UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

Weak Diffusion Limit of Real-Time GARCH Models: The Role of Current Return Information

Yashuang (Dexter) Ding
University of Cambridge

Abstract

We prove that Real-time GARCH (RT-GARCH) models converge to the same type of stochastic differential equations as the standard GARCH models as the length of sampling interval goes to zero. The additional parameter of RT-GARCH can be interpreted as current information risk premium. We show RT-GARCH has the same limiting stationary distribution and shares the same asymptotic properties for volatility filtering and forecast as standard GARCH. Simulation results confirm the current information parameter decreases with the length of sampling interval and hence, GARCH and RT-GARCH models behave increasingly similar for high frequency data. Moreover, empirical results show the current information risk premium has increased significantly after the 2008 financial crisis for S&P 500 index returns.

Reference Details

20112	Cambridge Working Papers in Economics
2020/53	Cambridge-INET Working Paper Series
Published	25 November 2020
Key Words	GARCH, RT-GARCH, SV, diffusion limit, high frequency data
JEL Codes	C22, C32, C58
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

Weak Diffusion Limit of Real-Time GARCH Models: The Role of Current Return Information

Yashuang (Dexter) Ding*

University of Cambridge

Abstract

We prove that Real-time GARCH (RT-GARCH) models converge to the same type of stochastic differential equations as the standard GARCH models as the length of sampling interval goes to zero. The additional parameter of RT-GARCH can be interpreted as current information risk premium. We show RT-GARCH has the same limiting stationary distribution and shares the same asymptotic properties for volatility filtering and forecast as standard GARCH. Simulation results confirm the current information parameter decreases with the length of sampling interval and hence, GARCH and RT-GARCH models behave increasingly similar for high frequency data. Moreover, empirical results show the current information risk premium has increased significantly after the 2008 financial crisis for S&P 500 index returns.

Keywords: GARCH, RT-GARCH, SV, diffusion limit, high frequency data
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, Ekaterina Smetanina, 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.

Introduction

Volatility of financial asset returns has been an active research in economics and finance. There are generally two approaches to model ex-ante volatility: Most econometricians develop models that follow Engle's (1982) original idea of dynamic conditional variance (ARCH models). ARCH and its general form, GARCH type models regard volatility as past information only and share the same source of uncertainty as return process. These models are observation driven and are easy to implement using quasi-maximum likelihood (QML) with discretely sampled financial data. On the other hand, option pricing (e.g. Heston (1993)) and term structure of interest rates models (e.g. Longstaff and Schwartz (1992) and Fong and Vasicek (1991)) regard volatility as a latent variable driven by another innovation outside the return process. Itô's calculus provides many elegant analytical properties for these models. However, questions regarding how well these models fit financial data and the difficulty of estimation are the main drawbacks.

The main difference between these two approaches is whether volatility information is generated within the model itself. Using Stroock and Varadhan's (1979)'s diffusion approximation theorem, Nelson (1990) derives the weak convergence of GARCH type models to a system of stochastic differential equations (SDE). This theorem connects the two volatility modelling approaches. In the following papers, Nelson (1992) and Nelson and Foster (1995) provide a series of conditions under which a (possibly) misspecified GARCH models can provide consistent filter and forecast of volatility for high frequency data. Moreover, Nelson and Foster (1994) have developed the asymptotic filtering theory for univariate GARCH models. Subsequent weak convergence results have been derived for other extensions of GARCH type dynamic models (e.g. Fornari and Mele (1997), Ishida and Engle (2002) and Hafner et al. (2017)).

Empirical studies have remarked that by not using all available internal information, in particular the current return, ARCH type models make an inefficient use of information for volatility forecasting (e.g. Hansen et al. (2012) and Politis (1995)). To address this, Smetanina (2017) proposes the Real-time GARCH (RT-GARCH) model to incorporate current return information in the volatility process. Specifically, the return and volatility are jointly modelled as

$$r_t = \sigma_t \epsilon_t, \tag{0.1}$$

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2 + \psi\epsilon_t^2, \quad (\alpha, \beta, \gamma, \psi) \geq 0, \tag{0.2}$$

where r_t is the (demeaned) return series, ϵ_t are i.i.d. random variables with density $f_\epsilon(\cdot)$ with first and second moments equal to 0 and 1, respectively. The model uses the squared current return innovation to feedback the current level of volatility. In doing so, the volatility process is no longer deterministic conditional on the information up to $t-1$.

However, it still retains the QML structure with conditional transition density function,

$$f_r(r|\mathcal{F}_{t-1}) = \frac{r}{d(r, b_{t-1}, \vartheta)\sqrt{b_{t-1}^2 + 4\psi r_t^2}} f_\epsilon(d(r, b_{t-1}, \vartheta)), \tag{0.3}$$

where $f_\epsilon(\cdot)$ is the pdf of ϵ_t, $\vartheta = (\alpha, \beta, \gamma, \psi)'$ and

$$d(r, b_{t-1}; \vartheta) = \text{sign}\,(r)\sqrt{\frac{\sqrt{b_{t-1}^2 + 4\psi r^2} - b_{t-1}}{2\psi}}, \tag{0.4}$$

$$b_{t-1} = \alpha + \beta\sigma_{t-1}^2 + \gamma r_{t-1}^2. \tag{0.5}$$

In particular, Smetanina and Wu (2019) show that the QML procedure still results in consistent and asymptotically normally distributed estimators. RT-GARCH can also be interpreted as a special case for discrete time stochastic volatility (SV) model where the innovations of return and volatility are uncorrelated but from the same source.[1] RT-GARCH has additional advatages over GARCH in that it responds faster to new shocks, captures time-varying conditional kurtosis of returns and thus, provides better fit to financial returns (see Smetanina's (2017)).

To formally define where in-between RT-GARCH lies with regard to other volatility models, and in particular, how RT-GARCH behaves relative to standard GARCH when the sampling frequency changes, the diffusion limit is needed. In this paper, we use the techniques of Nelson's (1990) diffusion approximation theorem to derive the diffusion limit of RT-GARCH. In contrast to standard GARCH, the volatility process is not independent of the return innovations and thus, cannot be separated when computing the limiting moments. We use the asymptotic results of realised power variations of semimartinagle due to Barndorff-Nielsen and Shephard (2003) to address this issue. As we will see, RT-GARCH converges weakly to the same type of SDEs as the standard GARCH with the added parameter characterising volatility risk premium of current information. It turns out that with the length of sampling interval goes to zero, this risk premium decreases and the RT-GARCH becomes 'closer' to GARCH in terms of limiting distribution and values of persistence parameters and conditional intercepts.[2] However, when using lower frequency data, the risk premium increases and thus, the two models diverge increasingly. The intuition is that if there is no discontinuity in price path, the volatility is almost a constant within an infinitely small interval. It follows immediately that the additional parameter ψ of RT-GARCH controls the scale of its limiting stationary distribution. The consistency of GARCH estimator for filtering and forecasting volatility (Nelson (1992), Nelson and Foster (1995)) extends directly to the case of RT-GARCH. This implies that

[1] Smetanina's (2017) claims that the two innovations are correlated with $\rho = 1$. However, this is not the case since $\mathbb{E}[\epsilon_t \epsilon_t^2] = 0$ for symmetric returns.

[2] For RT-GARCH, the conditional intercept term is $\alpha + \psi$ and for GARCH is α. The reason we call them conditional intercept is because this term represents the one step conditional expectation given the lagged state variable is zero.

the advantage of RT-GARCH lies primarily on its efficient use of current information when volatility movement is noticeable or the discretisation of high frequency data is non-negligible. These results also extend to RT-GARCH and RT-GARCH with student-t innovations.

The remainder of the paper is structured as follows. In section 2, we briefly review the main convergence results of Stroock and Varadhan (1979) and Nelson (1990). In section 3, we derive the diffusion limit of RT-GARCH with both normal and student-t innovations, as well as asymmetric RT-GARCH with leverage and feedback effects similar to the GJR-GARCH and thus, extend Nelson's (1990) diffusion approximation theorem. In section 4, we provide simulation results with varying frequencies to demonstrate convergence results for RT-GARCH and GARCH. We also fit both RT-GARCH and GARCH to daily, weekly and monthly data of S&P 500 index. Section 5 concludes. All proofs and derivations are in Appendix A and tables are in Appendix B.

1 Weak convergence of Markov processes to diffusion

In this section we present results, drawn largely from Stroock and Varadhan (1979) and Nelson (1990), on the weak convergence of a sequence of Markov processes to a diffusion. Define a sequence of processes: $({}_hX_{kh})_{n\times 1}$ for integers k, which are random step functions taking jumps at times h, $2h$, $3h$, and so on. Let $D([0,\infty], R^n)$ be the space of functions from $[0,\infty]$ into R^n that are right continuous with finite left limits endowed with the Skorohod metric. Let $\mathcal{F}_{kh}$ denote the sigma algebra generated by ${}_hX_0$ up to ${}_hX_{kh}$ for each $h > 0$ and $B(R^n)$ the Borel sets on R^n. Let v_h be a probability measure on $(R^n, B(R^n))$ and $\Pi_h(x,\cdot)$ be a transition function on R^n, i.e.:

(a) $\Pi_h(x,\cdot)$ is a probability measure on $(R^n, B(R^n))$ for all $x \in R^n$.

(b) $\Pi_h(x,\cdot)$ is $B(R^n)$ measurable for all $\Gamma \in B(R^n)$.

Let P_h be the probability measure on $D([0,\infty), R^n)$ such that

$$\mathbb{P}_h({}_hX_0 \in \Gamma) = v_h(\Gamma) \text{ for any } \Gamma \in B(R^n), \tag{1.1}$$

$$\mathbb{P}_h({}_hX_t = {}_hX_{kh}, kh \le t < (k+1)h) = 1, \tag{1.2}$$

$$\mathbb{P}_h({}_hX_{(k+1)h} \in \Gamma | \mathcal{F}_{kh}) = \Pi_h({}_hX_{kh}, \Gamma) \text{ a.s. under } \mathbb{P}_h \text{ for all } k \ge 0 \text{ and } \Gamma \in B(R^n). \tag{1.3}$$

At each jump time, (2.1) specifies the distribution of starting points and (2.3) the transition densities of the n-dimensional discrete time step process ${}_hX_{kh}$. (2.2) characterises the continuous time process by making ${}_hX_t$ a step function with jumps at $h, 2h, 3h$ and so on.

Finally, the limit diffusion is formed by making $({}_hX_t) \Rightarrow (X_t)$ as $h \downarrow 0$, where X_t is

the solution (weak) to

$$X_t = X_0 + \int_0^t m(X_s, s)ds + \int_0^t \Sigma(X_s, s)dW_{n,s}, \tag{1.4}$$

where $W_{n,t}$ is an n-dimensional standard Brownian motion. The convergence is achieved by making assumptions on ${}_hX_t$ as follows.

Assumption 1. ${}_hX_0 \Rightarrow X_0$ *as* $h \downarrow 0$, *where* X_0 *has probability measure* v_0.

Next, to match the first and second moments of the increments of discrete time process to its continuous time counterpart, suppose $m_h(x)$ and $\Omega_h(x)$ are well defined for all $x \in R^n$:

$$m_h(x) \equiv h^{-1}\mathbb{E}[{}_hX_{(k+1)h} - {}_hX_{kh}|\mathcal{F}_{kh}], \tag{1.5}$$

$$\Sigma_h(x) \equiv h^{-1}\mathbb{E}[({}_hX_{(k+1)h} - {}_hX_{kh})({}_hX_{(k+1)h} - {}_hX_{kh})^T|\mathcal{F}_{kh}], \tag{1.6}$$

where the expectations are taken under P_h.[3]

Assumption 2. *For every* $\eta > 0$,

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq\eta} \|m_h(x) - m(x)\| = 0, \tag{1.7}$$

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq\eta} \|\Sigma_h(x) - \Sigma(x)\| = 0. \tag{1.8}$$

To ensure sample path continuity,

Assumption 3. *For every* $\eta > 0$ *and all* $i = 1, 2, ..., n$,

$$\lim_{h\downarrow 0} \sup_{\|x\|\leq\eta} \mathbb{E}[({}_hX_{i,(k+1)h}) - {}_hX_{i,kh})^4|\mathcal{F}_{kh}] = 0, \tag{1.9}$$

where ${}_hX_{i,kh}$ *are the* i^{th} *element of* ${}_hX_{kh}$.

Assumption 4. *There is a distributionally unique (weak) solution to* (1.4).

Theorem 1.1 (Stroock-Varadhan). *Under Assumptions 1 - 4,* $({}_hX_t) \Rightarrow (X_t)$ *as* $h \downarrow 0$, *where " $\Rightarrow$ " denotes weak convergence, i.e. convergence in distribution.*

To customise the theorem for GARCH models, let $X_t \equiv [S_t^T, \sigma_t^T]^T$ where S_t is $n \times 1$ observed state variables and σ_t is $m \times 1$ latend variables. Then a sequence of rescaled GARCH processes (discrete) ${}_hX_{kh} \equiv [{}_hS_{kh}^T, {}_h\sigma_{kh}^T]^T$ converges weakly to an SDE system whose weak solution is X_t (see Nelson (1990)).

[3]Change (1.6) to centred moments will not affect the results.

2 Main Results

2.1 Diffusion limit of RT-GARCH with Gaussian innovations

Consider a sequence of $({}_hS_{kh}, {}_h\sigma^2_{kh})$ that depends on the length of interval between subsequent observations:[4]

$$
\begin{aligned}
&{}_hr_{kh} \equiv {}_hS_{kh} - {}_hS_{(k-1)h} = {}_h\sigma_{kh} \cdot {}_h\epsilon_{kh}, && (2.1)\\
&{}_h\sigma^2_{kh} = \alpha_h + \beta_h \cdot {}_h\sigma^2_{(k-1)h} + h^{-1}\gamma_h \cdot {}_h\sigma^2_{(k-1)h} \cdot {}_h\epsilon^2_{(k-1)h} + h^{-1}\psi_h \cdot {}_h\epsilon^2_{kh}, && (2.2)\\
&\mathbb{P}[({}_hS_0, {}_h\sigma^2_0) \in \Gamma] = v_h(\Gamma) \text{ for any } \Gamma \in B(R^2), && (2.3)
\end{aligned}
$$

where ${}_h\epsilon_{kh}$ are i.i.d. with mean zero and variance proportional to h and $\{v_h\}_{h\downarrow 0}$ satisfies Assumption 1. We first state the convergence result for ${}_h\epsilon_{kh}$ normally distributed. Consider continuous time processes $\{S_t, \sigma^2_t\}$ that satisfy a GARCH-type SDE system:

$$
\begin{aligned}
&dS_t = \sigma_t dW_{1,t}, && (2.4)\\
&d\sigma^2_t = (\mu - \theta\sigma^2_t)dt + \gamma\sigma^2_t dW_{2,t}, && (2.5)\\
&\mathbb{P}[(S_0, \sigma^2_0) \in \Gamma] = v_0(\Gamma) \text{ for any } \Gamma \in B(R^2), && (2.6)
\end{aligned}
$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions, independent of the initial values (S_0, σ^2_0).

Assumption 5. *Let the rescaled RT-GARCH parameters $\alpha_h, \beta_h, \gamma_h$ and ψ_h satisfy*

$$
\begin{aligned}
&\lim_{h\downarrow 0} h^{-1}(\alpha_h + \psi_h) = \mu, && (2.7)\\
&\lim_{h\downarrow 0} h^{-1}(\gamma_h + \beta_h - 1) = -\theta, && (2.8)\\
&\lim_{h\downarrow 0} \sqrt{2}h^{-1/2}\gamma_h = \gamma, && (2.9)
\end{aligned}
$$

where μ, θ and γ are the coefficients of (2.4) *and* (2.5)*, provided the limits exist and are finite.*

Theorem 2.1. *Suppose the initial points of* (2.1) *and* (2.2)*, $({}_hS_0, {}_h\sigma^2_0) \Rightarrow (S_0, \sigma^2_0)$, the initial points of* (2.4) *and* (2.5)*. Under Assumption* 5 *the RT-GARCH processes $({}_hS_t, {}_h\sigma^2_t) \Rightarrow (S_t, \sigma^2_t)$, where ${}_hS_t \equiv {}_hS_{kh}$, ${}_h\sigma^2_t \equiv {}_h\sigma^2_{kh}$ for $kh \le t < (k+1)h$, i.e.* (2.1) - (2.3) $\Rightarrow$ (2.4) - (2.6) *as $h \downarrow 0$.*

Theorem 2.1 shows RT-GARCH converges weakly to the same SDE system as standard GARCH with the additional parameter ψ_h entering the volatility drift term via its limit in (2.7) (see Nelson (1990) for GARCH diffusion limit). This tells us that although RT-GARCH has nonstandard conditional density, it is still within the class of GARCH models The intuition is that when the interval between subsequent observation (h) goes to zero,

[4]Here we only consider equally spaced samples.

the volatility is almost constant if the assumption of sample path continuity holds. That is, using the information up to time T, for every $\zeta > 0$ and $T > 0$, there exists, with probability one, a random $\Delta(t) > 0$ such that

$$\sup_{T-\Delta(t)\leq s<T} |\sigma_s^2 - \sigma_T^2| < \zeta. \tag{2.10}$$

Thus, RT-GARCH and GARCH are asymptotically equivalent as $h \downarrow 0$ for they share the same diffusion limit.

The result gives ψ another interpretation, that is, the volatility risk premium of current information. According to Girsanov theorem, change the measure of (2.5) an equivalent martingale measure involves only changing the drift term provided the existence of such measure and this change is called risk premium in finance literature. Specifically, let

$$\lambda(\sigma_t^2) = \frac{\psi}{\gamma\sigma_t^2}, \tag{2.11}$$

then the process $W_t^{\mathcal{R}} = W_t^P - \int_0^t \lambda(\sigma_s^2)ds$ is a standard Brownian motion under an equivalent martingale measure $\mathcal{R}$ to the physical measure P of (2.5), with the Radon-Nikodym derivative

$$\frac{d\mathcal{R}}{dP} = \mathcal{E}(\lambda(\sigma_t^2)), \tag{2.12}$$

where

$$\mathcal{E}(x) = \exp\left(\int_0^t x dW_s^P - \tfrac{1}{2}\int_0^t x^2 ds\right) \tag{2.13}$$

is the Doléans-Dade exponential. Thus, if the measure $\mathcal{R}$ exists, the procee σ_t^2 can be expressed as

$$d\sigma_t^2 = (\mu - \psi - \theta\sigma_t^2)dt + \gamma\sigma_t^2 dW_t^{\mathcal{R}}. \tag{2.14}$$

The additional parameter ψ in RT-GARCH can be interpreted as volatility risk premium of current information per unit standard deviation of volatility and the equivalent measure $\mathcal{R}$ can be thought of as the measure under which agents do not require compensation for not knowing the current return information and thus, this risk premium can be subtracted from the volatility drift term. The nonnegative restriction on ψ corresponds to the fact investors dislike uncertainty in volatility and are willing to exchange compensation for current return information. This form of risk premium is not affine in σ_t^2. However, it falls within the specification of extended affine models defined in Cheridito et al. (2007). Specifically, the total current information risk premium, $\sigma_t^2\lambda(\sigma_t^2)$, is a restricted case of the extended affine model, $\sigma_t^2\lambda(\sigma_t^2) = \lambda_0 + \lambda_1\sigma_t^2$, where $\lambda_0 = \psi/\gamma$ and $\lambda_2 = 0$. Cheridito et al. (2007) prove the existence of such equivalent martingale measure under the boundary nonattainment conditions. Specifically, since the process σ_t^2 satisfies the non-explosion condition defined in Nelson (1990), for a positive initial point, the existence of equivalent measure for RT-GARCH diffusion limit is guaranteed. The information risk premium parameter ψ can only be when fitting the data with RT-GARCH. In this

sense, RT-GARCH's efficient use of information is similar to the way of using options data to incorporate market risk neutral expectation on asset returns. This is, however, not to say the measure $\mathcal{R}$ is the risk neutral measure since there is no guarantee that the instantaneous drift term of risky asset equals to risk free rate under $\mathcal{R}$.[5]

In order to answer Smetanina's (2017) question about how the RT-GARCH relates to SV models, we next consider the discrete time (Euler Maruyama scheme) stochastic volatility analogue of (2.4) and (2.5):

$$ {}_h r_{kh} = {}_h\sigma_{kh} \cdot {}_h\epsilon_{kh}, \tag{2.15} $$

$$ {}_h\sigma^2_{(k+1)h} = {}_h\sigma^2_{kh} + (\mu - \theta \cdot {}_h\sigma^2_{kh})h + \gamma \cdot {}_h\sigma^2_{kh} \cdot {}_h z_{(k+1)h}, \tag{2.16} $$

for a sequence of h converge to 0, where ${}_h\epsilon_{kh}$ and ${}_h z_{kh}$ are both i.i.d. with zero mean and variance h and independent with each other. Thus, RT-GARCH is (asymptotically) equivalent to the discrete time SV model regardless of the correlation between innovations (and they are both asymptotically equivalent to standard GARCH model).

2.2 RT-GJR-GARCH with Gaussian innovations

For the correlated Brownian motions case, we need to add information asymmetry in the discrete time model similar to GJR-GARCH. Consider the rescaled sequence of RT-GJR-GARCH processes with leverage and feedback effects:

$$ {}_h r_{kh} \equiv {}_h S_{kh} - {}_h S_{(k-1)h} = {}_h\sigma_{kh} \cdot {}_h\epsilon_{kh}, \tag{2.17} $$

$$ \begin{aligned} {}_h\sigma^2_{kh} &= \alpha_h + \beta_h \cdot {}_h\sigma^2_{(k-1)h} + h^{-1}\gamma_h \cdot {}_h\sigma^2_{(k-1)h} \cdot {}_h\epsilon^2_{(k-1)h} \\ &\quad + h^{-1}\psi_h \cdot {}_h\epsilon^2_{kh} + h^{-1}\phi_h \cdot {}_h\sigma^2_{(k-1)h} \cdot ({}_h\epsilon^-_{(k-1)h})^2 + h^{-1}\eta_h \cdot ({}_h\epsilon^-_{kh})^2, \end{aligned} \tag{2.18} $$

$$ \mathbb{P}[({}_h S_0, {}_h\sigma^2_0) \in \Gamma] = v_h(\Gamma) \text{ for any } \Gamma \in B(R^2), \tag{2.19} $$

where ${}_h\epsilon^-_{kh}$ are i.i.d. half normally distributed, i.e., ${}_h\epsilon^-_{kh} \equiv {}_h\epsilon_{kh}\mathbb{1}_{\{{}_h\epsilon_{kh} \le 0\}}$ and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Next consider (S_t, σ^2_t) that satisfy

$$ dS_t = \sigma_t(\rho dW_{1,t} + \sqrt{1-\rho^2}dW_{2,t}), \tag{2.20} $$

$$ d\sigma^2_t = (\mu - \theta\sigma^2_t)dt + \gamma\sigma^2_t dW_{1,t}, \tag{2.21} $$

$$ \mathbb{P}[(S_0, \sigma^2_0) \in \Gamma] = v_0(\Gamma) \text{ for any } \Gamma \in B(R^2), \tag{2.22} $$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions, independent of the initial values (S_0, σ^2_0).

Assumption 6. *Let the rescaled RT-GARCH parameters $\alpha_h, \beta_h, \gamma_h, \psi_h, \phi_h$ and η_h satisfy*

$$ \lim_{h \downarrow 0} h^{-1}(\alpha_h + \psi_h + \tfrac{1}{2}\eta_h) = \mu, \tag{2.23} $$

[5]Note also that the market risk premium of volatility under risk neutral measure is usually negative.

$$\lim_{h\downarrow 0} h^{-1}(\gamma_h + \beta_h + \tfrac{1}{2}\phi_h - 1) = -\theta, \tag{2.24}$$

$$\lim_{h\downarrow 0} h^{-1}(2\gamma_h^2 + 2\phi_h\gamma_h + \tfrac{5}{4}\phi_h^2) = \gamma^2, \tag{2.25}$$

$$\lim_{h\downarrow 0} h^{-1/2}\sqrt{\tfrac{2}{\pi}}\phi_h = -\omega, \tag{2.26}$$

where $\omega \equiv \rho\gamma$ *and* $(\mu, \theta, \gamma, \rho)$ *are the coefficients of* (2.20) *and* (2.21)

Theorem 2.2. *Suppose the initial points of* (2.17) *and* (2.18), $({}_hS_0, {}_h\sigma_0^2)$ *converge to the initial points of* (2.20) *and* (2.21), (S_0, σ_0^2) *in distribution. Under Assumption* 6 *the RT-GJR-GARCH processes* (2.17) - (2.19) *converge to* (2.20) - (2.22) *in distribution as* $h \downarrow 0$.

Theorem 2.2 shows RT-GJR-GARCH further breaks down current information risk premium into those due to negative and positive return information. Since RT-GJR-GARCH is a generalisation of RT-GARCH, we will call both models the class of RT-GARCH models.

2.3 Stationary distribution of RT-GARCH models

A direct consequence of Theorems 2.1 and 2.2 is the limiting stationary distribution of RT-GARCH. Nelson (1990) has proved that in steady state, (2.21) has an inverse gamma distribution. Since RT-GARCH and GARCH converge to the same type of SDEs as $h \downarrow 0$, we have

Theorem 2.3. *Let* σ_t^2 *be generated by* (2.21). *If*

(a) the distribution of σ_0^2 *converges to the stationary distribution of* σ_t^2 *as* $h \downarrow 0$,

(b) the sequence of parameters $(\alpha_h, \beta_h, \gamma_h, \psi_h, \phi_h, \eta_h)_{h\downarrow 0}$ *satisfies Assumption 6,*

then in the discrete time system (2.17) *and* (2.18),

$${}_h\sigma_{kh}^2 \xrightarrow{d} \textit{Inverse-Gamma}(1 + 2\theta/\gamma^2, 2\mu/\gamma^2), \tag{2.27}$$

$$h^{-1/2}\sqrt{(2\theta + \gamma^2)/2\mu} \cdot {}_hr_{kh} \xrightarrow{d}$$
$$\textit{Skew-t}\left(0, \frac{\rho^2}{1 - \frac{2}{\pi}} + 1 - \rho^2, \frac{\rho}{\sqrt{(1 - \frac{2}{\pi})(1 - \rho^2)}}, 2 + \frac{4\theta}{\gamma^2}\right), \tag{2.28}$$

for any constant value of kh *as* $h \downarrow 0$, *where Skew-t*(a, b, c, n) *is the skewed Student-t distribution as defined in Azzalini and Capitanio (2003) with location parameter* a, *scale parameter* b, *shape parameter* c *and degree of freedom* n.

If there exists a $d > 0$ *such that* $\limsup \mathbb{E}[{}_h\sigma_0^{2d}] < \infty$ *and condition (b) is satisfied, then*

$${}_h\sigma_{kh}^2 \xrightarrow{d} \textit{Inverse-Gamma}(1 + 2\theta/\gamma^2, 2\mu/\gamma^2), \tag{2.29}$$

$$h^{-1/2}\sqrt{(2\theta+\gamma^2)/2\mu}\cdot {}_h r_{kh} \xrightarrow{d}$$
$$\textit{Skew-t}\left(0, \frac{\rho^2}{1-\frac{2}{\pi}}+1-\rho^2, \frac{\rho}{\sqrt{(1-\frac{2}{\pi})(1-\rho^2)}}, 2+\frac{4\theta}{\gamma^2}\right), \tag{2.30}$$

as $h \downarrow 0$ *and* $kh \to \infty$.

Remark 2.3.1. *All the parameters in RT-GARCH and its asymmetric form are restricted to be non-negative and satisfy weak stationarity conditions, therefore the diffusion limit parameters are also non-negative and within the support of inverse gamma distribution.*

Remark 2.3.2. *To our knowledge, the stationary distribution of return process* ${}_h r_{kh}$ *for GJR-GARCH has not been derived yet. Thus, Theorem* 2.3 *is novel for both GJR-GARCH and RT-GJR-GARCH since they share the same type of diffusion limit. For the symmetric RT-GARCH, that is* $\rho = 0$*, the distribution in* (2.28) *and* (2.30) *reduces to symmetric Student-t distribution with* $2 + 4\theta/\gamma^2$ *degrees of freedom as in Nelson's (1990) Theorem 2.3.*

Theorem 2.3 tells us that althougth RT-GARCH and GARCH have different conditional distributions in discrete time (Smetanina, 2017), they share the same stationary distribution as $h \downarrow 0$. The additional parameters ψ_h and η_h associated with current information enter the limiting distribution through its scale parameter $2\mu/\gamma^2$. That is, ψ_h and η_h contribute to how spread out the volatility is in steady state: The higher the current information risk premium, the more volatile the volatility stationary distribution since volatility will respond more rapidly to each new information. Moreover, the skew Student-t stationary distribution of return process implies heavy tails. Hansen (1994) first proposes the skew Student-t distribution to model heavy tails and asymmetry in conditional return distribution. RT-GARCH models therefore, can be seen as an alternative to Hansen's (1994) approach since RT-GJR-GARCH produces heavy tails and asymmetry in both conditional and conditional distributions while retaining a relatively simple expression for the conditional density function. The requirement of $\theta > 0$ is also indicative since the second moment of skew Student-t random variable exists if and only if the degree of freedom is larger than two.

2.4 RT-GARCH models with Student-t innovations

We next turn to RT-GARCH models with Student-t innovations due to the popularity of using Student-t innovations for standard GARCH models. Consider again the sequence of processes generated by (2.17) and (2.18) with ${}_h\epsilon_{kh}$ i.i.d. Student-t distributed with degree of freedom $\nu > 0$. We require a more restricted assumption on this innovation term.

Assumption 7. *The sequence of* $({}_h\epsilon_{kh})$ *are i.i.d. rescaled Student-t distributed with degree of freedom* $\nu > 8$ *and variance proportional to* h*, i.e.,* $var({}_h\epsilon_{kh}) = h\nu/(\nu-2)$.[6]

The requirement of more than 8 degrees of freedom is to ensure the limit of the first four moments exist (see Appendix A).

Assumption 8. *Let the parameters of rescaled RT-GARCH* (2.17) *and* (2.18) $\alpha_h, \beta_h, \gamma_h, \psi_h, \phi_h$ *and* η_h *satisfy*

$$\lim_{h\downarrow 0} h^{-1}[\alpha_h + \frac{\nu}{\nu-2}(\psi_h + \tfrac{1}{2}\eta_h)] = \mu, \tag{2.31}$$

$$\lim_{h\downarrow 0} h^{-1}[\beta_h + \frac{\nu}{\nu-2}(\gamma_h + \tfrac{1}{2}\phi_h) - 1] = -\theta, \tag{2.32}$$

$$\lim_{h\downarrow 0} h^{-1}\frac{\nu^2}{(\nu-4)(\nu-2)^2}[(2\nu-2)(\gamma_h^2 + \phi_h\gamma_h) + (\tfrac{5}{4}\nu - 2)\phi_h^2] = \gamma^2, \tag{2.33}$$

$$\lim_{h\downarrow 0} h^{-1/2}\frac{2\nu^{3/2}}{\sqrt{\pi}(\nu-3)(\nu-1)}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\gamma_h = -\omega, \tag{2.34}$$

where $(\omega \equiv \rho\gamma, \mu, \theta, \gamma, \rho)$ *are the coefficients of* (2.20) *and* (2.21) *and* $\Gamma(\cdot)$ *is the gamma function.*

Theorem 2.4. *Suppose the initial points of* (2.17) *and* (2.18), $({}_hS_0, {}_h\sigma_0^2)$ *converge to the initial points of* (2.20) *and* (2.21), (S_0, σ_0^2) *in distribution. Under Assumptions 7 and 8 the RT-GJR-GARCH processes* (2.17) - (2.19) *with rescaled Student-t innovations converge to* (2.20) - (2.22) *in distribution as* $h \downarrow 0$.

Theorem 2.4 is not surprising as the sum of any i.i.d. random variables with finite second moment can be approximated by a Brownian motion in increasingly finer partitions of a fixed interval by Donsker's theorem. The degree of freedom of Student-t innovations appears in both drift and diffusion terms. Thus, even in high frequency data, diffusion limit can still take into account heavy-tails of discretely sampled data.

2.5 Diffusion approximation with RT-GARCH

In light of Theorems 2.1, 2.2 and 2.4 we generalize Nelson's (1990) diffusion approximation theorem to incorporate current return information. First, define the SDE system:

$$dS_t = f(S_t, Y_t, t)dt + g(S_t, Y_t, t)dW_{1,t}, \tag{2.35}$$

$$dY_t = F(S_t, Y_t, t)dt + G(S_t, Y_t, t)dW_{2,t}, \tag{2.36}$$

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t}^T \end{bmatrix} = \begin{bmatrix} 1 & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix} dt \equiv \Omega dt, \tag{2.37}$$

where W_1 is a one-dimensional standard Brownian motion, W_2 is an n-dimensional Brownian motion and Ω is an $(n+1)\times(n+1)$ positive semi-definite matrix of rank two or less.

[6] See Lemma $A.3$ for the pdf of rescaled Student-t distribution.

$f(\cdot,\cdot,\cdot)$ and $g(\cdot,\cdot,\cdot)$ are real-valued, continuous scalar functions and $F(\cdot,\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot)$ are real-valued, continuous $n \times 1$ and $n \times n$ functions respectively. The initial points (S_0, Y_0) are random variables with joint probability measure v_0 and are independent of the Brownian motions. Let

$$b(s,y,t) \equiv \begin{bmatrix} f(s,y,t) \\ F(s,y,t) \end{bmatrix} \tag{2.38}$$

$$a(s,y,t) \equiv \begin{bmatrix} g(s,y,t)^2 & g(s,y,t)\Omega_{1,2}G(s,y,t)^T \\ G(s,y,t)\Omega_{2,1}g(s,y,t) & G(s,y,t)\Omega_{2,2}G(s,y,t)^T \end{bmatrix} \tag{2.39}$$

be an $(n+1) \times 1$ vector and an $(n+1) \times (n+1)$ matrix functions respectively.

Next define a sequence of step functions to approximate (2.35) – (2.37):

$$_hS_{kh} = {}_hS_{(k-1)h} + f(s,y,t)h + g(s,y,t)_hZ_{kh}, \tag{2.40}$$

$$_hY_{kh} = {}_hY_{(k-1)h} + F_1(s,y,t)h + G(s,y,t)_hZ^*_{(k-1)h} + F_2(t)_hZ^{**}_{kh}, \tag{2.41}$$

where $_hZ_{kh}$ are i.i.d. with $\mathbb{E}[_hZ_{kh}] = \mathbb{E}[_hZ^3_{kh}] = 0$, $\mathbb{E}[|_hZ_{kh}|^{2\delta}] < \mathcal{O}(h^\delta)$, for $0 < \delta \leq 4$,[7] and $_hZ^*_{kh}$ and $_hZ^{**}_{kh}$ are defined as:

$$_hZ^*_{kh} \equiv \begin{bmatrix} \theta_1 \cdot {}_hZ^2_{kh} + \phi_1(_hZ^-_{kh})^2 + \gamma_1 h \\ \dots \\ \theta_n \cdot {}_hZ^2_{kh} + \phi_n(_hZ^-_{kh})^2 + \gamma_n h \end{bmatrix}, \tag{2.42}$$

and

$$_hZ^{**}_{kh} \equiv [_hZ^2_{kh} + (_hZ^-_{kh})^2], \tag{2.43}$$

such that

$$\mathbb{E}\begin{bmatrix} _hZ_{kh} \\ _hZ^*_{kh} \end{bmatrix} [_hZ_{kh} \;\; _hZ^{*T}_{kh}] = \Omega h, \tag{2.44}$$

and finally,

$$F = F_1 + h^{-1}F_2\mathbb{E}[_hZ^*_{kh}]. \tag{2.45}$$

Theorem 2.5. *Let $_hS_t \equiv {}_hS_{kh}$ and $_hY_t \equiv {}_hY_{kh}$ for $kh \leq t < (k+1)h$. If $b(s,y,t)$ and $a(s,y,t)$ satisfy Assumption 4 and the starting points of (2.40) and (2.41), $(_hS_0, {}_hY_0^{'})$, converge to the measure v_0 as $h \downarrow 0$, then $(_hS_t, {}_hY_t^{'}) \Rightarrow (S_t, Y_t^{'})$ as $h \downarrow 0$.*

Remark 2.5.1. *$F_2(\cdot)$ can only take deterministic values or random variables independent of $_hZ^*_{kh}$ as argument. This is to ensure the convergence of observed power variations.*

Theorem 2.5 summarises the class of GARCH models (including RT-GARCH models) as diffusion approximation. It is clear this convergence is not unique since the drift

[7] We require the first and third moments equal to zero to ensure the return process is a martingale difference sequence, the even moments up to 8th power proportional to $\sqrt{h}$ to the respective powers to ensure the power variations converge to their expectations in $\mathcal{L}^2$.

term can be separated into a constant and an innovation term scaled by a deterministic function. As Nelson (1990) points out, the GARCH approximation requires only one innovation term in contrast to the Euler discretisation (two innovation terms). Moreover, we do not require global Lipschitz continuity in the diffusion parameter functions.

3 Consistent filtering and forecasting with misspecified RT-GARCH models

Another implication of RT-GARCH and GARCH sharing the same diffusion limit is that the consistency results of GARCH models for filtering and forecasting volatility can be extended to the RT-GARCH model. Specifically, we define the consistent filter of volatility as in Nelson (1992), i.e. for a sequence of processes $({}_hZ_{kh})$ which is the difference between the discretised volatility ${}_h\Sigma_t^2$ in (1.4) and the filtered volatility by RT-GARCH model $\mathbb{E}[{}_hr_{kh}^2|\mathcal{F}_{(k-1)h}]$ in (2.2). Consistent filtering requires as $h \downarrow 0$, $\|{}_hZ_t\| \to 0$ in probability for every $kh \leq t < (k+1)h$.

Formally, consider again the sequence of RT-GARCH processes (2.1) and (2.2). For simplicity we consider only the univariate case. The data-generating process is then

$$S_t = S_0 + \int_0^t m(S_u)dt + \int_0^t \Sigma(S_u)dW_u, \tag{3.1}$$

where $\Sigma(\cdot)$ is a real-valued continuous function and W_t is a one dimensional standard Brownian motion. Define

$${}_hZ_{kh} \equiv \Sigma({}_hS_{kh})^2 - \mathbb{E}[{}_hr_{kh}^2|\mathcal{F}_{(k-1)h}], \tag{3.2}$$

and $\mathbb{P}_h({}_hZ_t = {}_hZ_{kh}) = 1$ for all $kh \leq t < (k+1)h$.

Assumption 9. *for each $h > 0$, (3.1) generates $({}_hS_t)$ and satisfies Assumptions 1, 3 and 4 of section 2.*

Assumption 10. *for some $\delta > 0$, $\limsup_{h\downarrow 0} \mathbb{E}|{}_hZ_0|^{2+\delta} < \infty$.*

Assumption 11. *for every $\eta > 0$, there is an $\epsilon > 0$ such that*

$$\lim_{h\downarrow 0} \sup_{|x|<\eta} h^{-1}\mathbb{E}\big[|\Sigma^2({}_hS_{(k+1)h}) - \Sigma^2({}_hS_{kh})|^{2+\epsilon}\big|\mathcal{F}_{kh}\big] = 0, \tag{3.3}$$

$$\lim_{h\downarrow 0} \sup_{|x|<\eta} h^{-1}\mathbb{E}\big[|{}_hS_{(k+1)h} - {}_hS_{kh}|^{4+\epsilon}\big|\mathcal{F}_{kh}\big] = 0. \tag{3.4}$$

Theorem 3.1. *For some δ, $0 < \delta < 1$, let $\alpha_h, \beta_h, \gamma_h$ and ψ_h satisfy*

$$\alpha_h = o(h^\delta), \tag{3.5}$$

$$\psi_h = o(h^\delta), \tag{3.6}$$

$$1 - \beta_h - \gamma_h = o(h^\delta), \tag{3.7}$$

$$\gamma_h = h^\delta \gamma + o(h^\delta) \tag{3.8}$$

where $\alpha_h, \beta_h, \gamma_h$ and ψ_h are parameters in (2.1) *and* (2.2) *and γ is independent of h. If Assumptions 9 - 11 and Condition 3.3 in Nelson (1992) hold, then for each $t > 0$, $\|{}_hZ_t\| \to 0$ in probability as $h \downarrow 0$.*[8]

Remark 3.1.1. *If the data-generating process* (3.1) *is the diffusion limit of RT-GARCH, then all the Assumptions needed in the theorem are automatically satisfied. Nelson's (1992) theorem and Theorem* 3.1 *assumes more general diffusion process.*

If, In addition, we assume (2.1) and (2.2) correctly specify the functional form of the first two conditional moments of ${}_hS_t$ and ${}_h\Sigma_t^2$, then the forecast distribution generated by (2.1) and (2.2) also consistently estimates the forecast distribution generated by the true data generating process.[9] Formally,

Assumption 12. *For all $(s, y) \in \mathbb{R}^{n+m}$, $\hat{m}(s, y) = m(s, y)$ and $\hat{\Sigma}^2(s, y) = \Sigma^2(s, y)$, where $\hat{m}(\cdot, \cdot)$ and $\hat{\Sigma}^2(\cdot, \cdot)$ are first and second conditional moments of the diffusion limit of* (2.1) *and* (2.2)*.*

Theorem 3.2. *If Assumptions 9 - 12 and* (3.5) - (3.8) *are satisfied. Then:*
(a) For every $0 < \tau < \infty$, $({}_hS_t, {}_h\sigma_t^2)_{[\tau,\infty)}$ consistently estimates the forecast distribution of $(S_t, \Sigma_t^2)_{[\tau,\infty)}$.
(b) Let $G(s_1, s_2, y_1, y_2)$ be a continuous function from $\mathbb{R}^4$ into $\mathbb{R}^1$ satisfying

$$|g(s_1, s_2, y_1, y_2)| < A + B|s_1|^a|s_2|^b|y_1|^c|y_2|^d, \tag{3.9}$$

for finite, nonnegative A, B, a, b, c and d. Then the forecast moment function of RT-GARCH

$$G_h(g, s, y, \tau) = \mathbb{E}[g(\{{}_hS_t, {}_h\sigma_t^2\}_{[\tau,\infty)})|{}_h\mathcal{F}_t] \tag{3.10}$$

consistently estimate the moment forecast function generated by (3.1)*.*
(c) The stationary distributions of ${}_h\sigma_t^2$ and Σ_t^2 for each sufficiently small $h > 0$ exist, i.e., ${}_h\sigma_t^2 \Rightarrow {}_h\sigma_\infty^2$ and $\Sigma_t^2 \Rightarrow \Sigma_\infty^2$ as $t \to \infty$. Furthermore, ${}_h\sigma_\infty^2 \Rightarrow \Sigma_t^2$ as $h \downarrow 0$.

Together Theorems 3.1 and 3.2 show for high frequency data, RT-GARCH perform volatility filtering and forecasting as good as standard GARCH models in the sense that they both achieve consistency under the same regularity conditions. This can be seen from the one-step volatility forecast of RT-GARCH,

$$\mathbb{E}[{}_hr_{(k+1)h}^2|\mathcal{F}_{kh}] = h \cdot \mathbb{E}[{}_h\sigma_{(k+1)h}^2|\mathcal{F}_{kh}] + \psi_h(h^{-1}\mathbb{E}[{}_h\epsilon_{kh}^4] - h). \tag{3.11}$$

[8] In equation (3.12) of Nelson (1992), $\mu(x) = 0$ and $\Omega(x)$ is the $\Sigma(x)^2$ here. Adding non zero drift term to diffusion (3.1) will not change the result.

[9] See Nelson and Foster (1995) for detailed definition of consistent estimation of forecast distribution

Table 1: Parameter estimations of RT-GARCH and GARCH from diffusion

	RT-GARCH				GARCH		
	α_h	β_h	γ_h	ψ_h	α_h^g	β_h^g	γ_h^g
h = 0.004	0.0006	0.9696	0.0262	0.0029	0.0034	0.9686	0.0275
	(0.0005)	(0.0017)	(0.0018)	(0.0005)	(0.0002)	(0.0007)	(0.0006)
h = 0.02	0.0033	0.9396	0.0371	0.0152	0.0174	0.9285	0.0518
	(0.0028)	(0.0073)	(0.0069)	(0.0028)	(0.0015)	(0.0036)	(0.0027)
h = 0.1	0.0517	0.8436	0.0496	0.0322	0.0843	0.8249	0.0800
	(0.0129)	(0.0228)	(0.0113)	(0.0051)	(0.0132)	(0.0204)	(0.0094)

Note: Parameters are estimated using data generated by diffusion limit. The standard errors (in parentheses) are standard deviations across 100 sample paths.

If the persistence parameters of RT-GARCH and GARCH models are identical and $\alpha_h+\psi_h$ is identical to the GARCH constant term, then the one-step volatility forecasts of both models will also be identical as long as both models start at the same initial points since the second term of the right-hand side of (3.11) goes to zero as $h \downarrow 0$. The same applies to multi-step forecast since for both models

$$\mathbb{E}[{}_h r^2_{(k+j)h}|\mathcal{F}_{kh}] = h^{-1}\mathbb{E}[r^2] + h^{-1}(\beta_h + \gamma_h)\big(\mathbb{E}[{}_h r^2_{(k+j-1)h}|\mathcal{F}_{kh}] - \mathbb{E}[r^2]\big), \tag{3.12}$$

for all $j > 1$ and $\mathbb{E}[r^2]$ is the unconditional variance of returns.[10] In other words, RT-GARCH offers no advantage for volatility forecast over GARCH model in continuous time with negligible discretisation errors. The same arguments apply to the forecast distributions of both models.

This argument however, does not apply to asymptotic efficiency of filtering and forecasting especially when discretisation errors are not negligible. From this point of view, RT-GARCH can be regarded as superior in its more efficient use of information and better goodness of fit for conditional kurtosis of discretely observed data as noted in Smetanina (2017). It will be of particular interest to develop asymptotic distribution of volatility measurement errors under RT-GARCH similar to that of standard GARCH in Nelson and Foster (1994) and compare the asymptotic variances of both models. This is left for future researchers.

4 Simulations and empirical studies

4.1 Simulations

In this section we generate 100 sample paths from (2.4) and (2.5) using Euler's scheme for 1000 periods with discretisation interval $\Delta t = 1/500$. The parameters are set as: $\mu = 0.8, \theta = 0.9$ and $\gamma = 0.7$, which are typical for stock returns and sampled with three

[10] The unconditional variances for both models are also identical if the conditions for identical one-step volatility forecasts are satisfied.

Table 2: Diffusion parameters inferred by RT-GARCH and GARCH parameters

	RT-GARCH			GARCH		
	h=0.004	h=0.02	h=0.1	h=0.004	h=0.02	h=0.1
$\mu = 0.8$	0.8738	0.9251	0.8392	0.8557	0.8678	0.8430
	(0.0161)	(0.0301)	(0.0173)	(0.0058)	(0.0098)	(0.0190)
$\theta = 0.9$	1.0379	1.1668	1.0684	0.9702	0.9830	0.9510
	(0.0404)	(0.0974)	(0.0555)	(0.0097)	(0.0143)	(0.0262)
$\gamma = 0.7$	0.5862	0.3710	0.2217	0.6144	0.5183	0.3576
	(0.0145)	(0.1129)	(0.2313)	(0.0075)	(0.0337)	(0.1189)

Note: Diffusion parameters are obtained by plugging the GARCH and RT-GARCH parameters into (2.7) - (2.9). The mean squared errors are reported in parentheses.

different frequencies, $h = 1/10$, $1/50$ and $1/250$ which roughly correspond to monthly, weekly and daily frequencies, respectively in real world situation. We then fit GARCH and RT-GARCH with Gaussian innovations to each sample path.

Table 1 reports the estimated RT-GARCH and GARCH parameters. The current information parameter ψ_h increases with the length of sampling interval. The sum of α_h and ψ_h for RT-GARCH is almost identical to the α_h^g for GARCH at 1/250 frequency. So are the persistence parameters $(\beta_h + \gamma_h)$ of both models. The differences start to increase when the length of sampling interval increases. This confirms that when the length of sampling interval becomes increasingly finer, RT-GARCH and GARCH are asymptotically equivalent and only differ with the current information risk premium parameter separated from the volatility drift term. Note the current information parameter contains the error due to discretely sampling and thus, is still significant even for small h.

Table 2 reports the diffusion parameters inferred by both RT-GARCH and GARCH. The mean squared errors (MSE) for both models increase with the length of sampling interval. The MSE for RT-GARCH implied parameters are nearly twice as large as those implied by GARCH. This is not surprising since the current information is only asymptotically constant when $h \downarrow 0$. When h increases, this term becomes more stochastic and RT-GARCH diverges faster from its diffusion limit than GARCH due to this additional source of disturbances. This suggests RT-GARCH suffers larger discretisation bias than GARCH. Thus, if we were to use weak convergence results to estimate the parameters of an SV model, GARCH is preferred given its smaller MSE.[11] This, however, is not to say GARCH is superior in volatility modelling than RT-GARCH when the data generating process is its diffusion limit. Both models are essentially misspecified in this case and each model has its own advantage. Specifically, GARCH produces more accurate diffusion parameter estimates and RT-GARCH fits the tail distribution of the discretely observed data better in terms of conditional kurtosis.

[11] Note that the weak convergence does not imply consistent estimators of diffusion parameters (Wang, 2002)

Table 3: Parameter estimations of RT-GARCH and GARCH for S&P 500 index

	RT-GARCH				GARCH		
	α	β	γ	ψ	α^g	β^g	γ^g
Daily	$1.3046 * 10^{-4}$	0.8683	0.0988	0.0269	0.0199	0.8745	0.1095
	$(2.7135*10^{-4})$	(0.0103)	(0.0100)	(0.0032)	(0.0019)	(0.0073)	(0.0067)
Weekly	0.0245	0.8108	0.1138	0.2633	0.2608	0.7610	0.2027
	(0.0862)	(0.0443)	(0.0369)	(0.0646)	(0.0586)	(0.0234)	(0.0201)
Monthly	0.8092	0.7496	0.1592	0.6489	0.9394	0.7436	0.2123
	(0.8883)	(0.1034)	(0.0840)	(0.3760)	(0.7511)	(0.0844)	(0.0626)

Note: The sample size is 5030, 1043 and 239 for daily, weekly and monthly frequencies, respectively. The standard errors, calculated numerically, are given in parentheses.

4.2 Application

We now use the returns of S&P 500 index to examine the differences between RT-GARCH and GARCH models under different frequencies. The data spans from 04 January 2000 till 31 December 2019 and are sampled from daily, weekly and monthly frequencies. The returns are calculated using adjusted closing price at the end of each sampling interval. For the daily data, the persistence parameters of RT-GARCH and GARCH parameters are almost identical. The differences start to increase from daily to weekly and monthly data. This is consistent with our simulation results.

Similarly, we find the sum of the constant and current information parameters of RT-GARCH is close in value to the constant term of GARCH model for daily data, and the difference increases significantly for weekly and monthly data. Since GARCH model does not capture the current information, it treats the missing information as a constant term. Similar to the simulation results, the current information risk premium increases in magnitude with the length of sampling interval.

We also split the data into pre and post 2008 financial crisis to examine the change of current information effects on volatility. The results are presented in Table 4. Due to small sample sizes, we only perform separate estimations on daily data. The financial crisis is likely to have created structural breaks in volatility process as the parameters estimated from two data sets are very different.[12] Volatility is less persistent and the current information parameter ψ doubled in value after the crisis. This implies the 2008 financial crisis has changed the volatility structure in a way that current return information contributes to more variations in current level of volatility and the lagged level of volatility contributes slightly less compared to before the crisis. This can be due to investors' increasing aversion for information uncertainty since they require more risk premium in compensation after 2008 financial crisis. The standard GARCH model can only capture the decrease of volatility persistence but fails to account for the change due to the increasing importance of current return information risk premium. This provides another

[12]Formal test on structural break of volatility process would be difficult given its latent nature.

Table 4: RT-GARCH parameters for daily S&P 500 before and after 2008

	α	β	γ	ψ
04-Jan-2000 – 31-Jul-2008	$2.1781 * 10^{-4}$ $(5.4772*10^{-4})$	0.9210 (0.0137)	0.0645 (0.0129)	0.0147 (0.0034)
01-Aug-2008 – 30-Dec-2019	$4.6560 * 10^{-6}$ (0.0039)	0.8264 (0.0170)	0.1250 (0.0151)	0.0351 (0.0049)

Note: The sample sizes are 2157 and 2873 for pre- and post-crisis, respectively. The standard errors, calculated numerically, are given in parentheses.

advantage of using RT-GARCH over GARCH models in empirical applications.

5 Conclusion

In this paper, we have derived the diffusion limit of Smetanina's (2017) RT-GARCH model and extended Nelson's (1990) theorem to incorporate a broader range of GARCH type models for diffusion approximation. In doing so, we have answered the question where RT-GARCH stands in between GARCH and SV models and provide more theoretical evidence of advantages for using RT-GARCH to model discrete time volatility. First, since RT-GARCH and GARCH converge weakly to the same type of diffusion process, RT-GARCH performs at least as good as GARCH for data sampled at ultra high frequency. Moreover, both models provide consistent filters and estimators of volatility under mild conditions. Second, the additional parameter of RT-GARCH can be interpreted as current information risk premium and allows us to separate it from the volatility drift term. This risk premium also controls the scale of the limiting stationary distribution. On the other hand, if the data generating process is the diffusion limit, RT-GARCH suffers larger discretisation errors than GARCH and cautions need to be in place when using RT-GARCH to fit discretely sampled data. Given these results, we can formally define RT-GARCH models as a sub-class within the GARCH class.

GARCH type models encompass large variations and are relatively easy to implement in practice. RT-GARCH provides an alternative way of treating volatility as a stochastic process while retaining the elegant QML estimation procedure. While the results in this paper contribute and complement the theory of Smetanina's (2017) RT-GARCH and its relation with GARCH and SV models. In order to fully justify the use of RT-GARCH model, it would be useful to derive the asymptotic filtering theory of RT-GARCH to understand whether this added current information parameter helps reduce asymptotic variance of measurement error. It would also be interesting to derive the conditions under which RT-GARCH is the asymptotically optimal filter in the sence of Nelson and Foster (1994). These tasks await future research.

References

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 367–389.

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

Cheridito, P., D. Filipović, and R. L. Kimmel (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics 83*(1), 123–170.

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.

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

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review 35*(3), 705–730.

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.

Ishida, I. and R. F. Engle (2002). Modeling variance of variance: The square-root, the affine, and the CEV GARCH models. Department of finance working papers, New York University. `http://pages.stern.nyu.edu/~rengle/ishida.pdf`.

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.

Mykland, P. A. and L. Zhang (2006). Anova for diffusions and Itô processes. *The Annals of Statistics 34*(4), 1931–1963.

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

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 (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica 62*(1), 1–41.

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.

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. `https://sites.google.com/site/smetaninakatja/research`.

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

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

A Proofs

In this section we provide proofs of the main theorems in this paper. We suppose $kh \leq t < (k+1)h$ throughout this section. It is convenient to write the innovation terms ${}_h\epsilon_{kh}$ as increments of random variables ${}_hW_{kh} - {}_hW_{(k-1)h} \sim N(0,h)$. By Lévy's characterisation, ${}_hW_{kh}$ is a one dimensional standard Brownian motion.

See Stroock and Varadhan (1979) and Nelson (1990) for the proof of Theorem 1.1.

In proving the theorems in section 3, we need the following proposition:

Proposition 1. $\lim_{h\downarrow 0} h^{-1}{}_h\epsilon_{kh}^2 = \langle W, W\rangle'(t) = 1$,[13] *where W is a one dimensional standard Brownian motion and $\langle\cdot,\cdot\rangle'(t)$ is the time derivative of the quadratic variation of a stochastic process.*

Proof. Recall the definition of quadratic variation,

$$\langle W, W\rangle(t) \equiv \lim_{\|\Pi\|\to 0} \sum_{i=1}^{n} (W_{t_i} - W_{t_{i-1}})^2, \tag{A.1}$$

where Π ranges over the partition of the interval $[0,t]$ and the norm of the partition Π is the mesh. $\lim_{h\downarrow 0} h^{-1}{}_h\epsilon_{kh}^2$ is then the time derivative of the quadratic variation of a standard Brownian motion, which is equal to t. In other word, $\lim_{h\downarrow 0} h^{-1}{}_h\epsilon_{kh}^2$ converges to the quadratic variation of a standard Brownian motion per unit time. □

Lemma A.1 (Mykland and Zhang (2006))**.** *Let $\Pi = t_0, t_1, \ldots, t_n$ be a sequence of non-random partitions of interval $[0,t]$ and $\Delta t_i = t_{i+1} - t_i$, define the observed fourth-order variation for an Itô process X with a.s. bounded drift and $\langle X, X\rangle'$*

$$[X]^4(t) = \sum_{t_{i+1}\leq t} (X_{t_{i+1}} - X_{t_i})^4. \tag{A.2}$$

Then as $\|\Pi\| \to 0$,

$$\bar{\Delta t_i}^{-1}[X]^4(t) \to \int_0^T 3H'(u)(\langle X, X\rangle'(u))^2 du \tag{A.3}$$

uniformly in probability, where $\bar{\Delta t_i}$ is the average distance between successive observations T/n and $H(t) = \sum_{t_{i+1}\leq T}(\Delta t_i)^2/\bar{\Delta t_i}$.

Proposition 2. $\lim_{h\downarrow 0} h^{-2}{}_h\epsilon_{kh}^4 = 3\big(\langle W, W\rangle'(t)\big)^2 = 3$, *where W is a one dimensional standard Brownian motion.*

Proof. We assume equispaced observations throughout the paper. Thus, $\bar{\Delta t_i} = h$ and $H(t) = t$. Proposition 2 follows Lemma $A.1$ by taking the derivative of (A.3) w.r.t t. □

Remark A.1.1. *The convergence can be made stronger using $\mathcal{L}^2$ convergence argument for the fourth power variation of a scaled Brownian motion $W_t/t^{1/4}$.*

[13] Here the convergence is under the meaning of $\mathcal{L}^2$ convergence.

Proof of Theorem 2.1. The discrete time process (2.1) and (2.2) is a joint Markov chain. Moreover, under mild conditions it is geometrically moment contracting and there exists an a.s.-unique casual ergodic strictly stationary solution at its true parameters (Smetanina and Wu, 2019). Thus, to prove Theorem 2.2, it suffices to check Assumptions 1–4. Assumption 1 is already assumed in the theorem.

To verify Assumption 2, we first impose stationary conditions on the limit of the sequence of parameters. As $kh \to \infty$,

$$\mathbb{E}[r^2] = h\mathbb{E}[\sigma^2] + 2h\psi_h. \tag{A.4}$$

Plug into the unconditional expectation of (2.2),

$$\mathbb{E}[\sigma^2] = \alpha_h + (\beta_h + \gamma_h)\mathbb{E}[\sigma^2] + 2\psi_h\gamma_h + \psi_h. \tag{A.5}$$

This can only hold if and only if

$$\lim_{h\downarrow 0}(\beta_h + \gamma_h) = 1, \tag{A.6}$$

$$\lim_{h\downarrow 0}(\alpha_h + \psi_h + 2\psi_h\gamma_h) = 0. \tag{A.7}$$

Next we derive the limit of the increments per unit of time conditional on information at time $(k-1)h$. In contrast to standard GARCH, we have a smaller information set since the current volatility is no longer $\mathcal{F}_{k(h-1)}$-measurable, i.e., $\mathcal{F}_{k(h-1)}$ is the σ-algebra generated by $kh, {}_hS_0, ..., {}_hS_{k(h-1)}$, and ${}_h\sigma_0^2, ..., {}_h\sigma_{k(h-1)}^2$.[14]

$$\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})|\mathcal{F}_{(k-1)h}] = 0, \tag{A.8}$$

$$\begin{aligned}\mathbb{E}[h^{-1}({}_h\sigma_{kh}^2 - {}_h\sigma_{(k-1)h}^2)|\mathcal{F}_{(k-1)h}] = h^{-1}[&\alpha_h \\ &+ \psi_h + (\beta_h - 1 + h^{-1}\gamma_h \cdot {}_h\epsilon_{(k-1)h}^2){}_h\sigma_{(k-1)h}^2],\end{aligned} \tag{A.9}$$

Taking the limit and using Proposition 1 and (2.7) and (2.8) of Assumption 5,

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})|\mathcal{F}_{(k-1)h}] = 0, \tag{A.10}$$

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_h\sigma_{kh}^2 - {}_h\sigma_{(k-1)h}^2)|\mathcal{F}_{(k-1)h}] = \mu - \theta\sigma^2. \tag{A.11}$$

The second moment per unit time is given by

$$\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})^2|\mathcal{F}_{(k-1)h}] = \alpha_h + 3\psi_h + (\beta_h + h^{-1}\gamma_h \cdot {}_h\epsilon_{(k-1)h}^2){}_h\sigma_{(k-1)h}^2, \tag{A.12}$$

$$\begin{aligned}\mathbb{E}[h^{-1}({}_h\sigma_{kh}^2 - {}_h\sigma_{(k-1)h}^2)^2|\mathcal{F}_{(k-1)h}] = h^{-1}\Big(&\alpha_h^2 + 3\psi_h^2 + 2\alpha_h\psi_h + \big((\beta_h - 1)^2 \\ &+ h^{-1}\gamma_h^2 \cdot {}_h\epsilon_{(k-1)h}^4 + 2(\beta_h - 1)\gamma_h \cdot {}_h\epsilon_{(k-1)h}^2\big){}_h\sigma_{(k-1)h}^4 + \big(2(\alpha_h + \psi_h)(\beta_h - 1) \\ &+ 2\alpha_h\psi_h + \cdot{}_h\epsilon_{(k-1)h}^2 + 2h^{-1}\gamma_h\psi_h \cdot {}_h\epsilon_{(k-1)h}^2\big){}_h\sigma_{(k-1)h}^2\Big).\end{aligned} \tag{A.13}$$

[14] In standard GARCH the σ-algebra contains information up to ${}_h\sigma_{kh}^2$.

Using Propositions 1 and 2, the stationary conditions (A.6) and (A.7) and (2.9) of Assumption 5,

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_hS_{kh}-{}_hS_{(k-1)h})^2|\mathcal{F}_{(k-1)h}]=\sigma^2, \tag{A.14}$$

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_h\sigma^2_{kh}-{}_h\sigma^2_{(k-1)h})^2|\mathcal{F}_{(k-1)h}]=\gamma^2\sigma^4. \tag{A.15}$$

Finally, the cross-moment is given by

$$\mathbb{E}[h^{-1}({}_hS_{kh}-{}_hS_{(k-1)h})({}_h\sigma^2_{kh}-{}_h\sigma^2_{(k-1)h})|\mathcal{F}_{(k-1)h}]=h^{-2}\psi_h\mathbb{E}[{}_h\epsilon^3_{kh}\cdot{}_h\sigma_{kh}|\mathcal{F}_{(k-1)h}]. \tag{A.16}$$

Since ${}_h\epsilon_{kh}$ is symmetric around zero by assumption, cubic function is an odd function and σ is an even function of ϵ. It follows automatically that $\epsilon^3\sigma$ is symmetric around zero[15].

It is straightforward but tidious to verify the limits of the fourth moments go to zero since ${}_h\epsilon^6_{kh}=\mathcal{O}_p(h^3)$ and ${}_h\epsilon^8_{kh}=\mathcal{O}_p(h^4)$.

It remains to verify the distributional uniqueness of the diffusion limit. This follows directly from Nelson (1990) since RT-GARCH converges to the same SDEs as GARCH model. □

To prove Theorem 2.2 we need to derive the power variations of the negative increments. We slightly modify Barndorff-Nielsen and Shephard's (2003) limit theorem of power variations for a semimartingale.

Lemma A.2. *Let W be a standard Brownin motion, and denote*

$$\left([X]^-\right)^r(t)\equiv\sum_{t_{i+1}\le t}(X_{t_{i+1}}-X_{t_i})^r\mathbb{1}_{\{X_{t_{i+1}}-X_{t_i}<0\}},$$

where (t_i) are sequence of non-random partitions of interval $[0,t]$. That is, $\left([X]^-\right)^r(t)$ is the contribution of negative increments to the observed power variation of X. Then as $h\downarrow 0$, for all $r>0$,

$$h^{1-r/2}\left([{}_hW_t]^-\right)^r(t)\to\mathbb{E}[(u^-)^r]t \tag{A.17}$$

uniformly in probability, where $u^-=u\mathbb{1}_{\{u<0\}}$ and $u\sim N(0,1)$.

Proof. $(X_{t_{i+1}}-X_{t_i})^r\mathbb{1}_{\{X_{t_{i+1}}-X_{t_i}<0\}}$ has the same law as $u^-_{t_i}$, where u are i.i.d. standard normal. By symmetry, the contribution of the $\mathbb{E}[(u^-)^r]=\frac{1}{2}\mathbb{E}[|u|^r]$ for all $r\ge 0$. Set $A=0$, $H=1$ and $H^*=\int_0^t dt=t$ in the proof of Theorem 1 in Barndorff-Nielsen and Shephard (2003), we obtain

$$\left([W]^-\right)^r(t)\to\tfrac{1}{2}\mathbb{E}[|u|^r]t.$$

The result follows immediately. □

Remark A.2.1. *Replace with standard Brownian motion in Lemma A.2, we obtain the result in Proposition 2 since $\mathbb{E}[{}_h\epsilon^4_{kh}]=3h^2$.*

[15] One can also use the fact that ${}_h\epsilon^3_{kh}$ is of order $o_p(h^{3/2})$ and ${}_h\sigma_{kh}=o_p(h^{1/2})$ to argue the expression goes to 0 a.s..

Recall the moments of half normal distribution for all integer n,

$$\mathbb{E}[{}_h(\epsilon_{kh}^-)^{2n}] = \mathbb{E}[{}_h\epsilon_{kh}^{2n}], \tag{A.18}$$
$$\mathbb{E}[{}_h(\epsilon_{kh}^-)^{2n+1}] = n!h^{2n+1}/\sqrt{2\pi}. \tag{A.19}$$

Then a direct application of Lemma $A.2$ gives us the followings:

Proposition 3. *Let* ${}_h\epsilon_{kh}^- = {}_h\epsilon_{kh}^- \mathbb{1}_{\{{}_h\epsilon_{kh}<0\}}$, *then*

$$\lim_{h\downarrow 0} h^{-1}({}_h\epsilon_{kh}^-)^2 = \langle W, W\rangle^{-'}(t) = 1/2, \tag{A.20}$$
$$\lim_{h\downarrow 0} h^{-3/2}({}_h\epsilon_{kh}^-)^3 = -\sqrt{2/\pi}, \tag{A.21}$$
$$\lim_{h\downarrow 0} h^{-2}({}_h\epsilon_{kh}^-)^4 = 3\big(\langle W, W\rangle^{-'}(t)\big)^2 = 3/2, \tag{A.22}$$

where $\langle\cdot,\cdot\rangle^{'}(t)$ *is the time derivative of quadratic variation.*

Proof of Theorem 2.2. We only need to consider the moments regarding the additional asymmetric terms and their cross terms, the rest follows the same as the proof of Theorem 2.2.

In steady state, by symmetric distribution of returns, the unconditional variance is,

$$\mathbb{E}[\sigma^2] = \alpha_h + \psi_h + \tfrac{1}{2}\eta_h + \phi_h(\psi_h + \tfrac{1}{2}\eta_h) + (\beta_h + \gamma_h + \tfrac{1}{2}\phi_h)\mathbb{E}[\sigma^2]. \tag{A.23}$$

Thus, covariance stationatiry imposes conditions on the limits of the sequence of parameters,

$$\lim_{h\downarrow 0}(\beta_h + \gamma_h + \tfrac{1}{2}\phi_h) = 1, \tag{A.24}$$
$$\lim_{h\downarrow 0}(\alpha_h + \psi_h + \tfrac{1}{2}\eta_h + \phi_h(\psi_h + \tfrac{1}{2}\eta_h)) = 0. \tag{A.25}$$

The drifts per unit time are given by

$$\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})|\mathcal{F}_{(k-1)h}] = 0, \tag{A.26}$$
$$\begin{aligned}\mathbb{E}[h^{-1}({}_h\sigma_{kh}^2 - {}_h\sigma_{(k-1)h}^2)|\mathcal{F}_{(k-1)h}] = h^{-1}\Big(&\alpha_h + \tfrac{1}{2}\psi_h \\ &+ \big(\beta_h - 1 + h^{-1}\gamma_h \cdot {}_h\epsilon_{(k-1)h}^2 + h^{-1}\phi_h({}_h\epsilon_{(k-1)h}^-)^2\big){}_h\sigma_{(k-1)h}^2\Big)\end{aligned} \tag{A.27}$$

Using Propositions 1 and 3 and (2.23) and (2.24) of Assumption 6,

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})|\mathcal{F}_{(k-1)h}] = 0, \tag{A.28}$$
$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_h\sigma_{kh}^2 - {}_h\sigma_{(k-1)h}^2)|\mathcal{F}_{(k-1)h}] = \mu - \theta\sigma^2. \tag{A.29}$$

Similarly, for the second moments per unit time, we use Propositions 1 - 3, the sta-

tionary conditions (A.24) and (A.25) and (2.25) of Assumption 6,

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_hS_{kh}-{}_hS_{(k-1)h})^2|\mathcal{F}_{(k-1)h}]=\sigma^2, \tag{A.30}$$

$$\lim_{h\downarrow 0}\mathbb{E}[h^{-1}({}_h\sigma^2_{kh}-{}_h\sigma^2_{(k-1)h})^2|\mathcal{F}_{(k-1)h}]=\gamma^2\sigma^4. \tag{A.31}$$

Finally, by symmetric assumption of returns,

$$\mathbb{E}[{}_h\sigma_{kh}\cdot{}_h\epsilon^3_{kh}|\mathcal{F}_{(k-1)h}]=\mathbb{E}[{}_h\sigma_{kh}\cdot{}_h\epsilon_{kh}|\mathcal{F}_{(k-1)h}]=0. \tag{A.32}$$

Apply stationary conditions, the cross moment is given by

$$\begin{aligned}&\mathbb{E}[h^{-1}({}_hS_{kh}-{}_hS_{(k-1)h})({}_h\sigma^2_{kh}-{}_h\sigma^2_{(k-1)h})|\mathcal{F}_{(k-1)h}]=\\ &h^{-2}\eta_h\mathbb{E}[{}_h\sigma_{kh}({}_h\epsilon^-_{kh})^3|\mathcal{F}_{(k-1)h}]+h^{-2}\phi_h\mathbb{E}[{}_h\sigma^2_{(k-1)h}({}_h\epsilon^-_{(k-1)h})^2{}_h\sigma_{kh}\cdot{}_h\epsilon_{kh}|\mathcal{F}_{(k-1)h}].\end{aligned} \tag{A.33}$$

Since $\eta_h=\mathcal{O}(h)$ and $\mathbb{E}[{}_h\sigma_{kh}({}_h\epsilon^-_{kh})^3|\mathcal{F}_{(k-1)h}]=\mathcal{O}_p(h^2)$, the first term goes to 0 as $h\downarrow 0$. For the second term, ${}_h\sigma_{kh}\cdot{}_h\epsilon_{kh}={}_h\sigma_{(k-1)h}\cdot{}_h\epsilon_{(k-1)h}+o_p(h)$ for small h by sample path continuity.

$$\begin{aligned}&\mathbb{E}[h^{-1}({}_hS_{kh}-{}_hS_{(k-1)h})({}_h\sigma^2_{kh}-{}_h\sigma^2_{(k-1)h})|\mathcal{F}_{(k-1)h}]=\\ &h^{-2}\phi_h\cdot{}_h\sigma^3_{(k-1)h}({}_h\epsilon^-_{(k-1)h})^3+o_p(h)=-\sqrt{2/(h\pi)}\phi_h\cdot{}_h\sigma^3_{(k-1)h}+o_p(h),\end{aligned} \tag{A.34}$$

where in the last equality we use Proposition 3. Combine with (2.26) of Assumption 6, we obtain

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

For the expression of the correlation between return and volatility innovations, we take Cholesky decomposition of the covariance matrix

$$a(S,\sigma)\equiv\begin{bmatrix}\sigma^2 & \omega\sigma^3\\ \omega\sigma^3 & \gamma^2\sigma^4,\end{bmatrix} \tag{A.36}$$

and obtain

$$a(S,\sigma)^{1/2}=\begin{bmatrix}\sigma & 0\\ \rho\gamma\sigma^2 & \gamma\sqrt{1-\rho^2}\sigma^2,\end{bmatrix} \tag{A.37}$$

where $\rho=\omega/\gamma$. □

Proof of Theorem 2.3. See Nelson's (1990) Theorem 2.3 for the stationry distribution of ${}_h\sigma^2_{kh}$. From (2.20), for $kh\le t<(k+1)h$ and sufficiently small $h>0$, ${}_hr_{kh}$ can be approximated by

$${}_hS_{kh}-{}_hS_{(k-1)h}={}_h\sigma_{(k-1)h}[\rho(W_{1,kh}-W_{1,(k-1)h})+\sqrt{1-\rho^2}(W_{2,kh}-W_{2,(k-1)h})], \tag{A.38}$$

where $W_{1,kh}$ and $W_{2,kh}$ are innovation terms from (2.20) and (2.21). It is clear that $W_{2,kh}$ and ${}_h\sigma_{kh}$ are independent while $W_{1,kh}$ and ${}_h\sigma_{kh}$ are not. According to Theorem 3.2 in Nelson (1990), we can replace $W_{1,kh}$ by

$$Q_{kh}=(1-\tfrac{2}{\pi})^{-1/2}\sum_{j=1}^{k}\big(|W_{1,jh}-W_{1,(j-1)h}|-\sqrt{2/(\pi h)}\big), \tag{A.39}$$

This is because as $h \downarrow 0$,

$$(Q_{kh}, W_{1,kh}) \xrightarrow{d} W_t^*, \tag{A.40}$$

where W_t^* is a two-dimensional standard Brownian motion. That is, even if $W_{1,kh}$ and $|W_{1,kh}|$ are not independent, their partial sums in the limit are independent as $h \downarrow 0$. Hence, as $h \downarrow 0$, we can replace $W_{1,kh} - W_{1,(k-1)h}$ by

$$(1 - \tfrac{2}{\pi})^{-1/2} |W_{3,kh} - W_{3,(k-1)h}| \tag{A.41}$$

where $W_{3,kh}$ is a standard Brownian motion indepedent of $W_{1,kh}$ and $W_{2,kh}$. The law of $h^{-1/2}{}_h r_{kh}$, as $h \downarrow 0$, is equivalent to

$$\sigma_t(\rho/\sqrt{1 - \tfrac{2}{\pi}}|Z_1| + \sqrt{1-\rho^2} Z_2), \tag{A.42}$$

where Z_1 and Z_2 are bivariate standard normal random variable with zero correlation and are independent from σ_t. It can be shown that $\rho/\sqrt{1 - \frac{2}{\pi}}|Z_1| + \sqrt{1-\rho^2} Z_2$ is distributed as

$$\text{Skew-}N\left(0, \sqrt{\frac{\rho^2}{1-\frac{2}{\pi}} + 1 - \rho^2}, \frac{\rho}{\sqrt{(1-\frac{2}{\pi})(1-\rho^2)}}\right), \tag{A.43}$$

where Skew-$N(a, b, c)$ is the skewed normal distribution with location parameter a, scale parameter b and shape parameter c. Its probability density function (pdf) is

$$f_{SN}(x) = 2\phi(x-a; b)\Phi(c(x-a)/\sqrt{b}), \tag{A.44}$$

where $\phi(x; b)$ is the pdf of $N(0, b)$ and $\Phi(\cdot)$ is cumulative distribution function (cdf) of $N(0, 1)$. Since the stationary distribution of σ_t^2 has an inverse gamma distribution, by standard argument, we have

$$\frac{\gamma^2}{4\mu}\sigma^{-2} \sim \chi^2_{2+4\theta/\gamma^2}, \tag{A.45}$$

where $\chi^2_{2+4\theta/\gamma^2}$ is the chi-square distribution with $2 + 4\theta/\gamma^2$ degrees of freedom. Finally, combine (A.42), (A.43) and (A.45) we have

$$h^{-1/2}\sqrt{(2\theta+\gamma^2)/2\mu} \cdot {}_h r_{kh} \sim V^{-1/2} Y, \tag{A.46}$$

as $h \downarrow 0$, where

$$V \sim \chi^2_{2+4\theta/\gamma^2}/(2 + 4\theta/\gamma^2), \tag{A.47}$$

and Y is distributed as (A.43). (A.46) is the definition of skewed Student-t distribution in Azzalini and Capitanio (2003), that is,

$$V^{-1/2} Y \sim \text{Skew-}t(a, b, c, n), \tag{A.48}$$

where $a = 0$, $n = 2 + 4\theta/\gamma^2$ and

$$b = \frac{\rho^2}{1-\frac{2}{\pi}} + 1 - \rho^2, \qquad c = \frac{\rho}{\sqrt{(1-\frac{2}{\pi})(1-\rho^2)}}.$$

The probability density of Skew-$t(a, b, c, n)$ is

$$f_{St}(x) = 2t(x; a, b, n)T\left(c\frac{x-a}{\sqrt{b}}\sqrt{\frac{n+1}{(x-a)^2/b+n}}; n+1\right), \tag{A.49}$$

where $t(x; a, b, n)$ is the pdf of Student-t distribution with location a, scale b and n degrees of freedom and $T(y; n+1)$ is the cdf of standard Student-T distribution with $n+1$ degrees of freedom. When $\rho = 0$ the distribution of (A.46) reduces to $t(n)$. See Azzalini and Capitanio (2003) for more details on skew normal and Student-t distributions.

The proof of second part of the theorem, for $kh \to \infty$ and starting point not from its stationary distribution, follows exactly Nelson's (1990) proof of theorem 2.3. □

Lemma $A.1$ and Lemma $A.2$ hold whenever the sample path continuity is satisfied. Thus, to extend to the RT-GARCH with Student-t innovations, we need the conditions for its sample path continuity.

Proposition 4. *Processes with Student-t increments have continuous paths if the degree of freedom $\nu \geq 3$.*

Proof. For a Markov process, the sample path is a continuous functions of t, if for any $\epsilon > 0$,

$$\lim_{h\downarrow 0} h^{-1} \int_{|x-z|>\epsilon} p(x, t+h|y, t)dy = 0 \tag{A.50}$$

uniformly in x, t and h, where $p(\cdot|\cdot)$ is the transition density of random variable y. Equivalently, for a symmetric, zero mean density,

$$\lim_{h\downarrow 0} h^{-1}\big(1 - \int_{-\epsilon}^{\epsilon} p(x, t+h|y, t)dy\big) = 0. \tag{A.51}$$

For Student-t distribution with $\nu = 1$ (i.e. a Cauchy transition density),

$$\lim_{h\downarrow 0} h^{-1}\big(1 - \int_{-\epsilon}^{\epsilon} \frac{\sqrt{h}}{\pi((x-y)^2+h}dx\big) = \lim_{S\to\infty} 1 - \tfrac{2}{\pi}\arctan(\epsilon\sqrt{S}). \tag{A.52}$$

Expanding about $S = \infty$ we have,

$$\frac{2\sqrt{S}}{\epsilon\pi}(1 - \frac{1}{3\epsilon^2 S} + \frac{1}{5\epsilon^4 S^2}), \tag{A.53}$$

which has a dominant term $\sqrt{S}/\pi$ and the limit goes to infinity.

For $\nu = 2$,

$$\begin{aligned} &\lim_{h\downarrow 0} h^{-1}(1 - \int_{-\epsilon}^{\epsilon} \frac{\sqrt{2}}{4\sqrt{h}}(1 + \frac{x^2}{2h})^{-2/3}dx) = \\ &\lim_{S\to\infty} S(1 - \int_{-\epsilon}^{\epsilon} \frac{\sqrt{2S}}{4}(1 + \frac{Sx^2}{2})^{-2/3}dx) = \frac{1}{\epsilon^2}, \end{aligned} \tag{A.54}$$

which is not zero. So Student-t increments with $\nu = 2$ does not have continuous sample path.

For $\nu = 3$,

$$\lim_{h\downarrow 0} h^{-1}(1 - \int_{-\epsilon}^{\epsilon} \frac{2}{\pi\sqrt{3h}}(1+\frac{x^2}{3h})^{-2}dx) = \\ \lim_{S\to\infty} S\Big(1-2\frac{\arctan(\epsilon\sqrt{3S}/3)S\epsilon^2 + \epsilon\sqrt{3S} + 3\arctan(\epsilon\sqrt{3S}/3)}{\pi(\epsilon^2 S+3)}\Big). \tag{A.55}$$

Expanding about $S \to \infty$ shows the dominant term is $1/\sqrt{S}$,

$$\lim_{S\to\infty} \frac{4\sqrt{3}}{\pi\epsilon^3\sqrt{S}} + \frac{72\sqrt{3}}{5\pi\epsilon^5 S^{3/2}} + \mathcal{O}(S^{-5/2}) = 0. \tag{A.56}$$

This limit holds for all $\nu \geq 3$, and with ν increases, Student-t increments behave increasingly likely to Gaussian increments. □

Once the sample path continuity is established, we can directly apply Barndorff-Nielsen and Shephard's (2003) Theorem 1 for Student-t increments.

Lemma A.3. *Let ${}_h\epsilon_{(k+1)h} = {}_h\tau_{(k+1)h} - {}_h\tau_{kh}$ be a sequence of rescaled Student-t increments, i.e. the transition density is defined as*

$$f_\epsilon(x;\nu,h) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu h}}(1+\frac{x^2}{\nu h})^{-\frac{\nu+1}{2}} \tag{A.57}$$

If $\nu \geq 3$, then as $h \downarrow 0$ and $kh \leq t < (k+1)h$, for $0 < r \leq \nu$,

$$h^{1-r/2}[{}_h\tau_t]^r(t) \to \mathbb{E}[(\epsilon)^r]t \tag{A.58}$$

uniformly in probability, where ϵ is a standard Student-t.

Proof. By Proposition 4, sample path continuity is satisfied for Studnet-t increments. In Theorem 1 of Barndorff-Nielsen and Shephard (2003) replace A with 0, H with 1 and W with ϵ, the observed increments have the same law as $u_{t_i}/\sqrt{h}$, where u are i.i.d. standard Student-t. Provided the degree of freedom is such that all the moments smaller or equal to r exist, the rest follows directly from the proof of Barndorff-Nielsen and Shephard (2003) □

Corollary A.3.1. *Let ${}_h\epsilon^-_{(k+1)h} = ({}_h\tau_{(k+1)h} - {}_h\tau_{kh})\mathbb{1}_{\{{}_h\tau_{(k+1)h} - {}_h\tau_{kh}<0\}}$ be a sequence of negative rescaled Student-t increments, i.e. ${}_h\epsilon^-_{kh} = {}_h\epsilon_{kh}\mathbb{1}_{\{{}_h\epsilon_{kh}<0\}}$ and ${}_h\epsilon_{kh}/\sqrt{h}$ are i.i.d. standard Student-t. If $\nu \geq 3$, then as $h \downarrow 0$, for all $0 < r \leq \nu$,*

$$h^{1-r/2}\big([{}_h\tau_t]^-\big)^r(t) \to \mathbb{E}[(\epsilon^-)^r]t \tag{A.59}$$

uniformly in probability, where ϵ^- is standard half Student-t distributed.

Remark A.3.1. *One can also use $\mathcal{L}^2$ convergence argument to prove the power variation is proportional to square root of time to the respective power. However, we would require $\nu \geq 2r$ to ensure the higher moments exist and do not explode.*

We are now in the position to proof diffusion limit of RT-GARCH with Student-t innovations.

Proof of Theorem 2.4. Using the first four moments of Student-t and half Student-t random variables, we can establish the following uniform convergences in probability,

$$\lim_{h\downarrow 0} h^{-1}\epsilon_{kh}^2 = \langle \tau, \tau\rangle'(t) = \frac{\nu}{\nu-2}, \tag{A.60}$$

$$\lim_{h\downarrow 0} h^{-2}\epsilon_{kh}^4 = \frac{3\nu^2}{(\nu-4)(\nu-2)}, \tag{A.61}$$

$$\lim_{h\downarrow 0} h^{-1/2}\epsilon_{kh} = \lim_{h\downarrow 0} h^{-2/3}\epsilon_{kh}^3 = 0. \tag{A.62}$$

Similarly,

$$\lim_{h\downarrow 0} h^{-1}(\epsilon_{kh}^-)^2 = \langle \tau, \tau\rangle^{-\prime}(t) = \frac{\nu}{2(\nu-2)}, \tag{A.63}$$

$$\lim_{h\downarrow 0} h^{-2}(\epsilon_{kh}^-)^4 = \frac{3\nu^2}{2(\nu-4)(\nu-2)}, \tag{A.64}$$

$$\lim_{h\downarrow 0} h^{-3/2}(\epsilon_{kh}^-)^3 = \frac{2\nu^{\frac{3}{2}}}{\sqrt{\pi}(\nu-3)(\nu-1)}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}, \tag{A.65}$$

where $\Gamma(\cdot)$ is the gamma function. By Assumption 7, all these limits exist and are finite. Putting (A.60) - (A.65) in place of the proof of Theorem 2.2, we obtain the uniform convergence of first and second moments. It is straightforward but tedious to check that the fourth moments converge to zero as long as $\nu \geq 8$ by Assumption 7. This is to ensure $\mathbb{E}[{}_h\epsilon_{kh}^4]$ exist and is finite. The rest follows the same as the proof of Theorem 2.2. □

Proof of Theorem 2.5. We need only verify Assumptions 2 and 3. The drift matrix is

$$m_h(s,y,t) = \begin{bmatrix} f \\ F_1 + F_2 \end{bmatrix}, \tag{A.66}$$

Since $m_h(\cdot,\cdot,\cdot) = m(\cdot,\cdot,\cdot)$, (1.5) of Assumption 1 is satisfied. The diagonal elements of diffusion matrix are $hf^2 + g^2$ and $G\Omega_{2,2}G^T$. To calculate the covariance terms,

$$\begin{aligned}\mathbb{E}[h^{-1}({}_hS_{kh} - {}_hS_{(k-1)h})({}_hY_{kh} - {}_hY_{(k-1)h})^T|\mathcal{F}_{(k-1)h}] = \\ hfF_1^T + f\cdot {}_hZ_{(k-1)h}^{*T}G^T + h^{-1}f\cdot {}_hZ_{(k-1)h}^{**T}F_2^T + \mathbb{E}[g\cdot {}_hZ_{kh}F_1^T \\ + h^{-1}g\cdot {}_hZ_{kh}\cdot {}_hZ_{(k-1)h}^{*T}G^T + h^{-1}g\cdot {}_hZ_{kh}\cdot {}_hZ_{(k-1)h}^{**T}F_2^T|F_{(k-1)h}].\end{aligned} \tag{A.67}$$

Since we assume sample path continuity, $h^{-r/2}{}_hZ_{kh}^r \to \mathbb{E}[Z_t^r]$ and f, F_1, F_2, g and G are locally bounded, as $h \downarrow 0$,

$$(\text{A.67}) \to g\Omega_{2,1}G^T. \tag{A.68}$$

Thus,

$$\Sigma_h(s,y,t) \xrightarrow{h\downarrow 0} \begin{bmatrix} g^2 & G\Omega_{1,2}g \\ g\Omega_{2,1}G^T & G\Omega_{2,2}G^T \end{bmatrix}. \tag{A.69}$$

Finally,

$$h^{-1}\mathbb{E}\left[\begin{array}{c}(hf+g_hZ_{kh})^4\\(hF_1+G_hZ^*_{(k-1)h}+F_{2h}Z^{**}_{kh})^4\end{array}\middle|\mathcal{F}_{(k-1)h}\right]=\mathcal{O}_p(1), \tag{A.70}$$

uniformly on compacts. All the Assumptions for Theorem 1.1 are verified. □

Proof of Theorem 3.1. Under the conditions of (3.5) - (3.8), the filtering error process $({}_hZ_{kh})$ satisfy

$$h^{-1}\mathbb{E}[{}_hZ_{(k+1)h}-{}_hZ_{kh}|{}_hS_{kh}=s,{}_hZ_{kh}=z]=-h^{\delta-1}\gamma z+\mathcal{O}_p(1), \tag{A.71}$$

using Condition 3.4, Lemma A.1 of Nelson (1992) and $\mathbb{E}[\psi_h\cdot{}_h\epsilon^2_{kh}]=o(h^{\delta+1})$. The rest follows directly from the proof of Theorem 3.1 of Nelson (1992). □

Proof of Theorem 3.2. See the proofs of Theorems 2.4, 2.5, 2.6 and 3.1 of Nelson and Foster (1994). If in addition, the data generating process is its diffusion limit, i.e., (2.20) and (2.21), then Assumptions 9 - 12 are satisfied. The Lyapunov function needed to verify Nelson's (1992) Condition 3.3 can be $\omega(s,y)=K+f(s)|s|+f(y)\exp(|y|)$, where $f(x)\equiv\exp(-1/|x|)$ if $x\neq 0$ and 0 otherwise. $\omega(\cdot,\cdot)$ is arbitrarily continuously differentiable, nonnegative. To verify the partial differential inequality, use the fact that for large s and y, $\partial\omega(s,y)/\partial s\approx\text{sign}(s)$, $\partial^2\omega(s,y)/\partial s^2=0$, $\partial\omega(s,y)/\partial y\approx\text{sign}(y)\exp(|y|)$ and $\partial^2\omega(s,y)/\partial y^2\approx\exp(|y|)$. All the Assumptions and Conditions are verified and the results follow immediately. □