UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Diffusion Limits of Real-Time GARCH

Yashuang
(Dexter)
Ding

Abstract

We prove that the diffusion limit of Real-Time GARCH (RT-GARCH) exists if we introduce an auxiliary process to state the system in a Markovian form. The volatility in the diffusion limit follows an Ornstein-Uhlenbeck-type process which fails to be positive with probability one. Moreover, only a degenerate diffusion limit can render an almost surely positive volatility process. As a result, we call for caution when using RT-GARCH since it lacks compatibility with existing asset pricing theories. The result also provides a new insight into how different specifications for GARCH affect its diffusion limit.

Reference Details

CWPE	20112
Published	25 November 2020
Updated	19 November 2021

Key Words	GARCH, RT-GARCH, SV, diffusion limit
JEL Codes	C22, C32, C58

Website	www.econ.cam.ac.uk/cwpe

Diffusion Limit of Real-Time GARCH

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 prove that the diffusion limit of Real-Time GARCH (RT-GARCH) exists if we introduce an auxiliary process to state the system in a Markovian form. The volatility in the diffusion limit follows an Ornstein-Uhlenbeck-type process which fails to be positive with probability one. Moreover, only a degenerate diffusion limit can render an almost surely positive volatility process. As a result, we call for caution when using RT-GARCH since it lacks compatibility with existing asset pricing theories. The result also provides a new insight into how different specifications for GARCH affect its diffusion limit.

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

*I am grateful to my supervisor, Oliver Linton, for his valuable comments and suggestions, as well as his constant supports and encouragement. I would also like to thank Alexey Onatskiy, 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

There are generally two approaches to model the volatility of asset returns. GARCH-type models (e.g., Engle (1982), Bollerslev (1986) and Nelson (1991)) regard the volatility as solely determined by past observations. Stochastic volatility (SV) models (e.g., Heston (1993)), on the other hand, treat the volatility as a separate stochastic process. Nelson (1990) establishes the weak convergence of GARCH-type models to continuous time SV models when the length of the discrete time intervals between observations goes to zero. Duan (1997) extends the weak convergence for a wide class of GARCH-type models. In addition, Drost and Werker (1996) use the temporal aggregation properties to derive the diffusion limits for the class of weak GARCH-type models.

Politis (1995) and Breitung and Hafner (2016) argue that most GARCH-type and SV models make inefficient use of all available information for volatility estimation since they ignore the (arguably most important) information in the current observation. To address this issue, Smetanina (2017) proposes the RT-GARCH model which incorporates current return innovation into the volatility process. Specifically,

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

where $r_t \equiv S_t - S_{t-1} = \sigma_t \epsilon_t$ and ϵ_t are $i.i.d.(0,1)$ symmetric random variables with finite fourth moment. σ_t^2 is not deterministic conditional on the σ-algebra $\mathcal{F}_{t-1}$ generated by all available information up to time $t-1$. Setting $\gamma = 0$, (1.1) reduces to a special case of the contemporaneous stochastic autoregressive volatility (SARV) model of Andersen (1994). Therefore, RT-GARCH can be regarded as a hybrid of GARCH and SV models. Unlike most SV models, RT-GARCH has analytical expressions for both the likelihood function and conditional variance of returns. Separately, Breitung and Hafner (2016) propose a model where they use the log squared current return innovation to drive the log volatility process. Their model is closely related to E-GARCH and can be viewed as a special case of the contemporaneous exponential stochastic autoregressive volatility (E-SARV) model of Taylor (1994). However, their model is only useful for volatility nowcast. Forecast is not available in their model. Ding (2021) proposes the stochastic heteroskedastic autoregressive volatility (SHARV) model which allows for conditional heteroskedasticity in the volatility while incorporating the current observation into volatility estimate. He shows that SHARV retains all the advantages of RT-GARCH and Breitung and Hafner's (2016) model while having the usual GARCH diffusion limit.

While the properties of RT-GARCH and Breitung and Hafner's (2016) model are well studied in discrete time, it remains to derive their diffusion limits in order to understand their asymptotic properties in continuous time. In this paper, we employ the approach of Duan (1997) to derive the diffusion limit of RT-GARCH. As we will see, the diffusion limit only exists if we introduce an auxiliary process to state the system in a Markovian form. This is because the joint process (S_t, σ_t^2) under RT-GARCH is not Markov, only (r_t, σ_t^2) is

Markov. We show that the volatility in the diffusion limit follows an Ornstein-Uhlenbeck (OU)-type process. Since OU-type processes permit negative values, this is an undesirable feature. Moreover, the diffusion limit is difficult to interpret and lacks compatibility with existing asset pricing theories. As a result, we suggest that the SHARV model can address this issue since it has the usual GARCH diffusion limit.

The rest of the paper is structured as follows. In section 2, we derive the diffusion limit of RT-GARCH. In section 3, we discuss the specifications for the auxiliary process. Section 4 concludes. All proofs are presented in appendix A and the diffusion limit of Breitung and Hafner's (2016) model is included in appendix B.

2 Main result

We refer to section 2 of Nelson (1990) for a detailed discussion on the weak convergence of Markov chains to diffusion processes. 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 $r_{t-1} \equiv S_{t-1} - S_{t-2}$. This can be seen upon expressing (1.1) as an ARMA$(1,1)$ process,

$$\sigma_t^2 = \alpha + \psi(1+\kappa\gamma) + (\beta+\gamma)\sigma_{t-1}^2 + \psi(\epsilon_t^2 - 1) + \gamma z_{t-1}, \tag{2.1}$$

where $\kappa = \mathbb{E}\epsilon_t^4 - 1$ and $z_t = r_t^2 - \sigma_t^2 - \kappa\psi$. It is easy to see that z_t is a martingale difference sequence (MDS). In contrast, the joint process (S_t, σ_{t+1}^2) under GARCH is Markov and σ_{t+1}^2, which is $\mathcal{F}_t$-measurable, follows an AR(1) process. Naturally, we would expect the diffusion limit of (2.1) to have a continuous time ARMA$(1,1)$ structure. However, to our knowledge, there is no literature yet on the weak convergence of an ARMA$(1,1)$ process to a continuous time equivalence. As a result, we can only establish the weak convergence for the joint process (S_t, σ_t^2, r_t^2). Since the joint distribution of (S_t, σ_t^2) for RT-GARCH is not the same as that of (S_t, σ_{t+1}^2) for GARCH, the diffusion limit of RT-GARCH will not nest that of GARCH as a special case even though RT-GARCH nests GARCH by setting $\gamma = 0$ in discrete time. The situation is somewhat similar to the diffusion limit of GARCH(p,q) discussed in Duan (1997) where r_{t-i}^2 and σ_{t-j}^2 for $i,j > 1$ destroys the Markov structure. To state the system in a Markovian form, he specifies some auxiliary processes for each of the r_{t-i}^2 and σ_{t-j}^2 in the approximating GARCH process.[1] We will adopt his approach by specifying an auxiliary process for r_{t-1}^2. This will in turn, modify the rescaled RT-GARCH process.

Before deriving the diffusion limit, we first change some notations by letting $V_t \equiv \sigma_t^2$ and $R_t \equiv \sigma_t^2\epsilon_t^2$ for RT-GARCH for reasons of clarity which will become clear later. Next

[1] The approximation scheme of Duan (1997) differs from that of Nelson (1990) in that Nelson (1990) let the parameters vary with h in different rates while Duan (1997) scales the (conditionally) deterministic variables by h and the stochastic parts by $\sqrt{h}$ while keeping the parameters fixed. The results are albeit equivalent since they are different representations of the law of large numbers and functional central limit theorem.

we discuss why it is necessary to introduce an auxiliary process into RT-GARCH. Recall the rescaled GARCH process of Duan (1997),

$$\Delta_h S_{kh} = \sqrt{h}{}_h\sigma_{kh}\epsilon_{kh}, \tag{2.2}$$

$$\Delta_h\sigma^2_{(k+1)h} = h\alpha + h(\beta+\gamma-1)_h\sigma^2_{kh} + \sqrt{h}\gamma(\epsilon^2_{kh}-1)_h\sigma^2_{kh}, \tag{2.3}$$

where $\epsilon_{kh} \sim N(0,1)$ and $\Delta_h X_{kh} = {}_hX_{kh} - {}_hX_{(k-1)h}$. We use the left subscript to indicate that the state variables depend on the choice of h. The information set $\hat{\mathcal{F}}_{kh}$ for GARCH is generated by ${}_hS_0, ..., {}_hS_{kh}$ and ${}_h\sigma^2_0, ..., {}_h\sigma^2_{(k+1)h}$. It is easy to check that ${}_h\sigma^2_{kh} > 0$ almost surely for all $0 < h \leq 1$. Nelson (1990) and Duan (1997) show that under regularity conditions, as $h \downarrow 0$, the joint process $({}_hS_t, {}_h\sigma^2_t)$ which is obtained by setting ${}_hS_{kh} = {}_hS_t$ and ${}_h\sigma^2_{(k+1)h} = {}_h\sigma^2_t$ with probability one for $kh \leq t < (k+1)h$, converges weakly to the joint process (S_t, σ^2_t) which satisfies

$$dS_t = \sigma_t dW_{1,t} \tag{2.4}$$

$$d\sigma^2_t = (\alpha + (\beta+\gamma-1)\sigma^2_t)dt + \sqrt{2}\gamma\sigma^2_t dW_{2,t}, \tag{2.5}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent standard Brownian motions, independent of the initial points (S_0, σ^2_0). Since RT-GARCH nests GARCH as a special case, it is tempting to rescale RT-GARCH in a similar fashion:

$$\Delta_h V_{kh} = h(\alpha+\psi) + h(\beta+\gamma-1)_hV_{(k-1)h} + \sqrt{h}\gamma(\epsilon^2_{(k-1)h}-1)_hV_{(k-1)h} + \sqrt{h}\psi(\epsilon^2_{kh}-1). \tag{2.6}$$

For RT-GARCH, $\mathcal{F}_{kh}$ is generated by ${}_hS_0, ..., {}_hS_{kh}$ and ${}_hV_0, ..., {}_hV_{kh}$. (2.6) nests (2.3) as a special case by setting $\psi = 0$. Note that ${}_hV_{kh}$ fails to be positive with probability one since the term $h(\alpha+\psi) - \sqrt{h}\psi$ can be negative for some $0 \leq h < 1$. ${}_hV_{kh}$ is clearly not Markov, therefore, in order to apply the weak convergence results for Markov chains, we should treat ${}_hV_{kh}\epsilon^2_{kh}$ as a state variable rather than an error term. The scaling factor $\sqrt{h}$ for ${}_hV_{(k-1)h}(\epsilon^2_{(k-1)h}-1)$ is also problematic. This is because $h^{-1}\mathbb{E}[\Delta_hV_{kh}|\mathcal{F}_{(k-1)h}] \to \infty$ as $h \downarrow 0$ since $h^{-1/2}{}_hV_{(k-1)h}(\epsilon^2_{(k-1)h}-1) \to \infty$ as $h \downarrow 0$ conditional on $\mathcal{F}_{(k-1)h}$, that is, we cannot find a diffusion process whose drift term matches the limit of the conditional mean of Δ_hV_{kh} per unit time. Consequently, we have to rescale ${}_hV_{(k-1)h}(\epsilon^2_{(k-1)h}-1)$ by h on the right hand side (RHS) of (2.6) instead. In contrast, $\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h\sigma^2_{(k+1)h}|\hat{\mathcal{F}}_{(k-1)h}] = \alpha + (\beta+\gamma-1)_h\sigma^2_{kh}$ for GARCH. Even if we could find a continuous process V_t as the diffusion limit of ${}_hV_{kh}$, ${}_hV_{kh}\epsilon^2_{kh}$ would have a diffusion limit $V_t\epsilon^2_t$ which does not have a continuous sample path since ϵ^2_t is a pure jump process. As a result, we cannot apply the diffusion approximation theorem of Stroock and Varadhan (1979). To overcome this, we can replace ${}_hV_{kh}\epsilon^2_{kh}$ by a 'smoothed' version to ensure the sample path continuity.

Having discussed the issues with rescaling RT-GARCH, we proceed by introducing an auxiliary process ${}_hR_{kh}$ which is a function of ${}_hV_{kh}\epsilon^2_{kh}$ to replace the squared return in the volatility process. Let the joint process $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ be given by

$$\mathbb{P}\big(({}_hS_0, {}_hV_0, {}_hR_0) \in \Gamma\big) = v_h(\Gamma) \text{ for any } \Gamma \in B(\mathbb{R}^3), \tag{2.7}$$

$$\Delta_h S_{kh} = \sqrt{h}\sqrt{|{}_hV_{kh}|}\epsilon_{kh}, \tag{2.8}$$

$$\Delta_h V_{kh} = h(\alpha + \psi) + h(\beta - 1){}_hV_{(k-1)h} + h\gamma \cdot {}_hR_{(k-1)h} + \sqrt{h}\psi(\epsilon_{kh}^2 - 1), \tag{2.9}$$

$$\lim_{h\downarrow 0} \mathbb{E}[(\Delta_h R_{kh})^i | {}_hV_{(k-1)h} = V, {}_hR_{(k-1)h} = R] = b_i(V, R) \text{ for } i = 1, 2, \tag{2.10}$$

$$\lim_{h\downarrow 0} \mathbb{E}[\Delta_h V_{kh}\Delta_h R_{kh} | {}_hV_{(k-1)h} = V, {}_hR_{(k-1)h} = R] = b_3(V, R), \tag{2.11}$$

$$\lim_{h\downarrow 0} \mathbb{E}[\Delta_h S_{kh}\Delta_h R_{kh} | \mathcal{F}_{(k-1)h}] = 0, \tag{2.12}$$

$$\lim_{h\downarrow 0} \mathbb{E}[(\Delta_h R_{kh})^{2+\delta} | \mathcal{F}_{(k-1)h}] = 0 \text{ for some } \delta > 0, \tag{2.13}$$

where $\epsilon_{kh} \sim N(0, 1)$, $B(\mathbb{R}^n)$ denote the Borel sets on $\mathbb{R}^n$, $v_h(\cdot)$ is a probability measure on $(\mathbb{R}^3, B(\mathbb{R}^3))$ and the functions

$$b(S, V, R) \equiv \begin{bmatrix} 0 \\ \alpha + \psi + (\beta - 1)V + \gamma R \\ b_1(V, R) \end{bmatrix}, \tag{2.14}$$

$$a(S, V, R) \equiv \begin{bmatrix} |V| & 0 & 0 \\ 0 & 2\psi^2 & b_3(V, R) \\ 0 & b_3(V, R) & b_2(V, R) \end{bmatrix}, \tag{2.15}$$

satisfy the nonexplosion condition of either Nelson (1990) or Hafner et al. (2017) or both and one or more of conditions A, B, C and D of Nelson (1990). In addition, we require that ${}_1R_k = {}_1V_k\epsilon_k^2$, that is, the auxiliary process is exactly the return process when $h = 1$ to keep in-line with RT-GARCH in discrete time. However, for $h < 1$, (2.9) is not exactly a rescaled version of (2.6) since ${}_hR_{kh} \neq {}_hV_{kh}\epsilon_{kh}^2$ in general. Consequently, (2.9) no longer nests GARCH when $\psi = 0$. Note that we have taken the absolute value of ${}_hV_{kh}$ on the RHS of (2.8) to ensure that ${}_hS_{kh}$ is real-valued. Regardless of the exact specification of $\Delta_h R_{kh}$, the process ${}_hV_{kh}$ will always converge to the same type of diffusion process since ${}_hR_{(k-1)h}$ is fixed at time $(k-1)h$. Since our primary interest is the diffusion limit of ${}_hV_{kh}$, the arbitrariness in the choice of ${}_hR_{kh}$ is not a major issue.

Theorem 2.1. *Let $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ satisfy (2.7) - (2.10). Let ${}_hS_t = {}_hS_{kh}$, ${}_hV_t = {}_hV_{kh}$ and ${}_hR_t = {}_hR_{kh}$ with probability one for $kh \leq t < (k+1)h$. If $({}_hS_0, {}_hV_0, {}_hR_0) \xrightarrow{d} (S_0, V_0, R_0)$ as $h \downarrow 0$, then $({}_hS_t, {}_hV_t, {}_hR_t) \Rightarrow (S_t, V_t, R_t)$ as $h \downarrow 0$, where '$\Rightarrow$' denotes the weak convergence and the joint process (S_t, V_t, R_t) satisfies*

$$dS_t = \sqrt{|V_t|}dW_{1,t} \tag{2.16}$$

$$dV_t = \big(\alpha + \psi + (\beta - 1)V_t + \gamma R_t\big)dt + \sqrt{2}\psi dW_{2,t}, \tag{2.17}$$

$$dR_t = b_1(V_t, R_t)dt + \sqrt{b_2(V_t, R_t)}(\rho_t dW_{2,t} + \sqrt{1 - \rho_t^2}dW_{3,t}), \tag{2.18}$$

$$\mathbb{P}\big((S_0, V_0, R_0) \in \Gamma\big) = v_0(\Gamma) \; \textit{for any } \Gamma \in B(R^3), \tag{2.19}$$

where $W_{1,t}$*,* $W_{2,t}$ *and* $W_{3,t}$ *are three independent standard Brownian motions, independent of* (S_0, V_0, R_0) *and* $\rho_t = b_3(V_t, R_t)/(\sqrt{2}\psi\sqrt{b_2(V_t, R_t)})$*.*

From (2.16), V_t follows an OU-type with an explanatory variable R_t. Therefore, V_t fails to be positive with probability one. Consequently, we have to take the absolute value of V_t to ensure that S_t is real-valued. It is not surprising that (2.17) does not nest the GARCH diffusion as a special case since (2.9) does not nest (2.3) as a special case except when $h = 1$. By the Meyer-Tanaka formula (Protter, 2004), we have $d|V_t| = \text{sign}(V_t)dV_t + dL_t^0(V_t)$ where $L_t^0(V_t)$ is the local time of V_t at zero. Regardless of the specification of ${}_hR_{kh}$, the process ${}_hV_{kh}$ will always converge weakly to an OU-type process as long as the diffusion limit exists and $\psi \neq 0$. The diffusion process in Theorem 2.1 is difficult to interpret and lacks compatibility with existing asset pricing theories.

To understand why the diffusion limit of RT-GARCH cannot nest GARCH as a special case, recall that the randomness comes from ${}_hr_{kh}^2$ for GARCH, whereas for RT-GARCH, ${}_hr_{kh}^2$ is (conditional) deterministic. For GARCH, it is the joint process $({}_hS_{kh}, {}_h\sigma^2_{(k+1)h})$, not $({}_hS_{kh}, {}_h\sigma^2_{kh})$, that converges weakly to (S_t, σ_t^2) given in (2.4) and (2.5) as $h \downarrow 0$.[2] In fact, since ${}_h\sigma^2_{kh}$ is $\mathcal{F}_{(k-1)h}$-measurable, the conditional distribution of $({}_hS_{kh}, {}_h\sigma^2_{kh})$ degenerates to $({}_hS_{kh}, \sigma^2)$ for some σ^2 which coincides with RT-GARCH when $\psi = 0$. Therefore, the diffusion limit of the joint process $({}_hS_{kh}, {}_h\sigma^2_{(k+1)h})$ is not the same as that of $({}_hS_{kh}, {}_h\sigma^2_{kh})$ since the weak convergence concerns the entire sample path. A better way to understand this is to consider a different parameterisation of GARCH used by Corradi (2000):

$$\Delta_h S_{kh} = \sqrt{h}\,{}_h\sigma_{(k-1)h}\epsilon_{kh}, \tag{2.20}$$

$$\Delta_h \sigma^2_{kh} = h\alpha + h(\beta + \gamma - 1)\,{}_h\sigma^2_{(k-1)h} + \sqrt{h}\gamma(\epsilon^2_{kh} - 1)\,{}_h\sigma^2_{(k-1)h}. \tag{2.21}$$

In this case, ${}_h\sigma^2_{kh}$ is $\mathcal{F}_{kh}$-measurable and $\sqrt{h}\,{}_h\sigma^2_{kh}$ is the conditional variance of returns at time kh instead of $(k-1)h$. The joint process $({}_hS_{kh}, {}_h\sigma^2_{kh})$ converges weakly to the usual GARCH diffusion as $h \downarrow 0$ by proposition 2.1 of Corradi (2000). However, RT-GARCH does not nest this specification of GARCH as a special case even when $h = 1$ since $\Delta_1 S_t$ in (2.20) is scaled by ${}_1\sigma_{t-1}$ not ${}_1\sigma_t$.

3 Specifications for the auxiliary process

We next consider some specifications for the auxiliary process ${}_hR_{kh}$ using the approach of Duan (1997). Recall the auxiliary system of a GARCH(p, q) process given in (51) of Duan (1997), $\Phi_{t+1} = A + (B + C_t)\phi_t + D_t$, and the corresponding rescaled GARCH(p, q)

[2] To be more rigorous, for GARCH, it is the joint process $({}_hS_t, {}_h\sigma_t^2)$ where ${}_hS_{kh} = {}_hS_t$ and ${}_h\sigma^2_{(k+1)h} = {}_h\sigma_t^2$ with probability one for $kh \leq t < (k+1)h$ that converges weakly to (S_t, σ_t^2) as $h \downarrow 0$.

process given in (53) of Duan (1997),

$$\Delta_h\Phi_{(k+1)h} = h(A+\mathbb{E}D_{kh}) + \big(h(B+\mathbb{E}C_{kh}-I)+\sqrt{h}(C_{kh}-\mathbb{E}C_{kh})\big)_h\Phi_{kh}+\sqrt{h}(D_{kh}-\mathbb{E}D_{kh}),$$

where I is the identity matrix. For example, a GARCH$(1,2)$ process is given by

$$\begin{bmatrix}\phi_{t+1}\\ \phi_t\epsilon_t^2\end{bmatrix} = \begin{bmatrix}\alpha\\ 0\end{bmatrix} + \left(\begin{bmatrix}\beta & \gamma_2\\ 0 & 0\end{bmatrix} + \begin{bmatrix}\gamma_1\epsilon_t^2 & 0\\ \epsilon_t^2 & 0\end{bmatrix}\right)\begin{bmatrix}\phi_t\\ \phi_{t-1}\epsilon_{t-1}^2\end{bmatrix},$$

where $\phi_t \equiv \sigma_t^2$. The rescaled GARCH$(1,2)$ is then given by

$$\begin{bmatrix}\Delta_h\phi_{(k+1)h}\\ \Delta_h\phi_{kh}\epsilon_{kh}^2\end{bmatrix} = h\begin{bmatrix}\alpha\\ 0\end{bmatrix} + \left(h\begin{bmatrix}\beta+\gamma_1-1 & \gamma_2\\ 1 & -1\end{bmatrix} + \sqrt{h}\begin{bmatrix}\gamma_1(\epsilon_{kh}^2-1) & 0\\ \epsilon_{kh}^2-1 & 0\end{bmatrix}\right)\begin{bmatrix}{}_h\phi_{kh}\\ {}_h\phi_{(k-1)h}\epsilon_{(k-1)h}^2\end{bmatrix}.$$

Using this approach, we can treat RT-GARCH similar to a GARCH$(1,2)$ process since r_{t-1}^2 is $\mathcal{F}_{t-1}$-measurable in both models. Let $\Phi_{t+1}^T = (V_t, R_t)$, $A^T = (\alpha, 0)$, $C_t = 0_{2,2}$,

$$B = \begin{bmatrix}\beta & \gamma\\ 0 & 0\end{bmatrix} \text{ and } D_t = \begin{bmatrix}\psi\epsilon_t^2\\ V_t\epsilon_t^2\end{bmatrix},$$

then the corresponding rescaled RT-GARCH is given by

$$\begin{aligned}\begin{bmatrix}\Delta_h V_{kh}\\ \Delta_h R_{kh}\end{bmatrix} &= h\begin{bmatrix}\alpha+\psi\\ \mathbb{E}[{}_hV_{kh}\epsilon_{kh}^2|\mathcal{F}_{(k-1)h}]\end{bmatrix} + h\begin{bmatrix}\beta-1 & \gamma\\ 0 & -1\end{bmatrix}\begin{bmatrix}{}_hV_{(k-1)h}\\ {}_hR_{(k-1)h}\end{bmatrix}\\ &\quad+\sqrt{h}\begin{bmatrix}\psi(\epsilon_{kh}^2-1)\\ {}_hV_{kh}\epsilon_{kh}^2-\mathbb{E}[{}_hV_{kh}\epsilon_{kh}^2|\mathcal{F}_{(k-1)h}]\end{bmatrix}.\end{aligned} \tag{3.1}$$

It is easy to check that (3.1) becomes (1.1) upon setting $h=1$ and $k=t$.

Theorem 3.1. *Let* $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ *satisfy* (2.7) – (2.8) *and* (3.1). *Let* ${}_hS_t = {}_hS_{kh}$, ${}_hV_t = {}_hV_{kh}$ *and* ${}_hR_t = {}_hR_{kh}$ *with probability one for* $kh \le t < (k+1)h$. *If* $({}_hS_0, {}_hV_0, {}_hR_0) \xrightarrow{d} (S_0, V_0, R_0)$ *as* $h \downarrow 0$, *then* $({}_hS_t, {}_hV_t, {}_hR_t) \Rightarrow (S_t, V_t, R_t)$ *as* $h \downarrow 0$ *and the joint process* (S_t, V_t, R_t) *satisfies*

$$dS_t = \sqrt{|V_t|}dW_{1,t} \tag{3.2}$$

$$dV_t = \big(\alpha+\psi+(\beta-1)V_t+\gamma R_t\big)dt + \sqrt{2}\psi dW_{2,t}, \tag{3.3}$$

$$dR_t = (V_t - R_t)dt + \sqrt{2}V_t dW_{2,t}, \tag{3.4}$$

$$\mathbb{P}\big((S_0, V_0, R_0) \in \Gamma\big) = v_0(\Gamma) \text{ for any } \Gamma \in B(R^3), \tag{3.5}$$

where $W_{1,t}$ *and* $W_{2,t}$ *are two independent standard Brownian motions, independent of* (S_0, V_0, R_0).

Corradi (2000) shows that there exists an alternative approximation scheme which leads to a degenerate diffusion limit for GARCH. Specifically, Nelson (1990) and Duan (1997) scale the stochastic parts of the approximating process by $\sqrt{h}$. If we scale the stochastic parts by h instead of $\sqrt{h}$, we end up with a degenerate diffusion limit.[3] We next show that this approximation scheme also leads to a degenerate diffusion limit for RT-GARCH. Moreover, we show that the volatility is positive with probability one for both the approximating process and diffusion limit under this scheme.

Let the joint process $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ be given by

$$\Delta_h S_{kh} = \sqrt{h}\sqrt{{}_hV_{kh}}\epsilon_{kh}, \tag{3.6}$$

$$\begin{bmatrix} \Delta_h V_{kh} \\ \Delta_h R_{kh} \end{bmatrix} = h \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + h \begin{bmatrix} \beta - 1 & \gamma \\ 0 & -1 \end{bmatrix} \begin{bmatrix} {}_hV_{(k-1)h} \\ {}_hR_{(k-1)h} \end{bmatrix} + h \begin{bmatrix} \psi\epsilon_{kh}^2 \\ {}_hV_{kh}\epsilon_{kh}^2 \end{bmatrix}, \tag{3.7}$$

with initial points given in (2.7). It is easy to check that (3.7) becomes (1.1) upon setting $h = 1$ and $k = t$. Moreover, both ${}_hV_{kh}$ and ${}_hR_{kh}$ are positive almost surely. As a result, we do not need to take the absolute value of ${}_hV_{kh}$ on the RHS of (3.6).

Theorem 3.2. *Let $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ satisfy (2.7) and (3.6) – (3.7). Let ${}_hS_t = {}_hS_{kh}$, ${}_hV_t = {}_hV_{kh}$ and ${}_hR_t = {}_hR_{kh}$ with probability one for $kh \leq t < (k+1)h$. If $({}_hS_0, {}_hV_0, {}_hR_0) \xrightarrow{d} (S_0, V_0, R_0)$ as $h \downarrow 0$, then $({}_hS_t, {}_hV_t, {}_hR_t) \Rightarrow (S_t, V_t, R_t)$ as $h \downarrow 0$ and the joint process (S_t, V_t, R_t) satisfies*

$$dS_t = \sqrt{V_t}dW_t \tag{3.8}$$

$$dV_t = \big(\alpha + \psi + (\beta - 1)V_t + \gamma R_t\big)dt, \tag{3.9}$$

$$dR_t = (V_t - R_t)dt, \tag{3.10}$$

$$\mathbb{P}\big((S_0, V_0, R_0) \in \Gamma\big) = v_0(\Gamma) \text{ for any } \Gamma \in B(R^3), \tag{3.11}$$

where W_t is a standard Brownian motions, independent of (S_0, V_0, R_0).

In appendix B, we show that Breitung and Hafner's (2016) model converges weakly to the same diffusion limit as the E-GARCH model without leverage effect. Therefore, Breitung and Hafner's (2016) model has a more theoretically appealing diffusion limit compared to RT-GARCH but at the expense that volatility forecasts are not available for their model. In contrast, the stochastic heteroskedastic autoregressive volatility (SHARV) model of Ding (2021) retains all the advantages of RT-GARCH while having the same diffusion limit as (GJR-)GARCH. As a result, this paper provides additional theoretical justifications for the preference of SHARV over RT-GARCH.

4 Conclusion

In this paper, we have derived the diffusion limit of RT-GARCH. In doing so, we have answered the question where RT-GARCH stands in between GARCH and SV models.

[3] In this case, the law of large number applies to the stochastic part, instead of the functional central limit theorem.

The diffusion limit of RT-GARCH is not well-defined unless we introduce an auxiliary process and consequently, it does not nest GARCH. Moreover, the diffusion limit of RT-GARCH lacks compatibility with existing asset pricing theories and the volatility fails to be positive with probability one unless it comes from a degenerate diffusion. Therefore, it is hard to justify RT-GARCH as the true data generating process. We suggest that the SHARV model can serve as the remedy for the issues with RT-GARCH. We point out that since RT-GARCH follows an ARMA$(1,1)$ process, it would be useful to establish the weak convergence of an ARMA$(1,1)$ to a continuous time analogue. In this way, we no longer need to introduce an auxiliary process to state RT-GARCH in a Markov system.

A Proofs

Throughout this section we assume $kh \leq t < (k+1)h$ for each $0 < h \leq 1$ unless specified otherwise.

Proof of Theorem 2.1. The joint process $({}_hS_{kh}, {}_hV_{kh}, {}_hR_{kh})$ is Markov. Therefore, to prove Theorem 2.1, it suffices to verify Assumptions 1–4 of Nelson (1990). Assumption 3 of the convergence in distribution of initial points is already assumed in the theorem.

To verify Assumption 1, note that $\sqrt{|{}_hV_{kh}|}\epsilon_{kh}$ is an odd function of ϵ_{kh}. It then follows immediately that $\sqrt{|{}_hV_{kh}|}\epsilon_{kh}$ is an MDS. Therefore,

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

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

and the limit of $h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}|\mathcal{F}_{kh}]$ is given in (2.10). Denote ${}_h\hat{V}_{kh} \equiv h(\alpha+\psi) + h(\beta - 1){}_hV_{kh} + h\gamma \cdot {}_hR_{kh} = \mathcal{O}(h)$. since $\epsilon_{kh}^2 \geq 0$ with probability one,

$$\mathbb{E}[|{}_hV_{(k+1)h}|\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = \mathbb{E}[|({}_hV_{kh} + {}_h\hat{V}_{kh})\epsilon^2_{(k+1)h} + \sqrt{h}\psi(\epsilon^4_{(k+1)h} - \epsilon^2_{(k+1)h})||\mathcal{F}_{kh}].$$

By the triangular inequality,

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

Therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h S_{(k+1)h})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\mathbb{E}[|{}_hV_{(k+1)h}|\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = |V|. \tag{A.3}$$

For ${}_hV_{kh}$,

$$\begin{aligned} \lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h V_{(k+1)h})^2|\mathcal{F}_{kh}] &= \lim_{h\downarrow 0}\big(2\psi^2 + h(\alpha+\psi)^2 + h(\beta-1)^2{}_hV^2_{kh} + h\gamma^2{}_hR^2_{kh} \\ &+ 2h(\alpha+\psi)(\beta-1){}_hV_{kh} + 2h(\alpha+\psi)\gamma\cdot{}_hR_{kh} + 2h\gamma(\beta-1){}_hV_{kh}\cdot{}_hR_{kh}\big) = 2\psi^2. \end{aligned} \tag{A.4}$$

The limit of $h^{-1}\mathbb{E}[(\Delta_h R_{(k+1)h})^2|\mathcal{F}_{kh}]$ is given in (2.10). Finally,

$$\begin{aligned}\mathbb{E}[\Delta_h V_{(k+1)h}\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] = ({}_h\hat{V}_{kh} - \sqrt{h}\psi)\mathbb{E}[\sqrt{|{}_h V_{(k+1)h}|}\epsilon_{(k+1)h}|\mathcal{F}_{kh}] \\ + \sqrt{h}\psi\mathbb{E}[\sqrt{|{}_h V_{(k+1)h}|}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}].\end{aligned}$$

Since $\sqrt{|{}_h V_{kh}|}\epsilon^i_{kh}$ are odd functions of ϵ_{kh} for all odd i, they are all MDS and therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h V_{(k+1)h}\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] = 0. \tag{A.5}$$

The limits of $h^{-1}\mathbb{E}[\Delta_h V_{(k+1)h}\Delta_h R_{(k+1)h}|\mathcal{F}_{kh}]$ and $h^{-1}\mathbb{E}[\Delta_h S_{(k+1)h}\Delta_h R_{(k+1)h}|\mathcal{F}_{kh}]$ are given in (2.11) and (2.12). (A.3) - (A.5) and (2.10) - (2.12) imply the following instantaneous covariance matrix of $a(S, V, R)$ which is given in (2.15). Taking the Cholesky decomposition of $a(S, V, R)$, we obtain the following diffusion matrix of (S_t, V_t, R_t):

$$\begin{bmatrix} \sqrt{|V|} & 0 & 0 \\ 0 & \sqrt{2}\psi & 0 \\ 0 & \rho\sqrt{b_2(V,R)} & \sqrt{(1-\rho^2)b_2(V,R)} \end{bmatrix},$$

where $\rho = b_3(V, R)/(\sqrt{2}\psi\sqrt{b_2(V, R)})$.

It is straightforward though tedious to check that the limits of the conditional fourth moments of $\Delta_h S_{kh}$ and $\Delta_h V_{kh}$ per unit time are zero. Together with (2.13), Assumptions 1 and 2 of Nelson (1990) are verified.

Finally, the distributional uniqueness of the weak solution to (2.16) - (2.19) is satisfied by the assumptions on the coefficient matrix $a(S, V, R)$ and $b(S, V, R)$. Assumption 4 of Nelson (1990) is thus verified. □

Proof of Theorem 3.1. We need only to update the limits of the first two conditional moments of $\Delta_h R_{(k+1)h}$ per unit time as well as its cross moment with the increments of the other two state variables per unit time. First note that

$$\mathbb{E}[{}_h V_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = {}_h V_{kh} + 2\sqrt{h}\psi + h\big(\alpha + \psi + (\beta - 1)_h V_{kh} + \gamma\cdot{}_h R_{kh}\big) = {}_h V_{kh} + \mathcal{O}(\sqrt{h}). \tag{A.6}$$

Therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}|\mathcal{F}_{kh}] = V - R. \tag{A.7}$$

Similarly it is easy but tedious to show that

$$\mathbb{E}[{}_h V^2_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] = 3{}_h V^2_{kh} + \mathcal{O}(\sqrt{h}).$$

Therefore,

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h R_{(k+1)h})^2|\mathcal{F}_{kh}] &= \lim_{h\downarrow 0}\Big(\mathbb{E}[{}_h V^2_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] - {}_h V^2_{kh} + \mathcal{O}(\sqrt{h})\Big) \\ &= \lim_{h\downarrow 0}\Big(2\cdot{}_h V^2_{kh} + \mathcal{O}(\sqrt{h})\Big) = 2V^2.\end{aligned} \tag{A.8}$$

For the cross moments, first note that

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big(({}_hV_{kh} - {}_hR_{kh})\mathbb{E}[\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] \\ + \mathbb{E}[\sqrt{|{}_hV_{(k+1)h}|}{}_hV_{(k+1)h}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}] - {}_hV_{kh}\mathbb{E}[\sqrt{|{}_hV_{(k+1)h}|}\epsilon_{(k+1)h}|\mathcal{F}_{kh}]\Big) = 0, \tag{A.9}$$

using (A.1), (A.6) and the fact that the last two terms in the second equality of (A.9) are odd functions of $\epsilon_{(k+1)h}$ and therefore, are MDS. It is easy to show that

$$\mathbb{E}[{}_hV_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] = 3\cdot {}_hV_{kh} + \mathcal{O}(\sqrt{h}). \tag{A.10}$$

Therefore,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}\Delta_h V_{(k+1)h}|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big(\psi(\mathbb{E}[{}_hV_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] \\ - \mathbb{E}[{}_hV_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh} + \mathcal{O}(h)] = 2\psi V, \tag{A.11}$$

using (A.6) and (A.10).

(A.8), (A.9), (A.11) together with (A.3) - (A.5) imply the following instantaneous covariance matrix

$$\begin{bmatrix} |V| & 0 & 0 \\ 0 & 2\psi^2 & 2\psi V \\ 0 & 2\psi V & 2V^2 \end{bmatrix}, \tag{A.12}$$

Taking the Cholesky decomposition of (A.12), we obtain the following diffusion matrix,

$$\begin{bmatrix} \sqrt{|V|} & 0 & 0 \\ 0 & \sqrt{2}\psi & 0 \\ 0 & \sqrt{2}V & 0 \end{bmatrix}. \tag{A.13}$$

(A.13) together with (A.7) imply (3.4).

It is straightforward to check that the limits of the conditional fourth moments of $\Delta_h R_{kh}$. Finally, the existence and uniqueness of strong solution to the diffusion system (3.2) - (3.5) follows exactly the argument of Theorem 3 of Duan's (1997). Theorem 3.1 then follows. □

Proof of Theorem 3.2. It suffices to update the first two conditional moments of $\Delta_h V_{(k+1)h}$ and $\Delta_h R_{(k+1)h}$ as well as their cross moments per unit time under the new approximation scheme. (A.2) is still valid. Since

$$\mathbb{E}[{}_hV_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}] = {}_hV_{kh} + h(3\psi + (\beta-1){}_hV_{kh} + \gamma\cdot {}_hR_{kh}) = {}_hV_{kh} + \mathcal{O}(h),$$

(A.7) also holds. For the conditional second moments,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h V_{(k+1)h})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\big(2h\psi^2 + h(\alpha+\psi)^2 + h(\beta-1)^2{}_hV^2_{kh} + h\gamma^2{}_hR^2_{kh} \\ + 2h(\alpha+\psi)(\beta-1){}_hV_{kh} + 2h(\alpha+\psi)\gamma\cdot {}_hR_{kh} + 2h\gamma(\beta-1){}_hV_{kh}\cdot {}_hR_{kh}\big) = 0. \tag{A.14}$$

and

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h R_{(k+1)h})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big(h\mathbb{E}[{}_hV^2_{(k+1)h}\epsilon^4_{(k+1)h}|\mathcal{F}_{kh}] + h\cdot {}_hR^2_{kh} \\ - 2h\mathbb{E}[{}_hV_{(k+1)h}\epsilon^2_{(k+1)h}|\mathcal{F}_{kh}]{}_hR_{kh}\Big) = 0. \tag{A.15}$$

Finally, for the cross moments,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big(- \mathbb{E}[\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}]{}_hR_{kh} \\ + \sqrt{h}\mathbb{E}[\sqrt{|{}_hV_{(k+1)h}|}{}_hV_{(k+1)h}\epsilon^3_{(k+1)h}|\mathcal{F}_{kh}]\Big) = 0, \tag{A.16}$$

and

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h R_{(k+1)h}\Delta_h V_{(k+1)h}|\mathcal{F}_{kh}] = 0. \tag{A.17}$$

Theorem 3.2 then follows immediately. □

B Diffusion limit of Breitung and Hafner (2016)

Breitung and Hafner's (2016) model (without leverage effect) is given by

$$r_t \equiv S_t - S_{t-1} = \sigma_t\xi_t, \tag{B.1}$$

$$\log\sigma^2_t = \alpha + \beta\log\sigma^2_{t-1} + \phi z_t, \tag{B.2}$$

where ξ_t are i.i.d.$(0,1)$ and $z_t = \log\xi^2_t - \mathbb{E}\log\xi^2_t$. Compared to RT-GARCH and SHARV, Breitung and Hafner's (2016) model does not give an analytical expression for the conditional variance of returns since $\exp(\phi\log\xi^2_t) \neq \xi^{2\phi}_t$ unless ϕ is an integer. Consequently, volatility forecasts are not available for their model.[4] Therefore, (B.1) - (B.2) is less appealing than RT-GARCH and SHARV from an empirical point of view. As Breitung and Hafner (2016) point out, the main purpose of (B.1) - (B.2) is for volatility nowcasting. Unlike RT-GARCH, their model does not nest E-GARCH as a special case. Define the rescaled joint process $({}_hS_{kh}, \log{}_h\sigma^2_{kh})$ as follows,

$${}_hS_{kh} = {}_hS_{(k-1)h} + \sqrt{h}{}_h\sigma_{kh}\xi_{kh}, \tag{B.3}$$

$$\log{}_h\sigma^2_{kh} = \log{}_h\sigma^2_{(k-1)h} + h\alpha + h(\beta-1)\log{}_h\sigma^2_{(k-1)h} + \sqrt{h}\phi z_{kh}, \tag{B.4}$$

where $\xi_{kh} \sim N(0,1)$. It is straightforward to check (B.3) - (B.4) become (B.1) - (B.2) by setting $h = 1$ and $k = t$. Moreover, $\log\xi^2_{kh}$ follow the log-chi square distribution defined in Pav (2015) with one degree of freedom. The first moment is given by $\log 2 + \varphi(1/2)$, where $\varphi(\cdot)$ is the digamma function, i.e., the derivative of the log gamma function. The cumulant generating function of $\log\xi$ is given by

$$K(n) = n\log 2 + \log\Gamma(1/2+n) - \log\Gamma(1/2), \tag{B.5}$$

[4] If we restrict $\phi = 1$, then $\mathbb{E}[\exp(\phi\log\xi^2_t)] = 1$. However, empirical estimates of Breitung and Hafner (2016) suggest ϕ is much smaller than 1.

where $\Gamma(\cdot)$ is the gamma function.

Define $C = \phi\mathbb{E}[\log \xi_{kh}^2]$. Using the fact that ${}_h r_{kh}$ is an MDS (Breitung and Hafner, 2016) and z_{kh} has zero mean,

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta_h S_{(k+1)h}|\mathcal{F}_{kh}] = 0, \tag{B.6}$$

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[\Delta \log {}_h\sigma_{(k+1)h}^2|\mathcal{F}_{kh}] = \alpha + (\beta - 1)\log\sigma^2, \tag{B.7}$$

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

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h S_{(k+1)h})^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big\{ &\exp\Big(\log {}_h\sigma_{kh}^2 + h\big(\alpha + (\beta-1)\log {}_h\sigma_{kh}^2\big) - \sqrt{h}C\Big) \\ &\cdot \mathbb{E}[\exp(\sqrt{h}\phi \log\xi_{(k+1)h}^2)\xi_{(k+1)h}^2|\mathcal{F}_{kh}]\Big\}.\end{aligned} \tag{B.8}$$

Expanding $\exp(\sqrt{h}\phi\log\xi_{(k+1)h}^2)$ inside the expectation on the RHS of (B.8) and using the fact $\mathbb{E}[|\log^n\xi^2|] < \infty$ for all integer $n > 0$,[5] we have by Cauchy–Schwarz inequality,

$$\mathbb{E}[|\xi^2\log^n\xi^2|] \leq \sqrt{\mathbb{E}\xi^4\mathbb{E}[\log^{2n}\xi^2]} < \infty.$$

Therefore, (B.8) becomes

$$\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta_h S_{(k+1)h})^2|\mathcal{F}_{kh}] = \sigma^2. \tag{B.9}$$

Similarly,

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta \log {}_h\sigma_{(k+1)h}^2)^2|\mathcal{F}_{kh}] = \lim_{h\downarrow 0}\Big\{ &h\big(\alpha^2 + (\beta-1)^2\log^2 {}_h\sigma_{kh}^2\big) + \phi^2\mathbb{E}z_{(k+1)h}^2 \\ &+ \sqrt{h}\big(2\alpha(\beta-1)\log {}_h\sigma_{kh}^2 + 2\alpha\phi\mathbb{E}z_{(k+1)h} \\ &+ 2(\beta-1)\phi\log {}_h\sigma_{kh}^2\mathbb{E}z_{(k+1)h}]\big)\Big\} = \phi^2\mathbb{E}z^2,\end{aligned} \tag{B.10}$$

where $\mathbb{E}z^2$ can be calculated from the cumulant generating function (B.5) and is finite.

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

$$\begin{aligned}\lim_{h\downarrow 0} h^{-1}\mathbb{E}[(\Delta(\log {}_h\sigma_{(k+1)h}^2)({}_h S_{(k+1)h})|\mathcal{F}_{kh}] &= \phi\mathbb{E}[{}_h\sigma_{(k+1)h}\xi_{(k+1)h}\log\xi_{(k+1)h}^2|\mathcal{F}_{kh}] \\ &+ \lim_{h\downarrow 0}\sqrt{h}\big(\alpha + (\beta-1)\log {}_h\sigma_{kh}^2\big)\mathbb{E}[{}_h\sigma_{(k+1)h}\xi_{(k+1)h}|\mathcal{F}_{kh}].\end{aligned} \tag{B.11}$$

Since ${}_hS_{(k-1)h} - {}_hS_{kh}$ is an MDS, the second term on the RHS of (B.11) is zero. For the first term, note ${}_h\sigma_{(k+1)h}\xi_{(k+1)h}\log\xi_{(k+1)h}^2$ is an odd function of a symmetric around zero random variable $\xi_{(k+1)h}$. Therefore, it is also an MDS and (B.11) equals zero.

It is easy to verify the fourth moments per unit time go to zero as $h \downarrow 0$. Therefore, if $({}_hS_0, \log {}_h\sigma_0^2) \Rightarrow (S_0, \log\sigma_0^2)$, then the joint process $({}_hS_{kh}, \log {}_h\sigma_{kh}^2)$ converges weakly to a diffusion process which satisfies

$$dS_t = \sigma_t dW_{1,t}, \tag{B.12}$$

[5]This is evident from the cumulant generating function (B.5) since the digamma function is infinitely differentiable and the m-th derivative is the polygamma function of order m.

$$d \log \sigma_t^2 = \big(\alpha + (\beta - 1)\log \sigma_t^2\big)dt + \phi\sqrt{\mathbb{E}z^2}dW_{2,t}, \tag{B.13}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent Brownian motions. The distributional uniqueness is satisfied upon noting (B.12) - (B.13) are equivalent to (3.18) - (3.19) of Nelson (1990), the diffusion limit of E-GARCH and $\phi\sqrt{\mathbb{E}z^2} < \infty$.

References

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

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

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.

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

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.

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.

Pav, S. E. (2015). Moments of the log non-central chi-square distribution. `https://arxiv.org/abs/1503.06266`.

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.

Stroock, D. W. and S. R. S. Varadhan (1979). *Multidimensional diffusion processes.* Springer Verlag, Berlin.

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