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,

\[ \sigma_t^2 = \alpha + \beta \sigma_{t-1}^2 + \gamma r_{t-1}^2 + \psi \epsilon_t^2, \]  

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 \( \mathbb{E} \epsilon^4 < \infty \). The process \( \sigma_t^2 \) 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

---

1The 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, $\sigma_t^2$, 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 $\sigma_t^2$, 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 $(r_t, \sigma_t^2)$ satisfies

$$r_t = \sigma_t \epsilon_t,$$

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

where $\epsilon_t$ 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 $\sigma_t^2$ 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,

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

where

$$
\Phi_0 = \alpha + \psi_1 + \frac{1}{2} \eta, \quad (2.4)
$$

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

and

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

is a martingale difference sequence (MDS) with conditional variance

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

where $\kappa = E \epsilon_t^4 - 1$. By definition, the RHS of (2.7) is the conditional variance of $\sigma_t^2$ at time $t - 1$. We call $E[z_t^2 | \mathcal{F}_{t-1}]$ 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 $\sigma_t^2$ are not independent and thus, we can not forecast $\sigma_{t+n}^2$ for $n > 1$ directly. We present the final expression of $\sigma_t^2$ 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,

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

as $h \downarrow 0$ for $kh \leq t < (k+1)h$, where $[x]$ denotes the largest integral part less than or equal to $x$. Since $E[\sigma_t^2 | \mathcal{F}_{t-1}]$ is a quadratic function of $\sigma_{t-1}^2$, (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($p, q, l$):

$$
\sigma_t^2 = \alpha + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} (\gamma_j + \phi_j I_{t-j}) r_{t-j}^2 + \epsilon_t^2 \sum_{j=1}^{l} (\psi_1 + \psi_{j+1} \sigma_{t-j}^2 + \eta I_t), \quad (2.9)
$$
where $I_t = 1_{(r_t<0)}$ and $1_{(\cdot)}$ 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

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

(2.10)

using a first order Taylor expansion, where

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

(2.11)

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

(2.12)

The weights on past squared returns, $(a_{t-1-j} + \eta I_{t-j})/b_{t-1-j}$, are more flexible than 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 $\sigma_t^2$ 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 $\epsilon_{t-j}$. 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

$$
E[r_{t+1}^2|\mathcal{F}_t] = \bar{\alpha} + \frac{1}{2}\bar{\beta} \left( \bar{b} + \sqrt{\bar{b}^2 + 4\bar{\alpha} r_t^2 + 4\eta(r_t^-)^2} \right) + \gamma r_t^2 + \phi(r_t^-)^2,
$$

(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 $\bar{\alpha} = \psi_1 + \psi_2 \bar{\sigma}_2$ with $\bar{\sigma}_2$, $\bar{r}_2$ and $\bar{r}_2$ being the unconditional levels of $\sigma_t^2$, $r_t^2$ 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
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 $\epsilon_t$.

Assumption 1. Let $\epsilon_t$ 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 $\sigma$-algebra generated by $r_0, ..., r_t$. We now present the stationarity conditions for $r_t$ and $\sigma_t^2$.

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

\[
\begin{align*}
\mathbb{E}\log |\beta + (\gamma + \psi_2)\epsilon_0^2 + \phi(\epsilon_0^2)| < 0, \\
\mathbb{E}(\log |\alpha + \psi_1\epsilon_0^2 + \eta(\epsilon_0^2)|)' > 0,
\end{align*}
\]

where $(x)^+ = \max(x, 0)$ and $\epsilon_0$ is the starting point of $\epsilon_t$ endowed with probability measure $\mathbb{P}(\sigma_0^2, \epsilon_0) \in \Gamma = v_0(\Gamma)$ for any $\Gamma \in B(R^2)$,

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

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

\[
\beta + \psi_2 + \gamma + \frac{1}{2}\phi + \kappa\psi_2(\gamma + \frac{1}{2}\phi) < 1,
\]

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

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

and

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

Moreover,

\[
\mathbb{E}(r_t^-)^2 = \frac{1}{2}\mathbb{E}r_t^2 + \frac{1}{4}\eta\mathbb{E}\epsilon_t^4.
\]

**Remark 1.** Since $\mathbb{P}(r_t < 0) = \mathbb{P}(r_t > 0) = 0.5$ because $\epsilon_t$ 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 $\epsilon_t$. \[\text{7}\]
Assumption 2. In addition to Assumption 1, let $\mathbb{E} \xi_t^4 < \infty$.

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

$$\xi_4 < 1,$$

$$\xi_1 + \xi_4 - \xi_1 \xi_4 + \xi_2 \xi_3 < 1,$$

where

$$\xi_1 = \left( \beta^2 + \psi_2^2 u_4 + 2 \beta \psi_2 + (\gamma^2 + \frac{1}{2} \phi^2 + \gamma \phi) (u_4 + \psi_2^2 (u_8 - u_4^2) + 2 \beta \psi_2 (u_6 - u_4)) \right), \tag{3.10}$$

$$\xi_2 = (2 \gamma + \phi) \left( \beta + \psi_2 (1 + (\gamma^2 + \frac{1}{2} \phi^2 + \gamma \phi) (u_6 - u_4)) \right), \tag{3.11}$$

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

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

with $u_n = \mathbb{E} \xi_t^n$ for $n \geq 1$ and $\kappa = u_4 - 1$. Then the process $(r_t, \sigma_t)$ 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 $\epsilon_t$ satisfy Assumption 1 and $(r_t, \sigma_t^2)$ be generated by (2.1) and (2.2). Let $\theta = (\alpha, \beta, \gamma, \phi, \psi_1, \psi_2, \eta)^T \geq 0$. Then the transition density of the return process is given by

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

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

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

and

$$d_2(y, \sigma_{t-1}^2; \theta) = \begin{cases} \text{sign}(y) \sqrt{d_1(y, \sigma_{t-1}^2; \theta) - b_{t-1}}, & \text{for } (\psi_1, \psi_2, \eta)^T \neq 0, \\ y/\sqrt{b_{t-1}}, & \text{for } (\psi_1, \psi_2, \eta)^T = 0. \end{cases} \tag{3.16}$$

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

8
Finally, the first-order approximation of the n-th conditional moment of \( y \), for \( n \in \mathbb{Z}^+ \), is given by

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

(3.18)

where \( \mathbb{E}[\cdot] \) on the right hand side is taken w.r.t. \( \epsilon_t \).

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

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

(3.19)

which is clearly not zero. When \( \epsilon_t \) 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

$$\frac{\mathbb{E}[r_t^4|F_{t-1}]}{\mathbb{E}[r_t^2|F_{t-1}]^2} = \frac{b_{t-1}^2 \mathbb{E}e_t^4 + (2a_{t-1} + \eta)b_{t-1} \mathbb{E}e_t^2 + (a_{t-1}^2 + \frac{1}{2}\eta^2 + \eta a_{t-1})\mathbb{E}e_t^3}{(b_{t-1} + (a_{t-1} + \frac{1}{2}\eta)\mathbb{E}e_t^2)^2},$$

(3.20)

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

$$\frac{\mathbb{E}[r_t^3|F_{t-1}]}{\mathbb{E}[r_t^2|F_{t-1}]^{3/2}} \approx \frac{3\eta b_{t-1} \mathbb{E}(e_t^-)^5}{(b_{t-1} + (a_{t-1} + \frac{1}{2}\eta)\mathbb{E}e_t^2)^{3/2}},$$

(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 $\epsilon_t$ satisfy Assumption 1 and in addition, $\mathbb{E}e_t^6 < \infty$. Let $(r_t, \sigma_t^2)$ be generated by (2.1) and (2.2). The QML estimator $\hat{\theta}$ of the true parameter $\theta_0$ is given by $\hat{\theta} = \arg\max_{\theta \in \Theta} L_T(\theta)$, where $L_T(\theta)$ is the quasi-log-likelihood function of $r_t$, that is,

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

(3.22)

where

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

(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 $\theta_0 \in \Theta^o$, where $\Theta^o$ is the interior of the parameter space $\Theta$. Then,

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

(3.24)

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

$$\Sigma_1 = \mathbb{E}_{\theta_0} \left[ \frac{\partial l_t(\theta_0)}{\partial \theta_0} \frac{\partial l_t(\theta_0)}{\partial \theta_0'} \right] \quad \text{and} \quad \Sigma_2 = -\mathbb{E}_{\theta_0} \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta_0 \partial \theta_0'} \right].$$

(3.25)

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

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

(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 $\theta_0$, 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 $\theta_0$, the modified QLR statistics is given by $-2(L_T(\hat{\theta}_{-1}) - L_T(\hat{\theta})) / \hat{\kappa}$, where $\hat{\kappa}$ is a consistent estimator of $\kappa$. 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 $\sigma_t^2$ and conditional variance $\operatorname{Var}[r_t|\mathcal{F}_{t-1}]$. 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 $\operatorname{Var}[r_t|\mathcal{F}_{t-1}]$. 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 $\sigma_t^2$ simply the volatility. Readers should bear in mind that the true conditional variance is $\mathbb{E}[(r_{t+n} - \mathbb{E}[r_{t+n}|\mathcal{F}_t])^2|\mathcal{F}_t] \approx \mathbb{E}[r_{t+n}^2|\mathcal{F}_t]$ for all $n \geq 1$. The volatility filtering equation is given by

$$\sigma_t^2 = \frac{1}{2}b_{t-1} + \frac{1}{2} \sqrt{b_{t-1}^2 + 4\alpha_{t-1}^2 + 4\eta(r_{t-1}^2)^2}. \quad (3.27)$$

The one-step volatility forecast is given by

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

and the one-step conditional variance forecast is given by

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

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

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

and for conditional variance,

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

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

**Theorem 3.6.** Let $\epsilon_t$ satisfy Assumption 1 and $(r_t, \sigma_t^2)$ 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

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

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 $\kappa = \mathbb{E}r_t^4 - 1$.

The initial conditions for (3.32), $\mathbb{E}[\sigma_{t+n}^2|\mathcal{F}_t]$ 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

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

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 $\Phi_1$ and $\Phi_2$. 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 $(r_t^2, \sigma_t^2)$ can be expressed as a VARMA(2,1) process. Specifically,

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

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

\footnote{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

<table>
<thead>
<tr>
<th></th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>ψ₁</th>
<th>ψ₂</th>
<th>η</th>
<th>ϕ</th>
<th>BIC</th>
<th>QLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>0.0000</td>
<td>0.8856</td>
<td>0.0000</td>
<td>0.0022</td>
<td>0.0212</td>
<td>0.0423</td>
<td>0.1257</td>
<td>14214</td>
<td>54.54</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0086)</td>
<td>(0.0016)</td>
<td>(0.0017)</td>
<td>(0.0076)</td>
<td>(0.0063)</td>
<td>(0.0135)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.8807</td>
<td>0.0332</td>
<td>0.0000</td>
<td>0.0454</td>
<td>0.0593</td>
<td></td>
<td>14315</td>
<td>58.80</td>
</tr>
<tr>
<td></td>
<td>(0.0019)</td>
<td>(0.0100)</td>
<td>(0.0018)</td>
<td>(0.0018)</td>
<td>(0.0082)</td>
<td>(0.0082)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.8810</td>
<td>0.0201</td>
<td>0.0097</td>
<td>0.0872</td>
<td></td>
<td>-</td>
<td>14424</td>
<td>37.33</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0100)</td>
<td>(0.0112)</td>
<td>(0.0033)</td>
<td>(0.0136)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DJIA</td>
<td>0.0000</td>
<td>0.8886</td>
<td>0.0000</td>
<td>0.0027</td>
<td>0.0289</td>
<td>0.0358</td>
<td>0.1112</td>
<td>13956</td>
<td>44.59</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0082)</td>
<td>(0.0009)</td>
<td>(0.0013)</td>
<td>(0.0068)</td>
<td>(0.0057)</td>
<td>(0.0137)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.8861</td>
<td>0.0245</td>
<td>0.0000</td>
<td>0.0528</td>
<td>0.0526</td>
<td></td>
<td>14036</td>
<td>53.54</td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0085)</td>
<td>(0.0094)</td>
<td>(0.0010)</td>
<td>(0.0104)</td>
<td>(0.0073)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.8807</td>
<td>0.0166</td>
<td>0.0098</td>
<td>0.0905</td>
<td></td>
<td>-</td>
<td>14135</td>
<td>38.51</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0098)</td>
<td>(0.0110)</td>
<td>(0.0025)</td>
<td>(0.0121)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPM</td>
<td>0.0117</td>
<td>0.8885</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0762</td>
<td>0.0212</td>
<td>0.0516</td>
<td>19936</td>
<td>11.72</td>
</tr>
<tr>
<td></td>
<td>(0.0093)</td>
<td>(0.0127)</td>
<td>(0.0100)</td>
<td>(0.0097)</td>
<td>(0.0113)</td>
<td>(0.0125)</td>
<td>(0.0123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.9051</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0872</td>
<td>0.0434</td>
<td></td>
<td>19951</td>
<td>7.80</td>
</tr>
<tr>
<td></td>
<td>(0.0090)</td>
<td>(0.0109)</td>
<td>(0.0087)</td>
<td>(0.0101)</td>
<td>(0.0110)</td>
<td>(0.0121)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0145</td>
<td>0.8919</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1052</td>
<td></td>
<td>-</td>
<td>19958</td>
<td>78.30</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.0113)</td>
<td>(0.0087)</td>
<td>(0.0096)</td>
<td>(0.0119)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAPL</td>
<td>0.0000</td>
<td>0.9118</td>
<td>0.0083</td>
<td>0.0054</td>
<td>0.0633</td>
<td>0.0420</td>
<td>0.0188</td>
<td>22241</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>(0.0069)</td>
<td>(0.0092)</td>
<td>(0.0068)</td>
<td>(0.0108)</td>
<td>(0.0094)</td>
<td>(0.0181)</td>
<td>(0.0110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.9148</td>
<td>0.0109</td>
<td>0.0000</td>
<td>0.0670</td>
<td>0.0542</td>
<td></td>
<td>22235</td>
<td>6.03</td>
</tr>
<tr>
<td></td>
<td>(0.0061)</td>
<td>(0.0091)</td>
<td>(0.0073)</td>
<td>(0.0062)</td>
<td>(0.0091)</td>
<td>(0.0158)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.9143</td>
<td>0.0114</td>
<td>0.0203</td>
<td>0.0691</td>
<td></td>
<td>-</td>
<td>22239</td>
<td>41.94</td>
</tr>
<tr>
<td></td>
<td>(0.0093)</td>
<td>(0.0096)</td>
<td>(0.0075)</td>
<td>(0.0112)</td>
<td>(0.0101)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.0000</td>
<td>0.9348</td>
<td>0.0000</td>
<td>0.0005</td>
<td>0.0619</td>
<td>0.0015</td>
<td>0.0000</td>
<td>4684.4</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td>(0.0143)</td>
<td>(0.0090)</td>
<td>(0.0026)</td>
<td>(0.0187)</td>
<td>(0.0017)</td>
<td>(0.0073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.9348</td>
<td>0.0000</td>
<td>0.0005</td>
<td>0.0619</td>
<td>0.0015</td>
<td></td>
<td>4676.3</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>(0.0019)</td>
<td>(0.0162)</td>
<td>(0.0066)</td>
<td>(0.0025)</td>
<td>(0.0155)</td>
<td>(0.0016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.9340</td>
<td>0.0000</td>
<td>0.0013</td>
<td>0.0626</td>
<td></td>
<td>-</td>
<td>4669.0</td>
<td>44.79</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0170)</td>
<td>(0.0092)</td>
<td>(0.0021)</td>
<td>(0.0152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
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. $\alpha$ 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

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>DJIA</th>
<th>JPM</th>
<th>AAPL</th>
<th>EURUSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE pMCS</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
<td>MSE</td>
</tr>
<tr>
<td>GARCH</td>
<td>1.1825</td>
<td>0.001</td>
<td>1.2002</td>
<td>0.001</td>
<td>4.2540</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>1.2448</td>
<td>0.006</td>
<td>1.2399</td>
<td>0.004</td>
<td>4.1033</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>1.0434</td>
<td>0.006</td>
<td>1.0706</td>
<td>0.004</td>
<td>3.9953</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.9371</td>
<td>0.006</td>
<td>0.9853</td>
<td>0.004</td>
<td>3.9159</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.8598</td>
<td>1.000*</td>
<td>0.9248</td>
<td>1.000*</td>
<td>3.8439</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.9794</td>
<td>0.006</td>
<td>1.0047</td>
<td>0.004</td>
<td>3.7512</td>
</tr>
</tbody>
</table>

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

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), $\hat{M}_{95\%}^\ast$, as defined in Hansen et al. (2011) are reported in Table 2. 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 $\hat{M}_{95\%}^\ast$. 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. 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 $\text{Var} [\sigma_t^2 | \mathcal{F}_{t-1}]$. 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 RV. Recall from Barndorff-Nielsen and Shephard (2003), RV has asymptotic distribution:

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

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

---

4The 95 percentile model confidence set, $\hat{M}_{95\%}^\ast$, 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.

5From the author’s own experience at fixed income trading desk.
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 are always outside. For JPM and AAPL stock returns, ART-GJR-GARCH 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

<table>
<thead>
<tr>
<th></th>
<th>Sk&amp;P 500</th>
<th>DJIA</th>
<th>JPM</th>
<th>AAPL</th>
<th>EURUSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>pMCS</td>
<td>MSE</td>
<td>pMCS</td>
<td>MSE</td>
</tr>
<tr>
<td>1-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.3958</td>
<td>0.001</td>
<td>0.7569</td>
<td>0.005</td>
<td>2.2491</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.4005</td>
<td>0.001</td>
<td>0.7623</td>
<td>0.005</td>
<td>2.2477</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.3595</td>
<td>0.049</td>
<td>0.7070</td>
<td>0.074*</td>
<td>2.2072</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.3461</td>
<td>0.160*</td>
<td>0.7024</td>
<td>0.113*</td>
<td>2.2581</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.3211</td>
<td>1.000*</td>
<td>0.6660</td>
<td>1.000*</td>
<td>2.2262</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.3465</td>
<td>0.318*</td>
<td>0.6816</td>
<td>0.630*</td>
<td>2.1916</td>
</tr>
<tr>
<td>2-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.4643</td>
<td>0.002</td>
<td>0.8523</td>
<td>0.011</td>
<td>2.3930</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.5008</td>
<td>0.001</td>
<td>0.9021</td>
<td>0.003</td>
<td>2.4217</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.4301</td>
<td>0.017</td>
<td>0.8029</td>
<td>0.031</td>
<td>2.4022</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.3963</td>
<td>0.217*</td>
<td>0.7687</td>
<td>0.091*</td>
<td>2.3664</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.3831</td>
<td>1.000*</td>
<td>0.7446</td>
<td>1.000*</td>
<td>2.3388</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.4431</td>
<td>0.062*</td>
<td>0.8073</td>
<td>0.074*</td>
<td>2.3593</td>
</tr>
<tr>
<td>5-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.5675</td>
<td>0.001</td>
<td>0.9657</td>
<td>0.008</td>
<td>2.5346</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.6343</td>
<td>0.001</td>
<td>1.0493</td>
<td>0.002</td>
<td>2.5756</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.5294</td>
<td>0.002</td>
<td>0.9054</td>
<td>0.018</td>
<td>2.5873</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.4755</td>
<td>0.945*</td>
<td>0.8519</td>
<td>0.274*</td>
<td>2.4669</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.4749</td>
<td>1.000*</td>
<td>0.8386</td>
<td>1.000*</td>
<td>2.4574</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.5714</td>
<td>0.002</td>
<td>0.9380</td>
<td>0.018</td>
<td>2.4900</td>
</tr>
<tr>
<td>10-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.6171</td>
<td>0.006</td>
<td>1.0128</td>
<td>0.005</td>
<td>2.6216</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.6832</td>
<td>0.006</td>
<td>1.1039</td>
<td>0.005</td>
<td>2.6644</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.5647</td>
<td>0.006</td>
<td>0.9207</td>
<td>0.020</td>
<td>2.6744</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.5125</td>
<td>1.000*</td>
<td>0.8862</td>
<td>0.379*</td>
<td>2.5078</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.5175</td>
<td>0.804*</td>
<td>0.8699</td>
<td>1.000*</td>
<td>2.4936</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.6219</td>
<td>0.006</td>
<td>0.9822</td>
<td>0.016</td>
<td>2.5440</td>
</tr>
<tr>
<td>15-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.6588</td>
<td>0.001</td>
<td>1.0479</td>
<td>0.003</td>
<td>2.7354</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.7308</td>
<td>0.001</td>
<td>1.1515</td>
<td>0.003</td>
<td>2.8084</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.5963</td>
<td>0.001</td>
<td>0.9485</td>
<td>0.017</td>
<td>2.8462</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.5410</td>
<td>1.000*</td>
<td>0.9105</td>
<td>0.561*</td>
<td>2.5689</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.5555</td>
<td>0.528*</td>
<td>0.8963</td>
<td>1.000*</td>
<td>2.5587</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.6751</td>
<td>0.001</td>
<td>1.0221</td>
<td>0.003</td>
<td>2.6550</td>
</tr>
</tbody>
</table>

Note: pMCS are the p-values of Model Confidence Set of Hansen et al. (2011). The models marked with * fall in the model confidence set $M_{95\%}$. 


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

\[ \sum_{h=1}^{[t/h]} r_{kh}^2 \rightarrow P \int_0^t \sigma_s^2 ds + \sum_{i=1}^{N_t} J_i^2, \]  \hspace{1cm} (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),

\[ \frac{\pi}{2} \sum_{k=1}^{[t/h]-1} |r_{kh}|^2 \left| r_{(k+1)h} \right|^2 \rightarrow P \int_0^t \sigma_s^2 ds, \]  \hspace{1cm} (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

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>DJIA</th>
<th>JPM</th>
<th>AAPL</th>
<th>EURUSD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PMCS</td>
<td>MSE</td>
<td>PMCS</td>
<td>MSE</td>
</tr>
<tr>
<td>1-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.4082</td>
<td>0.001</td>
<td>0.7113</td>
<td>0.002</td>
<td>1.9901</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.4222</td>
<td>0.001</td>
<td>0.7331</td>
<td>0.002</td>
<td>2.0117</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.3695</td>
<td>0.032</td>
<td>0.6579</td>
<td>0.062*</td>
<td>1.9593</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.3494</td>
<td>0.240*</td>
<td>0.6412</td>
<td>0.169*</td>
<td>1.9363</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.3282</td>
<td>1.000*</td>
<td>0.6106</td>
<td>1.000*</td>
<td>1.9134</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.3629</td>
<td>0.240*</td>
<td>0.6420</td>
<td>0.294*</td>
<td>1.9041</td>
</tr>
<tr>
<td>2-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.4762</td>
<td>0.002</td>
<td>0.8007</td>
<td>0.022</td>
<td>2.1284</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.5233</td>
<td>0.002</td>
<td>0.8661</td>
<td>0.001</td>
<td>2.1777</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.4392</td>
<td>0.017</td>
<td>0.7474</td>
<td>0.022</td>
<td>2.1487</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.3991</td>
<td>0.392*</td>
<td>0.7036</td>
<td>0.239*</td>
<td>2.0416</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.3901</td>
<td>1.000*</td>
<td>0.6845</td>
<td>1.000*</td>
<td>2.0325</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.4606</td>
<td>0.017</td>
<td>0.7616</td>
<td>0.189*</td>
<td>2.0660</td>
</tr>
<tr>
<td>5-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.5760</td>
<td>0.001</td>
<td>0.9075</td>
<td>0.006</td>
<td>2.2912</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.6494</td>
<td>0.001</td>
<td>1.0003</td>
<td>0.004</td>
<td>2.3434</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.5352</td>
<td>0.001</td>
<td>0.8426</td>
<td>0.010</td>
<td>2.3586</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.4754</td>
<td>1.000*</td>
<td>0.7814</td>
<td>0.369*</td>
<td>2.1511</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.4784</td>
<td>0.794*</td>
<td>0.7693</td>
<td>1.000*</td>
<td>2.1516</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.5827</td>
<td>0.017</td>
<td>0.8813</td>
<td>0.010</td>
<td>2.2134</td>
</tr>
<tr>
<td>10-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.6268</td>
<td>0.000</td>
<td>0.9548</td>
<td>0.002</td>
<td>2.3902</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.6992</td>
<td>0.000</td>
<td>1.0555</td>
<td>0.002</td>
<td>2.4605</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.5718</td>
<td>0.000</td>
<td>0.8670</td>
<td>0.007</td>
<td>2.4921</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.5133</td>
<td>1.000*</td>
<td>0.8153</td>
<td>0.531*</td>
<td>2.2062</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.5223</td>
<td>0.625*</td>
<td>0.8029</td>
<td>1.000*</td>
<td>2.2064</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.6350</td>
<td>0.000</td>
<td>0.9269</td>
<td>0.007</td>
<td>2.2985</td>
</tr>
<tr>
<td>15-step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>0.6688</td>
<td>0.000</td>
<td>0.9896</td>
<td>0.001</td>
<td>2.4916</td>
</tr>
<tr>
<td>GJR-GARCH</td>
<td>0.7445</td>
<td>0.000</td>
<td>1.0999</td>
<td>0.001</td>
<td>2.5803</td>
</tr>
<tr>
<td>RT-GARCH</td>
<td>0.6038</td>
<td>0.001</td>
<td>0.8851</td>
<td>0.004</td>
<td>2.6605</td>
</tr>
<tr>
<td>ART-GARCH</td>
<td>0.5421</td>
<td>1.000*</td>
<td>0.8393</td>
<td>0.673*</td>
<td>2.2627</td>
</tr>
<tr>
<td>ART-GJR-GARCH</td>
<td>0.5607</td>
<td>0.412*</td>
<td>0.8290</td>
<td>1.000*</td>
<td>2.2647</td>
</tr>
<tr>
<td>ART-GJR-GARCH-F</td>
<td>0.6877</td>
<td>0.000</td>
<td>0.9655</td>
<td>0.001</td>
<td>2.3924</td>
</tr>
</tbody>
</table>

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

19
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 $\sigma_t^2$ 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


Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in


Proof of Theorem 3.1. By Assumption 1, the process \((\epsilon_t)\) is strictly stationary and ergodic. Thus, \((r_t)\) is strictly stationary if and only if \((\sigma_t)\) is strictly stationary. By repeated substitution, the process \(\sigma^2_t\) 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 \(\epsilon_0\) or \(\epsilon_{-t}\) for \(t \to -\infty\) are in the steady state. Let’s assume (3.3) and express \(\sigma^2_t\) in terms of a stochastic difference equation

\[
\sigma^2_{t+1} = A_{t+1} \sigma^2_t + B_{t+1},
\]

(A.1)
where $A_t$ and $B_t$ are given by

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

The sequences $A_t$ and $B_t$ are measurable functions of $\epsilon_t$ 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

$$P(A_0 = 0) > 0 \quad (A.4)$$

or

$$E \log |A_0| < 0, \quad (A.5)$$
$$E(\log |B_0|)^+ < \infty, \quad (A.6)$$

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 $\sigma^2_t$ are positive with probability one, we have

$$E \mathbb{1}_{(r_t < 0)} = 0.5,$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function. Using contemporaneous independence of $r_s^2, \sigma^2_s$ and $\epsilon_t$ for $s < t$, it is straightforward to show

$$E[(r_{t+n})^2 | F_t] = \frac{1}{2} E[r_{t+n}^2 | F_t] + \frac{1}{4} E \epsilon_t^4 \eta, \quad (A.7)$$

for all $n \geq 1$. Thus,

$$E[(r_t)^2] = \frac{1}{2} E[r_t^2] + \frac{1}{4} E \epsilon_t^4 \eta. \quad (A.8)$$

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

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

where $\kappa = E \epsilon_t^4 - 1$. 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). \qed

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

$$E(r_t^4) = \frac{1}{2} E r_t^4 + c_1, \quad (A.11)$$
$$E[(r_t)^2 \sigma_t^2] = \frac{1}{4} E[r_t^2 \sigma_t^2] + c_2, \quad (A.12)$$
Thus, \( \xi = \frac{1}{2} \eta \left( \frac{1}{2} \eta + \psi_1 + \psi_2 \right) u_8 + (\alpha + \beta + \gamma + \phi) u_6 + (\alpha + \beta + \gamma + \phi) u_5 \). A direct calculation using (A.11) and (A.12) leads to

\[
\mathbb{E} r_1^4 = (u_4 + \psi_2^2(u_8 - u_3^2) + 2 \beta \psi_2(u_6 - u_4) ) \mathbb{E} \sigma_1^4 + \psi_2(u_6 - u_4)(2\gamma + \phi) \mathbb{E}[\sigma_1^2 \sigma_2^2] + f_1(\dot{r}_2, \dot{\sigma}_2; \theta), \tag{A.13}
\]

where

\[
f_1(\dot{r}_2, \dot{\sigma}_2; \theta) = (u_8 - u_4^2)(\psi_1^2 + \frac{1}{2} \eta^2 + \psi_1 \eta) + (u_6 - u_4)(\alpha \eta + \frac{1}{2} u_4 \phi^2 \psi_1 \eta + 2 \phi \psi_2 c_2 \\
+ \frac{1}{2} u_4 \phi^2 \eta^2) + (\psi_2(u_8 - u_4^2)(\eta + 2 \psi_1) + \beta (u_6 - u_4)(2 \psi_1 + \eta)) \dot{\sigma}_2 \\
+ (u_6 - u_4)(2 \psi_1 + \eta) (2 \psi_1 + \eta) \dot{r}_2, \tag{A.14}
\]

with \( \dot{r}_2 = \mathbb{E} r_1^2 \) and \( \dot{\sigma}_2 = \mathbb{E} \sigma_1^2 \). Plugging (A.13) to the expression of \( \mathbb{E} \sigma_1^4 \),

\[
\mathbb{E} \sigma_1^4 = \xi_1 \mathbb{E} \sigma_1^4 + \xi_2 \mathbb{E}[r_1^2 \sigma_1^2] + (\gamma^2 + \frac{1}{2} \phi^2 + \gamma \phi) f_1(\dot{r}_2, \dot{\sigma}_2; \theta) + f_2(\theta), \tag{A.15}
\]

where \( \xi_1 \) and \( \xi_2 \) are given in (3.10) and (3.11) and

\[
f_2(\theta) = \phi(c_1 \phi + 2 \gamma) + 2 c_2 (\beta + \phi) + (\psi_1^2 + \frac{1}{2} \eta^2 + \psi_1 \eta) \\
+ \phi^2 (\frac{1}{4} \alpha + \psi_1 + \frac{1}{4} \eta) u_4 + \alpha (2 \psi_1 + \eta). \tag{A.16}
\]

We next calculate \( \mathbb{E}[r_1^2 \sigma_1^2] = \mathbb{E}[\sigma_1^4 \epsilon_t^2] \). By (A.11) and (A.12), a direct calculation leads to

\[
\mathbb{E}[r_1^2 \sigma_1^2] = \xi_3 \mathbb{E} \sigma_1^4 + \xi_4 \mathbb{E}[r_1^2 \sigma_1^2] + f_3(\dot{r}_2, \dot{\sigma}_2; \theta), \tag{A.17}
\]

where \( \xi_3 \) and \( \xi_4 \) are given in (3.12) and (3.13) and

\[
f_3(\dot{r}_2, \dot{\sigma}_2; \theta) = \eta (u_6 - u_4)(\frac{1}{2} \eta + \psi_1) + \kappa (\alpha (2 \psi_1 + \eta) + u_4 \phi^2 \eta(\frac{1}{2} \psi_1 + \frac{1}{4} \eta) + 2 \phi \psi_2 c_2) \\
+ (\phi (u_6 - u_4) (2 \phi + \eta) + \kappa (2 \alpha \phi_2 + 2 \beta \psi_1 + \beta \eta)) \dot{\sigma}_2 \\
+ \kappa (2 \psi_1 + \eta)(\gamma + \frac{1}{2} \psi) \dot{r}_2. \tag{A.18}
\]

Thus, \( \xi_4 < 1 \) is necessary for \( r_t \sigma_t \) to be covariance stationary. Solving for \( \mathbb{E}[r_1^2 \sigma_1^2] \) from (A.17) and plugging into (A.15),

\[
\mathbb{E} \sigma_1^4 = \frac{(\gamma^2 + \frac{1}{2} \phi^2 + \gamma \phi) f_1(\dot{r}_2, \dot{\sigma}_2; \theta) + f_2(\theta) + f_3(\dot{r}_2, \dot{\sigma}_2; \theta)/(1 - \xi_4)}{1 - \xi_1 - \xi_1 \xi_4 - \xi_2 \xi_3}. \tag{A.19}
\]

(A.19) is only valid iff (3.9) is satisfied. Substitute (A.19) into (A.17) then into (A.13), we obtain the expression for \( \mathbb{E} r_1^4 \).

\[\square\]

**Proof of Theorem 3.4.** The proof follows exactly the proof of Theorem 1 of Smetanina and Wu (2019) by replacing \( \psi \) with \( a_{t-1} + \eta \mathbb{I}_{(\theta < 0)} \) where \( \mathbb{I}_{(\cdot)} \) 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.

\[\square\]

**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^2_t, \sigma^2_t) is still geometrically moment contracting with moment coefficient 1 + \delta for some \delta > 0 since a_{t-1} is F_{t-1}—measurable and ergodic as long as \sigma^2_{t-1} is ergodic. Therefore, there exists an a.s.-unique causal ergodic strictly stationary solution to (2.1) and (2.2) at \theta. The derivative process \partial \sigma^2_t(\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^2_t(\theta)/\partial \theta, \partial l_t(\theta)/\partial \theta and V_{\theta_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,

\[ E[\sigma^2_{t+n}|F_t] = \alpha + \psi_1 + \frac{1}{2}\eta + \frac{1}{4}u_4\phi\eta + (\beta + \psi_2)E[\sigma^2_{t-n-1}|F_t] + (\gamma + \frac{1}{2}\phi)E[r^2_{t+n-1}|F_t], \]  

(A.20)

where \( u_4 = E\epsilon^4_t \). Rearranging terms we obtain,

\[ E[r^2_{t+n-1}|F_t] = \frac{1}{1 + \frac{1}{2}\phi}E[\sigma^2_{t+n}|F_t] - \frac{\alpha + \psi_1 + \frac{1}{2}\eta + \frac{1}{4}u_4\phi\eta}{\gamma + \frac{1}{2}\phi} - \frac{\beta + \psi_2}{\gamma + \frac{1}{2}\phi} E[\sigma^2_{t+n-1}|F_t]. \]  

(A.21)

Expand (A.20) for another lag and substitute \( E[r^2_{t+n-1}|F_t] \) with (A.21),

\[ E[\sigma^2_{t+n}|F_t] = \alpha + \psi_1 + \frac{1}{2}\eta + \frac{1}{4}\phi u_4 + \kappa(\gamma + \frac{1}{2}\phi)(\psi_1 + \frac{1}{2}\eta) + (\beta + \gamma + \psi_2 + \frac{1}{2}\phi)E[\sigma^2_{t+n-1}|F_t] + \kappa\psi_2(\gamma + \frac{1}{2}\phi)E[\sigma^2_{t+n-2}|F_t], \]  

(A.22)

where \( \kappa = u_4 - 1 \). Upon examining (3.5) in Theorem 3.2, the numerator of the expression for \( E\sigma^2_t \) 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).

\[ \square \]

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 \( \sigma^2_t \) in (2.2) and \( r^2_t \) by direct calculation. \( z_t \) is an MDS by Assumption 1. \( e_t \) is an MDS because \( E[r^2_t|F_{t-1}] = E[\sigma^2_t|F_{t-1}] + \kappa(\psi_1 + \psi_2\sigma^2_{t-1}). \) \[ \square \]
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
Figure 9: Filtered volatility of volatility of ART-GJR-GARCH.