.[4] We use the last 2514 observations and for EUR/USD, last 3119 observations (exchange rate is traded over-the-counter and on Sundays) spanning from 04 January 2010 to 31 December 2019 for evaluation. The better goodness-of-fit directly results in smaller MSE for all ART-GARCH models over the benchmark models. Moreover, only the ART-GARCH models fall in the $\hat{\mathcal{M}}^*_{95\%}$. For S&P 500, DJIA and EUR/USD, ART-GJR-GARCH is the best model, while for JPM and AAPL, ART-GJR-GARCH-F has the smallest MSE. Volatility filtering is still an important tool for ex-post volatility estimation since other consistent ex-post estimators like RV or in general, RM, rely on the availability of high frequency data. For relatively new and exotic products, for example inflation-linked bonds and emerging market currencies, trading activities are still infrequent and thus, high frequency data are not always available.[5] In such situation, we can only rely on filtering techniques to estimate volatility from low frequency data.

The ART-GARCH models also contain information about $\mathrm{Var}[\sigma_t^2|\mathcal{F}_{t-1}]$. It can be filtered out according to (2.7) once we obtain the filtered volatility process itself. We plot the filtered volatility of volatility of S&P 500 index returns in Figure 4 against the rescaled RQ. Recall from Barndorff-Nielsen and Shephard (2003), RV has asymptotic distribution:

$$\frac{\sum_{k=1}^{[t/h]} r_{kh}^2 - \int_0^t \sigma_s^2 ds}{\sqrt{\frac{2}{3}\sum_{k=1}^{[t/h]} r_{kh}^4}} \xrightarrow{d} N(0,1), \tag{4.1}$$

as $h \downarrow 0$, where $[x]$ denotes the largest integer part less than or equal to x. The de-

[4] The 95 percentile model confidence set, $\hat{\mathcal{M}}^*_{95\%}$, is the set that contains the best models with probability 0.95 in terms of loss functions, see Hansen et al. (2011) for more details.

[5] From the author's own experience at fixed income trading desk.

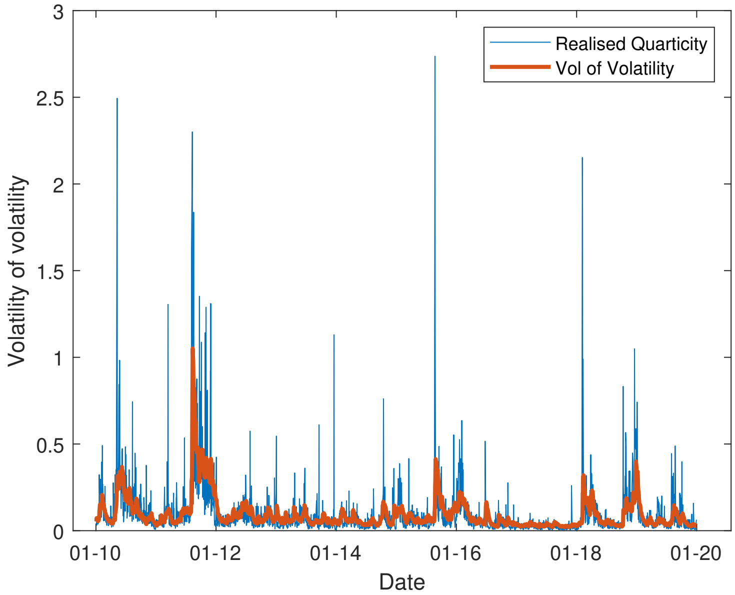


Figure 4: Filtered volatility of volatility and rescaled RQ of S&P 500 index.

nominator of the LHS of (4.1), rescaled RQ, is a consistent estimator of IQ which is the asymptotic variance of RV. The ideal estimator for the volatility of volatility would be the realised volatility of RV. However, for such estimator one would require ultra high frequency data to calculate the RV of each intraday squared return. Such dataset would not only be of severely limited availability but also subject to large microstructure noises (see Zhang et al. (2005)). As a result, we use the asymptotic theory for RV, (4.1), to evaluate the filtered volatility of volatility. The filtered volatility of volatility path is generally in-line with rescaled RQ. Figures for the volatility of volatility of DJIA, JPM, AAPL and EUR/USD can be found in appendix B.

4.3 Out-of-sample volatility forecasts comparison

Finally, we compare the out-of-sample 1-, 2-, 5-, 10- and 15-step volatility forecasts. From Table 3, the three ART-GARCH models consistently outperform the benchmark models. For S&P 500 and DJIA indices, ART-GJR-GARCH dominates 1-, 2-, 5-step volatility forecasts in terms of smallest MSE. For 10- and 15-step forecasts, ART-GARCH is preferred for S&P 500 while ART-GJR-GARCH is still the best model for DJIA. In terms of MCS, ART-GARCH and ART-GJR-GARCH always fall in the 95% MCS while GARCH and GJR-GARCH are always outside. For JPM and AAPL stock returns, ART-GJR-GARCH-F has the smallest MSE for 1-step forecast, ART-GJR-GARCH outperforms the others from 2-step forecast onward except for AAPL, where ART-GARCH performs the best for 10- and 15-step forecasts. Similarly, ART-GARCH and ART-GJR-GARCH are always in the 95% MCS. ART-GJR-GARCH-F is also in the 95% MCS for all except 15-step forecast. For EUR/USD exchange rate, the differences across all models are very small. However, all three ART-GARCH models are still in the 95% MCS and consistently

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

	S&P 500		DJIA		JPM		AAPL		EURUSD	
	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
1-step										
GARCH	0.3958	0.001	0.7569	0.005	2.2491	0.059*	4.6462	0.020	0.0742	1.000*
GJR-GARCH	0.4005	0.001	0.7623	0.005	2.2477	0.119*	4.5942	0.042	0.0756	0.166*
RT-GARCH	0.3595	0.049	0.7070	0.074*	2.2072	0.637*	4.5992	0.059*	0.0755	0.166*
ART-GARCH	0.3461	0.160*	0.7024	0.113*	2.2581	0.070*	4.4424	0.059*	0.0745	0.711*
ART-GJR-GARCH	0.3211	1.000*	0.6660	1.000*	2.2262	0.255*	4.4111	0.267*	0.0749	0.166*
ART-GJR-GARCH-F	0.3465	0.318*	0.6816	0.630*	2.1916	1.000*	4.3965	1.000*	0.0749	0.166*
2-step										
GARCH	0.4643	0.002	0.8523	0.011	2.3930	0.191*	4.8988	0.028	0.0763	1.000*
GJR-GARCH	0.5008	0.001	0.9021	0.003	2.4217	0.173*	4.8815	0.041	0.0772	0.726*
RT-GARCH	0.4301	0.017	0.8029	0.031	2.4022	0.191*	4.9255	0.041	0.0786	0.025
ART-GARCH	0.3963	0.217*	0.7687	0.091*	2.3664	0.191*	4.6783	0.202*	0.0768	0.726*
ART-GJR-GARCH	0.3831	1.000*	0.7446	1.000*	2.3488	1.000*	4.6641	1.000*	0.0771	0.204*
ART-GJR-GARCH-F	0.4431	0.062*	0.8073	0.074*	2.3593	0.569*	4.6725	0.476*	0.0771	0.204*
5-step										
GARCH	0.5675	0.001	0.9657	0.008	2.5436	0.075*	5.1718	0.014	0.0759	0.968*
GJR-GARCH	0.6343	0.001	1.0493	0.002	2.5726	0.065*	5.1638	0.014	0.0767	0.199*
RT-GARCH	0.5294	0.002	0.9054	0.018	2.5873	0.065*	5.2670	0.013	0.0779	0.066*
ART-GARCH	0.4755	0.945*	0.8519	0.274*	2.4669	0.247*	4.9021	0.525*	0.0759	1.000*
ART-GJR-GARCH	0.4749	1.000*	0.8365	1.000*	2.4574	1.000*	4.8952	1.000*	0.0762	0.604*
ART-GJR-GARCH-F	0.5714	0.002	0.9380	0.018	2.4900	0.233*	4.9149	0.254*	0.0762	0.199*
10-step										
GARCH	0.6171	0.006	1.0128	0.005	2.6216	0.023	5.4353	0.021	0.0691	1.000*
GJR-GARCH	0.6832	0.006	1.1039	0.005	2.6644	0.006	5.4470	0.021	0.0701	0.565*
RT-GARCH	0.5647	0.006	0.9297	0.020	2.6744	0.023	5.6242	0.006	0.0718	0.038
ART-GARCH	0.5125	1.000*	0.8862	0.379*	2.5078	0.122*	5.1173	1.000*	0.0691	0.953*
ART-GJR-GARCH	0.5175	0.804*	0.8699	1.000*	2.4936	1.000*	5.1258	0.416*	0.0694	0.565*
ART-GJR-GARCH-F	0.6219	0.006	0.9822	0.016	2.5440	0.117*	5.1589	0.103*	0.0694	0.565*
15-step										
GARCH	0.6588	0.001	1.0479	0.003	2.7354	0.002	5.7037	0.028	0.0698	1.000*
GJR-GARCH	0.7308	0.001	1.1515	0.003	2.8084	0.001	5.7287	0.028	0.0710	0.268*
RT-GARCH	0.5963	0.001	0.9485	0.017	2.8462	0.001	6.0316	0.000	0.0731	0.036
ART-GARCH	0.5410	1.000*	0.9105	0.561*	2.5689	0.310*	5.3624	1.000*	0.0698	0.942*
ART-GJR-GARCH	0.5555	0.528*	0.8963	1.000*	2.5587	1.000*	5.3747	0.476*	0.0701	0.268*
ART-GJR-GARCH-F	0.6751	0.001	1.0221	0.003	2.6550	0.007	5.4251	0.036	0.0701	0.617*

Note: p_{MCS} are the p-values of Model Confidence Set of Hansen et al. (2011). The models marked with $*$ fall in the model confidence set $\hat{\mathcal{M}}^*_{95\%}$.

outperform RT-GARCH. The best performing model alternates between GARCH and ART-GARCH. This is consistent with Figure 8 that asymmetric models do not result in better goodness-of-fit and leverage effect parameters are insignificant from Table 1. We also use another robust loss function, QLIKE, for the forecast evaluations and obtain similar rankings. For reasons of brevity, we do not report the QLIKE results here. Overall, ART-GARCH models consistently outperform RT-GARCH, GJR-GARCH and GARCH models.

Since volatility forecast performance depends partially on the choice of volatility proxy, we next use BPV to assess the forecast performance. In the absence of jumps and microstructure noises, both RV and BPV are consistent estimators of IV. However, if jumps are present, RV is not robust since

$$\sum_{k=1}^{[t/h]} r_{kh}^2 \xrightarrow{p} \int_0^t \sigma_s^2 ds + \sum_{i=1}^{N_t} J_i^2, \tag{4.2}$$

as $h \downarrow 0$ for $kh \leq t < (k+1)h$, where (N_t) is a finite activity counting process and J_i are nonzero random variables that represent the infrequent jumps in the price process. Various studies have confirmed empirically that jumps are present in asset prices (see for example, Barndorff-Nielsen and Shephard (2006), Bollerslev et al. (2007), Aït-Sahalia and Jacod (2009), among others). Often, we are interested in the continuous part of volatility for reasons including risk management and portfolio allocation purposes. All GARCH-type and SV models are designed to model the continuous part of volatility. However, using RV to evaluate the forecast performance may result in inconsistency due to the presence of jumps. Recall from Barndorff-Nielsen and Shephard (2004),

$$\frac{\pi}{2} \sum_{k=1}^{[t/h]-1} |r_{kh}||r_{(k+1)h}| \xrightarrow{p} \int_0^t \sigma_s^2 ds, \tag{4.3}$$

as $h \downarrow 0$ for $kh \leq t < (k+1)h$. We compute the rescaled BPV according to the LHS of (4.3) and report the forecast evaluation in Table 4.

The rankings are similar to those using RV as volatility proxy for all assets except EUR/USD where ART-GARCH models now have the best forecast performance by a clear margin. Moreover, the reductions of MSE are more profound for ART-GARCH models than for the benchmark models across all assets. This suggests that ART-GARCH models and are more robust to jumps. This is evident from the heavy-tails of ART-GARCH models in the conditional density. The differences in MSE between RV and BPV as volatility proxies are more profound for stocks and exchange rates than indices. This is intuitive since individual stocks have jumps associated with both market conditions and idiosyncratic characteristics, for example, earning announcements (see Maheu and Mccurdy (2004)). On the other hand, exchange rates are affected by economic fundamentals and central banks' announcements of both sides, adding more potentials for jump events.

Table 4: Out-of-sample volatility forecasts comparison using BPV as volatility proxy

	S&P 500		DJIA		JPM		AAPL		EURUSD	
	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}	MSE	p_{MCS}
1-step										
GARCH	0.4082	0.001	0.7113	0.002	1.9901	0.002	4.8117	0.013	0.0575	0.111*
GJR-GARCH	0.4222	0.001	0.7331	0.002	2.0117	0.002	4.7627	0.029	0.0589	0.061*
RT-GARCH	0.3695	0.032	0.6579	0.062*	1.9593	0.149*	4.7801	0.029	0.0574	0.222*
ART-GARCH	0.3494	0.240*	0.6412	0.169*	1.9363	0.149*	4.5532	0.080*	0.0561	1.000*
ART-GJR-GARCH	0.3282	1.000*	0.6106	1.000*	1.9134	0.669*	4.5249	0.448*	0.0563	0.222*
ART-GJR-GARCH-F	0.3629	0.240*	0.6420	0.294*	1.9041	1.000*	4.5143	1.000*	0.0563	0.222*
2-step										
GARCH	0.4762	0.002	0.8007	0.022	2.1284	0.026	5.0634	0.016	0.0596	0.295*
GJR-GARCH	0.5233	0.002	0.8661	0.001	2.1777	0.000	5.0501	0.021	0.0607	0.295*
RT-GARCH	0.4392	0.017	0.7474	0.022	2.1487	0.026	5.1107	0.024	0.0606	0.003
ART-GARCH	0.3991	0.392*	0.7036	0.239*	2.0416	0.000	4.7899	0.389*	0.0585	1.000*
ART-GJR-GARCH	0.3901	1.000*	0.6845	1.000*	2.0325	1.000*	4.7794	1.000*	0.0586	0.374*
ART-GJR-GARCH-F	0.4606	0.017	0.7616	0.189*	2.0660	0.390*	4.7922	0.389*	0.0586	0.374*
5-step										
GARCH	0.5760	0.001	0.9075	0.006	2.2912	0.000	5.3290	0.011	0.0596	0.139*
GJR-GARCH	0.6494	0.001	1.0003	0.004	2.3434	0.000	5.3276	0.011	0.0606	0.139*
RT-GARCH	0.5352	0.001	0.8426	0.010	2.3586	0.000	5.4536	0.009	0.0605	0.007
ART-GARCH	0.4754	1.000*	0.7814	0.369*	2.1511	1.000*	5.0085	0.907*	0.0581	1.000*
ART-GJR-GARCH	0.4784	0.794*	0.7693	1.000*	2.1516	0.954*	5.0072	1.000*	0.0582	0.585*
ART-GJR-GARCH-F	0.5827	0.001	0.8813	0.010	2.2134	0.025	5.0323	0.150*	0.0582	0.585*
10-step										
GARCH	0.6268	0.000	0.9548	0.002	2.3902	0.000	5.6129	0.008	0.0534	0.219*
GJR-GARCH	0.6992	0.000	1.0555	0.002	2.4605	0.000	5.6316	0.006	0.0544	0.219*
RT-GARCH	0.5718	0.000	0.8670	0.007	2.4921	0.000	5.8512	0.001	0.0550	0.013
ART-GARCH	0.5133	1.000*	0.8153	0.531*	2.2062	1.000*	5.2442	1.000*	0.0521	0.535*
ART-GJR-GARCH	0.5223	0.625*	0.8029	1.000*	2.2064	0.988*	5.2593	0.166*	0.0521	0.535*
ART-GJR-GARCH-F	0.6350	0.000	0.9269	0.007	2.2985	0.004	5.2993	0.053*	0.0520	1.000*
15-step										
GARCH	0.6688	0.000	0.9896	0.001	2.4916	0.000	5.8798	0.009	0.0541	0.289*
GJR-GARCH	0.7445	0.000	1.0999	0.001	2.5803	0.000	5.9106	0.007	0.0552	0.173*
RT-GARCH	0.6038	0.001	0.8851	0.004	2.6605	0.000	6.2716	0.000	0.0562	0.006
ART-GARCH	0.5421	1.000*	0.8393	0.673*	2.2627	1.000*	5.4898	1.000*	0.0528	0.289*
ART-GJR-GARCH	0.5607	0.412*	0.8290	1.000*	2.2647	0.845*	5.5110	0.193*	0.0527	0.289*
ART-GJR-GARCH-F	0.6877	0.000	0.9655	0.001	2.3924	0.000	5.5671	0.016	0.0527	1.000*

Note: p_{MCS} are the p-values of Model Confidence Set of Hansen et al. (2011). The models marked with $*$ fall in the model confidence set $\hat{\mathcal{M}}^*_{95\%}$.

5 Conclusion

In this paper we have proposed a new class of model which builds on Smetanina's (2017) RT-GARCH model by allowing heteroskedasticity in the conditional variance of volatility process. In doing so, we are able to simultaneously model both volatility and volatility of volatility from the observed return process. This is important since the volatility of volatility is widely regarded an additional source of risk. Our model has the advantage of computational tractability since only one filter is needed for both latent variables and estimation is conducted in the usual QML framework with analytical quasi-likelihood function.

Empirical studies show that incorporating volatility of volatility is important in order to (i) obtain heavier-tails of the conditional density of returns and better goodness-of-fit, (ii) filter volatility ex-post more efficiently, (iii) forecast volatility ex-ante more accurately for multiple forecast horizons.

We finish by suggesting some further researches. A natural extension is to incorporate RM in the model in the fashion of Hansen et al. (2012). Since one of the reasons to use a quadratic function of σ_t^2 for the volatility of volatility stems from the asymptotic distribution (2.8) of RV. We can replace $\psi_1+\psi_2\sigma_{t-1}^2$ by a function of RQ or other consistent estimator of IQ. Another possible extension is to specify a separate latent process for the volatility of volatility in a GARCH fashion. This enables the volatility of volatility to have its own dynamics which can be very different from volatility itself. Finally, extension to multivariate case can allow more flexible covariance structure across assets and their volatility.

References

Aït-Sahalia, Y. and J. Jacod (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics 37*(1), 184–222.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard (2003). Realized power variation and stochastic volatility models. *Bernoulli Society for Mathematical Statistics and Probability 9*(2), 243–265.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–37.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T., F. Diebold, and T. Andersen (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics 89*(4), 701–720.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009). A discrete-time model for daily s & p500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics 150*(2), 151 – 166. Recent Development in Financial Econometrics.

Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability 18*(1), 211–220.

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

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

Ding, Y. (2020). Weak diffusion limit of real-time GARCH models: the role of current return information. Cambridge Working Paper in Economics CWPE20112, University of Cambridge. Available at: `http://www.econ.cam.ac.uk/research/cwpe-abstracts?cwpe=20112`.

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

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

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

Fong, H. G. and O. A. Vasicek (1991). Fixed-income volatility management. *Journal of Portfolio Management 17*(4), 41–46.

Francq, C. and J.-M. Zakoïan (2009). Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association 104*(485), 313–324.

Francq, C. and J.-M. Zakoïan (2010). *GARCH models: structure, statistical inference, and financial applications* (1st ed.). Wiley.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1986). On the relation between the expected value and the volatility of nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

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

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

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Heston, S. L. (1993). A closed form solutions for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies 6*(2), 327–343.

Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Economstrics 147*(1), 47–59.

Linton, O. (2019). *Financial Econometrics: Models and Methods*, Chapter 11, pp. 358–421. Cambridge University Press.

Longstaff, A. and E. Schwartz (1992). Interest rate volatility and the term structure: a two-factor general equilibrium model. *Journal of Finance 47*(4), 1259–1282.

Maheu, J. M. and T. H. Mccurdy (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance 59*(2), 755–793.

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

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

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), 61–90.

Park, Y.-H. (2015). Volatility-of-volatility and tail risk hedging returns. *Journal of Financial Markets 26*, 38 – 63.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160*(1), 246–256.

Podolskij, M. and M. Vetter (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and Their Applications 119*(9), 2803–2831.

Shepherd, N. (2005). *Stochastic volatility selected readings / edited by Neil Shephard.* Advanced Texts in Econometrics. Oxford; New York: Oxford University Press.

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

Smetanina, E. and W. B. Wu (2019). Asymptotic theory for QMLE for the Real-time GARCH(1,1) model. Working paper, University of Chicago. Available at: `https://sites.google.com/site/smetaninakatja/research`.

Stout, W. F. (1974). *Almost sure convergence.* Probability and mathematical statistics. New York; London: Academic Press.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance 4*(2), 183–204.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.

A Proofs

Proof of Theorem 3.1. By Assumption 1, the process (ϵ_t) is strictly stationary and ergodic. Thus, (r_t) is strictly stationary if and only if (σ_t) is strictly stationary. By repeated substitution, the process σ_t^2 is essentially a stochastic difference equation with stationary coefficients. In order to obtain the strictly stationarity condition, we need to either assume the trivial $\sigma-$algebra $\mathcal{F}_0$ and the probability measure v_0 associated with it or assume the process extends infinitely into the past. These two approaches are identical if we assume ϵ_0 or ϵ_{-t} for $t \to -\infty$ are in the steady state. Let's assume (3.3) and express σ_t^2 in terms of a stochastic difference equation

$$\sigma_{t+1}^2 = A_{t+1}\sigma_t^2 + B_{t+1}, \tag{A.1}$$

where A_t and B_t are given by

$$A_t = \beta + \gamma\epsilon_{t-1}^2 + \phi(\epsilon_{t-1}^-)^2 + \psi_2\epsilon_t^2, \tag{A.2}$$
$$B_t = \alpha + \psi_1\epsilon_t^2 + \eta(\epsilon_t^-)^2. \tag{A.3}$$

The sequences A_t and B_t are measurable functions of ϵ_t and are thus, strictly stationary and ergodic as well as the joint sequence (A_t, B_t) by Theorem 3.5.8 of Stout (1974). We can then apply Theorem 1 of Brandt (1986) upon making the usual assumption that $\epsilon_0 = \epsilon_{-1}$ and conclude that the process (A.1) is strictly stationary iff

$$\mathbb{P}(A_0 = 0) > 0 \tag{A.4}$$

or

$$\mathbb{E}\log|A_0| < 0, \tag{A.5}$$
$$\mathbb{E}(\log|B_0|)^+ < \infty, \tag{A.6}$$

where $x^+ = \max(0, x)$. Plugging in the expressions for A_0 and B_0, we obtain (3.1) and (3.2). We also require A_0 and B_0 not equal to zero. □

Proof of Theorem 3.2. By Assumption 1 and σ_t^2 are positive with probability one, we have $\mathbb{E}\mathbb{1}_{(r_t<0)} = 0.5$, where $\mathbb{1}_{(\cdot)}$ is the indicator function. Using contemporaneous independence of r_s^2, σ_s^2 and ϵ_t for $s < t$, it is straightforward to show

$$\mathbb{E}[(r_{t+n}^-)^2|\mathcal{F}_t] = \tfrac{1}{2}\mathbb{E}[r_{t+n}^2|\mathcal{F}_t] + \tfrac{1}{4}\mathbb{E}\epsilon_t^4\eta, \tag{A.7}$$

for all $n \geq 1$. Thus,

$$\mathbb{E}[(r_t^-)^2] = \tfrac{1}{2}\mathbb{E}[r_t^2] + \tfrac{1}{4}\mathbb{E}\epsilon_t^4\eta. \tag{A.8}$$

Taking unconditional expectation on both sides of (2.2) and assuming σ_t^2 and r_t^2 are weakly stationary, we obtain

$$\mathbb{E}\sigma_t^2 = \alpha + \psi_1 + \tfrac{1}{2}\eta + \tfrac{1}{4}\mathbb{E}\epsilon_t^4\phi\eta + (\beta + \psi_2)\mathbb{E}\sigma_t^2 + (\gamma + \tfrac{1}{2}\phi)\mathbb{E}r_t^2, \tag{A.9}$$
$$\mathbb{E}r_t^2 = (1 + \psi_2\kappa)\mathbb{E}\sigma_t^2 + (\psi_1 + \tfrac{1}{2}\eta_1)\kappa, \tag{A.10}$$

where $\kappa = \mathbb{E}\epsilon_t^4 - 1$. Plugging (A.10) into (A.9), we obtain (3.5). This is only valid iff the denominator and numerator are both positive and finite. We obtain the condition for covariance stationarity (3.4). Plugging (3.5) into (A.10), we obtain (3.6). □

Proof of Theorem 3.3. By the same argument as in the proof of Theorem 3.2, it is straightforward to show

$$\mathbb{E}(r_t^-)^4 = \tfrac{1}{2}\mathbb{E}r_t^4 + c_1, \tag{A.11}$$
$$\mathbb{E}[(r_t^-)^2\sigma_t^2] = \tfrac{1}{2}\mathbb{E}[r_t^2\sigma_t^2] + c_2, \tag{A.12}$$

where $c_1 = \frac{1}{2}\eta\big((\frac{1}{2}\eta + \psi_1 + \psi_2)u_8 + (\alpha + \beta + \gamma + \phi)u_6\big)$ and $c_2 = \frac{1}{2}\eta\big((\frac{1}{2}\eta + \psi_1 + \psi_2)u_6 + (\alpha + \beta + \gamma + \phi)u_4\big)$. A direct calculation using (A.11) and (A.12) leads to

$$\mathbb{E}r_t^4 = \big(u_4 + \psi_2^2(u_8 - u_4^2) + 2\beta\psi_2(u_6 - u_4)\big)\mathbb{E}\sigma_t^4 + \psi_2(u_6 - u_4)(2\gamma + \phi)\mathbb{E}[r_t^2\sigma_t^2] + f_1(\bar{r}_2, \bar{\sigma}_2; \theta), \quad \text{(A.13)}$$

where

$$\begin{aligned} f_1(\bar{r}_2, \bar{\sigma}_2; \theta) = {} & (u_8 - u_4^2)(\psi_2^2 + \tfrac{1}{2}\eta^2 + \psi_1\eta) + (u_6 - u_4)(\alpha\eta + \tfrac{1}{2}u_4\phi^2\psi_1\eta + 2\phi\psi_2 c_2 \\ & + \tfrac{1}{4}u_4\phi^2\eta^2) + \big(\psi_2(u_8 - u_4^2)(\eta + 2\psi_1) + \beta(u_6 - u_4)(2\psi_1 + \eta)\big)\bar{\sigma}_2 \\ & + (u_6 - u_4)(2\psi_1 + \eta)(\gamma + \tfrac{1}{2}\phi)\bar{r}_2, \end{aligned} \quad \text{(A.14)}$$

with $\bar{r}_2 = \mathbb{E}r_t^2$ and $\bar{\sigma}_2 = \mathbb{E}\sigma_t^2$. Plugging (A.13) to the expression of $\mathbb{E}\sigma_t^4$,

$$\mathbb{E}\sigma_t^4 = \xi_1\mathbb{E}\sigma_t^4 + \xi_2\mathbb{E}[r_t^2\sigma_t^2] + (\gamma^2 + \tfrac{1}{2}\phi^2 + \gamma\phi)f_1(\bar{r}_2, \bar{\sigma}_2; \theta) + f_2(\theta), \quad \text{(A.15)}$$

where ξ_1 and ξ_2 are given in (3.10) and (3.11) and

$$\begin{aligned} f_2(\theta) = {} & \phi\big(c_1(\phi + 2\gamma) + 2c_2(\beta + \phi)\big) + \big(\psi_1^2 + \tfrac{1}{2}\eta^2 + \psi_1\eta \\ & + \phi^2\eta(\tfrac{1}{2}\alpha + \psi_1 + \tfrac{1}{4}\eta)\big)u_4 + \alpha(2\psi_1 + \eta). \end{aligned} \quad \text{(A.16)}$$

We next calculate $\mathbb{E}[r_t^2\sigma_t^2] \equiv \mathbb{E}[\sigma_t^4\epsilon_t^2]$. By (A.11) and (A.12), a direct calculation leads to

$$\mathbb{E}[r_t^2\sigma_t^2] = \xi_3\mathbb{E}\sigma_t^4 + \xi_4\mathbb{E}[r_t^2\sigma_t^2] + f_3(\bar{r}_2, \bar{\sigma}_2; \theta), \quad \text{(A.17)}$$

where ξ_3 and ξ_4 are given in (3.12) and (3.13) and

$$\begin{aligned} f_3(\bar{r}_2, \bar{\sigma}_2; \theta) = {} & \eta(u_6 - u_4)(\tfrac{1}{2}\eta + \psi_1) + \kappa\big(\alpha(2\psi_1 + \eta) + u_4\phi^2\eta(\tfrac{1}{2}\psi_1 + \tfrac{1}{4}\eta) + 2\phi\psi_2 c_2\big) \\ & + \big(\phi_2(u_6 - u_4)(2\phi_1 + \eta) + \kappa(2\alpha\psi_2 + 2\beta\psi_1 + \beta\eta)\big)\bar{\sigma}_2 \\ & + \kappa(2\psi_1 + \eta)(\gamma + \tfrac{1}{2}\phi)\bar{r}_2, \end{aligned} \quad \text{(A.18)}$$

Thus, $\xi_4 < 1$ is necessary for $r_t\sigma_t$ to be covariance stationary. Solving for $\mathbb{E}[r_t^2\sigma_t^2]$ from (A.17) and plugging into (A.15),

$$\mathbb{E}\sigma_t^4 = \frac{(\gamma^2 + \frac{1}{2}\phi^2 + \gamma\phi)f_1(\bar{r}_2, \bar{\sigma}_2; \theta) + f_2(\theta) + f_3(\bar{r}_2, \bar{\sigma}_2; \theta)/(1 - \xi_4)}{1 - \xi_1 - \xi_1 + \xi_1\xi_4 - \xi_2\xi_3}. \quad \text{(A.19)}$$

(A.19) is only valid iff (3.9) is satisfied. Substitude (A.19) into (A.17) then into (A.13), we obtain the expression for $\mathbb{E}r_t^4$. □

Proof of Theorem 3.4. The proof follows exactly the proof of Theorem 1 of Smetanina (2017) by replacing ψ with $a_{t-1} + \eta\mathbb{1}_{(y<0)}$ where $\mathbb{1}_{(\cdot)}$ is the indicator function. Since a_{t-1} is $\mathcal{F}_{t-1}-$measurable, we obtain the transition density (3.14). Furthermore, since $\mathbb{E}\epsilon_t^{2n} = \frac{1}{2}\mathbb{E}(\epsilon_t^-)^{2n}$ for all integers $n \geq 1$, we obtain the conditional moments approximation in (3.18) by making the same substitution. □

Proof of Theorem 3.5. The proof follows step by step of Smetanina and Wu (2019) by making the same substitution in the proof of Theorem 3.4. Crucially, the joint process

(r_t^2, σ_t^2) is still geometrically moment contracting with moment coefficient $1+\delta$ for some $\delta > 0$ since a_{t-1} is $\mathcal{F}_{t-1}$−measurable and ergodic as long as σ_{t-1}^2 is ergodic. Therefore, there exists an a.s.-unique causal ergodic strictly stationary solution to (2.1) and (2.2) at θ. The derivative process $\partial\sigma_t^2(\theta)/\partial\theta$ has an a.s.-unique strictly stationary and ergodic solution and is also geometrically moment contracting with the moment coefficient $1+\delta$ for some small $\delta > 0$. Therefore, Theorem 3.5 follows directly from Theorem 7 of Smetanina and Wu (2019). The exact expressions for $\partial\sigma_t^2(\theta)/\partial\theta$, $\partial l_t(\theta)/\partial\theta$ and V_{θ_0} are much more involved and we leave them for future research. □

Proof of Theorem 3.6. The filtering equation (3.27) can be obtained by solving the quartic equation of (2.1) and the one-step forecast equations (3.28) and (3.29) are from straightforward calculations. By twice repeated substitution we obtain the two-step forecast equations (3.30) and (3.31). For multi-step forecast, let $n > 2$ for any integer n, using (A.7) we have,

$$\mathbb{E}[\sigma_{t+n}^2|\mathcal{F}_t] = \alpha + \psi_1 + \tfrac{1}{2}\eta + \tfrac{1}{4}u_4\phi\eta + (\beta+\psi_2)\mathbb{E}[\sigma_{t-n-1}^2|\mathcal{F}_t] + (\gamma + \tfrac{1}{2}\phi)\mathbb{E}[r_{t+n-1}^2|\mathcal{F}_t], \tag{A.20}$$

where $u_4 = \mathbb{E}\epsilon_t^4$. Rearranging terms we obtain,

$$\mathbb{E}[r_{t+n-1}^2|\mathcal{F}_t] = \frac{1}{\gamma+\frac{1}{2}\phi}\mathbb{E}[\sigma_{t+n}^2|\mathcal{F}_t] - \frac{\alpha+\psi_1+\frac{1}{2}\eta+\frac{1}{4}u_4\phi\eta}{\gamma+\frac{1}{2}\phi} - \frac{\beta+\psi_2}{\gamma+\frac{1}{2}\phi}\mathbb{E}[\sigma_{t+n-1}^2|\mathcal{F}_t]. \tag{A.21}$$

Expand (A.20) for another lag and substitute $\mathbb{E}[r_{t+n-1}^2|\mathcal{F}_t]$ with (A.21),

$$\begin{aligned}\mathbb{E}[\sigma_{t+n}^2|\mathcal{F}_t] &= \alpha + \psi_1 + \tfrac{1}{2}\eta + \tfrac{1}{4}\phi\eta u_4 + \kappa(\gamma+\tfrac{1}{2}\phi)(\psi_1+\tfrac{1}{2}\eta) \\ &\quad + (\beta+\gamma+\psi_2+\tfrac{1}{2}\phi)\mathbb{E}[\sigma_{t+n-1}^2|\mathcal{F}_t] + \kappa\psi_2(\gamma+\tfrac{1}{2}\phi)\mathbb{E}[\sigma_{t+n-2}^2|\mathcal{F}_t],\end{aligned} \tag{A.22}$$

where $\kappa = u_4 - 1$. Upon examining (3.5) in Theorem 3.2, the numerator of the expression for $\mathbb{E}\sigma_t^2$ is the intercept term of the right hand side of (A.22). Combining (3.5) and (A.22), we obtain (3.32). Similarly, we can derive (3.33) using (3.6). □

Derivation of (3.34). We only need to prove z_t and e_t are both MDS. It is straightforward to verify (3.34) reduces to σ_t^2 in (2.2) and r_t^2 by direct calculation. z_t is an MDS by Assumption 1. e_t is an MDS because $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}] + \kappa(\psi_1 + \psi_2\sigma_{t-1}^2)$. □

B Additional Figures

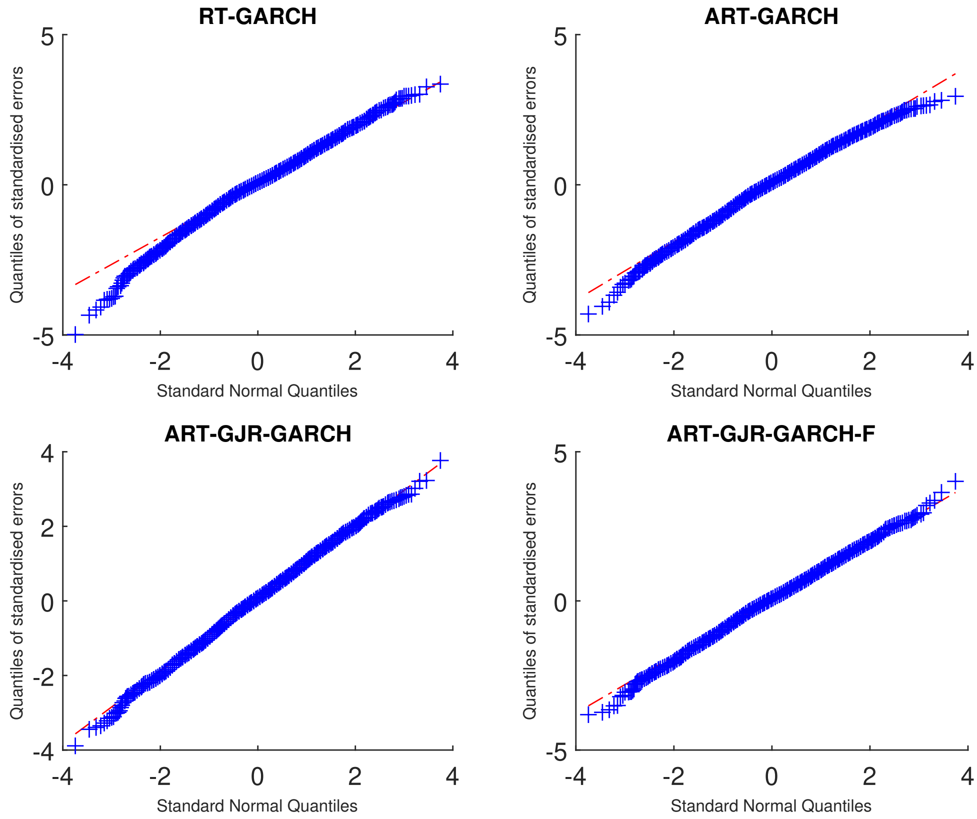


Figure 5: QQ plots of the standardised returns for DJIA index

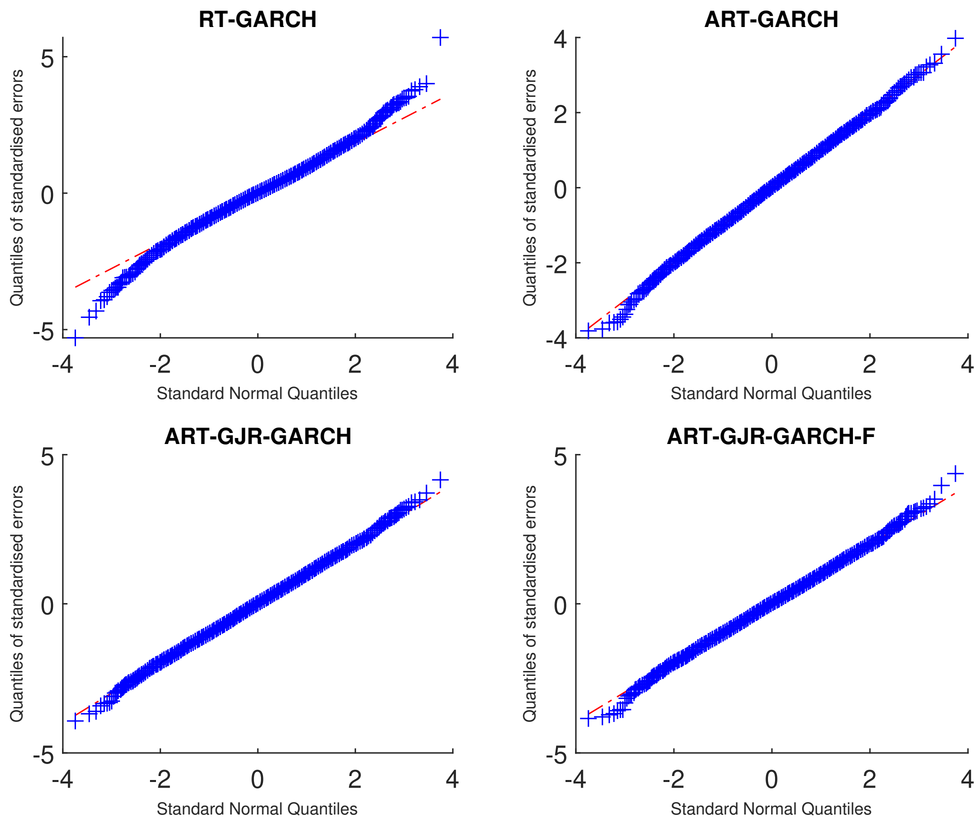


Figure 6: QQ plots of the standardised returns for JPM stock

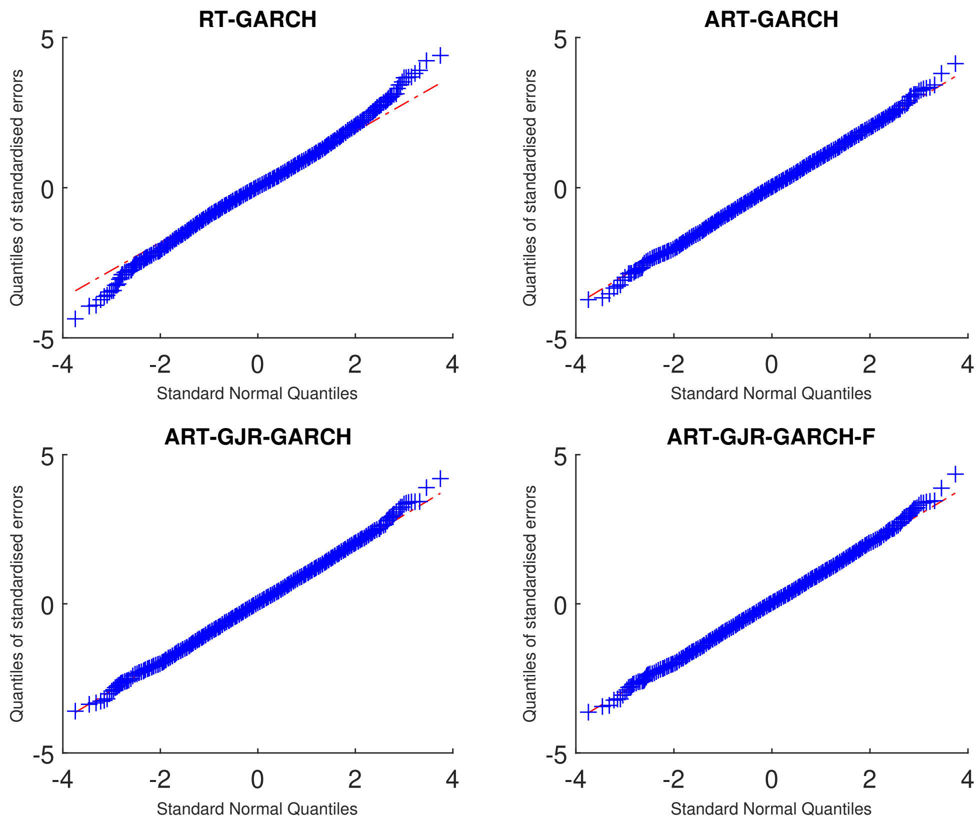


Figure 7: QQ plots of the standardised returns for AAPL stock

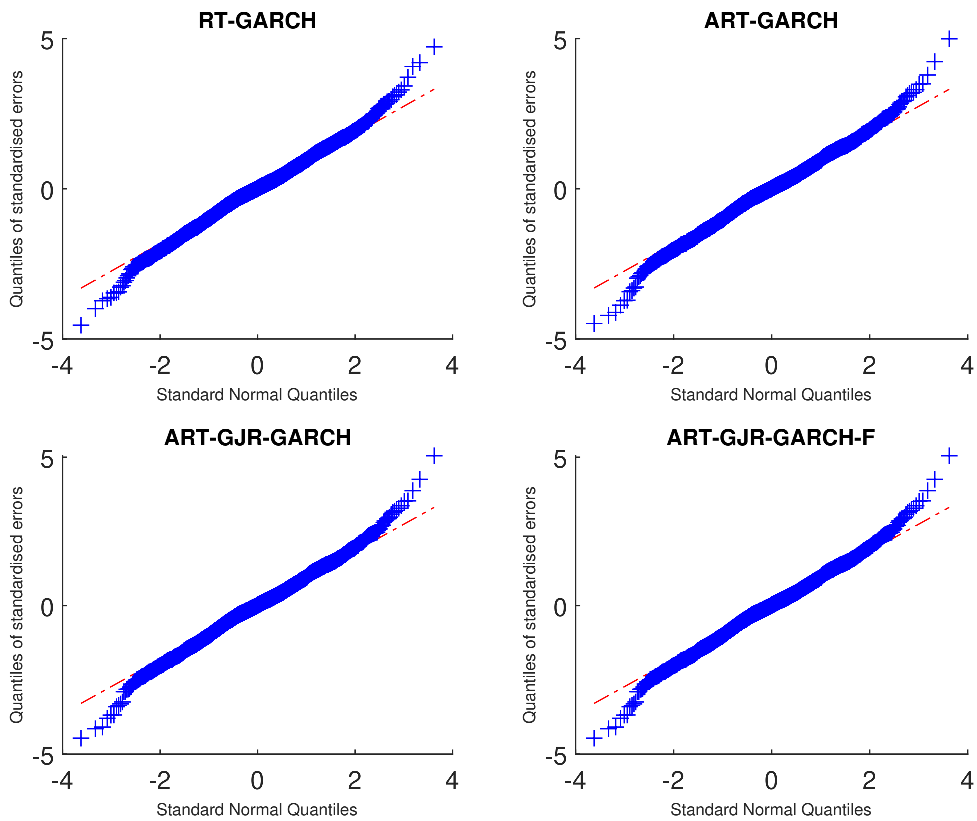
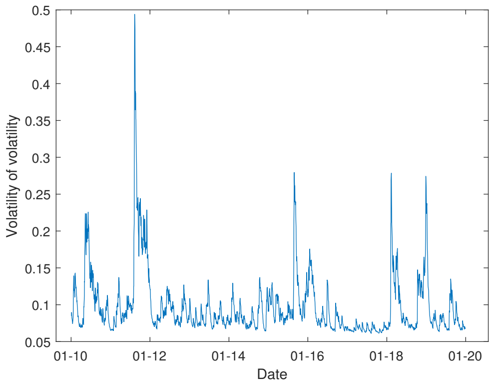
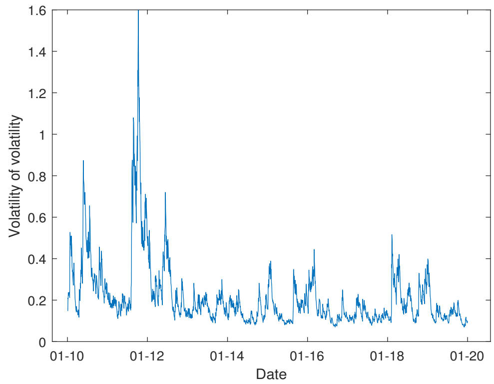


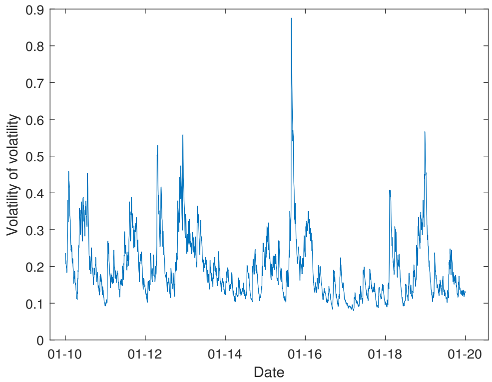
Figure 8: QQ plots of the standardised returns for EUR/USD exchange rate



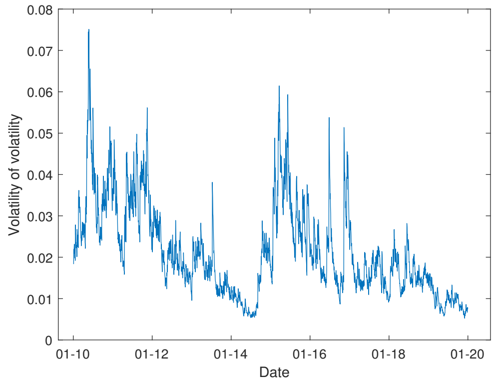
(a) DJIA.



(b) JPM.



(c) AAPL.



(d) EUR/USD.

Figure 9: Filtered volatility of volatility of ART-GJR-GARCH.