UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Weak Diffusion Limits of Two Real-Time GARCH-type Models

Yashuang
(Dexter)
Ding

Abstract

We derive the diffusion limits of two Real-Time GARCH (RT-GARCH)-type models. We show the diffusion limit of the original RT-GARCH model fails to guarantee with probability one positive volatility unless it is degenerate. Consequently, we propose a novel square-root stochastic heteroskedastic autoregressive volatility (SQ-SHARV) model that builds upon the idea of RT-GARCH while maintaining the usual GARCH diffusion limit. As a result, we call for caution when using RTGARCH since it lacks compatibility with existing asset pricing theories. On the contrary, SQ-SHARV combines the advantages of both RT-GARCH and GARCH models.

Reference Details

CWPE	20112
Published	25 November 2020
Updated	28 February 2021
Key Words	GARCH, RT-GARCH, SV, diffusion limit
JEL Codes	C22, C32, C58
Website	www.econ.cam.ac.uk/cwpe

Weak Diffusion Limits of Two Real-Time GARCH-type Models

Yashuang (Dexter) Ding*

University of Cambridge

Abstract

We derive the diffusion limits of two Real-Time GARCH (RT-GARCH)-type models. We show the diffusion limit of the original RT-GARCH model fails to guarantee with probability one positive volatility unless it is degenerate. Consequently, we propose a novel square-root stochastic heteroskedastic autoregressive volatility (SQ-SHARV) model that builds upon the idea of RT-GARCH while maintaining the usual GARCH diffusion limit. As a result, we call for caution when using RT-GARCH since it lacks compatibility with existing asset pricing theories. On the contrary, SQ-SHARV combines the advantages of both RT-GARCH and GARCH models.

Keywords: GARCH, RT-GARCH, SV, diffusion limit
JEL classification: C22, C32, C58

*I thank the editor, Silvia Gonçalves, the associate editor and two anonymous referees for helpful suggestions. 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, Tacye Hong, Tristan Hennig and Christian Tien for useful comments. All remaining errors are mine. Address 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 of financial returns has long been an active research area. There are generally two approaches to model volatility ex-ante. GARCH-type models (Engle (1982), Bollerslev (1986), Nelson (1991), Glosten et al. (1993), Hansen et al. (2012), among others) regard volatility as determined solely by past information. GARCH models are popular largely due to their simplicity. Stochastic volatility (SV) models (Heston (1993), Fong and Vasicek (1991), Longstaff and Schwartz (1992), among others), on the other hand, assume that volatility is driven by a different innovation term. Itô's calculus provides many elegant analytical properties for continuous time SV models. However, questions regarding how well SV models fit the data and the difficulty of estimation are still present.

Nelson (1990) links these two approaches by providing the weak convergence results of GARCH-type models to continuous time SV models. Nelson (1992) and Nelson and Foster (1995) prove misspecified GARCH-type models can still consistently filter and forecast volatility when the data generating process is a diffusion or near-diffusion process under mild conditions. The weak convergence results have been extended to other GARCH-type models (Duan (1997), Fornari and Mele (1997), Hafner et al. (2017), among others). On the other hand, Corradi (2000) shows there exist alternative conditions that lead to degenerate diffusion limits. In addition, Drost and Werker (1996) use the temporal aggregation properties discussed in Drost and Nijman (1993) to derive the diffusion limits of the class of weak GARCH models.

Empirical studies have remarked that by not using all available internal information, in particular the current return, GARCH-type models make inefficient use of information (see for example, Politis (1995) and Hansen et al. (2012)). To address this, Smetanina (2017) proposes the RT-GARCH model which incorporates current return innovation into the volatility process. Specifically,

$$r_t \equiv S_t - S_{t-1} = \sigma_t \epsilon_t, \tag{1.1}$$

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \psi\epsilon_t^2,\ (\alpha, \beta, \gamma, \psi)' \geq 0, \tag{1.2}$$

where ϵ_t are i.i.d. random variables symmetric around zero with unit variance and finite fourth moment. The model uses ϵ_t^2 to drive the volatility process. In doing so, σ_t^2 is no longer deterministic conditional on the information up to time $t-1$. By construction, $\psi^2\mathbb{E}[(\epsilon_t^2 - 1)^2]$ is the conditional variance of σ_t^2 at time $t-1$. When $\gamma = 0$, RT-GARCH reduces to a special case of the contemporaneous version of the square-root stochastic autoregressive volatility (SQ-SARV) model of Meddahi and Renault (2004).[1] Thus, RT-GARCH is in fact a hybrid of GARCH and SV models. Unlike the general SV models, it has analytical expressions for both the likelihood function and the conditional variance of

[1] The original SQ-SARV model in Meddahi and Renault (2004) uses the lagged volatility, σ_{t-1}, instead of σ_t to drive the return process, therefore, it is called lagged SQ-SARV.

return. Smetanina and Wu (2019) derive the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) of RT-GARCH.

The idea of RT-GARCH is similar to that of Breitung and Hafner (2016), where they incorporate the current return innovation into the log volatility process. Breitung and Hafner's (2016) model is closely related to Nelson's (1991) E-GARCH model and can be viewed as a special case of the contemporaneous exponential stochastic autoregressive volatility (E-SARV) model defined in Taylor (1994).

To formally define where in-between RT-GARCH lies with respect to other volatility models, we need to derive its diffusion limit. In this paper, we use the approximation theorems of Nelson (1990) and Duan (1997) to derive the diffusion limit of RT-GARCH. As we will see, the volatility process of RT-GARCH converges weakly to an Ornstein-Uhlenbeck (OU) process. This is in direct contrast to the diffusion limit of GARCH. The reason is that the joint process of the (log) price and volatility under RT-GARCH is no longer Markov. Since the OU process permits negative values of volatility, this is an undesirable feature of RT-GARCH. Consequently, we propose a new model that builds on the idea of RT-GARCH and show it converges to the usual GARCH diffusion. We call this model the square-root stochastic heteroskedastic autoregressive volatility (SQ-SHARV) model. This new model has an additional feature of allowing conditional heteroskedasticity in the variance of volatility, which enables us to model volatility and the volatility of volatility jointly.

The remainder of the paper is structured as follows. In section 2, we derive the diffusion limit of RT-GARCH. In section 3, we introduce the SQ-SHARV model and derive its diffusion limit. Section 4 concludes. All proofs are in Appendix A.

2 Diffusion limit of RT-GARCH

For a detailed discussion on the weak convergence of Markov chains to diffusion processes, see section 2 of Nelson (1990). Unfortunately, the joint process (S_t, σ_t^2) under RT-GARCH is not Markov since σ_t^2 is $\mathcal{F}_t-$measurable and depends on $r_{t-1} \equiv S_{t-1} - S_{t-2}$, where $\mathcal{F}_t$ is the information set up to time t. Duan (1997) proposes a Markov form of GARCH(p,q) in order to derive its diffusion limit. We apply the same principle by introducing an auxiliary process. Specifically, let the rescaled joint process $({}_h\sigma_{kh}^2, {}_hr_{kh}^2)$ be given by

$$ {}_h\sigma_{kh}^2 = {}_h\sigma_{(k-1)h}^2 + h(\alpha+\psi) + h(\beta-1){}_h\sigma_{(k-1)h}^2 + h\gamma_h r_{(k-1)h}^2 + \sqrt{h}\psi(\epsilon_{kh}^2 - 1), \tag{2.1} $$

$$ {}_hr_{kh}^2 = {}_hr_{(k-1)h}^2 + h({}_h\sigma_{(k-1)h}^2 - {}_hr_{(k-1)h}^2) + \sqrt{h}({}_h\sigma_{kh}^2\epsilon_{kh}^2 - {}_h\sigma_{(k-1)h}^2), \tag{2.2} $$

$$ \mathbb{P}\big(({}_h\sigma_0^2, {}_hr_0^2) \in \Gamma\big) = v_h(\Gamma) \text{ for any } \Gamma \in B(R^2), \tag{2.3} $$

where $\epsilon_{kh} \sim N(0,1)$, $B(R^n)$ denote the Borel sets on R^n. It is immediate that (2.1) becomes (1.2) and $r_t^2 = \sigma_t^2\epsilon_t^2$ by setting $h = 1$ and $k = t$.

Theorem 2.1. *Let $({}_h\sigma^2_{kh}, {}_hr^2_{kh})$ satisfy (2.1) - (2.3). If $({}_h\sigma^2_0, {}_hr^2_0) \Rightarrow (\sigma^2_0, r^2_0)$, where '$\Rightarrow$' denotes weak convergence in distribution, then $({}_h\sigma^2_t, {}_hr^2_t) \Rightarrow (\sigma^2_t, r^2_t)$, where ${}_h\sigma^2_t \equiv {}_h\sigma^2_{kh}$ and ${}_hr^2_t \equiv {}_hr^2_{kh}$ for all $kh \leq t < (k+1)h$ as $h \downarrow 0$ and the joint process (σ^2_t, r^2_t) satisfies*

$$d\sigma^2_t = \big(\alpha + \psi + (\beta - 1)\sigma^2_t + \gamma r^2_t\big)dt + \sqrt{2}\psi dW_t, \tag{2.4}$$

$$dr^2_t = (2\psi + \sigma^2_t - r^2_t)dt + \sqrt{2}\sigma^2_t dW_t, \tag{2.5}$$

$$\mathbb{P}\big((\sigma^2_0, r^2_0) \in \Gamma\big) = v_0(\Gamma) \textit{ for any } \Gamma \in B(R^2), \tag{2.6}$$

where W_t is a standard Brownian motion.

Theorem 2.1 states that ${}_h\sigma^2_{kh}$, conditional on the auxiliary process ${}_hr^2_{kh}$, converges to an OU process. It is immediate that (2.4) fails to guarantee almost surely positive σ^2_t. Moreover, we can only establish weak convergence for the joint process $({}_h\sigma^2_{kh}, {}_hr^2_{kh})$, not for $({}_hS_{kh}, {}_h\sigma^2_{kh}, {}_hr^2_{kh})$. This is because the diffusion limit $dS_t = \sigma_t dB_t$, that we would have obtained for ${}_hS_{kh} - {}_hS_{(k-1)h} = \sqrt{h}_h\sigma_{kh}\epsilon_{kh}$, is not well defined for $\sigma^2_t < 0$ and real-valued S_t. This is in contrast to the diffusion limit of GARCH model even though RT-GARCH nests GARCH in discrete time. To understand this, recall the rescaled GARCH process $({}_h\tilde{S}_{kh}, {}_h\tilde{\sigma}^2_{(k+1)h})$ defined in section 2.3 of Nelson (1990). The joint process of the (log) price and volatility alone is Markov. In fact, it is ${}_h\tilde{\sigma}^2_{(k+1)h}$, not ${}_h\tilde{\sigma}^2_{kh}$, that converges weakly to σ^2_t since ${}_h\tilde{\sigma}^2_{(k+1)h}$ is $\mathcal{F}_{kh}-$measurable. Therefore, although similar in structure in discrete time, RT-GARCH and GARCH have distinct diffusion limits.

It is still possible to establish joint convergence of $({}_hS_{kh}, {}_h\sigma^2_{kh}, {}_hr^2_{kh})$ by redefining the rescaled joint process as follows,

$${}_hS_{kh} - {}_hS_{(k-1)h} = \sqrt{h}\sqrt{|{}_h\sigma^2_{kh}|}\epsilon_{kh}, \tag{2.7}$$

$${}_h\sigma^2_{kh} = {}_h\sigma^2_{(k-1)h} + h(\alpha + \psi) + h(\beta - 1){}_h\sigma^2_{(k-1)h} + h\gamma|{}_hr^2_{(k-1)h}| + \sqrt{h}\psi(\epsilon^2_{kh} - 1), \tag{2.8}$$

$${}_hr^2_{kh} = {}_hr^2_{(k-1)h} + h({}_h\sigma^2_{(k-1)h} - {}_hr^2_{(k-1)h}) + \sqrt{h}({}_h\sigma^2_{kh}\epsilon^2_{kh} - {}_h\sigma^2_{(k-1)h}), \tag{2.9}$$

$$\mathbb{P}\big(({}_hS_0, {}_h\sigma^2_0, {}_hr^2_0) \in \Gamma\big) = v_h(\Gamma) \text{ for any } \Gamma \in B(R^3). \tag{2.10}$$

It can be shown that $|r^2_t| = (S_t - S_{t-1})^2 = |\sigma^2_t|\epsilon^2_t$ and (2.8) becomes (1.2) by setting $h = 1$ and $k = t$. Then as $h \downarrow 0$, we obtain the following diffusion limit:

$$dS_t = \sqrt{|\sigma^2_t|}dB_t \tag{2.11}$$

$$d\sigma^2_t = \big(\alpha + \psi + (\beta - 1)\sigma^2_t + \gamma|r^2_t|\big)dt + \sqrt{2}\psi dW_t, \tag{2.12}$$

$$dr^2_t = (2\psi + \sigma^2_t - r^2_t)dt + \sqrt{2}\sigma^2_t dW_t, \tag{2.13}$$

$$\mathbb{P}\big((S_0, \sigma^2_0, r^2_0) \in \Gamma\big) = v_0(\Gamma) \text{ for any } \Gamma \in B(R^3), \tag{2.14}$$

where B_t and W_t are two independent standard Brownian motions. To obtain the dynamics of $|\sigma^2_t|$ and $|r^2_t|$, we can use the Meyer-Tanaka formula (Protter, 2004). Specifically,

$$d|X_t| = \text{sign}(X_t)dX_t + dL^0_t(X_t), \tag{2.15}$$

where $L_t^0(X_t)$ is the local time of a semimartingale X_t at zero, i.e.,

$$L_t^0(X_t) = \lim_{h \downarrow 0} \frac{1}{2h} \int_0^t \mathbb{1}_{(-h<X_s<h)} d\langle X, X\rangle_s, \tag{2.16}$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function and $\langle X, X\rangle_t$ is the quadratic variation of X_t.

It may seem odd at first glance that RT-GARCH, which guarantees $\sigma_t^2 > 0$ with probability one in discrete time, converges to an OU process that fails to guarantee $\sigma_t^2 > 0$ with probability one. By examining (2.8), it is immediate that ${}_h\sigma_{kh}^2$ fails to be positive with probability one because of the term $h(\alpha + \psi) - \sqrt{h}\psi$ in (2.8). For small h, these terms will eventually become negative. To ensure almost surely positive σ_t^2 in the diffusion limit, we consider a degenerate case in which the innovation term affects σ_t^2 via $|r_t^2|$ only. This can be achieved by rescaling the term $\psi(\epsilon_{kh}^2 - 1)$ in (2.8) by a factor of h instead of $\sqrt{h}$. This is similar to the degenerate case of GARCH considered by Corradi (2000), although in our case, σ_t^2 is still random because of $|r_t^2|$.[2] For the rest of the paper, we assume the probability measures $v_h(\cdot)$ and $v_0(\cdot)$ are such that $\mathbb{P}({}_h\sigma_0^2 > 0) = \mathbb{P}(\sigma_0^2 > 0) = 1$ and $\mathbb{P}({}_hr_0^2 \geq 0) = \mathbb{P}(r_0^2 \geq 0) = 1$.

Theorem 2.2. *Let* ${}_hS_{kh} - {}_hS_{(k-1)h} = \sqrt{h}{}_h\sigma_{kh}\epsilon_{kh}$. *Let* ${}_h\sigma_{kh}^2$ *and* ${}_hr_{kh}^2$ *satisfy* (2.8) *and* (2.9) *while replacing* $\sqrt{h}\psi(\epsilon_{kh}^2 - 1)$ *with* $h\psi(\epsilon_{kh}^2 - 1)$ *in* (2.8). *If* $({}_hS_0, {}_h\sigma_0^2, {}_hr_0^2) \Rightarrow (S_0, \sigma_0^2, r_0^2)$, *where* $({}_hS_0, {}_h\sigma_0^2, {}_hr_0^2)$ *satisfies* (2.10), *then* $({}_hS_{kh}, {}_h\sigma_{kh}^2, {}_hr_{kh}^2) \Rightarrow (S_t, \sigma_t^2, r_t^2)$ *for all* $kh \leq t < (k+1)h$ *as* $h \downarrow 0$ *and the joint process* (S_t, σ_t^2, r_t^2) *satisfies*

$$dS_t = \sigma_t dB_t, \tag{2.17}$$

$$d\sigma_t^2 = \big(\alpha + \psi + (\beta - 1)\sigma_t^2 + \gamma|r_t^2|\big)dt, \tag{2.18}$$

$$dr_t^2 = (\sigma_t^2 - r_t^2)dt + \sqrt{2}\sigma_t^2 dW_t, \tag{2.19}$$

$$\mathbb{P}\big((S_0, \sigma_0^2, r_0^2) \in \Gamma\big) = v_0(\Gamma) \text{ for any } \Gamma \in B(R^3), \tag{2.20}$$

where B_t *and* W_t *are two independent standard Brownian motions.*

Remark 2.2.1. *We can further degenerate the process in Theorem* 2.2 *by replacing* $\sqrt{h}({}_h\sigma_{kh}^2\epsilon_{kh}^2 - {}_h\sigma_{(k-1)h}^2)$ *with* $h({}_h\sigma_{kh}^2\epsilon_{kh}^2 - {}_h\sigma_{(k-1)h}^2)$ *in the approximating process* (2.9). *In this case,* ${}_hr_{kh}^2 \geq 0$ *a.s. for all* $0 \leq h \leq 1$ *and* $k \geq 1$. *Therefore, we can replace* $h\gamma|{}_hr_{(k-1)h}^2|$ *with* $h\gamma{}_hr_{(k-1)h}^2$ *in* (2.8). *Then as* $h \downarrow 0$, *we obtain the following diffusion limit:*

$$d\sigma_t^2 = \big(\alpha + \psi + (\beta - 1)\sigma_t^2 + \gamma r_t^2\big)dt, \tag{2.21}$$

$$dr_t^2 = (\sigma_t^2 - r_t^2)dt, \tag{2.22}$$

together with (2.17) *and* (2.20).

[2] We thank the referees to point out the degenerate case. Note the difference between the approximation schemes in Duan (1997) and Nelson (1990) and therefore, Corradi (2000) is that, Duan (1997) rescales the steps of the discrete time chain, while Nelson (1990) rescales their parameters. The conclusion is albeit equivalent.

The weak convergence results of Theorems 2.1 and 2.2 show that RT-GARCH is less appealing than GARCH. The diffusion limit of RT-GARCH is hard to interpret and does not align well with existing asset pricing theories. Moreover, it is not straightforward to derive the stationary distributions of σ_t^2 and r_t. On the other hand, it can be shown that the Breitung and Hafner's (2016) model converges to the same diffusion limit as the (symmetric) E-GARCH model of Nelson (1991) (see the supplementary material). The exponential link function in their model ensures the volatility process to be positive with probability one. The aim of diffusion limits is to provide justification for using discrete time volatility models to estimate unobserved volatility, which is often assumed to be generated by a diffusion process. It is therefore, crucial to use a model that aligns well with existing asset pricing theories. In section 3, we propose a new model that is based upon the idea of RT-GARCH while preserving the usual GARCH diffusion limit.

We now briefly discuss the temporal aggregation of RT-GARCH. Note RT-GARCH does not fall into the class of weak GARCH models defined in Drost and Nijman (1993) since σ_t^2 is not $\mathcal{F}_{t-1}-$measurable. As a result, the temporal aggregation of RT-GARCH should be considered in the fashion of SV models. Consider a special case with $\gamma = 0$. Let $h = 1$ and $k = t$, the discrete time processes (2.1) - (2.2) then become

$$r_t = \sigma_t \epsilon_t, \tag{2.23}$$

$$\sigma_t^2 = \alpha + \beta \sigma_{t-1}^2 + \psi \epsilon_t^2. \tag{2.24}$$

For simplicity, we have dropped the left subscript of $h = 1$ in the state variables. Since r_t is driven by σ_t instead of σ_{t-1} in (2.23), we can not directly apply the temporal aggregation result of SQ-SARV in Meddahi and Renault (2004), where they use σ_{t-1} to drive the return process.

Assumption 1. *Let* $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = \mathbb{E}[\epsilon_t^3|\mathcal{F}_{t-1}] = 0$, $\mathbb{E}[\epsilon_t^2|\mathcal{F}_{t-1}] = 1$ *and* $\mathbb{E}[\epsilon_t^4|\mathcal{F}_{t-1}] < \infty$.

We state the temporal aggregation result for flow variables only since return is a flow variable, i.e. $r_{(m)t} = \sum_{j=0}^{m-1} r_{t-j}$ for any integer $m > 1$.

Theorem 2.3. *Let* ϵ_t *satisfy Assumption* 1*, then the joint process* (2.23) - (2.24) *is closed under temporal aggregation with the filtration* $\mathcal{F}_{(m)t}$ *generated by aggregated returns* r_0, r_m, ..., r_t. *That is, the joint process* $(r_{(m)t}, \sigma_{(m)t}^2)$ *satisfies*

$$r_{(m)t} = \sigma_{(m)t} \epsilon_{(m)t}, \tag{2.25}$$

$$\sigma_{(m)t}^2 = \alpha_{(m)} + \beta_{(m)} \sigma_{(m)t-m}^2 + \psi_{(m)} \epsilon_{(m)t}^2, \tag{2.26}$$

where

$$\sigma_{(m)t}^2 = \sum_{j=0}^{m-1} \sigma_{t-j}^2 + \psi \sum_{j=0}^{m-2} \Big(\epsilon_{t-j}^2 \sum_{k=0}^{m-2-j} \beta^k / \beta^{m-1-j} \Big), \tag{2.27}$$

$$\alpha_{(m)} = m\alpha \sum_{j=0}^{m-1} \beta^j, \tag{2.28}$$

$$\beta_{(m)} = \beta^m, \tag{2.29}$$

$$\psi_{(m)} = \psi \sum_{j=0}^{m-1} \beta^j \sum_{j=0}^{m-1} 1/\beta^j, \tag{2.30}$$

and the aggregated innovation term is given by,

$$\epsilon_{(m)t} = \text{sign}(r_{(m)t})\sqrt{\frac{\sum_{j=0}^{m-1} \beta^j \epsilon_{t-j}^2}{\sum_{t=0}^{m-1} \beta^j}}. \tag{2.31}$$

For RT-GARCH with lagged squared return, we need to define a new class of weak RT-GARCH models in order to follow the arguments of Drost and Nijman (1993). We leave this for future research.

3 Diffusion limit of SQ-SHARV

We next propose a new model based upon the idea of RT-GARCH while preserving the usual GARCH diffusion. We consider the augmented RT-GARCH (ART-GARCH) model proposed by Ding (2021). The model is motivated by the conditionally stochastic nature of σ_t^2 in RT-GARCH, which allows us to introduce conditional heteroskedasticity in the variance of σ_t^2. By construction, ART-GARCH jointly models volatility and the volatility of volatility. There is a growing literature emphasizing the importance of volatility of volatility as an additional risk factor (see Ding (2021) for a review). We next show that a special case of ART-GARCH converges weakly to the same diffusion limit as GARCH. Specifically, let the joint process (S_t, σ_t^2) be given by

$$r_t \equiv S_t - S_{t-1} = \sigma_t \epsilon_t, \tag{3.1}$$

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + (\psi + \eta\sigma_{t-1}^2)\epsilon_t^2, \tag{3.2}$$

where ϵ_t satisfy the same conditions as in RT-GARCH and $(\alpha, \beta, \psi, \eta)' \geq 0$. (3.1) - (3.2) are obtained by setting $\gamma = 0$, i.e., excluding r_{t-1}^2 in the ART-GARCH model.[3] We call this reduced form model the square-root stochastic heteroskedastic autoregressive (SQ-SHARV) model. Straightforward calculation shows σ_t^2 depends on r_t^2 non-linearly:

$$\sigma_t^2 = 0.5(\alpha + \beta\sigma_{t-1}^2) + 0.5\sqrt{(\alpha + \beta\sigma_{t-1}^2)^2 + 4(\psi + \eta\sigma_{t-1}^2)r_t^2}. \tag{3.3}$$

Since σ_{t-1}^2 is $\mathcal{F}_{t-1}-$measurable, we still have an analytical expression for the conditional density of r_t. It is easy to see that $\text{Var}[\sigma_t^2|\mathcal{F}_{t-1}] = 2(\psi + \eta\sigma_{t-1}^2)^2$ is the conditional variance of volatility at time $t-1$ and the conditional variance of return is given by $\mathbb{E}[r_t^2|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_t^2|\mathcal{F}_{t-1}] + 2\psi + 2\eta\sigma_{t-1}^2 = \alpha + 3\psi + (\beta + 3\eta)\sigma_{t-1}^2$ for $\epsilon_t \sim N(0,1)$. All

[3] The full specification of ART-GARCH is given by $\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + (\psi + \eta\sigma_{t-1}^2)\epsilon_t^2$.

the statistical properties of the general ART-GARCH model in section 3 of Ding (2021) can be directly applied to SQ-SHARV by setting $\gamma = 0$. Ding (2021) shows that ART-GARCH significantly improves the volatility filtering and forecasting over RT-GARCH and GARCH and the standardised residuals $\hat{\epsilon}_t$ under ART-GARCH are close to Gaussian. In addition, we conduct the nonstandard quasi-likelihood ratio test of Francq and Zakoïan (2009) for the hypothesis $H_0 : \gamma = 0$ to compare SQ-SHARV to the general ART-GARCH.[4] We obtain the p-values of 0.0826 for Dow Jones Industrial Average index returns and 0.1998 for Apple Inc. stock returns. The results are in favour of the reduced form SQ-SHARV model.

Under SQ-SHARV, the joint process (S_t, σ_t^2) is Markov and we do not need to introduce auxiliary processes. Therefore, we use Nelson's (1990) approach by letting the parameters of the approximating processes vary with h. Specifically, let the joint rescaled process $({}_hS_{kh}, {}_h\sigma_{kh}^2)$ be given by

$$
\begin{aligned}
{}_hr_{kh} &\equiv {}_hS_{kh} - {}_hS_{(k-1)h} = {}_h\sigma_{kh} \cdot {}_h\epsilon_{kh}, && (3.4)\\
{}_h\sigma_{kh}^2 &= \alpha_h + \beta_h \cdot {}_h\sigma_{(k-1)h}^2 + h^{-1}(\psi_h + \eta_h \cdot {}_h\sigma_{(k-1)h}^2)_h\epsilon_{kh}^2, && (3.5)\\
\mathbb{P}\big(({}_hS_0, {}_h\sigma_0^2) \in \Gamma\big) &= v_h(\Gamma) \text{ for any } \Gamma \in B(R^2), && (3.6)
\end{aligned}
$$

where ${}_h\epsilon_{kh} \sim N(0, h)$. Note in (3.4) - (3.6) we use the left subscript to indicate ${}_h\epsilon_{kh}$ depend on the choice of h while in section 2, ϵ_{kh} without the left subscript are standard normal and do not depend on h.

Assumption 2. *Let the sequence $(\alpha_h, \beta_h, \psi_h, \eta_h)'$ satisfy*

$$
\begin{aligned}
&\lim_{h\downarrow 0} h^{-1}(\alpha_h + \psi_h) = \mu, && (3.7)\\
&\lim_{h\downarrow 0} h^{-1}(\beta_h + \eta_h - 1) = -\theta, && (3.8)\\
&\lim_{h\downarrow 0} 2h^{-1}\eta_h^2 = \lambda^2. && (3.9)
\end{aligned}
$$

Theorem 3.1. *Let $({}_hS_{kh}, {}_h\sigma_{kh}^2)$ satisfy (3.4) - (3.6). If $({}_hS_0, {}_h\sigma_0^2) \Rightarrow (S_0, \sigma_0^2)$, then under Assumption 2, $({}_hS_t, {}_h\sigma_t^2) \Rightarrow (S_t, \sigma_t^2)$, where ${}_hS_t \equiv {}_hS_{kh}$ and ${}_h\sigma_t^2 \equiv {}_h\sigma_{kh}^2$ for all $kh \leq t < (k+1)h$ as $h \downarrow 0$ and the joint process (S_t, σ_t^2) satisfies*

$$
\begin{aligned}
dS_t &= \sigma_t dW_{1,t}, && (3.10)\\
d\sigma_t^2 &= (\mu - \theta\sigma_t^2)dt + \lambda\sigma_t^2 dW_{2,t}, && (3.11)\\
\mathbb{P}\big((S_0, \sigma_0^2) \in \Gamma\big) &= v_0(\Gamma) \textit{ for any } \Gamma \in B(R^2), && (3.12)
\end{aligned}
$$

where $W_{1,t}$ and $W_{2,t}$ are two independent standard Brownian motions.

Remark 3.1.1. *The non-negative constraints on the sequence of parameters prevent ψ_h*

[4] The test is nonstandard since under the null, γ is located on the boundary of the parameter space and the statistics is based on quasi-likelihood functions, see Francq and Zakoïan (2009).

from being of order $\mathcal{O}(\sqrt{h})$. On the other hand, if we choose $\eta_h = \mathcal{O}(h)$ instead of $\mathcal{O}(\sqrt{h})$, we will obtain $d\sigma_t^2 = (\mu - \theta\sigma_t^2)dt$, which is the degenerate case discussed in Corradi (2000).

Theorem 3.1 shows that SQ-SHARV shares the same diffusion limit and thus, stationary distribution as GARCH. Moreover, since $\sqrt{2}h^{-1}(\psi_h + \eta_h \cdot {}_h\sigma_{kh}^2)_h\epsilon_{(k+1)h}^2$ converges to $\lambda\sigma_t^2 dW_{2,t}$, $2(\psi_h + \eta_h \cdot {}_h\sigma_{kh}^2)^2$ consistently estimates the quadratic variation of σ_t^2 between time t to $t+h$, i.e. $\lambda^2 \int_t^{t+h} \sigma_s^4 ds$. Therefore, SQ-SHARV provides consistent estimators for both integrated volatility and integrated volatility of volatility. In the GARCH case, even though $\sqrt{2}h^{-1}\tilde{\gamma}_h \cdot {}_h\tilde{\sigma}_{kh}^2$ also converges to the instantaneous volatility of volatility, in discrete time, the conditional variance of ${}_h\tilde{\sigma}_{(k+1)h}^2$ given $\mathcal{F}_{kh}$ is zero. For RT-GARCH, the volatility process does not have a well defined diffusion limit at all (or its diffusion limit is degenerate). This is the main advantage of SQ-SHARV over RT-GARCH and GARCH models. The diffusion limit of the general ART-GARCH model nests those of RT-GARCH and SQ-SHARV. We leave the detailed derivation for future research.

The volatility of volatility acts as a risk premium in the volatility process (3.11). Specifically, let u_1 and u_2 be the contributions of ψ_h and η_h to μ and θ in (3.11) in the limit, respectively. Then, the risk premium takes the form

$$g(\sigma_t^2) = \frac{u_1}{\lambda\sigma_t^2} + \frac{u_2}{\lambda}. \tag{3.13}$$

To see how this risk premium changes the probability measure of the volatility process, define an equivalent martingale measure $\mathcal{R}$ to the physical measure P under which the returns are observed, with the Radon-Nikodym derivative

$$\frac{d\mathcal{R}}{dP} = \mathcal{E}\Big(\int_0^t g(\sigma_s^2) dW_{2,s}^P \Big), \tag{3.14}$$

where $W_{2,t}^P$ is a standard Brownian motion under the P measure and

$$\mathcal{E}(x_t) = \exp\big(x_t - \tfrac{1}{2}\langle x, x\rangle_t\big) \tag{3.15}$$

is the Doléans-Dade exponential. $W_{2,t}^{\mathcal{R}} = W_{2,t}^P - \int_0^t g(\sigma_s^2)ds$ is a standard Brownian motion under the equivalent measure $\mathcal{R}$. Under this measure, the volatility process is given by

$$d\sigma_t^2 = \big(\mu - u_1 - (\theta + u_2)\sigma_t^2\big)dt + \lambda\sigma_t^2 dW_{2,t}^{\mathcal{R}}. \tag{3.16}$$

The non-negativity constraints on the parameters result in a non-negative risk premium in (3.13). This is intuitive since we would expect a positive correlation between volatility and the volatility of volatility. Note this risk premium is due to the volatility of volatility risk and should not be confused with the volatility risk premium, which is defined as the excess of implied volatility over realised volatility in option pricing theory.

We can also add a first order risk premium in the return process similar to the GARCH-in mean model (the risk premium in the volatility process is a second order risk premium). Specifically, let the rescaled return process satisfy

$${}_hr_{kh} \equiv {}_hS_{kh} - {}_hS_{(k-1)h} = h \cdot c \cdot {}_h\sigma_{(k-1)h}^2 + {}_h\sigma_{kh} \cdot {}_h\epsilon_{kh}. \tag{3.17}$$

Note we use ${}_h\sigma^2_{(k-1)h}$ instead of ${}_h\sigma^2_{kh}$ in the mean process. This is because we would not be able to obtain an analytical expression for the conditional density of ${}_hr_{kh}$ otherwise. The additional term will add a drift term in dS_t in the diffusion limit. Specifically,

$$dS_t = c\sigma_t dt + \sigma_t dW_{1,t}. \tag{3.18}$$

The proof follows exactly that of Theorem 3.1 and it is left as an exercise to the reader. To add the leverage effect in SQ-SHARV, we require $\mathbb{E}[{}_hr_{kh} \cdot {}_h\sigma^2_{kh}|\mathcal{F}_{(k-1)h}] \neq 0$ since ${}_h\sigma^2_{kh}$ is no longer $\mathcal{F}_{(k-1)h}-$measurable. In this case, ${}_hr_{kh}$ is no longer a martingale difference sequence (MDS). Consequently, we can not obtain analytical expressions for the conditional odd moments of ${}_hr_{kh}$. We may impose stronger assumptions on the parameter sequence to approximate these moments (see Theorem 3.4 of Ding (2021) for a first-order approximation of the conditional moments of ${}_hr_{kh}$).

Wang (2002) points out SV models are not asymptotically equivalent to GARCH models in the LeCam sense despite sharing the same diffusion limit. This is due to their different noise propagation systems. In contrast, the distributions and likelihood functions of both SQ-SHARV and GARCH models are completely determined by $({}_h\epsilon_{kh}, {}_h\epsilon^2_{kh})$ for all $k, h > 0$. Therefore, it is possible that SQ-SHARV and GARCH are asymptotically equivalent in the LeCam sense. This is a rather heuristic argument and awaits formal discussions in future research.

We close this section with a discussion on the temporal aggregation of SQ-SHARV. The full model is not closed under temporal aggregation. However, a special case when $\alpha = \psi = 0$ is closed. In this case, the discrete time process (3.2) becomes

$$\sigma_t^2 = \beta\sigma^2_{t-1} + \eta\sigma^2_{t-1}\epsilon_t^2. \tag{3.19}$$

We still have conditional heteroskedasticity in the variance of volatility.

Theorem 3.2. *If ϵ_t satisfy Assumption 1, then (3.1) and (3.19) are closed under temporal aggregation with the filtration $\mathcal{F}_{(m)t}$ generated by aggregated returns $r_0, r_m, ..., r_t$. That is, the joint process $(r_{(m)t}, \sigma^2_{(m)t})$ satisfies*

$$r_{(m)t} = \sigma_{(m)t}\epsilon_{(m)t}, \tag{3.20}$$

$$\sigma^2_{(m)t} = \beta_{(m)}\sigma^2_{(m)t-m} + \eta_{(m)}\sigma^2_{(m)t-m}\epsilon^2_{(m)t}, \tag{3.21}$$

where

$$\sigma^2_{(m)t} = \sum_{j=0}^{m-1}\sigma^2_{t-j} + \eta\sum_{j=0}^{m-2}\frac{\sum_{k=0}^{m-2-j}\beta^k}{\beta^{m-1-j}}\sigma^2_{t-1-j}\epsilon^2_{t-j}, \tag{3.22}$$

$$\beta_{(m)} = \beta^m, \tag{3.23}$$

$$\eta_{(m)} = \eta\sum_{j=1}^{m}\binom{m}{k}\beta^{m-k}\eta^{k-1}. \tag{3.24}$$

4 Conclusion

In this paper, we have derived the diffusion limits of RT-GARCH and SQ-SHARV models. In doing so, we have answered the question of where RT-GARCH stands in between GARCH and SV models and argued RT-GARCH should be used with caution. This is because the diffusion limit of RT-GARCH fails to generate with probability one positive volatility unless it is degenerate. In light of this, we have proposed the SQ-SHARV model to address these issues. This paper therefore, provides additional theoretical justification for the SQ-SHARV and its general form, ART-GARCH model proposed by Ding (2021).

GARCH-type models encompass large variations and are relatively easy to implement in practice. RT-GARCH-type models provide an alternative way of treating volatility as a stochastic process while retaining the elegant QML estimation procedure. While the results call for caution when using RT-GARCH for volatility modelling, the structure of RT-GARCH opens up a whole new area: introducing the volatility of volatility as an additional source of risk. To our knowledge, the consideration of the volatility of volatility in SQ-SHARV is novel. However, the parametric specification of the volatility of volatility in SQ-SHARV is quite simple; we can specify a separate latent process to describe the dynamic of this source of risk. We leave this for future research.

A Proofs

In this section we present proofs of the main theorems of this paper. Throughout this section we assume $kh \leq t < (k+1)h$ for each $h > 0$ unless specified otherwise.

Proof of Theorem 2.1. The joint process $({}_h\sigma^2_{kh}, {}_hr^2_{kh})$ which satisfies (2.1) - (2.3) is Markov. Therefore, to prove Theorem 2.1, it suffices to verify Assumptions 1–4 of Nelson (1992). Assumption 3 of the convergence of initial points is already assumed in the theorem.

To verify Assumption 1, we need to match the first two conditional moments of the approximating process to those of the diffusion process. The conditional means per unit time are given by

$$h^{-1}\mathbb{E}[{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}|\mathcal{F}_{kh}] = \alpha + \psi + (\beta - 1){}_h\sigma^2_{kh} + \gamma_h r^2_{kh}, \tag{A.1}$$

$$h^{-1}\mathbb{E}[{}_hr^2_{(k+1)h} - {}_hr^2_{kh}|\mathcal{F}_{kh}] = {}_h\sigma^2_{kh} - {}_hr^2_{kh} + h^{-1/2}\big(\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] - {}_h\sigma^2_{kh}\big). \tag{A.2}$$

Substituting (2.1) into the expectation on the RHS of (A.2),

$$\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = {}_h\sigma^2_{kh} + 2\sqrt{h}\psi + h(\alpha + \psi) + h(\beta - 1){}_h\sigma^2_{kh} + h\gamma_h r^2_{kh} \tag{A.3}$$

Therefore, the drift terms of the diffusion limit are given by

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}|\mathcal{F}_{kh}] = \alpha + \psi + (\beta - 1)\sigma^2 + \gamma r^2, \tag{A.4}$$

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_hr^2_{(k+1)h} - {}_hr^2_{kh}|\mathcal{F}_{kh}] = 2\psi + \sigma^2 - r^2. \tag{A.5}$$

Similarly, the limits of the second moments per unit time are given by

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} \big(&2\psi^2 + h(\alpha+\psi)^2 + h(\beta-1)^2{}_h\sigma^4_{kh} \\ &+ h\gamma^2{}_hr^4_{kh} + 2h(\alpha+\psi)(\beta-1){}_h\sigma^2_{kh} + 2h(\alpha+\psi)\gamma_h r^2_{kh} \\ &+ 2h(\beta-1)\gamma_h\sigma^2_{kh}\cdot{}_hr^2_{kh}\big) = 2\psi^2,\end{aligned} \tag{A.6}$$

and

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hr^2_{(k+1)h} - {}_hr^2_{kh})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} \big(&h({}_h\sigma^2_{kh} - {}_hr^2_{kh})^2 + {}_h\sigma^4_{kh} \\ &+ 2\sqrt{h}(\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] - {}_h\sigma^2_{kh})({}_h\sigma^2_{kh} - {}_hr^2_{kh}) \\ &+ \mathbb{E}[{}_h\sigma^4_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] - 2\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}]{}_h\sigma^2_{kh}\big).\end{aligned} \tag{A.7}$$

By (A.3), $\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = {}_h\sigma^2_{kh} + \mathcal{O}(\sqrt{h})$. Taking the square of (2.1), multiplying by $\epsilon^4_{(k+1)h}$ and taking the conditional expectation, we obtain $\mathbb{E}[{}_h\sigma^4_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] = 3{}_h\sigma^4_{kh} + \mathcal{O}(\sqrt{h})$. Therefore, (A.7) becomes,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hr^2_{(k+1)h} - {}_hr^2_{kh})^2|\mathcal{F}_{kh}] = 2\sigma^4. \tag{A.8}$$

Finally, the limit of the cross moment per unit time is given by

$$\begin{aligned}&\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})({}_hr^2_{(k+1)h} - {}_hr^2_{kh})|\mathcal{F}_{kh}] \\ &= \lim_{h\downarrow 0} \big(\psi(\mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] - \mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}]) + \mathcal{O}(h)\big) = 2\psi\sigma^2.\end{aligned} \tag{A.9}$$

(A.6), (A.8) and (A.9) imply the following instantaneous covariance matrix of (σ^2_t, r^2_t):

$$\begin{bmatrix} 2\psi^2 & 2\psi\sigma^2 \\ 2\psi\sigma^2 & 2\sigma^4 \end{bmatrix}. \tag{A.10}$$

Using the Cholesky decomposition to take the square root of (A.10), we obtain the following diffusion matrix of (σ^2_t, r^2_t):

$$\begin{bmatrix} \sqrt{2}\psi & 0 \\ \sqrt{2}\sigma^2 & 0 \end{bmatrix}. \tag{A.11}$$

Therefore, the correlation ρ between the innovation terms of σ^2_t and r^2_t is one. In other words, σ^2_t and r^2_t share the same innovation term W_t.

It is straightforward to verify the limits of the fourth moments per unit time are zero. Thus, Assumptions 1 and 2 of Nelson (1990) are verified.

It remains to verify the distributional uniqueness of the diffusion limit. (2.4) - (2.6) satisfy Lipschitz and Growth conditions and therefore, there exists a unique square integrable solution to this stochastic differential equation (SDE) system. Assumptions 1 - 4 of Nelson (1990) are verified and Theorem 2.1 follows.

For the derivation of the revised diffusion limit (2.11) - (2.14), we need to calculate the limits of the first two conditional moments of ${}_hS_{(k+1)h} - {}_hS_{kh}$ per unit time, as well as its cross moments with the increments of other state variables. For the other state variables, only $\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}|\mathcal{F}_{kh}]$ will be affected with γr^2 replaced by $\gamma|r^2|$ in (A.4). (A.5), (A.6) and (A.8) continue to hold. Since $\sqrt{|{}_h\sigma^2_{kh}|}\epsilon_{kh}$ is an odd function of ϵ_{kh}, it follows immediately $\sqrt{|{}_h\sigma^2_{kh}|}\epsilon_{kh}$ is an MDS and therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_hS_{(k+1)h} - {}_hS_{kh}|\mathcal{F}_{kh}] = 0. \tag{A.12}$$

The limit of the conditional second moment is more involved. First denote

$${}_hA_{kh} \equiv h(\alpha+\psi) + h(\beta-1){}_h\sigma^2_{kh} + h\gamma|{}_hr^2_{kh}| = \mathcal{O}(h). \tag{A.13}$$

Since $\epsilon^2_{(k+1)h} \geq 0$ a.s., by triangular inequality,

$$\begin{aligned} |{}_h\sigma^2_{kh} + {}_hA_{kh}| - \sqrt{h}\psi\mathbb{E}[|\epsilon^4_{(k+1)h} - \epsilon^2_{(k+1)h}|] &\leq \mathbb{E}[|{}_h\sigma^2_{(k+1)h}|\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] \\ &\leq |{}_h\sigma^2_{kh} + {}_hA_{kh}| + \sqrt{h}\psi\mathbb{E}[|\epsilon^4_{(k+1)h} - \epsilon^2_{(k+1)h}|]. \end{aligned} \tag{A.14}$$

Since $\mathbb{E}[|\epsilon^4_{(k+1)h} - \epsilon^2_{(k+1)h}|] < \infty$ a.s., (A.14) reduces to

$$\mathbb{E}[|{}_h\sigma^2_{(k+1)h}|\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = |{}_h\sigma^2_{kh}| + \mathcal{O}(\sqrt{h}), \tag{A.15}$$

for small h. Therefore, the limit of the second moment per unit time is given by

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})^2|\mathcal{F}_{kh}] = |\sigma^2|. \tag{A.16}$$

The limits of the cross moments with the increments of other state variables are zero. For their derivations, see the proof of Theorem 2.2 below. Therefore, the correlation between the innovation terms of S_t and σ^2_t (and r^2_t) is zero.

For the distributional uniqueness of the weak solution of (2.11) - (2.14), we can apply the standard induction argument in the proof of Theorem 3 of Duan (1997). Since r^2_t and σ^2_t do not depend on S_t, this argument is still valid. □

Proof of Theorem 2.2. It is straightforward to check ${}_h\sigma^2_{kh} > 0$ for all $0 \leq h \leq 1$ and $k \geq 1$ as long as ${}_h\sigma^2_0 > 0$. Therefore, ${}_h\sigma_{kh}$ is well defined for real-valued ${}_h\sigma^2_{kh}$. The limits of the conditional means per unit time are given by

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}|\mathcal{F}_{kh}] = \alpha + \psi + (\beta-1)\sigma^2 + \gamma|r^2|. \tag{A.17}$$

Since

$$\begin{aligned} \mathbb{E}[{}_h\sigma^2_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] &= {}_h\sigma^2_{kh} + 2h\psi + h(\alpha+\psi) + h(\beta-1){}_h\sigma^2_{kh} + h\gamma|{}_hr^2_{kh}| \\ &= {}_h\sigma^2_{kh} + \mathcal{O}(h), \end{aligned} \tag{A.18}$$

we obtain

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_hr^2_{(k+1)h} - {}_hr^2_{kh}|\mathcal{F}_{kh}] = \sigma^2 - r^2. \tag{A.19}$$

The limit of the second moment per unit time of ${}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}$ is given by

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} \big(&h2\psi^2 + h(\alpha+\psi)^2 + h(\beta-1)^2{}_h\sigma^4_{kh} \\ &+ h\gamma^2|{}_hr^2_{kh}|^2 + 2h(\alpha+\psi)(\beta-1){}_h\sigma^2_{kh} + 2h\gamma(\alpha+\psi)|{}_hr^2_{kh}| \\ &+ 2h\gamma(\beta-1){}_h\sigma^2_{kh}|{}_hr^2_{kh}|\big) = 0.\end{aligned} \tag{A.20}$$

(A.8) in the proof of Theorem 2.1 continues to hold.

Since ${}_h\sigma_{kh}\epsilon_{kh}$ is an odd function of ϵ_{kh}, by symmetry of ϵ_{kh}, it follows immediately ${}_h\sigma_{kh}\epsilon_{kh}$ is an MDS. Therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_hS_{(k+1)h} - {}_hS_{kh}|\mathcal{F}_{kh}] = 0. \tag{A.21}$$

The second moment per unit time is given by,

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} \big(&{}_h\sigma^2_{kh} + h(\alpha+\psi) \\ &+ h(\beta-1){}_h\sigma^2_{kh} + h\gamma|{}_hr^2_{kh}| + 2h\psi\big) = \sigma^2.\end{aligned} \tag{A.22}$$

Using the MDS property of ${}_h\sigma_{kh}\epsilon_{kh}$ again, the two cross moments are given by

$$h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})|\mathcal{F}_{kh}] = \sqrt{h}\psi\mathbb{E}[{}_h\sigma_{(k+1)h}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}], \tag{A.23}$$

$$h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})({}_hr^2_{(k+1)h} - {}_hr^2_{kh})|\mathcal{F}_{kh}] = \mathbb{E}[{}_h\sigma^3_{(k+1)h}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}]. \tag{A.24}$$

Since ${}_h\sigma_{kh}\epsilon_{kh}$ is symmetric around zero, $\mathbb{E}[{}_h\sigma^3_{(k+1)h}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}] = 0$. The same applies to ${}_h\sigma_{kh}\epsilon^3_{kh}$ because it is an odd function of ϵ_{kh}. Therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})|\mathcal{F}_{kh}] = 0, \tag{A.25}$$

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})({}_hr^2_{(k+1)h} - {}_hr^2_{kh})|\mathcal{F}_{kh}] = 0. \tag{A.26}$$

The distributional uniqueness follows exactly the argument in the proof of Theorem 2.1. Since $\mathbb{P}({}_h\sigma^2_{kh} > 0) = 1$ for all $0 \le h \le 1$ and $k \ge 1$, the weak convergence in distribution ensures $\mathbb{P}(\sigma^2_t > 0) = 1$. □

Proof of Theorem 2.3. It suffices to consider the case for $m = 2$, the general case follows by induction. Let

$$r_{(2)t} = \sigma_{(2)t}\epsilon_{(2)t}, \tag{A.27}$$

$$\sigma^2_{(2)t} = \alpha_{(2)} + \beta_{(2)}\sigma^2_{(2)t-2} + \psi_{(2)}\epsilon^2_{(2)t}, \tag{A.28}$$

where

$$\alpha_{(2)} = 2\alpha(\beta+1), \tag{A.29}$$

$$\beta_{(2)} = \beta^2, \tag{A.30}$$

$$\psi_{(2)} = \psi(\beta+1)(1/\beta+1), \tag{A.31}$$

and the aggregated innovation term is given by,

$$\epsilon_{(2)t} = \text{sign}(r_{(2)t})\sqrt{(\epsilon_t^2 + \beta\epsilon_{t-1}^2)/(1+\beta)}. \tag{A.32}$$

Define $\sigma^2_{(2)t} \equiv \sigma_t^2 + \sigma_{t-1}^2 + \psi/\beta\epsilon_t^2$. By repeated substitution, we obtain (A.28). Since $\mathbb{E}[\epsilon_t^2 + \beta\epsilon_{t-1}^2|\mathcal{F}_{(2)t-2}] = \mathbb{E}[\mathbb{E}[\epsilon_t^2|\mathcal{F}_{t-1}] + \beta\mathbb{E}[\epsilon_{t-1}^2|\mathcal{F}_{t-2}]|\mathcal{F}_{(2)t-2}] = 1+\beta$ by law of iterated expectation, $\mathbb{E}[\epsilon^2_{(2)t}|\mathcal{F}_{(2)t-2}] = 1$. Similarly, we can show $r_{(2)t}$ is an MDS with respect to the aggregated filtration $\mathcal{F}_{(2),t-2}$ by law of iterated expectation and the fact r_t is a symmetric MDS w.r.t. $\mathcal{F}_{t-1}$. Therefore, the aggregated innovation term $\epsilon_{(2)t}$ is a symmetric MDS w.r.t $\mathcal{F}_{(2),t-2}$ since $\text{sign}(r_{(2)t}) = \pm 1$ with equal probabilities and $\epsilon_t^2 + \beta\epsilon_{t-1}^2 \geq 0$ with probability one. □

Proof of Theorem 3.1. The distributional uniqueness follows directly the argument of section 2.3 of Nelson (1990) since it is the same as GARCH diffusion. Therefore, we only need to verify Assumptions 1 and 2 of Nelson (1990). ${}_hr_{kh}$ is still an MDS under SQ-SHARV in the absence of contemporaneous leverage effect. Thus, the limits of the first moments per unit time are given by

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_hS_{(k+1)h} - {}_hS_{kh}|\mathcal{F}_{kh}] = 0, \tag{A.33}$$

and

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[{}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} h^{-1}\big(\alpha_h + \psi_h + (\beta_h - 1 + \eta_h){}_h\sigma^2_{kh}\big) = \mu - \theta\sigma, \tag{A.34}$$

by (3.7) and (3.8) of Assumption 2.

The limits of the second moments per unit time are given by

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_hS_{(k+1)h} - {}_hS_{kh})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0} \alpha_h + 3\psi_h + (\beta_h - 1 + \eta_h){}_h\sigma^2_{kh} + {}_h\sigma^2_{kh} = \sigma^2, \tag{A.35}$$

using (3.7) and (3.8) of Assumption 2 again and

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})^2|\mathcal{F}_{kh}] &= \lim_{h\downarrow 0} h^{-1}\Big((\alpha_h + \psi_h)^2 + (\beta_h - 1 + \eta_h)^2{}_h\sigma^4_{kh} + 2\psi_h^2 \\ &\quad + 2\eta_h^2 \cdot {}_h\sigma^4_{kh} + 2(\alpha_h + \psi_h)(\beta_h - 1 + \eta_h){}_h\sigma^2_{kh} + 4\psi_h\eta_h \cdot {}_h\sigma^2_{kh}\Big) = 2\lambda^2\sigma^4,\end{aligned} \tag{A.36}$$

using (3.7) - (3.9) of Assumption 2 and the fact that the non-negativity constraints on α_h and ψ_h requires both of them to be of order $\mathcal{O}(h)$ in order to satisfy (3.7).

Finally, the cross moment is given by

$$\begin{aligned}&\lim_{h\downarrow 0} h^{-1}\mathbb{E}[({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})({}_hS_{(k+1)h} - {}_hS_{kh})|\mathcal{F}_{kh}] \\ &= \lim_{h\downarrow 0} h^{-2}\Big(\psi_h\mathbb{E}[{}_h\sigma_{(k+1)h} \cdot {}_h\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}] + \eta_h\mathbb{E}[{}_h\sigma_{(k+1)h} \cdot {}_h\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}]{}_h\sigma^2_{kh}\Big).\end{aligned} \tag{A.37}$$

We have argued in the proof of Theorem 2.2 that both terms inside the limit of (A.37) are MDS. Therefore, (A.37) equals zero. Since ${}_h\epsilon^6_{kh} = \mathcal{O}_p(h^3)$ and ${}_h\epsilon^8_{kh} = \mathcal{O}_p(h^4)$, it is straightforward to verify the fourth moments go to zero in the limit. The weak convergence in Theorem 3.1 then follows. □

Proof of Theorem 3.2. The exact expression for the aggregated innovation term is very complicated as it involves the sum of the products of different combinations of n elements of $(\epsilon_t^2, ...\epsilon_{t-m+1}^2)$ for $n = 1$ to $m-1$. To prove the theorem, it suffices to consider a particular number of m and the general case will follow by induction. Since the formulae are complicated, we consider the case when $m = 3$. Let

$$\sigma_{(3)t}^2 = \sigma_t^2 + \sigma_{t-1}^2 + \sigma_{t-2}^2 + \tfrac{\eta(1+\beta)}{\beta^2}\sigma_{t-1}^2\epsilon_t^2 + \tfrac{\eta}{\beta}\sigma_{t-2}^2\epsilon_{t-1}^2. \tag{A.38}$$

By repeated substitution, we obtain

$$\begin{aligned}\sigma_{(3)t}^2 &= \beta^3(\sigma_{t-3}^2 + \sigma_{t-4}^2 + \sigma_{t-5}^2 + \tfrac{\eta(1+\beta)}{\beta^2}\sigma_{t-4}^2\epsilon_{t-3}^2 + \tfrac{\eta}{\beta}\sigma_{t-5}^2\epsilon_{t-4}^2) \\ &\quad + \eta\tfrac{1+\beta+\beta^2}{\beta^2}(\sigma_{t-1}^2\epsilon_t^2 + \beta\sigma_{t-2}^2\epsilon_{t-1}^2 + \beta^2\sigma_{t-3}^2\epsilon_{t-2}^2).\end{aligned} \tag{A.39}$$

Substituting for σ_{t-1}^2 and σ_{t-2}^2 again and using (A.38),

$$\begin{aligned}\sigma_{(3)t}^2 &= \beta^3\sigma_{(3)t-3}^2 + \tfrac{\eta(1+\beta+\beta^2)}{\beta^2}\sigma_{t-3}^2\big(\beta^2(\epsilon_t^2 + \epsilon_{t-1}^2 + \epsilon_{t-2}^2) \\ &\quad + \beta\eta(\epsilon_t^2\epsilon_{t-1}^2 + \epsilon_{t-1}^2\epsilon_{t-2}^2 + \epsilon_{t-3}^2\epsilon_t^2) + \eta^2\epsilon_t^2\epsilon_{t-1}^2\epsilon_{t-2}^2).\end{aligned} \tag{A.40}$$

Finally, expanding σ_{t-3}^2 again and using (A.38),

$$\begin{aligned}\sigma_{(3)t}^2 &= \beta^3\sigma_{(3)t-3}^2 + \eta\sigma_{(3)t-3}^2\big(\beta^2(\epsilon_t^2 + \epsilon_{t-1}^2 + \epsilon_{t-2}^2) \\ &\quad + \beta\eta(\epsilon_t^2\epsilon_{t-1}^2 + \epsilon_{t-1}^2\epsilon_{t-2}^2 + \epsilon_{t-3}^2\epsilon_t^2) + \eta^2\epsilon_t^2\epsilon_{t-1}^2\epsilon_{t-2}^2\big).\end{aligned} \tag{A.41}$$

Note the term involving ϵ^2 in (A.41) is the sum of all the different combinations of $1, 2, 3$ elements of $(\epsilon_t^2, \epsilon_{t-1}^2, \epsilon_{t-2}^2)$. Denote this term by $f_3(\epsilon_t^2)$. Since ϵ_t is an MDS with unit conditional variance, using binomial theorem, we obtain

$$\mathbb{E}[f_3(\epsilon_t^2)|\mathcal{F}_{(3)t-3}] = \sum_{k=1}^{3}\binom{3}{k}\beta^{3-k}\eta^{k-1} = 3\beta^2 + 3\beta\eta + \eta^2. \tag{A.42}$$

Dividing $f_3(\epsilon_t^2)$ by the RHS of (A.42), we obtain a term with unit conditional mean. Therefore, the aggregated squared innovation is given by

$$\epsilon_{(3)t}^2 = \frac{\beta^2(\epsilon_t^2 + \epsilon_{t-1}^2 + \epsilon_{t-2}^2) + \beta\eta(\epsilon_t^2\epsilon_{t-1}^2 + \epsilon_{t-1}^2\epsilon_{t-2}^2 + \epsilon_{t-3}^2\epsilon_t^2) + \eta^2\epsilon_t^2\epsilon_{t-1}^2\epsilon_{t-2}^2}{3\beta^2 + 3\beta\eta + \eta^2}. \tag{A.43}$$

The expression for $\eta_{(3)}$ is obtained by multiplying η by the conditional mean of $f_3(\epsilon_t^2)$. The aggregated innovation term can then be defined as

$$\epsilon_{(3)t} = \mathrm{sign}(r_{(3)t})\sqrt{\epsilon_{(3)t}^2}. \tag{A.44}$$

Since return is an MDS, it is easy to check the conditional mean of $\epsilon_{(3)t}$ is zero. □

References

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

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

Corradi, V. (2000). Reconsidering the continuous time limit of the GARCH(1,1) process. *Journal of Econometrics 96*(1), 145 – 153.

Ding, Y. (2021). Augmented Real-Time GARCH: A joint model for returns, volatility and volatility of volatility. Cambridge Working Paper in Economics CWPE2112, University of Cambridge. Available at: http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2112.pdf.

Drost, F. C. and T. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica 61*(4), 909–927.

Drost, F. C. and B. J. Werker (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics 74*(1), 31 – 57.

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

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

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

Fornari, F. and A. Mele (1997). Weak convergence and distributional assumptions for a general class of nonlinear ARCH models. *Econometric Reviews 16*(2), 205–229.

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

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Hafner, C. M., S. Laurent, and F. Violante (2017). Weak diffusion limits of dynamic conditional correlation models. *Econometric Theory 33*(3), 691–716.

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

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

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

Meddahi, N. and E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics 119*(2), 355 – 379.

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

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

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

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

Politis, D. N. (1995). Model-free versus model-based volatility prediction. *Journal of Financial Econometrics 67*(3), 358–359.

Protter, P. E. (2004). *Stochastic integration and differential equations* (2 ed.). Applications of mathematics, 21. Berlin; London: Springer.

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

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

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

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Annals of Statistics 30*(3), 754–783.