

CAMBRIDGE WORKING PAPERS IN ECONOMICS  
JANEWAY INSTITUTE WORKING PAPERS

# Conditional Heteroskedasticity in the Volatility of Asset Returns

Yashuang  
(Dexter) Ding  
University of  
Cambridge

## Abstract

We propose a new class of conditional heteroskedasticity in the volatility (CHV) models which allows for time-varying volatility of volatility in the volatility of asset returns. This class nests a variety of GARCH-type models and the SHARV model of Ding (2021). CH-V models can be seen as a special case of the stochastic volatility of volatility model. We then introduce two examples of CH-V in which we specify a GJR-GARCH and an E-GARCH processes for the volatility of volatility, respectively. We also show a novel way of introducing the leverage effect of negative returns on the volatility through the volatility of volatility process. Empirical study confirms that CH-V models have better goodness-of-fit and out-of-sample volatility and Value-at-Risk forecasts than common GARCH-type models.

## Reference Details

2179	Cambridge Working Papers in Economics
2021/11	Janeway Institute Working Paper Series
Published	08 November 2021
Key Words	GARCH, SHARV, volatility, volatility of volatility, forecasting
JEL Codes	C22, C32, C53, C58, G17
Websites	<a href="http://www.econ.cam.ac.uk/cwpe">www.econ.cam.ac.uk/cwpe</a> <a href="http://www.inet.econ.cam.ac.uk/working-papers">www.inet.econ.cam.ac.uk/working-papers</a>

# Conditional Heteroskedasticity in the Volatility of Asset Returns

Yashuang (Dexter) Ding\*

*Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue,  
Cambridge CB3 9DD, United Kingdom of Great Britain and Northern Ireland*

## Abstract

We propose a new class of conditional heteroskedasticity in the volatility (CH-V) models which allows for time-varying volatility of volatility in the volatility of asset returns. This class nests a variety of GARCH-type models and the SHARV model of Ding (2021). CH-V models can be seen as a special case of the stochastic volatility of volatility model. We then introduce two examples of CH-V in which we specify a GJR-GARCH and an E-GARCH processes for the volatility of volatility, respectively. We also show a novel way of introducing the leverage effect of negative returns on the volatility through the volatility of volatility process. Empirical study confirms that CH-V models have better goodness-of-fit and out-of-sample volatility and Value-at-Risk forecasts than common GARCH-type models.

**Keywords:** GARCH, SHARV, volatility, volatility of volatility, forecasting

**JEL classification:** C22, C32, C53, C58, G17

---

\*Corresponding author.

*Email address:* yd274@cam.ac.uk.

# 1 Introduction

Volatility of asset returns has been an active research area following the seminal works of Engle (1982) and Bollerslev (1986). Most GARCH-type models define the volatility as the conditional variance of returns based on past observations (see Francq and Zakoïan, 2010 for a survey of GARCH-type models). The increasing availability of high frequency data in recent decades has motivated the use of the daily realised measures (RM), which are consistent estimators of the daily integrated volatility (IV), as volatility proxies. This makes the estimation easier since the volatility is no longer latent in these models. Since then, various authors have found evidence of volatility clustering in the RM (e.g., Corsi et al., 2008 and Bollerslev et al., 2009) which implies the presence of conditional heteroskedasticity in the volatility. This casts doubt on the GARCH-type specification since the volatility is conditionally deterministic. Ding (2021) argues that failing to take into account the time-varying volatility of volatility can result in a measurement error in the volatility estimates. Moreover, GARCH-type models ignore the (arguably) most important information in the current observation which makes it difficult to justify the accuracy of the volatility estimates (Ding, 2021). On the other hand, models based on RM rely exclusively on the high frequency data which is usually not available for asset managers with medium horizons. In addition, these models are subject to the discretisation error and market microstructure noise stemming from estimating the RM. Stochastic volatility (SV) models are natural candidates to address these issues.<sup>1</sup> However, the difficulty in estimation and inference therein is hard to overcome.

To capture the time-varying volatility of volatility and make use of the current observation in the volatility estimate, Ding (2021) proposes the stochastic heteroskedastic autoregressive volatility (SHARV) model. Specifically, the volatility process is given by

$$\sigma_t^2 = \beta\sigma_{t-1}^2 + (\alpha + \psi\sigma_{t-1}^2)\epsilon_t^2, \quad (1.1)$$

where  $\epsilon_t \equiv r_t/\sigma_t$  are *i.i.d.*  $(0, 1)$  random variables. He shows that SHARV along with its asymmetric form, ASHARV, can capture all stylised facts of financial returns even when  $\epsilon_t \sim N(0, 1)$ . Moreover, he shows that both models have better goodness-of-fit and out-of-sample volatility and Value-at-Risk (VaR) forecasts than GARCH-type models. The conditional variance of  $\sigma_t^2$  is given by  $\mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}] = (\alpha + \beta\sigma_{t-1}^2)^2\mathbb{E}\epsilon_t^4$  provided  $\mathbb{E}\epsilon_t^4 < \infty$ . The dynamics of the volatility of volatility is completely determined by the lagged level of volatility. This is quite restrictive. Ideally, we would want the volatility of volatility to follow a separate process which in turn, drives the volatility process. This would lead to the so called stochastic volatility of volatility (SVV) model where both the volatility and volatility of volatility are latent processes.

---

<sup>1</sup>Only the contemporaneous SV models make use of the current observation in the volatility estimates. This is captured by the correlation between return and volatility innovations. See Taylor (1994) for the definitions of contemporaneous and lagged SV models.

Apart from directly modelling the volatility of volatility, another way to alleviate the heteroskedasticity in the volatility is to model the log volatility. This can be explained by the fact that the asymptotic variance of the log realised volatility (RV), which is given by  $2 \int_0^t \sigma_s^4 ds / (\int_0^t \sigma_s^2 ds)^2$  (Barndorff-Nielsen and Shephard, 2003), is much more stable than the asymptotic variance of RV,  $2 \int_0^t \sigma_s^4 ds$ . However, this can only partially alleviate the heteroskedasticity in the volatility. To see this, consider a simple diffusion process,

$$d\sigma_t^2 = (a - b\sigma_t^2)dt + v_t dL_t,$$

where  $L_t$  is a (homogenous) Lévy process and  $v_t$  is an almost surely positive and locally square-integrable process. By Itô's lemma, the log volatility satisfies

$$d \log \sigma_t^2 = \left( \frac{a}{\sigma_t^2} - b \right) dt - \frac{v_t^2}{2\sigma_t^4} d\langle L \rangle_t + \frac{v_t}{\sigma_t^2} dL_t, \quad (1.2)$$

where  $\langle L \rangle_t$  is the quadratic variation of  $L_t$ . Clearly, only when  $v_t$  is proportional to  $\sigma_t^2$  and  $L_t$  is a Brownian motion can we eliminate the heteroskedasticity in  $\log \sigma_t^2$ . Moreover, most exponential-SV models assume an autoregressive structure for  $\log \sigma_t^2$ , which is subject to misspecification if  $\sigma_t^2$  has an autoregressive structure.<sup>2</sup> This can be seen from (1.2) where the drift term does not depend on  $\log \sigma_t^2$  explicitly. On the other hand, if we assume

$$d \log \sigma_t^2 = (a - b \log \sigma_t^2) dt + v_t dL_t,$$

then again by Itô's lemma,

$$d\sigma_t^2 = (a - b \log \sigma_t^2) \sigma_t^2 dt + \frac{1}{2} v_t^2 \sigma_t^2 d\langle L \rangle_t + v_t \sigma_t^2 dL_t,$$

and  $\sigma_t^2$  still has an autoregressive structure. Therefore, modelling  $\sigma_t^2$  directly while taking into account the time-varying volatility of volatility is less prone to misspecifications than modelling  $\log \sigma_t^2$ .

Compared to SV models, literature on SVV models without using high frequency data are relatively scarce. Barndorff-Nielsen and Veraart (2012) propose a volatility modulated non-Gaussian Ornstein-Uhlenbeck (OU) SVV model. In their model, the volatility follows a non-Gaussian OU process driven by a Lévy subordinator and the volatility of volatility is captured by one or more stochastic components of the Lévy subordinator. In particular, they discuss the cases of stochastic proportional and stochastic time change in the Lévy subordinator. Moreover, they show that the so-called leverage effect can be captured by the correlation between the volatility of volatility and asset price processes.<sup>3</sup> Meanwhile, Huang et al. (2019) propose a model where both the volatility and volatility of volatility are driven by Brownian motions. Using VIX and VVIX spot and option data, they show

---

<sup>2</sup>For example, in the stochastic autoregressive volatility (SARV) model,  $\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \gamma u_t$ .

<sup>3</sup>The Lévy subordinator is intrinsically independent from the Brownian motion term in the asset price process. Therefore, in the non-Gaussian OU process with stochastic volatility of volatility, the leverage effect can only be introduced by adding a jump component in the asset price process.

that the fluctuations in the volatility of volatility are not directly related to the level of volatility itself. However, the volatility in their model fails to be positive with probability one which makes it hard to justify their model to be the true data generating process.

All above-mentioned models are in continuous time settings. To our knowledge, discrete time SVV models have not yet been studied. In this paper, we propose a new class of conditional heteroskedasticity in the volatility (CH-V) models which includes SHARV and several GARCH-type models as special cases. Despite having two latent processes, we show that CH-V models have analytical expressions for the likelihood function which makes the parameter estimation and statistical inference straightforward. Moreover, we show a novel way of introducing the leverage effect via the volatility of volatility. Subsequently, we introduce two examples of CH-V models where the volatility of volatility follows a GJR- and E-GARCH processes, respectively. We call them the GARCH-V models. Empirical evidence shows that GARCH-V models have slightly better goodness-of-fit than SHARV. For volatility forecasts, while the difference between GARCH-V and SHARV is not significant, both models have much more accurate forecasts than other GARCH-type models. This confirms the importance of modelling the time-varying volatility of volatility. For VaR forecasts, GARCH-V models produce more accurate results than SHARV.

The rest of this paper is organised as follows: In section 2 we introduce the class of CH-V models. In section 3 we introduce the GJR- and E-GARCH-V models. In section 4 we present the empirical analysis and section 5 concludes. All proofs and derivations can be found in appendix A.

## 2 Conditional heteroskedasticity in the volatility

We now introduce the class of univariate CH-V models which nests SHARV and several GARCH-type models as special cases. We only consider the first-order case, extension to higher order is straightforward. Specifically, let  $\mathcal{F}_t$  denote the  $\sigma$ -algebra generated by all available information up to time  $t$ . Let the return and its volatility processes satisfy

$$r_t = \mu_{t-1} + \sigma_t \epsilon_t, \quad (2.1)$$

$$\sigma_t^2 = b_{t-1} + (a_{t-1} + c_{t-1} \mathbb{1}_{(\epsilon_t < 0)}) \epsilon_t^2, \quad (2.2)$$

where  $\epsilon_t \sim i.i.d.(0, 1)$  with  $\mathbb{E}\epsilon_t^4 < \infty$  and  $\mathbb{1}_{(\cdot)}$  is the indicator function.  $\mu_t$ ,  $a_t$ ,  $b_t$  and  $c_t$  are all  $\mathcal{F}_t$ -measurable functions with  $b_t > 0$ ,  $a_t$  and  $c_t \geq 0$  with probability one. The term  $c_{t-1} \epsilon_t^2 \mathbb{1}_{(\epsilon_t < 0)}$  captures the skewness in the conditional density of  $r_t$  and the leverage effect. This leverage effect is contemporaneous, i.e., the correlation between the current negative returns and the current level of volatility. The lagged leverage effect, i.e., the correlation between the lagged negative returns and the current level of volatility, can be included in the term  $b_{t-1}$ . See Ding (2021) for more discussions on the contemporaneous and lagged

leverage effects. For SHARV, set  $\mu_t = 0$ ,  $b_t = \beta\sigma_t^2$ ,  $a_t = \alpha + \psi\sigma_t^2$  and  $c_t = 0$  while for ASHARV, reset  $\mu_t = \mu\sigma_t$  and  $c_t = \omega + \phi\sigma_t^2$ .<sup>4</sup> When  $a_t = 0$ ,  $c_t = 0$  and  $b_t = \alpha + \beta\sigma_t^2 + \gamma r_t^2$ , we obtain GARCH. We can relax the *i.i.d.* assumption on  $\epsilon_t$  by requiring  $\epsilon_t|\mathcal{F}_{t-1} \sim (0, 1)$  with  $\mathbb{E}[\epsilon_t^4|\mathcal{F}_{t-1}] < \infty$ . This results in a semi-strong CH-V model similar to the semi-strong GARCH of Drost and Nijman (1993). Most results of this paper continue to hold for the semi-strong case. Moreover, if we are only interested in the forecast of  $\sigma_t^2$ , we do not need the assumption of  $\mathbb{E}\epsilon_t^4 < \infty$ . The finite fourth moment condition is merely to ensure that  $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}] < \infty$ . For convenience, we will keep the *i.i.d.*(0, 1) and finite fourth moment assumptions in place. Empirical evidence in section 4 shows that the standardised return residuals for all three CH-V models that we consider in this paper are close to Gaussian for indices and individual stocks. Therefore, these assumptions can be empirically justified.

When  $a_t$  and/or  $c_t$  are not constants, the process  $\sigma_t^2$  exhibits conditional heteroskedasticity. Ding (2021) argues that ignoring the time-varying volatility of volatility results in a measurement in the volatility estimate. The volatility is no longer defined as the conditional variance of returns unless it degenerates to GARCH-type. Since financial returns are known to be heavy-tailed, the conditional variance is not a good measurement of the tail risk. The specification of the volatility process in CH-V can capture the time-varying tail index in the conditional density of returns. To see this, for simplicity, consider the symmetric CH-V where  $\mu_t = c_t = 0$  for all  $t$ . The conditional kurtosis

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

is larger than 3 when  $b_{t-1}^2(\mathbb{E}\epsilon_t^4 - 3) + 2a_{t-1}b_{t-1}(\mathbb{E}\epsilon_t^6 - 3\mathbb{E}\epsilon_t^4) + a_{t-1}^2(\mathbb{E}\epsilon_t^8 - 3\mathbb{E}\epsilon_t^4) > 0$ . For Gaussian  $\epsilon_t$ , this condition is satisfied for all  $a_{t-1} > 0$ . Therefore,  $a_{t-1}$ , and  $c_{t-1}$  in the case of asymmetric CH-V, not only captures the conditional heteroskedasticity in  $\sigma_t^2$  but also controls for the conditional kurtosis in  $r_t$ . For asymmetric CH-V, it is not difficult to see that  $\mathbb{E}[\sigma_t\epsilon_t|\mathcal{F}_{t-1}] < 0$  which in turn, induces a time-varying conditional skewness in  $r_t$ . The direction of the skewness is hard to determine analytically since  $\mathbb{E}[\sigma_t\epsilon_t|\mathcal{F}_{t-1}]$  does not have an analytical expression in general.

Ding (2021) argues that by ignoring the current observation in the volatility estimate, GARCH-type models disregard the (arguably) most important information which makes it difficult to justify the accuracy of volatility estimates produced by these models. On the other hand, the volatility filtering equation for CH-V is given by

$$\sigma_t^2 = 1/2\left(b_{t-1} + \sqrt{b_{t-1}^2 + 4(a_{t-1} + c_{t-1}\mathbb{1}_{(\epsilon_t < 0)})(r_t - \mu_{t-1})^2}\right).$$

In addition, the non-quadratic response of  $\sigma_t^2$  to  $r_t$  has the advantage of down-weighting the influence of large observations on the volatility estimates. See Harvey (2013) for the importance of down-weighting outliers in volatility estimation.

---

<sup>4</sup>We can reduce the number of parameters for ASHARV by specifying  $\sigma_t^2 = \beta\sigma_{t-1}^2 + (\alpha + \psi\sigma_{t-1}^2)(\epsilon_t^2 + \rho\epsilon_t^2\mathbb{1}_{(\epsilon_t < 0)})$ .

The restrictions on  $\mu_{t-1}$ ,  $a_{t-1}$ ,  $b_{t-1}$  and  $c_{t-1}$  are very weak. This allows for an abundance of functional forms for these terms that can be used to capture the persistence of both the volatility and volatility of volatility, lagged leverage effect, long-run and short-run effects as well as the addition of exogenous explanatory variables. If we require the return process to be a martingale difference sequence (MDS) as required under the efficient market hypothesis, we need to impose  $\mathbb{E}\epsilon_t^3 = 0$  and  $c_{t-1} = 0$ . CH-V can be viewed as a special case of the contemporaneous SV models in which the volatility is driven by the squared (contemporaneous) return innovation. The conditional density of returns is given in the following theorem.

**Theorem 2.1.** *Let  $(r_t, \sigma_t^2)$  satisfy (2.1)–(2.2). Let  $\tilde{r}_t \equiv r_t - \mu_{t-1}$  and  $\theta$  be the parameter vector. For  $y \neq 0$ , the conditional density of  $\tilde{r}_t | \mathcal{F}_{t-1}$  is given by*

$$f_r(y | \mathcal{F}_{t-1}) = \frac{y}{d_{1,t-1}(y; \theta) d_{2,t-1}(y; \theta)} f_\epsilon(d_{2,t-1}(y; \theta)), \quad (2.3)$$

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

$$d_{1,t-1}(y; \theta) = \sqrt{b_{t-1}^2 + 4a_{t-1}y^2 + 4c_{t-1}(y^-)^2}, \quad (2.4)$$

$$d_{2,t-1}(y; \theta) = \begin{cases} \text{sign}(y) \sqrt{\frac{d_{1,t-1}(y; \theta) - b_{t-1}}{2a_{t-1} + 2c_{t-1}\mathbb{1}_{(y < 0)}}}, & \text{if } (a_{t-1}, c_{t-1}) \neq 0 \\ y / \sqrt{b_{t-1}}, & \text{if } (a_{t-1}, c_{t-1}) = 0 \end{cases} \quad (2.5)$$

and  $\epsilon_t = d_{2,t-1}(\tilde{r}_t; \theta_0)$  at the true parameter vector  $\theta_0$ . For  $y = 0$ ,

$$\lim_{y \rightarrow 0} f_r(y | \mathcal{F}_{t-1}) = \frac{1}{\sqrt{b_{t-1}}} f_\epsilon(0). \quad (2.6)$$

The conditional cumulative distribution function (CDF) of  $\tilde{r}_t$  is given by  $F_r(y | \mathcal{F}_{t-1}) = F_\epsilon(d_{2,t-1}(y; \theta))$ , where  $F_\epsilon(\cdot)$  is the CDF of  $\epsilon_t$ .

**Remark 1.** *If we replace the i.i.d. assumption on  $\epsilon_t$  by  $\epsilon_t | \mathcal{F}_{t-1} \sim (0, 1)$ , then we need to replace  $f_\epsilon(\cdot)$  by  $f_{\epsilon_t | \mathcal{F}_{t-1}}(\cdot)$  in (2.3) and (2.6).*

The 1-step ahead  $p\%$  VaR forecast can be obtained by inverting the conditional probability  $\mathbb{P}(r_{t+1} \leq -\text{VaR}_{t+1} | \mathcal{F}_t)$  given in Theorem 2.1,

$$\text{VaR}_{t+1} = -\mu_t + \sqrt{(a_t + c_t \mathbb{1}_{(F_\epsilon^{-1}(p) < 0)})(F_\epsilon^{-1}(p))^4 + b_t(F_\epsilon^{-1}(p))^2}, \quad (2.7)$$

where  $F_\epsilon^{-1}(p)$  is the  $p\%$  quantile of  $\epsilon_t$ . See section 3 of Ding (2021) for the derivation.

We next discuss the stationarity conditions for CH-V models. The case when  $\sigma_t^2$  has an autoregressive structure will be of particular interest. Since the indicator function of a measurable set is a measurable function and therefore,  $\mathbb{1}_{(\epsilon_t < 0)}$  is strictly stationary and ergodic. Therefore, without loss of generality, we assume  $c_t = 0$  for all  $t$  and  $\sigma_t^2$  satisfies

$$\sigma_t^2 = \phi_t \sigma_{t-1}^2 + \tilde{a}_{t-1} \epsilon_t^2, \quad (2.8)$$

where  $\phi_t > 0$ ,  $\tilde{a}_t \geq 0$  almost surely,  $\phi_t$  is a measurable function of  $\epsilon_t$  and  $\tilde{a}_t$  is contemporaneously independent of  $\sigma_t^2$ . Setting  $\phi_t = \beta + \psi\epsilon_t^2$  and  $\tilde{a}_{t-1} = \alpha$ , we obtain SHARV. Before we state strict stationarity conditions, we need a further assumption on  $\epsilon_t$ .

**Assumption 1.**  $\epsilon_t^2$  is non-degenerate ( $\epsilon_t$  thus need to be different from scaled symmetric Bernoulli or denerate random variables).

**Theorem 2.2.** Let  $(r_t, \sigma_t^2)$  be generated (2.1) and (2.8) and  $\epsilon_t$  satisfy Assumption 1. Let  $\hat{a}_t \equiv \tilde{a}_{t-1}\epsilon_t^2$  be strictly stationary and ergodic. Define  ${}_u\sigma_t^2 \equiv \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j}\hat{a}_{t-k}$ , where  $\prod_{j=0}^{-1} \phi_{t-j} = 1$ . If

$$\mathbb{E} \log \phi_t < 0 \text{ and } \mathbb{E}(\log \tilde{a}_t)^+ < \infty, \quad (2.9)$$

where  $x^+ = \max(x, 0)$ , then  ${}_u\sigma_t^2$  is the unique strictly stationary and ergodic solution of (2.8) given  $\phi_t$  and  $\hat{a}_t$ . Furthermore,  ${}_u\sigma_t^2$  converges absolutely almost surely and  $|\sigma_t^2 - {}_u\sigma_t^2| \rightarrow 0$  almost surely for an arbitrary random starting point  $\sigma_0^2$  defined on the same probability space as  $\phi_t$  and  $\hat{a}_t$ . In particular,  $\sigma_t^2 \xrightarrow{d} {}_u\sigma_t^2$  as  $t \rightarrow \infty$ . Moreover, condition (2.9) is the necessary condition for the results to hold.

**Remark 2.** Theorem 2.2 also nests the strict stationarity conditions for GARCH if we set  $\phi_t = \beta + \gamma\epsilon_{t-1}^2$  and  $\hat{a}_t = \alpha$  and for E-GARCH by setting  $\sigma_t^2 = \log h_t^2$ ,  $\phi_t = \beta$  and  $\hat{a}_t = \omega + \alpha(|\epsilon_{t-1}| - \mathbb{E}|\epsilon_{t-1}|) + \gamma\epsilon_{t-1}$ .

**Corollary 2.2.1.** If  $\mathbb{E} \log \phi_t \geq 0$ , then  $\sigma_t^2 \rightarrow \infty$  almost surely as  $t \rightarrow \infty$ . If we relax the i.i.d. assumption and assume only strictly stationary and ergodic  $\epsilon_t$ , then with an additional assumption (Assumption A2 of Linton et al., 2010), we have the same conclusion.

Finally, If  $0 \leq \mathbb{E}\tilde{a}_t = \tilde{a} < \infty$  and  $0 < \mathbb{E}\phi_t = \phi < 1$ , where  $\tilde{a}$  and  $\phi$  are constants and do not depend on  $t$ , then  $\sigma_t$  is weakly stationary.

### 3 GARCH in the volatility

In this section, we discuss two examples of CH-V.

#### 3.1 GJR-GARCH-V

In the first example, we specify a GJR-GARCH process for  $\tilde{a}_{t-1}$ . Let  $(r_t, \sigma_t^2, v_t)$  satisfy

$$r_t = \sigma_t \epsilon_t, \quad (3.1)$$

$$\sigma_t^2 = \phi\sigma_{t-1}^2 + v_t\epsilon_t^2, \quad (3.2)$$

$$v_t = \omega + \beta v_{t-1} + (\alpha + \gamma \mathbb{1}_{(\epsilon_{t-1} < 0)})v_{t-1}\epsilon_{t-1}^2, \quad (3.3)$$

where  $\phi > 0$  and  $(\omega, \alpha, \beta, \gamma)' \geq 0$ . We call (3.1)–(3.3) the GJR-GARCH-V model. It is easy to see that  $r_t$  is an MDS if  $\epsilon_t$  are symmetric. Hence similar to SHARV, we generally

cannot allow for skewness in  $\epsilon_t$  and  $\mathbb{E}[r_t|\mathcal{F}_{t-1}] = 0$  simultaneously. Although symmetry of  $\epsilon_t$  is quite a strong restriction, skewness is usually negligible for short horizon returns (Neuberger and Payne, 2020). Thus for convenience, we will assume  $\mathbb{E}\epsilon_t^3 = 0$  for the rest of the paper in addition to the *i.i.d.* and moment assumptions in section 2. Note that for long horizon returns, our model can still generate skewness because of the leverage effect coupled with volatility persistence. We collect all assumptions on  $\epsilon_t$  and state as follows,

**Assumption 2.** Let  $\epsilon_t$  be *i.i.d.*(0, 1) with  $\mathbb{E}\epsilon_t^3 = 0$  and  $\mathbb{E}\epsilon_t^4 < \infty$ .

At first glance, (3.3) does not have the familiar GARCH form. However, upon noticing that  $v_{t-1}\epsilon_{t-1}^2 = \sigma_{t-1}^2 - \phi\sigma_{t-2}^2$ , it is immediate that  $v_t$  is driven by the volatility residuals. Note that in (3.3) we have modelled the conditional standard deviation  $v_t$  instead of the conditional variance  $v_t^2$ . This is because in order to forecast  $\sigma_{t+n}^2$  for  $n > 1$ , we need to forecast  $v_{t+n}$  first.

We can express (3.1)–(3.3) as a VARMA(1,1) process. Specifically,

$$\begin{bmatrix} r_t^2 \\ \sigma_t^2 \\ v_t \end{bmatrix} = \begin{bmatrix} \omega\iota_4 \\ \omega \\ \omega \end{bmatrix} + \begin{bmatrix} \phi & 0 & \xi\iota_4 - \kappa\phi \\ 0 & \phi & \xi \\ 0 & 0 & \xi \end{bmatrix} \begin{bmatrix} r_{t-1}^2 \\ \sigma_{t-1}^2 \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_t \\ e_t \\ u_t \end{bmatrix} + \begin{bmatrix} -\phi & \alpha\iota_4 & \gamma\iota_4 \\ 0 & \alpha & \gamma \\ 0 & \alpha & \gamma \end{bmatrix} \begin{bmatrix} z_{t-1} \\ e_{t-1} \\ u_{t-1} \end{bmatrix}, \quad (3.4)$$

where  $\iota_4 = \mathbb{E}\epsilon_t^4$ ,  $\kappa = \iota_4 - 1$ ,  $\xi = \beta + \alpha + 1/2\gamma$  and  $(z_t, e_t, u_t)'$  are all MDS given by

$$z_t = r_t^2 - \sigma_t^2 - \kappa v_t, \quad (3.5)$$

$$e_t = v_t(\epsilon_t^2 - 1), \quad (3.6)$$

$$u_t = v_t(\epsilon_t^2 \mathbb{1}_{(\epsilon_t < 0)} - \frac{1}{2}). \quad (3.7)$$

Unlike the first-order GARCH and SHARV where the volatility follows an AR(1) process, the volatility in GJR-GARCH-V follows an ARMA(1,1) process. Moreover, the volatility of volatility is no longer solely determined by the lagged level of volatility but follows an AR(1) process instead and affects the current level of volatility. This allows us to analyse the volatility and volatility of volatility separately. The leverage effect is captured by the term  $\gamma \mathbb{1}_{(\epsilon_{t-1} < 0)} v_{t-1} \epsilon_{t-1}^2$  in (3.3), which transfers into (3.2) through  $v_t$ . This demonstrates a novel way of introducing the leverage effect via the volatility of volatility process. Note that in GJR-GARCH-V, we have the lagged leverage effect instead of the contemporaneous one in ASHARV. It is straightforward to model the contemporaneous leverage effect by including an additional term  $\rho v_t \epsilon_t^2 \mathbb{1}_{(\epsilon_t < 0)}$  in (3.2). In doing so,  $r_t$  is no longer an MDS. In this paper, we focus on the case when  $\mathbb{E}[r_t|\mathcal{F}_{t-1}] = 0$  and assume only the lagged leverage effect. We leave the simultaneous treatment of the contemporaneous and lagged leverage effects for future research.

The conditional variance of returns is given by  $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}] = \phi\sigma_{t-1}^2 + v_t\mathbb{E}\epsilon_t^4$  and the conditional variance of the volatility is given by  $\text{Var}(\sigma_t^2|\mathcal{F}_{t-1}) = v_t^2(\mathbb{E}\epsilon_t^4 - 1)$ . The strict stationarity and ergodicity conditions are given in the following proposition.

**Proposition 3.1.** *The processes (3.2) and (3.3) have unique strictly stationary and ergodic solutions if and only if  $\phi < 1$  and  $\mathbb{E} \log (\beta + (\alpha + \gamma \mathbb{1}_{(\epsilon_t < 0)}) \epsilon_t^2) < 0$ .*

The weak stationarity conditions are given by  $\phi < 1$  and  $\beta + \alpha + 1/2\gamma < 1$ . The unconditional level of  $\sigma_t^2$  is given by  $\mathbb{E}\sigma_t^2 = \mathbb{E}v_t/(1 - \phi)$  where  $\mathbb{E}v_t = \omega/(1 - \beta - \alpha - 1/2\gamma)$ . Finally, if  $v_t$  follows an integrated GJR-GARCH process, i.e.  $\beta + \alpha + 1/2\gamma = 1$ ,  $\sigma_t^2$  and  $v_t$  can still be strictly stationary provided the conditions in Proposition 3.1 are satisfied. On the other hand, the volatility process itself cannot follow an integrated process, that is,  $\phi$  cannot be equal to 1 if we want to preserve the strict stationarity of  $\sigma_t^2$ .

### 3.2 E-GARCH-V

We next introduce another example of CH-V in which the volatility of volatility follows an E-GARCH process. Let  $(r_t, \sigma_t^2, v_t)$  satisfy

$$r_t = \sigma_t \epsilon_t, \tag{3.8}$$

$$\sigma_t^2 = \phi \sigma_{t-1}^2 + v_t \epsilon_t^2, \tag{3.9}$$

$$\log v_t = \omega + \beta \log v_{t-1} + \gamma \epsilon_{t-1} + \alpha(|\epsilon_{t-1}| - \mathbb{E}|\epsilon_{t-1}|). \tag{3.10}$$

We call (3.8)–(3.10) the E-GARCH-V model. To ensure that  $\sigma_t^2 > 0$  almost surely, we only require that  $\phi > 0$ . Clearly,  $\mathbb{E}(\log v_t)^+ < \infty$  if  $|\beta| < 1$ , together with  $\phi < 1$  we have the strict stationarity conditions. These conditions coincide with the weak stationarity conditions and are easy to verify in practice. Moreover, E-GARCH-V allows the volatility of volatility to exhibit oscillatory behaviour since no positivity constraint is imposed on its parameters. This in turn, allows for oscillations in the volatility. Note that we can model either the log conditional variance or log conditional standard deviation of  $\sigma_t^2$  in (3.10) since  $\mathbb{E} \log v_t^2 = 2\mathbb{E} \log v_t$  for  $v_t > 0$ . To be consistent with GJR-GARCH-V, we proceed with modelling the log conditional standard deviation of  $\sigma_t^2$ , i.e.,  $\log v_t$ .

The disadvantage of E-GARCH-V, similar to E-GARCH, is that the multi-step ahead volatility forecasts directly depends on the distributional assumption of  $\epsilon_t$  since  $\mathbb{E} \exp(\epsilon_t)$  depends on the distribution of  $\epsilon_t$ . Empirical evidence in section 4.2 suggests that  $\epsilon_t$  are close to Gaussian for indices and individual stocks. Therefore, we assume  $\epsilon_t \sim N(0, 1)$  for volatility forecasts purposes for E-GARCH-V. On the other hand,  $\epsilon_t$  for E-GARCH are far from Gaussian and exhibit heavy tails. It is therefore, more reasonable to justify the distributional (Gaussian) assumption of  $\epsilon_t$  for E-GARCH-V than for E-GARCH empirically. This is important for multi-step ahead volatility forecasts. As Nelson (1991) points out, the moments of E-GARCH do not typically exist for heavy-tailed  $\epsilon_t$ . The moment of  $\gamma \epsilon_t + \alpha(|\epsilon_t| - \mathbb{E}|\epsilon_t|)$  is given in Theorems A1.1 and A1.2 of Nelson (1991). Note that the case of constant volatility under E-GARCH-V is not trivial. This is because in order for  $v_t = 0$  for all  $t$ , we would require  $\log v_t = -\infty$  for all  $t$ . If we want to nest this case, we can replace  $v_t \epsilon_t^2$  by  $\eta v_t \epsilon_t^2$  in (3.9). If  $\phi = 1$  and  $\eta = 0$ , we obtain the case of constant

Table 1: Parameter estimates for GJR- and E-GARCH-V

	$\phi$	$\omega$	$\beta$	$\alpha$	$\gamma$	BIC
S&P 500	0.8731 (0.0088)	0.0119 (0.0048)	0.8830 (0.0296)	0.0213 (0.0087)	0.1826 (0.0476)	23221
	0.8745 (0.0087)	-0.0169 (0.0068)	0.9712 (0.0096)	0.2024 (0.0370)	-0.1234 (0.0289)	23228
DJIA	0.8744 (0.0093)	0.0127 (0.0049)	0.8743 (0.0307)	0.0061 (0.0024)	0.2035 (0.0513)	23030
	0.8763 (0.0095)	-0.0206 (0.0093)	0.9670 (0.0122)	0.2039 (0.0426)	-0.1403 (0.0339)	23036
JPM	0.8859 (0.0090)	0.0075 (0.0041)	0.9325 (0.0119)	0.0294 (0.0104)	0.0706 (0.0197)	28425
	0.8870 (0.0090)	0.0022 (0.0014)	0.9930 (0.0024)	0.1402 (0.0236)	-0.0426 (0.0119)	28443
AAPL	0.9057 (0.0083)	0.0029 (0.0013)	0.9598 (0.0003)	0.0355 (0.0053)	0.0085 (0.0103)	30331
	0.9078 (0.0089)	0.0025 (0.0013)	0.9960 (0.0028)	0.0909 (0.0253)	-0.0047 (0.0091)	30351
EUR/USD	0.9429 (0.0096)	0.0002 (0.0005)	0.9632 (0.0213)	0.0238 (0.0314)	0.0249 (0.0160)	10403
	0.9468 (0.0084)	-0.0111 (0.0128)	0.9953 (0.0048)	0.0808 (0.0306)	-0.0278 (0.0098)	10403

*Note:* The first row for each asset refers to GJR-GARCH-V, followed by E-GARCH-V. The standard errors in parentheses are calculated numerically. The BIC of SHARV for all above assets are 23254, 23071, 28437, 30335 and 10395, respectively. The BIC of E-GARCH and GJR-GARCH are larger in values than those of the three models for all assets, meaning CH-V-type models are always preferred over GARCH-type models.

volatility. Since conditional heteroskedasticity is a stylised fact of asset returns, we will proceed with (3.9) without the additional parameter  $\eta$  for simplicity.

## 4 Empirical analysis

### 4.1 Data description

Our primary sample consists of daily open-to-close returns of S&P 500, Dow Jones Industrial Average (DJIA), JPMorgan Chase & Co. (JPM) and Apple Inc. (AAPL) spanning from 3 January 2000 to 31 December 2020 and EUR/USD exchange rate from 2 January 2009 to 31 December 2020. We use the bipower variation (BPV) of Barndorff-Nielsen and Shephard (2003) calculated from 5-min intraday returns as volatility proxy to eliminate the effect of large jumps on volatility. This is important since COVID-19 is likely to have created large jumps in asset returns during 2020.

### 4.2 In-sample analysis

The Gaussian quasi-maximum likelihood estimates (QMLE) for GJR- and E-GARCH-V are reported in Table 1. Note that we have annualised the log returns and imposed the

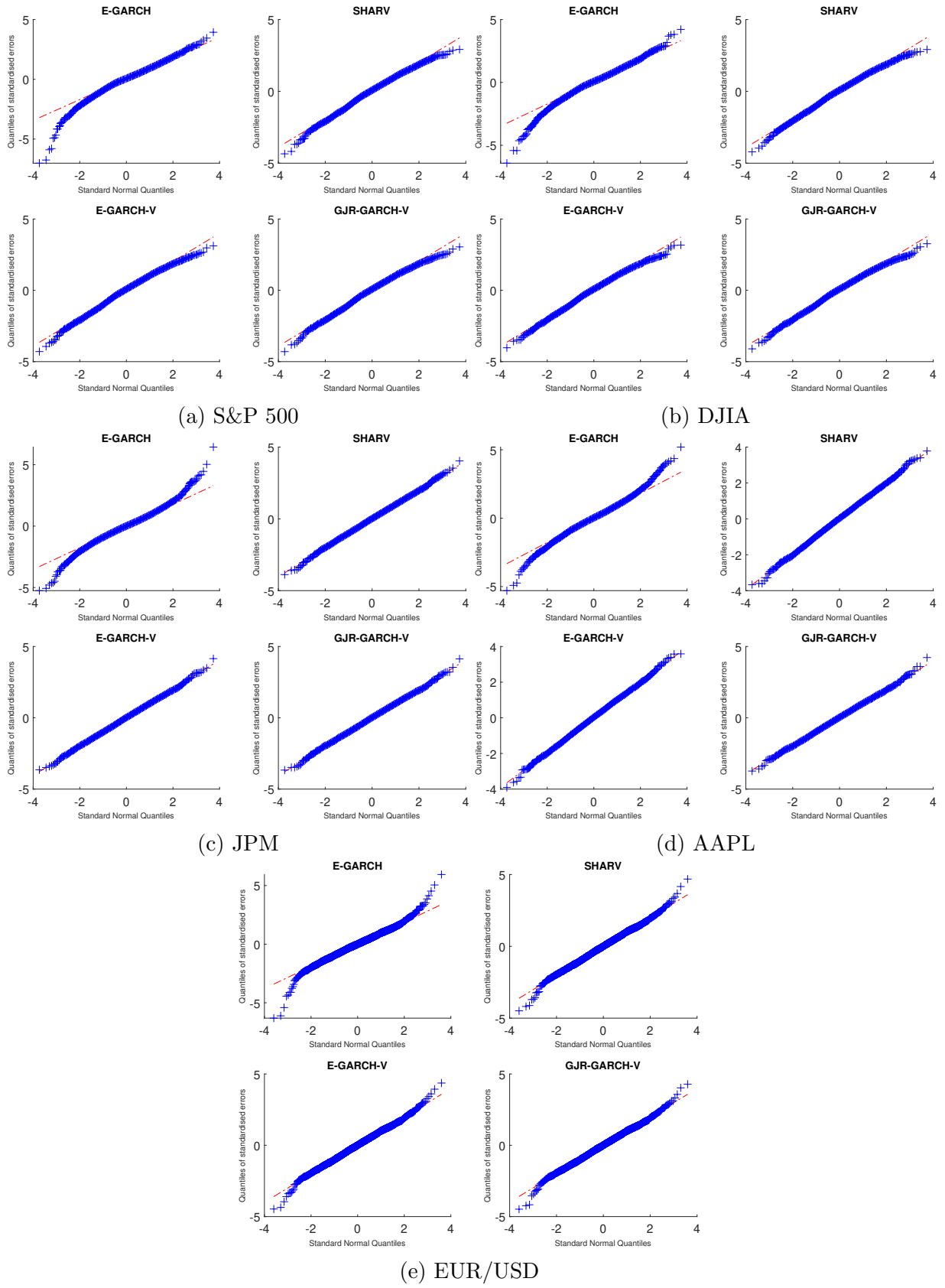


Figure 1: QQ plots of the standardised return residuals for the full sample

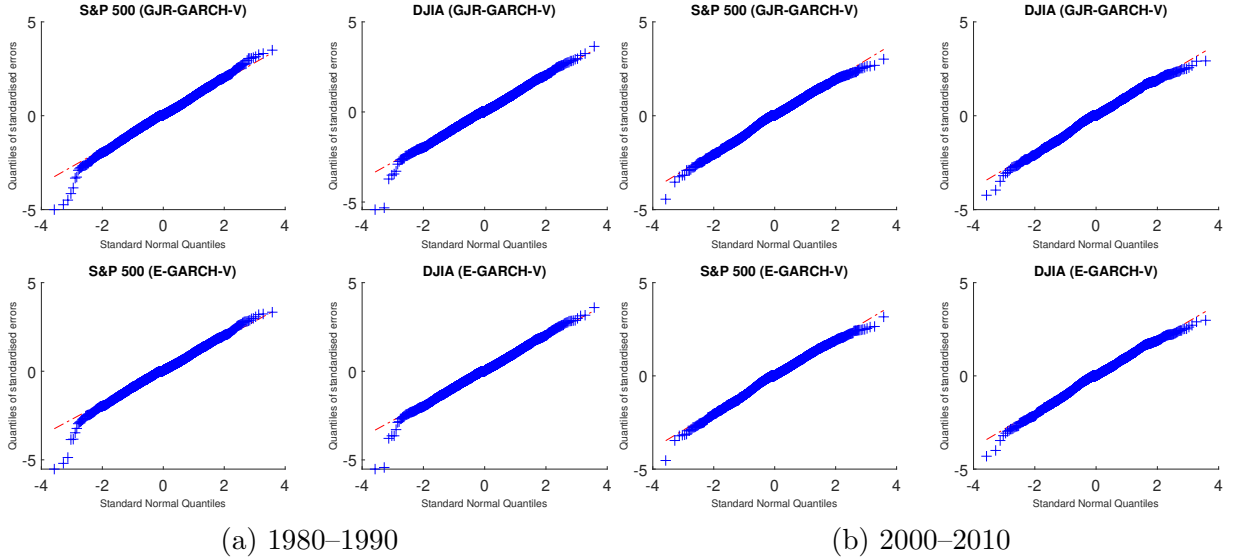


Figure 2: QQ plots of the standardised return residuals for 1980–1990 and 2000–2010

weak stationarity conditions. For brevity, we do not report the parameter estimates for SHARV and GARCH. In terms of BIC, the three CH-V models are all preferred over GJR- and E-GARCH. For the two GARCH-V models, both the volatility and volatility of volatility are highly persistent. The leverage effects are significant for indices and JPM but less so for AAPL and EUR/USD. Recall from section 5 of Ding (2021), the contemporaneous leverage effect is directly ruled out by QML for EUR/USD. This demonstrates that there is a distinction between the contemporaneous and lagged leverage effects which deserves a more detailed analysis for future research.

The QQ plots of the standardised return residuals for CH-V and E-GARCH are shown in Figure 1. The standardised return residuals for the three CH-V models are almost identical to standard normal in terms of their quantiles except for EUR/USD. For EUR/USD, the three CH-V models still have better goodness-of-fit than E-GARCH. Within the three CH-V models, the differences are marginal.

The (nearly) Gaussianity of standardised return residuals has important implications. Clark (1973) argues that the non-Gaussianity and heavy-tails of asset returns arises from the random news arrival process and the difference between clock time and market time scales. Therefore, conditional on the correct news arrival process, we should be able to recover the Gaussianity of asset returns. The QQ plots in Figure 1 seem to suggest that both SHARV and GARCH-V correctly capture the news arrival process for indices and individual stocks in the period of 2000–2020. We next consider the periods of 1980–1990 and 2000–2010 which cover the 1987 stock market crash and the latest financial crisis, respectively. The QQ plots for SHARV are similar to those for GARCH-V while for E-GARCH, the standardised return residuals are far from Gaussian. Therefore for brevity, we do not report the QQ plots for them. It is clear from Figure 2 that  $\epsilon_t$  are still close to Gaussian for GARCH-V in the period of 2000–2010. However, in the period of 1980–

Table 2: Comparison of standard errors of QMLE and MLE for GARCH-V for JPM

	$\phi$	$\omega$	$\beta$	$\alpha$	$\gamma$
GJR-GARCH-V					
$(\Sigma_1^{-1})_{ii}$	0.0087	0.0021	0.0090	0.0122	0.0151
$(\Sigma_2^{-1})_{ii}$	0.0088	0.0028	0.0099	0.0110	0.0165
$(\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})_{ii}$	0.0090	0.0041	0.0119	0.0104	0.0197
E-GARCH-V					
$(\Sigma_1^{-1})_{ii}$	0.0084	0.0014	0.0017	0.0178	0.0103
$(\Sigma_2^{-1})_{ii}$	0.0087	0.0014	0.0020	0.0199	0.0107
$(\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})_{ii}$	0.0090	0.0014	0.0024	0.0234	0.0119

Note:  $\Sigma_1 = \mathbb{E}_{\theta_0} \left[ \frac{\partial l_t(\theta_0)}{\partial \theta_0} \frac{\partial l_t(\theta_0)}{\partial \theta_0'} \right]$  and  $\Sigma_2 = -\mathbb{E}_{\theta_0} \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta_0 \partial \theta_0'} \right]$  where  $l_t(\theta_0)$  is the quasi-log-likelihood function evaluated at the true parameter vector. The standard errors are calculated from the square root of the diagonal elements of these matrices.

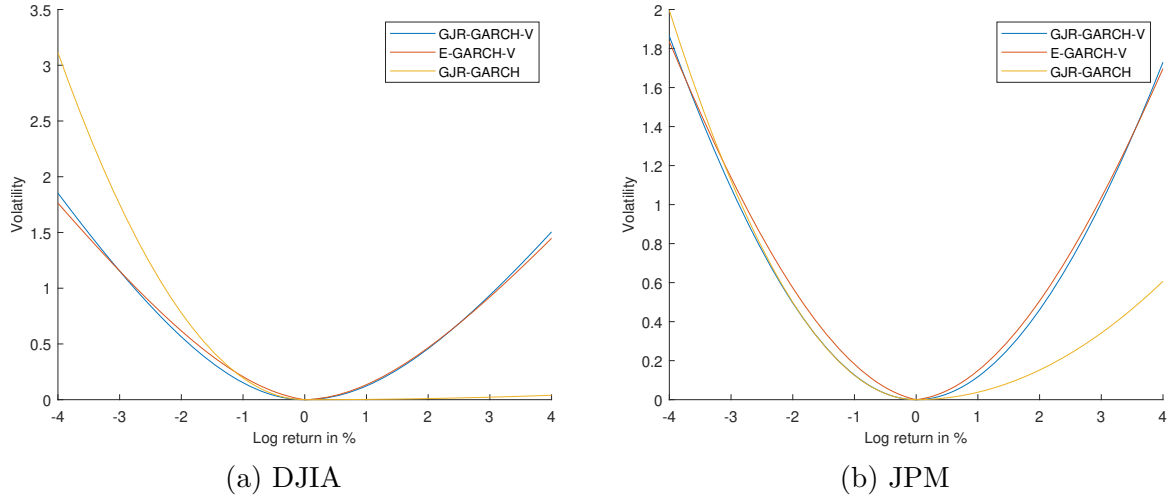


Figure 3: News impact curves of GJR-GARCH-V, E-GARCH-V and GJR-GARCH

1980, the left tail of  $\epsilon_t$  starts to deviate from Gaussian although it is still much closer to Gaussian than E-GARCH. Events like 1987 are relatively rare, we can therefore include a jump component in the return process similar to the GARCH-Jump model of Maheu and McCurdy (2004) and see whether the standardised return residuals will be even closer to Gaussian in the period of 1980–1990. We leave this for future research.

Another implication concerns the efficiency of the Gaussian QMLE which approaches that of the MLE since  $\epsilon_t$  are close to Gaussian for CH-V. This can be seen by comparing the standard errors of the QMLE and the MLE. Ding (2021) has compared the standard errors for SHARV and ASHARV. Therefore, we only compare the standard errors for the two GARCH-V models for JPM for brevity. From Table 2, it is clear that the standard errors of the Gaussian QMLE are almost identical to those of the MLE.

We next consider the news impact curves. Since the class of CH-V models does not nest E-GARCH, we use GJR-GARCH as benchmark for comparison. From Figure 3, it is clear that the magnitude of the leverage effect is much smaller for GARCH-V models than

Table 3: Volatility nowcast comparison using BPV as volatility proxy

	S&P 500		DJIA		JPM		AAPL		EUR/USD	
	MSE	$p_{MCS}$	MSE	$p_{MCS}$	MSE	$p_{MCS}$	MSE	$p_{MCS}$	MSE	$p_{MCS}$
GJR-GARCH	0.1328	0.000	0.1384	0.000	0.3060	0.003	0.3632	0.000	0.0229	0.000
E-GARCH	0.1234	0.000	0.1303	0.000	0.3156	0.000	0.3745	0.000	0.0221	0.000
SHARV	0.0988	0.670*	0.1083	0.776*	0.2776	1.000*	0.2918	0.752*	0.0187	0.135*
GJR-GARCH-V	0.0978	0.670*	0.1067	0.781*	0.2798	0.753*	0.2896	1.000*	0.0184	1.000*
E-GARCH-V	0.0968	1.000*	0.1060	1.000*	0.2845	0.021	0.2944	0.000	0.0187	0.327*

Note:  $p_{MCS}$  are the p-values of the model confidence set of Hansen et al. (2011). Models marked with \* fall in the 95% model confidence set.

for GJR-GARCH. This is because the leverage effect in GARCH-V is introduced via the volatility of volatility whose magnitude is often much smaller than that of the volatility itself. We can include another leverage effect term in the volatility process to increase its magnitude. We can use this additional term to capture either the contemporaneous leverage effect in the fashion of ASHARV (see Ding, 2021) or the lagged leverage effect in the fashion of GJR-GARCH. The simultaneous treatment of the contemporaneous and lagged leverage effects is of particular interest. We leave this for future research.

To show that CH-V models have more accurate volatility nowcasts, we next compare the mean square errors (MSE) and the 95 percentile model confidence set (MCS)  $\hat{\mathcal{M}}_{95\%}^*$  of Hansen et al. (2011) in Table 3. Clearly, CH-V models consistently outperform GARCH-type models since they are the only models in the 95% MCS. For individual stocks, E-GARCH-V is outside the 95% MCS. However, it still outperforms GJR- and E-GARCH according to the superior predictive ability (SPA) test of Hansen (2005). For brevity, we do not report the SPA test results. We then plot the filtered volatility ( $\sigma_t$ ) and volatility of volatility ( $\sqrt{\mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}]}$ ) of E-GARCH-V for DJIA and JPM for the year of 2020 in Figure 4. The dynamics of the volatility is generally in-line with that of the volatility of volatility while the volatility of volatility exhibits more fluctuations than the volatility. This is in contrast to SHARV where the current level of volatility of volatility is entirely determined by the lagged level of volatility. For GARCH-V, the volatility of volatility has its own dynamics which feeds back into the current level of volatility.

### 4.3 Forecast evaluation

For out-of-sample volatility and VaR forecasts, we divide the sample into the estimation and forecast periods. The forecast period contains the last 1500 observations. We employ an expanding window and update the estimation for every 50 observations. For volatility forecast comparison, we use the QLIKE loss function which is robust to imperfect volatility proxies as shown by Patton (2011). From Table 4, it is evident that CH-V consistently outperform GJR- and E-GARCH since they are the only models in the 95% MCS for all assets except for JPM where all models are in the 95% MCS. Within the class of CH-V, E-GARCH-V is outside the 95% MCS for AAPL and EUR/USD. However, even

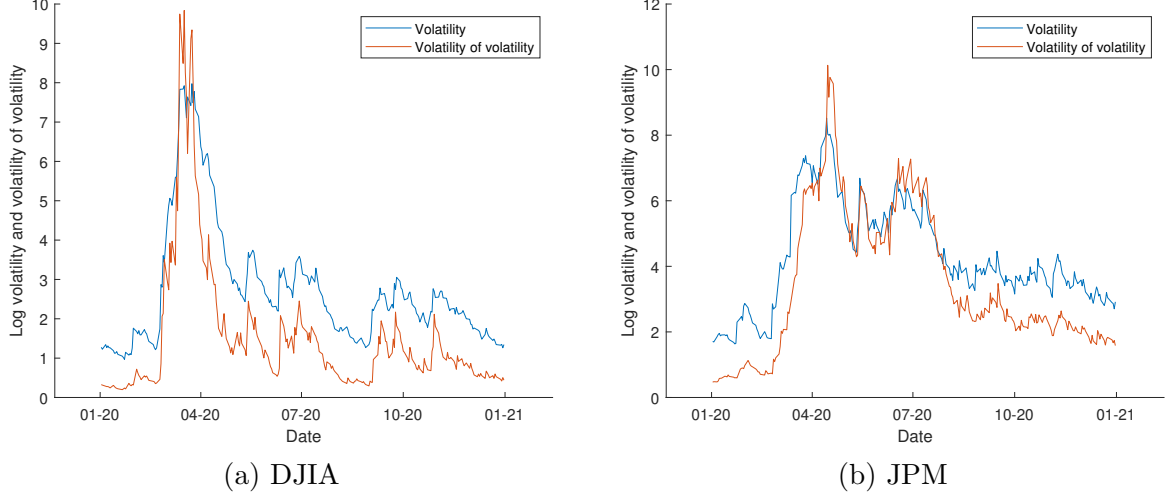


Figure 4: Filtered volatility ( $\sigma_t$ ) and volatility of volatility ( $\sqrt{\mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}]}$ ) of E-GARCH-V

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

	S&P 500		DJIA		JPM		AAPL		EUR/USD	
	QLIKE	$p_{MCS}$	QLIKE	$p_{MCS}$	QLIKE	$p_{MCS}$	QLIKE	$p_{MCS}$	QLIKE	$p_{MCS}$
1-step										
GJR-GARCH	0.1069	0.000	0.1102	0.000	0.0616	0.506*	0.0780	0.000	0.0530	0.000
E-GARCH	0.0998	0.000	0.1041	0.000	0.0627	0.436*	0.0780	0.000	0.0497	0.001
SHARV	0.0832	1.000*	0.0837	1.000*	0.0600	1.000*	0.0667	0.090*	0.0431	1.000*
GJR-GARCH-V	0.0853	0.126*	0.0864	0.067*	0.0618	0.436*	0.0643	1.000*	0.0451	0.065*
E-GARCH-V	0.0847	0.199*	0.0861	0.094*	0.0620	0.436*	0.0708	0.000	0.0474	0.004
5-step										
GJR-GARCH	0.1469	0.000	0.1526	0.000	0.0790	0.260*	0.0958	0.000	0.0584	0.001
E-GARCH	0.1366	0.000	0.1423	0.000	0.0802	0.260*	0.0971	0.000	0.0546	0.035
SHARV	0.1176	1.000*	0.1212	1.000*	0.0762	1.000*	0.0837	0.404*	0.0499	1.000*
GJR-GARCH-V	0.1222	0.057*	0.1262	0.055*	0.0796	0.376*	0.0818	1.000*	0.0527	0.071*
E-GARCH-V	0.1210	0.158*	0.1244	0.221*	0.0798	0.260*	0.0908	0.000	0.0548	0.035

Note:  $p_{MCS}$  are the p-values of the model confidence set of Hansen et al. (2011). Models marked with \* fall in the 95% model confidence set.

in this case, E-GARCH-V still outperforms GJR- and E-GARCH according to the SPA test results. This confirms the importance of the current observation and time-varying volatility of volatility in volatility forecasts. On the other hand, the specifications of the volatility of volatility in SHARV and GARCH-V only result in marginal differences. This is because the dynamics of the volatility and volatility of volatility are largely in-line as seen in Figure 4, therefore, a function of the lagged level of volatility can provide a good approximation of the volatility of volatility as in the case of SHARV.

Finally, we compare 1-step ahead 1% and 5% VaR forecasts. We use the conditional coverage test of Christoffersen (1998) to evaluate the accuracy and report the p-values in Table 5. Clearly, only GARCH-V have correct conditional coverage for all assets and for both percentiles. Moreover, E-GARCH-V is the only model with correct conditional coverage for 1% VaR forecast for S&P 500. The results are consistent with Ding (2021) where he finds that both the time-varying volatility of volatility and leverage effect are

Table 5: P-values of the conditional coverage test for 1-step ahead VaR forecasts

	S&P 500		DJIA		JPM		AAPL		EUR/USD	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
GJR-GARCH	0.008	0.730*	0.002	0.325*	0.123*	0.074*	0.227*	0.025	0.006	0.004
E-GARCH	0.000	0.902*	0.008	0.658*	0.123*	0.086*	0.049	0.031	0.006	0.018
SHARV	0.004	0.500*	0.000	0.063*	0.817*	0.166*	0.812*	0.021	0.570*	0.053*
GJR-GARCH-V	0.007	0.876*	0.065*	0.146*	0.812*	0.123*	0.710*	0.124*	0.680*	0.135*
E-GARCH-V	0.065*	0.858*	0.340*	0.189*	0.710*	0.299*	0.812*	0.132*	0.447*	0.001

*Note:* The conditional coverage test has an asymptotic  $\chi^2$  distribution with 2 degrees of freedom (Christoffersen, 1998). Models marked with \* are those we fail to reject the null at 95% confidence level, that is, the proportion of failures is consistent with the VaR confidence level and failures on consecutive time periods are independent.

important for VaR forecasts. We point out that the leverage effects in GARCH-V in the paper are lagged leverage effects, the analysis of the contemporaneous leverage effect is left for future research.

## 5 Conclusion

In this paper, we have proposed a new class of CH-V models which nests both SHARV and several GARCH-type models. We have discussed the importance of modelling the volatility of volatility and derived some statistical properties of CH-V models including the conditional density of returns and the stationarity conditions. Subsequently, we have introduced two examples of CH-V: GJR- and E-GARCH-V. We have also demonstrated a novel way of modelling the leverage effect through the volatility of volatility process. Empirical analysis showed that the standardised return residuals of CH-V are close to Gaussian for indices and individual stocks. Finally, we concluded that CH-V have more accurate volatility nowcasts and forecasts than GARCH-type models. Within the CH-V, GARCH-V models have more accurate VaR forecasts than SHARV.

Further works are still needed to complete the theory and practice of CH-V models. It remains to develop a multivariate version and an asymptotic theory for the Gaussian QMLE. It would be useful to derive the diffusion limit of the two GARCH-V models to establish the link to asset pricing theories. For empirical studies, it would be helpful to consider the simultaneous treatment of the contemporaneous and lagged leverage effects and the impacts on volatility and VaR forecasts. Finally, using heavy-tailed innovations and adding a jump component in the return process would provide more insights into different ways of modelling excess kurtosis in financial returns.

## Acknowledgements

I am grateful to my supervisor, Oliver Linton, for his valuable suggestions and support. I thank Mark Salmon for helpful discussions. The Cambridge Trust and the Cambridge-

INET provided financial support.

## A Proof of main theorems

*Proof of Theorem 2.1.* Since  $b_t > 0$  with probability one, when  $(a_t, c_t)' = 0$ , the conditional density of  $\tilde{r}_t$  can be obtained by a change of variable argument. Therefore, we focus on the case when either  $a_t > 0$  or  $c_t > 0$  or both. In this case,  $\sigma_t^2$  is strictly increasing in  $\epsilon_t$ , so is  $\tilde{r}_t$ . Therefore, for each  $y$ , there exists a unique  $\bar{\epsilon}$ , such that for all  $\epsilon_t \leq \bar{\epsilon}$ ,

$$\tilde{r}_t = \sqrt{b_{t-1} + (a_{t-1} + c_{t-1} \mathbb{1}_{(\epsilon_t < 0)}) \epsilon_t^2} \leq y, \quad (\text{A.1})$$

holds and for all  $\epsilon_t \geq \bar{\epsilon}$ , the direction of the inequality sign flips in (A.1). The rest follows exactly the proof of Theorem 3.1 of Ding (2021).  $\square$

*Proof of Theorem 2.2.* By repeated substitution for  $\sigma_t^2$  in (2.8), we obtain

$$\sigma_t^2 = \prod_{j=0}^{t-1} \phi_{t-j} \sigma_0^2 + \sum_{k=0}^{t-1} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}, \quad (\text{A.2})$$

where

$$\mathbb{P}((\hat{r}_0, \sigma_0^2) \in \Gamma) = \nu_0(\Gamma) \text{ for any } \Gamma \in B(R^2), \quad (\text{A.3})$$

where  $B(R^n)$  denote the Borel sets on  $R^n$  and  $\nu_0(\cdot)$  is a probability measure such that  $\mathbb{P}(0 < \sigma_0^2 < \infty) = 1$ .

We first show that  ${}_u\sigma_t^2 < \infty$  almost surely. Since  $\phi_t$  is a measurable function of  $\epsilon_t$ , it is strictly stationary and ergodic. By the strong law of large numbers for strictly stationary and ergodic processes, we have

$$\limsup_{k \rightarrow \infty} k^{-1} \left( \sum_{j=0}^{k-1} \log \phi_{t-j} + \log \hat{a}_{t-k} \right) < 0 \quad \text{a.s.},$$

if and only if  $\mathbb{E} \log \phi_t < 0$  and  $\mathbb{E}(\log \hat{a}_t)^+ < \infty$ . Therefore,

$$\limsup_{k \rightarrow \infty} \left( \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k} \right)^{1/k} < 1 \quad \text{a.s..}$$

By Cauchy's root test, the series  $\sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}$  converges absolutely almost surely. Condition  $\mathbb{E}(\log \hat{a}_t)^+ < \infty$  is equivalent to  $-\infty \leq \mathbb{E} \log \hat{a}_t < \infty$  by Lebesgue integrability. By definition,  $\mathbb{E} \log \hat{a}_t = \mathbb{E}[\log \tilde{a}_t + \log \epsilon_t^2]$ . Since  $\mathbb{E} \epsilon_t^2 = 1$ , we have  $\mathbb{E} \log \epsilon_t^2 \leq \log \mathbb{E} \epsilon_t^2 = 0$  by Jensen's inequality. Therefore,  $\mathbb{E}(\log \hat{a}_t)^+ < \infty$  if and only if  $\mathbb{E}(\log \tilde{a}_t)^+ < \infty$ , which is the second part of condition (2.9) in Theorem 2.2. The sequence  $\prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}$  is a measurable function of strictly stationary and ergodic sequences, hence it is also strictly stationary and ergodic and so is the sequence  $s_n \equiv \sum_{k=0}^n \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}$  for all finite  $n$ . Since  $s_n$  is increasing in  $n$ , we have  $\sup_n s_n = s_{\infty} = \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k} = {}_u\sigma_t^2$  which

is measurable by Theorem 20, Chapter 3 of Royden (1963). Therefore,  ${}_u\sigma_t^2$  is strictly stationary and ergodic with a well defined probability measure.

We next show  $\sigma_t^2$  converges almost surely to  ${}_u\sigma_t^2$ . To see this,

$$\sigma_t^2 - {}_u\sigma_t^2 = \prod_{j=0}^{t-1} \phi_{t-j} \sigma_0^2 - \sum_{k=t}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}.$$

Using the strong law of large numbers and condition (2.9) again, we have

$$\prod_{j=0}^{k-1} \phi_{t-j} = \left( \exp \left( \frac{1}{k} \sum_{j=0}^{k-1} \log \phi_{t-j} \right) \right)^k \xrightarrow{k \rightarrow \infty} 0 \quad \text{a.s..} \quad (\text{A.4})$$

Therefore,  $\sigma_t^2 - {}_u\sigma_t^2 \rightarrow 0$  almost surely as  $t \rightarrow \infty$ . Since  ${}_u\sigma_t^2$  is strictly stationary and ergodic,  ${}_u\sigma_t^2$  has the same law as  ${}_u\sigma_0^2$ . The almost sure convergence implies convergence in distribution, i.e.,  $\sigma_t^2 \xrightarrow{d} {}_u\sigma_t^2$  as  $t \rightarrow \infty$ . Moreover,

$$\begin{aligned} \phi_{t+1} \cdot {}_u\sigma_t^2 + \hat{a}_{t+1} &= \phi_{t+1} \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k} + \hat{a}_{t+1} \\ &= \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t+1-j} \hat{a}_{t+1-k} = {}_u\sigma_{t+1}^2. \end{aligned}$$

Thus  ${}_u\sigma_t^2$  is a proper strictly stationary and ergodic solution of (2.8).

We next show  ${}_u\sigma_t^2$  is the unique strictly stationary and ergodic solution. Assume  ${}_u\hat{\sigma}_t^2$  is another strictly stationary and ergodic solution of (2.8). Then

$${}_u\hat{\sigma}_t^2 - {}_u\sigma_t^2 = \phi_t({}_u\hat{\sigma}_{t-1}^2 - {}_u\sigma_{t-1}^2) = \dots = \prod_{j=0}^{k-1} \phi_{t-j}({}_u\hat{\sigma}_{t-k}^2 - {}_u\sigma_{t-k}^2).$$

By (A.4),  $\prod_{j=0}^{k-1} \phi_{t-j}({}_u\hat{\sigma}_{t-k}^2 - {}_u\sigma_{t-k}^2) \xrightarrow{k \rightarrow \infty} 0$  almost surely since both  ${}_u\hat{\sigma}_t^2$  and  ${}_u\sigma_t^2$  are strictly stationary and ergodic. Therefore,  ${}_u\hat{\sigma}_t^2 - {}_u\sigma_t^2 \rightarrow 0$  almost surely.

Finally, we prove the necessary part of the theorem. Suppose (2.8) has a unique strictly stationary and ergodic solution  ${}_u\sigma_t^2$ . By Cauchy's root test, the series  $\sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k}$  diverges almost surely if

$$\limsup_{k \rightarrow \infty} \left( \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k} \right)^{1/k} > 1 \quad \text{a.s..}$$

If  $\mathbb{E} \log \phi_t \geq 0$ , using the argument of Corollary 2.2.1 (see below), we have  $\sigma_t^2 \rightarrow \infty$  and thus,  ${}_u\sigma_t^2 \rightarrow \infty$ . If both  $\limsup_{t \rightarrow \infty} \log \hat{a}_t/t = \infty$  and  $\mathbb{E} \log \phi_t = -\infty$ , we define

$$k^{-1} \left( \sum_{j=0}^{k-1} \log \phi_{t-j} + \log \hat{a}_{t-k} \right) \downarrow 0, \quad \text{a.s. as } k \rightarrow \infty,$$

i.e., the limit approaches zero strictly from above. Then

$$\limsup_{k \rightarrow \infty} \left( \prod_{j=0}^{k-1} \phi_{t-j} \hat{a}_{t-k} \right)^{1/k} \downarrow 1 \quad \text{a.s. as } k \rightarrow \infty,$$

and  ${}_u\sigma_t^2 \rightarrow \infty$ . In both cases, the results contradict the fact (2.8) has a strictly stationary and ergodic solution. We conclude that condition (2.9) is the necessary condition.  $\square$

*Proof of Corollary 2.2.1.* Since  $\phi_t > 0$  and  $\hat{a}_t \geq 0$  almost surely, by (A.2)

$$\sigma_t^2 \geq \prod_{j=0}^{t-1} \phi_{t-j} \sigma_0^2,$$

where  $\sigma_0^2$  is defined in (A.3). Since  $\phi_t$  is a measurable function of the *i.i.d.* sequence  $\epsilon_t$ , it follows the same argument of Theorem 1 of Nelson (1990). For non *i.i.d.*  $\epsilon_t$ , it follows the same arguments of Lemma 1 and Theorem 1 together with Assumption A2 of Linton et al. (2010).  $\square$

*Derivation of (3.4).* It is straightforward to write  $v_t = \omega + \xi v_{t-1} + \alpha e_{t-1} + \gamma u_{t-1}$  since it is a (GJR-)GARCH process. By the symmetric assumption on  $\epsilon_t$  and the independence of  $\epsilon_t$  and  $v_t$ ,  $e_t$  and  $u_t$  are two MDS. Consequently, we can write  $\sigma_t^2 = v_t + \phi \sigma_{t-1}^2 + e_t$ . Substituting for  $v_t$ , we have

$$\sigma_t^2 = \omega + \phi \sigma_{t-1}^2 + \xi v_{t-1} + e_t + \alpha e_{t-1} + \gamma u_{t-1}. \quad (\text{A.5})$$

Adding and subtracting  $r_t^2$  and  $\phi r_{t-1}^2$  on the right hand side (RHS) of (A.5), we obtain

$$r_t^2 = \omega + \phi r_{t-1}^2 + r_t^2 - \sigma_t^2 - \phi(r_{t-1}^2 - \sigma_{t-1}^2) + \xi v_{t-1} + e_t + \alpha e_{t-1} + \gamma u_{t-1}. \quad (\text{A.6})$$

Since  $\mathbb{E}[r_t^2 | \mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t^2 | \mathcal{F}_{t-1}] + \kappa v_t$ ,  $z_t$  is an MDS. Substituting  $r_t^2 - \sigma_t^2 = z_t + \kappa v_t$  on the RHS of (A.6),

$$r_t^2 = \omega + \phi r_{t-1}^2 + \xi v_{t-1} + z_t + \kappa v_t - \phi z_{t-1} - \phi \kappa v_{t-1} + e_t + \alpha e_{t-1} + \gamma u_{t-1}. \quad (\text{A.7})$$

Substituting for  $v_t$  on the RHS of (A.7) and rearranging the terms, we obtain (3.4).  $\square$

*Proof of Proposition 3.1.* Condition  $\mathbb{E} \log (\beta + (\alpha + \gamma \mathbb{1}_{(\epsilon_t < 0)}) \epsilon_t^2) < 0$  is the strict stationarity condition for GJR-GARCH. By Theorem 2 of Nelson (1990),  $v_t$  has a unique strictly stationary and ergodic solution  ${}_u v_t$  which is a measurable function of  $\epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots$  and is given by

$${}_u v_t \equiv \omega \left( 1 + \sum_{k=1}^{\infty} \prod_{j=1}^k (\beta + (\alpha + \gamma \mathbb{1}_{(\epsilon_{t-j} < 0)}) \epsilon_{t-j}^2) \right).$$

Moreover,  ${}_u v_t < \infty$  almost surely and  $v_t - {}_u v_t \rightarrow 0$  as  $t \rightarrow \infty$  almost surely. Consequently,  $\mathbb{E}(\log {}_u v_t)^+ \leq \mathbb{E}({}_u v_t)^+ < \infty$ . By Assumption 1,  ${}_u v_t \epsilon_t^2$  is a measurable function of  $\epsilon_t^2, \epsilon_{t-1}^2, \dots$  and is therefore, strictly stationary and ergodic. All conditions of Theorem 2.2 are satisfied with  $\phi_t = \phi$  and  $\hat{a}_t = {}_u v_t \epsilon_t^2$ . Finally, define  ${}_u r_t \equiv {}_u \sigma_t \epsilon_t$ , where  ${}_u \sigma_t^2$  is the unique strictly stationary and ergodic solution to (3.3). It follows immediately that  ${}_u r_t < \infty$  almost surely and  ${}_u r_t$  is strictly stationary and ergodic since it is a measurable function of  $\epsilon_t, \epsilon_{t-1}, \dots$  as long as  $\epsilon_t^2$  is non-degenerate. Moreover, since  ${}_u \sigma_t^2 > 0$  almost surely, by continuous mapping theorem,  ${}_u r_t - r_t \rightarrow 0$  almost surely and  $r_t \xrightarrow{d} {}_u r_t$  as  $t \rightarrow \infty$ .  $\square$

## References

- 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 A. E. D. Veraart (2012). Stochastic volatility of volatility and variance risk premia. *Journal of Financial Econometrics* 11(1), 1–46.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009). A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics* 150(2), 151–166.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review* 39(4), 841–862.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41(1), 135–155.
- Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews* 27(1-3), 46–78.
- Ding, Y. D. (2021). A simple joint model for returns, volatility and volatility of volatility. *Journal of Econometrics*, forthcoming.
- Drost, F. C. and T. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica* 61(4), 909–927.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Francq, C. and J.-M. Zakoïan (2010). *GARCH models: structure, statistical inference, and financial applications* (1st ed.). Wiley.
- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics* 23(4), 365–380.
- 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.

- Huang, D., C. Schlag, I. Shaliastovich, and J. Thimme (2019). Volatility-of-volatility risk. *Journal of Financial and Quantitative Analysis* 54(6), 2423–2452.
- Linton, O., J. Pan, and H. Wang (2010). Estimation for a nonstationary semi-strong garch(1,1) model with heavy-tailed errors. *Econometric Theory* 26(1), 1–28.
- 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). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6(3), 318–334.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2), 347–370.
- Neuberger, A. and R. Payne (2020). The skewness of the stock market over long horizons. *The Review of Financial Studies* 34(3), 1572–1616.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160(1), 246–256.
- Royden, H. L. (1963). *Real analysis*. New York: Macmillan.
- Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4(2), 183–204.