The Continuous-Time Limit of Score-Driven Volatility Models

Giuseppe Buccheri* Fulvio Corsi† Franco Flandoli‡ Giulia Livieri§

July 27, 2020

Abstract

We provide general conditions under which a class of discrete-time volatility models driven by the score of the conditional density converges in distribution to a stochastic differential equation as the interval between observations goes to zero. We show that the form of the diffusion limit depends on: (i) the link function, (ii) the conditional second moment of the score, (iii) the normalization of the score. Interestingly, the properties of the stochastic differential equation are strictly entangled with those of the discrete-time counterpart. Score-driven models with fat-tailed densities lead to continuous-time processes with finite volatility of volatility, as opposed to fat-tailed models with a GARCH update, for which the volatility of volatility is explosive. We examine in simulations the implications of such results on approximate estimation and filtering of diffusion processes. An extension to models with a time-varying conditional mean and to conditional covariance models is also developed.

Keywords: Weak diffusion limits, Score-driven models, Student-t, General error distribution

JEL codes: C10, C58

*University of Rome Tor Vergata, Via Columbia 2, 001333 Rome. E-mail: giuseppe.buccheri@uniroma2.it

†University of Pisa and City University of London, Via C. Ridolfi, 10 56124 Pisa. E-mail: fulvio.corsi@unipi.it

‡Scuola Normale Superiore, P.zza dei Cavalieri 7, 56126 Pisa. E-mail: franco.flandoli@sns.it

§Scuola Normale Superiore, P.zza dei Cavalieri 7, 56126 Pisa. E-mail: giulia.livieri@sns.it

1 Introduction

Volatility modeling and forecasting is a topic of prominent interest in theoretical and applied finance. Since the seminal works of Engle (1982) and Bollerslev (1986), GARCH-type models have become very popular in the academic community and many extensions have been proposed over the years. Recently, there was a growing interest in score-driven volatility models, introduced by Creal et al. (2013) and Harvey (2013). Score-driven models provide a general framework for volatility modeling based on the score of the conditional density. This methodology has been proved to be optimal from an information theoretic criterion (Blasques et al., 2015) and encompasses several existing models, like the GARCH model of Bollerslev (1986) and the EGARCH model of Nelson (1991). In addition, it leads to new models with a different update rule for volatility. Thanks to their generality, score-driven models have been applied successfully in many different areas (see, among others, Creal et al., 2011; Barunik et al., 2016; Harvey and Lange, 2017, 2018; Blasques et al., 2018). However, nothing has been studied on the relation between the score-driven stochastic difference equations and the continuous-time stochastic differential equations that are usually found in the theoretical finance literature. This lack of knowledge is in sharp contrast with the literature on GARCH models, for which the continuous-time limit is known since the seminal work of Nelson (1990).

It is the purpose of the present paper to partially fill this gap. We provide a set of general conditions for the weak convergence toward an Itô diffusion of a class of score-driven volatility models based on scale family conditional densities. The latter include the Student-t and the General Error distributions, which are commonly employed when modeling volatility through score-driven models (see e.g. Creal et al., 2013; Harvey, 2013). We show that the form of the coefficients characterizing the diffusion limit is determined by the link function, the conditional second moment of the score, and the scaling quantity used to normalize the score. Interestingly, it turns out that the properties of the diffusion limit are strictly entangled with those of the corresponding discrete-time process. Compared to a GARCH update, the score of a fat-tailed density is less responsive to large returns. This different behavior is reflected into the diffusion limit process. In models with a GARCH update, the volatility of volatility of the diffusion limit can diverge as the density becomes more and more fat-tailed. In contrast, in score-driven models, the volatility of volatility of the diffusion limit remains finite, even for extremely fat-tailed densities. Similar results are obtained when comparing score-driven models with an exponential link function to the EGARCH model of Nelson (1991).

The conditions guaranteeing the weak converge of score-driven volatility models to diffusions are easy to verify in practice, since they are related to the existence and finiteness of the moments of the score and of the conditional density up to a certain order. Furthermore, we need conditions guaranteeing the existence and uniqueness of the solution of the limiting stochastic differential equation. We show that these conditions are satisfied for the class of score-driven models based on scale-family conditional densities. Our results can be regarded as a generalization of the well-known diffusion limit of Nelson (1990). In the case of the normal density, we exactly recover Nelson's limit. For non-normal densities, we obtain the limiting stochastic differential equation of a class of volatility models characterized by a different update rule. Extension of our results to models with a time-varying conditional mean and to conditional covariance models is also provided.

The likelihood of discrete-time observations generated by a continuous-time process is rarely available in closed form. Common techniques employed for parameter estimation of diffusion models include simulation-based methods (Gourieroux et al., 1993; Gallant and Tauchen, 1996), generalized method of moments (Hansen and Scheinkman, 1995; Duffie and Glynn, 2004), asymptotic expansions of the transition density (Ait-Sahalia, 2002; Ait-Sahalia and Yu, 2006). Similarly, the estimation of volatility in continuous-time models is a non-linear, infinite dimensional filtering problem which often requires numerical integration (see e.g. Kitagawa, 1987). The result of Nelson (1990) suggests that discrete-time GARCH models can be regarded as diffusion approximations. This is the main idea of "Quasi Approximate Maximum Likelihood" (QAML) (Engle and Lee, 1996; Barone-Adesi et al., 2005; Fornari and Mele, 2006; Stentoft, 2011), which recovers approximate continuous-time parameters from the estimated discrete-time parameters, and of misspecified GARCH filters, whose asymptotic optimality has been studied by Nelson (1992), Nelson and Foster (1994) and Nelson (1996).

We perform an extensive Monte-Carlo study to examine the implications of the recovered continuous-time limit on approximate estimation and filtering of diffusion models. We find three main results. First, models based on fat-tailed conditional densities provide better filtered estimates of the volatility of the underlying diffusion. These models are indeed more flexible in approximating the likelihood of the continuous-time model, which is generally non-normal. Second, and more importantly, among the models based on fat-tailed conditional densities, those driven by the score are significantly less biased in QAML estimation. This is due to the particular form of their continuous-time limit, which has non-explosive volatility of volatility. Finally, in estimating the volatility of the underlying diffusion, score-driven models perform better than fat-tailed models with a GARCH update, in line with the results of analogous experiments involving discrete-time stochastic volatility models (Koopman et al., 2016).

The literature on the convergence of discrete-time Markov sequences towards diffusion processes goes back to Stroock and Varadhan (1979), Kushner (1984) and Ethier and Kurtz (1986). In 1990, Nelson established a set of conditions guaranteeing such convergence; see Nelson (1990). Precisely, the convergence as the sampling interval gets arbitrarily small, at appropriate rates, of a set of conditional moments to a well defined limit is required in

Nelson's framework. In his work, he shows that the GARCH(1,1) model of Bollerslev (1986) and the AR(1) Exponential ARCH of Nelson (1991) converge to a continuous-time diffusion. The continuous-time limit of the GARCH(1,1) has been reconsidered in Corradi (2000) under different assumptions, and in Kallsen and Taqqu (1998) using different mathematical techniques. Instead, the limit of the non-linear ARCH model of Ding et al. (1993) has been determined in Fornari and Mele (1996), whereas that of the augmented GARCH has been examined by Duan (1997). Other more recent works related to the previous ones are those of Alexander and Lazar (2005) and Trifi (2006), where the diffusion limit of the weak GARCH, of the CEV-ARCH of Fornari and Mele (1996), and of the CMSV model of Hobson and Rogers (1998) and Jeantheau (2004) is determined. In addition, Brown et al. (2003) and Wang (2002) discuss the speed of convergence to the continuous time limit of GARCH-type models and its statistical implications, whereas Klüppelberg et al. (2004) develop a class of continuous time GARCH models. The problem of temporal aggregation of GARCH models is related to their continuous-time limit, and has been studied by Drost and Nijman (1993) and Drost and Werker (1996) among others. Finally, Hafner et al. (2017) derive the weak diffusion limits of a modified version of the dynamic conditional correlation (DCC) model of Engle (2002). The work of Hafner et al. (2017) is, to the best of our knowledge, the only one in a multivariate framework.

We proceed as follows. In Section 2 we present convergence results for score-driven scale family models in a general setting. Section 3 specializes the results of the previous section to models generated by a Student-t and a General Error distribution. Section 4 presents the Monte-Carlo results. Section 5 turns to the generalization of the results in Section 2 to location-scale family models (Subsection 5.1) and to multivariate models (Subsection 5.2). Section 6 concludes. All the technical results and proofs are reported in an online appendix.

In the following sections, when we refer to Assumptions 1-4 and conditions (**A1.1**) – (**A1.3**), we tacitly refer to the corresponding assumptions and conditions in Section A of the online appendix, where we gather a set of conditions for the weak convergence of a system of discrete-time stochastic difference equations to a system of stochastic differential equations (SDEs) (see the works of Stroock and Varadhan, 1979; Kushner, 1984; Ethier and Kurtz, 1986; Nelson, 1990).

2 Dynamic scale family models

Let $\{y_t\}_{t=1}^{n}$ denote a univariate time-series of financial log-returns and let us denote by $\mathcal{G}_t = \sigma(y_1, \ldots, y_t)$, the σ-algebra generated by observations up to and including time t. We assume that $\{y_t\}_{t=1}^n$ are sampled from the following conditional density function:

$$y_t|\mathcal{G}_{t-1} \overset{d}{\sim} p\left(y_t|\mathcal{G}_{t-1}; c_t, \Theta\right) \tag{2.1}$$

$$p\left(y_t|\mathcal{G}_{t-1}; c_t, \Theta\right) = \frac{1}{\sqrt{c_t}}\Psi\left(\frac{y_t}{\sqrt{c_t}}, \Theta\right) \tag{2.2}$$

where $c_t \in \mathbb{R}^+$ is a $\mathcal{G}_{t-1}-$measurable scale parameter, Θ denotes a set of static parameters and Ψ is a probability density function. The function $p\left(y_t|\mathcal{G}_{t-1}; c_t, \Theta\right)$ belongs to the class of scale family densities, implying that $\frac{y_t}{\sqrt{c_t}}|\mathcal{G}_{t-1} \overset{d}{\sim} \Psi\left(\cdot, \Theta\right)$. In Eq. (2.2), $c_t = \Lambda\left(\lambda_t\right)$, with $\Lambda : \mathbb{R} \to \mathbb{R}^+$ being a monotonic and differentiable function of a $\mathcal{G}_{t-1}-$measurable time-varying parameter $\lambda_t \in \mathbb{R}$. As it is common in the econometric literature, the function $\Lambda\left(\cdot\right)$ is refereed to as *link function*. The time-varying parameter λ_t is computed based on observations available up to time $t-1$. In particular, it obeys the following law of motion:

$$\lambda_t = \omega + \beta\lambda_{t-1} + \alpha u_{t-1} \tag{2.3}$$

where $\omega, \beta, \alpha \in \mathbb{R}$ and $u_{t-1} = u\left(y_{t-1}, \lambda_{t-1}\right)$ is a function depending on y_{t-1} and λ_{t-1}. The model in Eq. (2.2), equipped with the update rule in Eq. (2.3), is in the class of observation-driven models (Cox, 1981), meaning that parameters are pre-determined given past observations.

Let us assume that $p\left(y_t|\mathcal{G}_{t-1}; \Lambda(\lambda_t), \Theta\right)$ is differentiable with respect to λ_t. We set the function u_{t-1} in the driving mechanism in Eq. (2.3) as in score-driven models (Creal et al., 2013; Harvey, 2013), i.e. we assume that it is proportional to the score of the conditional density:

$$u_t = s\left(\lambda_t\right)\nabla_t, \qquad \nabla_t = \frac{\partial \log p\left(y_t|\mathcal{G}_{t-1}; \Lambda(\lambda_t), \Theta\right)}{\partial \lambda_t}, \tag{2.4}$$

where $s\left(\lambda_t\right)$ is a continuous and measurable function of λ_t and Θ collects the static parameters ω, β, α and other parameters appearing in the conditional density function. Different choices of $s\left(\lambda_t\right)$ lead to different dynamic scale models. In order to account for the curvature of the log-density function, Creal et al. (2013) set $s\left(\lambda_t\right)$ equal to a power of the inverse of the conditional Fisher information $\chi_t = \mathbb{E}\left[\nabla_t^2|\mathcal{G}_{t-1}\right]$, i.e. they set $s\left(\lambda_t\right) = \chi_t^{-\varphi}$, $\varphi \in [0,1]$. Standard choices for φ are $\varphi = 0, 1/2, 1$. The main advantage of the above formulation is that the time-varying parameter λ_t is updated by taking into account the full shape of the conditional density function $p\left(y_t|\mathcal{G}_{t-1}; \Lambda(\lambda_t), \Theta\right)$ (see discussions in Creal et al., 2013; Harvey, 2013). For instance, if the Student-t density is employed, the score undermines volatility forecasts when extremely large returns are observed, since they are likely to be due to outliers rather than to large changes in volatility. A similar mechanism is absent in the t-GARCH model of Bollerslev (1987), whose update rule is the same as in the GARCH, although based on the Student-t conditional density.

Using the results in Section A of the Appendix, we now derive weak diffusion limits for the general class of score-driven scale family models described by Eq. (2.1)-(2.4). Without loss of generality, we set $y_t = \sqrt{\Lambda(\lambda_t)}\epsilon_t$, where ϵ_t is distributed according to a scale family density with scale equal to one. We can thus write the cumulative log-return process $x_t = \sum_{i=1}^{t} y_i$ as:

$$\begin{aligned} x_t &= x_{t-1} + \sqrt{\Lambda\left(\lambda_t\right)}\,\epsilon_t && (2.5)\\ \lambda_{t+1} &= \omega + \beta\lambda_t + \alpha s\left(\lambda_t\right)\nabla_t && (2.6)\end{aligned}$$

In order to exploit the results in Section A of the Appendix, we assume a timestamp of length h and allow the static parameters in Eq. (2.3) to depend on h. Formally, for $k \in \mathbb{N}$, we write:

$$\begin{aligned} x_{kh}^{(h)} &= x_{(k-1)h}^{(h)} + \sqrt{\Lambda\left(\lambda_{kh}\right)}\epsilon_{kh}^{(h)} && (2.7)\\ \lambda_{(k+1)h}^{(h)} &= \omega_h + \beta_h\lambda_{kh}^{(h)} + \alpha_h s\left(\lambda_{kh}^{(h)}\right)\nabla_{kh}^{(h)} && (2.8)\end{aligned}$$

and:

$$\mathbb{P}\left[\left(x_0^{(h)}, \lambda_0^{(h)}\right) \in \Gamma\right] = \nu_h\left(\Gamma\right) \quad \text{for any} \quad \Gamma \in \mathcal{B}\left(\mathbb{R}^2\right), \qquad (2.9)$$

where $\left\{\epsilon_{kh}^{(h)}\right\}$ has scale $\sqrt{h}$ and ν_h is a sequence of probability measures on $(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right))$ satisfying Assumption 3. Together with the discrete-time processes $\left(x_{kh}^{(h)}, \lambda_{kh}^{(h)}\right)$, we consider the continuous-time processes $\left(x_t^{(h)}, \lambda_t^{(h)}\right)$ constructed as specified in Appendix, Section A, i.e. $x_t^{(h)} = x_{kh}^{(h)}$ and $\lambda_t^{(h)} = \lambda_{kh}^{(h)}$ for $kh \leq t < (k+1)\,h$.

Let $\mathcal{F}_{kh}^{(h)} = \sigma\left(x_0^{(h)}, \ldots, x_{(k-1)h}^{(h)}\right)$ denote the σ-algebra generated by observations of the discrete-time process defined by Eq. (2.7)-(2.9). We aim to determine under which conditions the continuous-time process $\left(x_t^{(h)}, \lambda_t^{(h)}\right)$ converges in distribution to an Itô process as h goes to zero. Let

$$\nabla_{kh}^{(h)} = \frac{\partial \log p\left(y_{kh}^{(h)}|\mathcal{F}_{kh}^{(h)}; \Lambda\left(\lambda_{kh}^{(h)}\right), \Theta\right)}{\partial\lambda_{kh}^{(h)}} \qquad (2.10)$$

be the score of the conditional density of log-returns $y_{kh}^{(h)} = x_{kh}^{(h)} - x_{(k-1)h}^{(h)}$ computed with respect to the time-varying parameter $\lambda_{kh}^{(h)}$. We denote by $\xi_{kh}^{(\ell)} = \mathbb{E}\left[\left(\nabla_{kh}^{(h)}\right)^{\ell}|\mathcal{F}_{kh}^{(h)}\right]$, $\ell \in \mathbb{N}$ the conditional moments of order ℓ of the score. Let z be a standardized random variable with p.d.f. Ψ and denote by $\zeta^{(\ell)} = \mathbb{E}\left[z^{\ell}\right]$, $\ell \in \mathbb{N}$ its moment of order ℓ. The following theorem, which is the main result of this section, shows that convergence in distribution is attained provided that the parameters $\omega_h, \beta_h, \alpha_h$ are well-behaved as h goes to zero, and the moments $\xi_{kh}^{(\ell)}$, $\zeta^{(\ell)}$ exist and are finite up to the fourth order.

Theorem 2.1. *Under the assumption that:*

$$\lim_{h\to 0} h^{-1}\omega_h = \omega \tag{2.11}$$
$$\lim_{h\to 0} h^{-1}(1-\beta_h) = \theta \tag{2.12}$$
$$\lim_{h\to 0} h^{-1}\alpha_h^2 = \alpha^2 \tag{2.13}$$

where ω, θ, $\alpha \in \mathbb{R}$, and under the assumption that the moments $\xi_{kh}^{(\ell)}$, $\zeta^{(\ell)}$ exist and are finite for $\ell \leq 4$, the continuous-time process $\left(x_t^{(h)}, \lambda_t^{(h)}\right)$ converges in distribution to the following Itô process as h goes to zero:

$$dx_t = \sqrt{\Lambda(\lambda_t)\,\zeta^{(2)}}\,dW_t^{(1)} \tag{2.14}$$
$$d\lambda_t = (\omega - \theta\lambda_t)\,dt + \alpha s(\lambda_t)\sqrt{\chi(\lambda_t)}\,dW_t^{(2)} \tag{2.15}$$

and

$$\mathbb{P}\left[(x_0, \lambda_0) \in \Gamma\right] = \nu_0(\Gamma) \quad \textit{for any} \quad \Gamma \in \mathcal{B}\left(\mathbb{R}^2\right) \tag{2.16}$$

where $\chi(\lambda_t) = \lim_{h\to 0}\xi_{kh}^{(2)}$, and $W_t^{(1)}$, $W_t^{(2)}$ are independent standard Brownian motions, independent of the initial values (x_0, λ_0).

The proof of the previous theorem consists of two main steps. In the first step, we show that, for the general class of scale family densities considered here, the conditional moments $\xi_{kh}^{(\ell)}$ do not depend on h; see Theorem 2.2. This result greatly simplifies the computation of the limits in (**A1.1**), (**A1.2**) and (**A1.3**). Indeed, if the conditional moments $\xi_{kh}^{(\ell)}$ are independent of h, they can be regarded as constants and the asymptotic behavior of the expressions in (**A1.1**), (**A1.2**) and (**A1.3**) is only determined by that of the static parameters $\omega_h, \beta_h, \alpha_h$. In addition, if $\xi_{kh}^{(\ell)}$ are independent of h, the limit of $\xi_{kh}^{(2)}$, which appears in the limiting SDE, trivially exists and is finite under the assumptions of Theorem 2.1. The second step consists in verifying that, under the hypothesis of Theorem (2.1), Assumptions 1-4 hold.

Theorem 2.2. *For the class of conditional scale family densities*

$$p\left(y_{kh}^{(h)} | \mathcal{F}_{(k-1)h}^{(h)}; \Lambda\left(\lambda_{kh}^{(h)}\right), \Theta\right) = \frac{1}{\sqrt{\Lambda\left(\lambda_{kh}^{(h)}\right)h}}\Psi\left(\frac{y_{kh}^{(h)}}{\sqrt{\Lambda\left(\lambda_{kh}^{(h)}\right)h}}, \Theta\right) \tag{2.17}$$

the moments $\xi_{kh}^{(\ell)} = \mathbb{E}\left[\left(\nabla_{kh}^{(h)}\right)^{\ell} | \mathcal{F}_{kh}^{(h)}\right]$ *are given by:*

$$\xi_{kh}^{(\ell)} = (-1)^{\ell} \left[\frac{1}{2}\frac{\Lambda'\left(\lambda_{kh}^{(h)}\right)}{\Lambda\left(\lambda_{kh}^{(h)}\right)}\right]^{\ell} \int_{-\infty}^{+\infty} \left(1 + \frac{\Psi'(z,\Theta)}{\Psi(z,\Theta)} z\right)^{\ell} \Psi(z,\Theta)\, dz \tag{2.18}$$

where $\Lambda'\left(\lambda_{kh}^{(h)}\right) = \frac{\partial \Lambda\left(\lambda_{kh}^{(h)}\right)}{\partial \lambda_{kh}^{(h)}}$ *and* $\Psi'(z) = \frac{\partial \Psi(z,\Theta)}{\partial z}$.

Proof. See Appendix C.1. □

Thanks to the result in Theorem 2.2, the asymptotic behavior of the expectations in (**A1.1**), (**A1.2**) and (**A1.3**) is only determined by the parameters $\omega_h, \beta_h, \alpha_h$. In what follows, we report explicit expressions for the limits of these expectations, computed under the hypothesis of Eq. (2.11)-(2.13). Detailed derivations of the latter are reported in Appendix C.2.

The drift per unit of time (Condition **A1.1**) is given by:

$$h^{-1}\mathbb{E}\left[\left(x_{kh}^{(h)} - x_{(k-1)h}^{(h)}\right) | \mathcal{F}_{kh}^{(h)}\right] = 0 \tag{2.19}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda_{(k+1)h}^{(h)} - \lambda_{kh}^{(h)}\right) | \mathcal{F}_{kh}^{(h)}\right] = \omega - \theta\lambda_{kh}^{(h)} \tag{2.20}$$

Note that, at this stage, the parameters ω and θ can vary over $\mathbb{R}$. Constraints guaranteeing that the scale remains positive with probability one depend on the link function and will be discussed in specific cases. The second moments per unit of time (Condition **A1.2**) are:

$$h^{-1}\mathbb{E}\left[\left(x_{kh}^{(h)} - x_{(k-1)h}^{(h)}\right)^2 | \mathcal{F}_{kh}^{(h)}\right] = \Lambda\left(\lambda_{kh}^{(h)}\right)\zeta^{(2)} \tag{2.21}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda_{(k+1)h}^{(h)} - \lambda_{kh}^{(h)}\right)^2 | \mathcal{F}_{kh}^{(h)}\right] = \alpha^2 s\left(\lambda_{kh}^{(h)}\right)^2 \xi_{kh}^{(2)} + o(1) \tag{2.22}$$

$$h^{-1}\mathbb{E}\left[\left(x_{kh}^{(h)} - x_{(k-1)h}^{(h)}\right)\left(\lambda_{(k+1)h}^{(h)} - \lambda_{kh}^{(h)}\right) | \mathcal{F}_{kh}^{(h)}\right] = 0 \tag{2.23}$$

As the function $s\left(\lambda_{kh}^{(h)}\right)$ is typically related to the inverse of $\xi_{kh}^{(2)}$, the expression in Eq. (2.22) is finite provided that the assumptions of Theorem 2.1 are satisfied. Finally, we need to prove that Condition (**A1.3**) holds. As in Nelson (1990), we choose $\delta = 2$ and set $\omega_h = h\omega$, $\alpha_h = h^{1/2}\alpha$ and $\beta_h = 1 - h\theta$. We obtain the following expressions for the fourth moments:

$$h^{-1}\mathbb{E}\left[\left(x_{kh}^{(h)} - x_{(k-1)h}^{(h)}\right)^4 | \mathcal{F}_{kh}^{(h)}\right] = h\Lambda\left(\lambda_{kh}^{(h)}\right)^2 \zeta^{(4)} \tag{2.24}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda_{kh}^{(h)} - \lambda_{(k-1)h}^{(h)}\right)^4 | \mathcal{F}_{kh}^{(h)}\right] = h\,\alpha^4 s\left(\lambda_{kh}^{(h)}\right)^4 \xi_{kh}^{(4)} + O\left(h^{\gamma}\right), \quad \gamma \geq 3/2. \tag{2.25}$$

As $\zeta^{(4)}$, $\xi_{kh}^{(4)}$ are finite by assumption, we conclude that both expressions converge to zero as h goes to zero. Thus, Assumption 1 is satisfied and the two coefficients $a(x,t)$, $b(x,t)$ are given by:

$$b\left(x,\lambda\right) \equiv \begin{bmatrix} 0 \\ \omega-\theta\lambda \end{bmatrix} \tag{2.26}$$

$$a\left(x,\lambda\right) = \begin{bmatrix} \Lambda\left(\lambda\right)\zeta^{(2)} & 0 \\ 0 & \alpha^2 s\left(\lambda\right)^2\chi(\lambda) \end{bmatrix} \tag{2.27}$$

Assumption 2 holds by setting $\sigma\left(x,\lambda\right)$ equal to the element-by-element square root of $a\left(x,\lambda\right)$. Note that the form of the coefficients of the limiting SDE is fully determined by the link function $\Lambda(\lambda)$, the conditional second moment of the score $\chi(\lambda)$, and the scaling quantity $s(\lambda)$ used to normalize the score. Note that, if $p\left(y_{kh}^{(h)}|\mathcal{F}_{(k-1)h}^{(h)};\Lambda\left(\lambda_{kh}^{(h)}\right),\Theta\right)$ is the normal and $\Lambda\left(\lambda_{kh}^{(h)}\right)=\lambda_{kh}^{(h)}$, Eq. (2.26), (2.27) reduce to the well-known GARCH diffusion limit of Nelson (1990). We report additional comments related to the result in Theorem 2.1 in the following two remarks.

Remark 1. *In the present paper, we assume, as in Nelson (1990), that the rate of convergence of* α_h *is* $h^{1/2}$. *Corradi (2000) recovered a weak diffusion limit for the GARCH by assuming a rate of convergence* h. *In the first case, one obtains a stochastic volatility model driven by two independent Brownian motions, whereas in the second case the continuous-time limit is a deterministic variance model. The limit result of Corradi (2000) has the advantage of preserving the number of the sources of randomness in the transition from the discrete-time process to the continuous-time process. However, the conditional volatility of volatility is zero. In this paper, we use the same rate of convergence of Nelson (1990) in all our computations.*

Remark 2. *The Euler-Maruyama (EM, henceforth) discretization of the SDE in Eq.* (2.14)-(2.15) *leads to a discrete-time stochastic volatility model characterized by two independent sources of noise, a fact that has already been observed in the GARCH literature (see, e.g., the discussion in Section 3 of Corradi, 2000). In addition, since the EM scheme weakly converges to the limiting SDE, it has by construction as limit model the one given in Eq.* (2.14)-(2.15). *Therefore, the EM discretization and the dynamic scale family models in Theorem 2.1 converge weakly to the a SDE of similar form, with the same order of convergence. They can thus be used interchangeably to simulate the limiting SDE. The fact that the two discrete-time models above converge to the same SDE may seem unnatural because of the different number of independent noises in the two models; however, in Appendix B, we provide a simple, although representative for our results, example showing that two independent Brownian motions may be obtained as weak limit from a single sequence of independent random variables in a very natural way.*

To conclude the proof of Theorem 2.1, we need to discuss the issues of finiteness of the process in finite intervals and uniqueness of the diffusion limit; see Assumption 4. The answer to these issues depends on the growth and regularity of the two coefficients $b(x, \lambda)$ and $a(x, \lambda)$ in Eq. (2.26), (2.27), as well as on the link function $\Lambda(\cdot)$. The next subsection specializes the general formulation in Eq. (2.5), (2.6) to two classes of models. More precisely, in the first class, the link function is the identity, i.e. $\Lambda(\lambda_t) = \lambda_t$, and we set $\lambda_t = \sigma_t^2$. In this class of models, the function $s(\sigma_t^2)$ is set equal to the inverse of the Fisher information matrix, and the product $s(\sigma_t^2)^2 \chi(\sigma_t^2)$ appearing in the coefficient $a(x, \sigma_t^2)$ turns out to be proportional to σ_t^4. In the second class of models, the link function is exponential, i.e. we set $\Lambda(\lambda_t) = \exp(2\lambda_t)$. In this case, the product $s(\lambda_t)^2 \chi(\lambda_t)$ turns out to be constant. These two classes include, but are not limited to, important examples of models used in the score literature, namely the t-GAS or Beta-t-GARCH model of Creal et al. (2013) and Harvey (2013), and the Beta-t-EGARCH model of Harvey (2013).

The following two theorems show that, in both classes, there is a unique adapted solution to Eq. (2.14), (2.15) and provide an explicit formula for it[1]. In particular, it turns out that the solution of the SDE is finite almost surely in compact sets.

Theorem 2.3 (Existence and uniqueness, $\Lambda(\lambda_t) = \lambda_t$). *For the following system of SDE:*

$$\begin{aligned} dx_t &= C_1(\Theta)\, \sigma_t dW_t^{(1)} \\ d\sigma_t^2 &= \left(\omega - \theta\, \sigma_t^2\right) dt + C_2(\Theta)\, \sigma_t^2 dW_t^{(2)}, \qquad \left(x_0, \sigma_0^2\right) = \left(x_0, \sigma_0^2\right) \end{aligned} \tag{2.28}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent Brownian motions independent of the initial values (x_0, σ_0^2), $\omega \geq 0$, $\theta \in \mathbb{R}$ constant parameters, and $C_1(\Theta)$ and $C_1(\Theta)$ positive constant functions (possibly) dependent on the vector of static parameters Θ, there is one and only one continuous adapted solution (x_t, σ_t^2) which is given by the following explicit formula:

$$\begin{aligned} x_t &= x_0 + \int_0^t e^{-\frac{1}{2}\left(\theta + \frac{1}{2} C_2(\Theta)^2\right)s + \frac{1}{2} C_2(\Theta) W_s^{(2)}} \sqrt{\left(\sigma_0^2 + \omega \int_0^s e^{\left(\theta + \frac{1}{2} C_2(\Theta)^2\right)r - C_2(\Theta) W_r^{(2)}} dr\right)} dW_s^{(1)} \\ \sigma_t^2 &= e^{-\left(\theta + \frac{1}{2} C_2(\Theta)^2\right)t + C_2(\Theta) W_t^{(2)}} \left(\sigma_0^2 + \omega \int_0^t e^{\left(\theta + \frac{1}{2} C_2(\Theta)^2\right)s - C_2(\Theta) W_s^{(2)}} ds\right) \end{aligned}$$

Proof. See Appendix C.3. □

[1] In principle, there are other possible choices for the link function and for the scaling quantity used to normalize the score. While the convergence result in Theorem 2.1 is general, the problem of showing the existence and uniqueness of the solution of the SDE depends on the form of its coefficients. For choices of $\Lambda(\lambda_t)$ and $s(\lambda_t)$ which lead to SDE's of different forms, we refer the reader to standard textbooks in stochastic analysis, e.g. Karatzas and Shreve (1991) and Protter (1992).

Theorem 2.4 (Existence and uniqueness, $\Lambda(\lambda_t) = e^{2\lambda_t}$). *For the following system of SDE:*

$$
\begin{aligned}
dx_t &= C_1\left(\Theta\right)\exp\left(\lambda_t\right)dW_t^{(1)} \\
d\lambda_t &= \left(\omega - \theta\,\lambda_t\right)dt + C_2\left(\Theta\right)dW_t^{(2)}, \qquad \left(x_0, \lambda_0\right) = \left(x_0, \lambda_0\right)
\end{aligned} \tag{2.29}
$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent Brownian motions independent of the initial values (x_0, σ_0^2), $\omega, \theta \in \mathbb{R}$ constant parameters, and $C_1\left(\Theta\right)$ and $C_2\left(\Theta\right)$ positive constant functions (possibly) dependent on the vector of static parameters Θ, there is one and only one continuous adapted solution (x_t, λ_t) which is given by the explicit formula:

$$
\begin{aligned}
x_t &= x_0 + C_1\left(\Theta\right)\int_0^t \exp\left(e^{-\theta s}\left(\Lambda_0 + \frac{\omega}{\theta}\left(1 - e^{\theta s}\right) + C_2\left(\Theta\right)\int_0^s e^{\theta r}dW_r^{(2)}\right)\right)dW_s^{(1)} \\
\lambda_t &= e^{-\theta t}\left(\lambda_0 + \frac{\omega}{\theta}\left(1 - e^{\theta t}\right) + C_2\left(\Theta\right)\int_0^t e^{\theta s}dW_s^{(2)}\right)
\end{aligned}
$$

Proof. See Appendix C.3. □

We point out that the framework of Nelson (1990) implicitly assumes that the update of volatility is of the same form for any data frequency. Drost and Nijman (1993) prove that the GARCH, in its strong form, is not invariant under temporal aggregation. They show that it is necessary to introduce the larger class of weak GARCH models to achieve closure under temporal aggregation. In proving their result, Drost and Nijman (1993) exploit the fact that the squared returns in a GARCH model follow an ARMA process. The temporal aggregation problem for GARCH is thus mapped into an equivalent problem for ARMA models, which is well-established in the literature. The scaled score of a non-Gaussian conditional density is generally different from the innovation process $y_t^2 - \sigma_t^2$ in the ARMA representation of GARCH models. Typically, it is a non-linear function of y_t^2 and σ_t^2; see e.g. Eq. (3.3). Contrary to the GARCH, the squared returns are not described by an ARMA process, and it is not possible to use the argument of Drost and Nijman (1993) to construct a larger class of models closed under temporal aggregation. However, score-driven volatility models can be regarded as particular instances of the class of square-root stochastic autoregressive volatility (SR-SARV) models of Meddahi and Renault (2004); see their Definition 2.1. Indeed, as the conditional expectation of the score is zero by construction, the variance process is a VAR(1) with respect to the filtration generated by past returns. The class of SR-SARV models is an extension of the class of weak GARCH models, and includes both GARCH-type (e.g. GARCH with leverage and skewness) and stochastic volatility models. In particular, Meddahi and Renault (2004) show that the process resulting from the temporal aggregation of a SR-SARV process is still a SR-SARV process; see their Proposition 2.2. This result implies that volatility models driven by the score of the conditional density aggregate in the class of SR-SARV process. To enclose the class of score-driven volatility models under temporal aggregation, it is thus necessary to consider a more general class of models, wider than the class of weak GARCH's. Essentially, this is due to the non-linear structure of the score, which differs from the simple GARCH innovation process.

3 Models based on Student-t and General Error distribution

In this section, we show some examples of weak diffusion limits of dynamic volatility models driven by the score of the conditional density. As in Creal et al. (2013) and Harvey (2013), we focus on models generated by a Student-t and a General Error distribution. We distinguish the class of models for which the link function is the identity, i.e. $\Lambda(\lambda_t) = \lambda_t$, and the class of models for which the link function is exponential, i.e. $\Lambda(\lambda_t) = e^{2\lambda_t}$. As underlined above, these two classes include important examples of volatility models used in the score literature. Using the same nomenclature in Harvey (2013), such models are dubbed Beta-t-GARCH and Gamma-GED-GARCH, or Beta-t-EGARCH and Gamma-GED-EGARCH, depending on the type of link-function. To avoid confusion with models with a GARCH update, like the t-GARCH of Bollerslev (1987) or the EGARCH model of Nelson (1991), we adopt here a slightly different nomenclature. If $\Lambda(\lambda_t) = \lambda_t$, we refer to them as Beta-t and Gamma-GED, whereas if $\Lambda(\lambda_t) = e^{2\lambda_t}$, we refer to them as Beta-t-E and Gamma-GED-E. Harvey (2013) discusses the statistical properties of these models and provides sufficient conditions for achieving consistency and asymptotic normality of the maximum likelihood estimator in the Beta-t-E and Gamma-GED-E.

Despite being characterized by different conditional densities and by different update rules, these models converge to a limiting SDE of similar form. For instance, both Beta-t and Gamma-GED converge to a continuous-time limit which has a form similar to that of the well-known GARCH diffusion. However, compared to Nelson's limit, models generated by fat-tails densities give rise to an SDE with lower volatility of volatility. This is due to the damping mechanism of the score, which undermines volatility forecasts in the presence of large returns, and generates less erratic volatility paths. Similarly, models generated by light-tails densities lead to an SDE with larger volatility of volatility. In this case the score overreacts to large returns and, in turn, generates more erratic volatility paths.

It is interesting to note that the opposite behavior is observed in models generated by non-normal densities but with a GARCH update. As it will be shown, the limiting SDE of the t-GARCH model of Bollerslev (1987) is characterized by a volatility of volatility which diverges as the number of degrees of freedom ν decreases. This is due to the absence of the above-mentioned damping mechanism. Similarly, the volatility of volatility of the limiting SDE of a GARCH with a GED density diverges when the distribution is fat-tailed, and becomes finite - but lower than that of the Gamma-GED diffusion - when the distribution is light-tailed. Similar results are recovered when examining models with an exponential link function.

We point out that the different behavior of the diffusive coefficient generated by score-driven models is likely to hold even for models constructed with a different choice for the link function and/or for the scaling quantity. Essentially, these results come from the computation of the conditional second moment of the process, which in turn relies on the particular structure of the score. As far as the score is different from the innovation process $y_t^2 - \Lambda(\lambda_t)$ that we have in the GARCH update - and this is generally true for non-normal models - the limiting SDE in the two models has a different diffusion coefficient. The latter is thus less affected by outliers when the distribution is fat-tailed and more affected when the distribution is light-tailed. Here, we limit ourselves to illustrate this result for a specific although representative class of score-driven models.

3.1 Case 1: $\Lambda(\lambda_t) = \lambda_t$

We set $\lambda_t = \sigma_t^2$. The general framework developed in Section 2 enables us to determine the coefficients of the limiting SDE by computing the function $s\left(\sigma_t^2\right)$, the conditional second moment $\chi(\sigma_t^2)$ of the score, and the second moment $\zeta^{(2)}$ of the standardized variable z. Weak existence and uniqueness, as well as the finiteness of the solution of the limiting SDE in compact sets, are guaranteed by Theorem 2.3. Since the moments of the score are independent of the scale parameter h, we consider innovations of unit scale in the measurement equation. This choice does not affect the form of the limiting SDE.

3.1.1 Beta-t

Let t_ν denote a standardized Student-t density and assume[2] $\nu > 4$. Consider the model:

$$y_t = \sigma_t \epsilon_t \tag{3.1}$$

$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\sigma_t^2 u_t \tag{3.2}$$

where $\epsilon_t = \left((\nu-2)/\nu\right)^{1/2} \tilde{\epsilon}_t$, $\tilde{\epsilon}_t \overset{d}{\sim} t_\nu$, $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and u_t is given by:

$$u_t = \left[\frac{(\nu+1)\, y_t^2}{(\nu-2)\,\sigma_t^2 + y_t^2} - 1\right] \tag{3.3}$$

[2]Such assumption is necessary to guarantee the existence of the conditional moments in Assumption 1. In real data the estimates of ν in score models based on Student-t density are typically slightly above 4; see for instance Creal et al. (2011).

The conditional density function is:

$$p\left(y_t|\mathcal{G}_{t-1};\sigma_t^2,\nu\right)=\frac{\Gamma(\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\left(\nu-2\right)\sigma_t^2}}\left[1+\frac{y_t^2}{\left(\nu-2\right)\sigma_t^2}\right]^{-\frac{(\nu+1)}{2}} \tag{3.4}$$

Setting $z_t=y_t/\sqrt{\sigma_t^2}$, we have $\Psi\left(z_t,\nu\right)=\frac{\Gamma(\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left[1+\frac{z_t^2}{\nu-2}\right]^{-\frac{(\nu+1)}{2}}$ and therefore:

$$\nabla_t=\frac{1}{2\sigma_t^2}\left[\frac{\left(\nu+1\right)y_t^2}{\left(\nu-2\right)\sigma_t^2+y_t^2}-1\right] \tag{3.5}$$

$$\chi\left(\sigma_t^2\right)=\frac{1}{2\sigma_t^4}\frac{\nu}{\nu+3} \tag{3.6}$$

$$\zeta^{(2)}=1 \tag{3.7}$$

The scaling quantity is $s\left(\sigma_t^2\right)=2\sigma_t^4$, and is proportional to the inverse of the Fisher information. Note also that the two fourth moments $\xi_{kh}^{(4)}$ and $\zeta^{(4)}$ exist and are finite under the assumption $\nu>4$. Using equations (2.26), (2.27), the limiting SDE of the Beta-t is:

$$dx_t=\sigma_t dW_t^{(1)} \tag{3.8}$$

$$d\sigma_t^2=\left(\omega-\theta\sigma_t^2\right)\,dt+\alpha\sqrt{2\left(\frac{\nu}{\nu+3}\right)}\sigma_t^2 dW_t^{(2)} \tag{3.9}$$

If $\nu\to\infty$, we obtain a result similar to that of Nelson (1990). The only difference is that the drift component in the log-price is zero, as we did not include for simplicity the conditional mean in the measurement equation. In Section 5.1, we consider the case of models with a time-varying conditional mean. Furthermore, the presence of the factor $\sqrt{2}$ is due to the slightly different choice of α_h, which is set as $\alpha_h=\alpha h^{1/2}$, whereas in Nelson (1990) it is set as $\alpha_h=\alpha\left(h/2\right)^{1/2}$. When ν is finite, the diffusion coefficient in Eq. (3.9), the so-called volatility of volatility, turns out to be smaller than that of the well-known GARCH diffusion, implying that the dynamics of volatility are less erratic. This is due to the fact that the score of the Student-t density is less sensitive to large returns.

It is interesting to investigate whether a similar result is found when computing the diffusion limit of the t-GARCH model of Bollerslev (1987). The latter is characterized by a Student-t conditional density, but the update rule for volatility is the same as in the GARCH. The t-GARCH reads:

$$y_t = \sigma_t\epsilon_t \tag{3.10}$$

$$\sigma_{t+1}^2 = \omega+\beta\,\sigma_t^2+\alpha\,\sigma_t^2\epsilon_t^2 \tag{3.11}$$

where $\epsilon_t=\left((\nu-2)/\nu\right)^{1/2}\tilde{\epsilon}_t$, $\tilde{\epsilon}_t\overset{d}{\sim}t_\nu$, $\omega>0$, $\alpha\geq0$, $\beta\geq0$. Furthermore, we assume $\nu>4$.

We work in the same framework of Section 2, i.e. we partition the time on a grid of length h, then construct the continuous-time process $\left(x_t^{(h)}, \sigma_t^{(h),2}\right)$ as described in Section A of the Appendix and allow the parameters ω_h, β_h, α_h to depend on h. Using the results in Section A, in Appendix D.1 we prove the following:

Theorem 3.1. *Under the assumption that:*

$$\lim_{h\to 0} h^{-1}\omega_h = \omega \tag{3.12}$$

$$\lim_{h\to 0} h^{-1}\left(1-\beta_h-\alpha_h\right) = \theta \tag{3.13}$$

$$\lim_{h\to 0} h^{-1}\alpha_h^2 = \alpha^2 \tag{3.14}$$

where $\omega > 0$, $\theta > 0$, $\alpha \geq 0$, *the continuous-time process* $\left(x_t^{(h)}, \sigma_t^{(h),2}\right)$ *constructed from a* t-GARCH *converges in distribution to the following Itô process as h goes to zero:*

$$dx_t = \sigma_t dW_t^{(1)} \tag{3.15}$$

$$d\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right)dt + \alpha\sqrt{2\left(\frac{\nu-1}{\nu-4}\right)}\sigma_t^2 dW_t^{(2)}, \tag{3.16}$$

where $W_t^{(1)}$, $W_t^{(2)}$ *are independent standard Brownian motions, independent of the initial values* (x_0, σ_0^2).

As in the Beta-t, if $\nu \to \infty$, we recover Nelson's limit. However, if ν is finite, the volatility of volatility of the t-GARCH diffusion is larger than that of the Beta-t. Figure (3.1.1) plots the diffusion coefficient $f_1(\nu) = \sqrt{2\left(\frac{\nu}{\nu+3}\right)}$ of the Beta-t and the diffusion coefficient $f_2(\nu) = \sqrt{2\left(\frac{\nu-1}{\nu-4}\right)}$ of the t-GARCH as a function of ν. As $\nu \downarrow 4$, the volatility of volatility of the t-GARCH diffusion diverges. This happens because outliers are entirely imputable to large changes in volatility. In contrast, as $\nu \downarrow 4$, the volatility of volatility of the Beta-t diffusion is finite. Outliers in the Beta-t are indeed more likely to be generated by the conditional density, rather than by large changes in the underlying volatility process.

3.1.2 Gamma-GED

Let GED(ν) denote a Generalized Error Distribution with shape parameter $\nu > 0$. Consider the model:

$$y_t = \sigma_t \epsilon_t \tag{3.17}$$

$$\sigma_{t+1}^2 = \omega + \beta\,\sigma_t^2 + \alpha\sigma_t^2 u_t \tag{3.18}$$

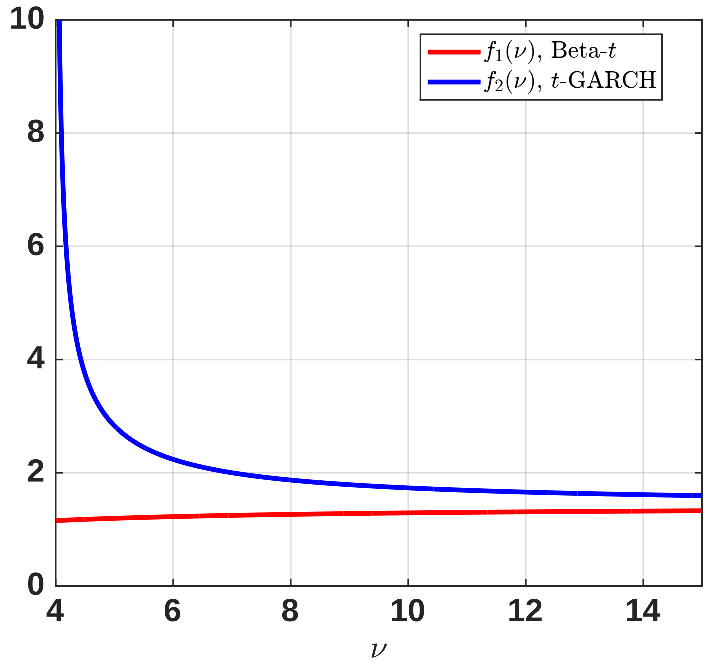

Figure 1: We report the diffusion coefficient $f_1(\nu) = \sqrt{2\left(\frac{\nu}{\nu+3}\right)}$ of the Beta-t and the diffusion coefficient $f_2(\nu) = \sqrt{2\left(\frac{\nu-1}{\nu-4}\right)}$ of the t-GARCH as a function of ν.

where $\epsilon_t \overset{d}{\sim} \text{GED}(\nu)$, $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and u_t is given by:

$$u_t = \frac{\nu}{2}\left|\frac{y_t}{\sqrt{\sigma_t^2}}\right|^{\nu} - 1 \tag{3.19}$$

The conditional density function is:

$$p\left(y_t|\mathcal{G}_{t-1};\sigma_t^2,\nu\right) = \frac{1}{2^{1+\frac{1}{\nu}}\sqrt{\sigma_t^2}\Gamma\left(1+\frac{1}{\nu}\right)}\exp\left(-\frac{1}{2}\left|\frac{y_t}{\sqrt{\sigma_t^2}}\right|^{\nu}\right) \tag{3.20}$$

Setting $z_t = y_t/\sqrt{\sigma_t^2}$, we have $\Psi\left(z_t,\nu\right) = \frac{1}{2^{1+\frac{1}{\nu}}\Gamma\left(1+\frac{1}{\nu}\right)}\exp\left(-\frac{1}{2}\left|z_t\right|^{\nu}\right)$ and therefore:

$$\nabla_t = \frac{1}{2\sigma_t^2}\left[\frac{\nu}{2}\left|\frac{y_t}{\sqrt{\sigma_t^2}}\right|^{\nu} - 1\right] \tag{3.21}$$

$$\chi\left(\sigma_t^2\right) = \frac{\nu}{4\sigma_t^4} \tag{3.22}$$

$$\zeta^{(2)} = \frac{4^{1/\nu}\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} \tag{3.23}$$

from which we have $s\left(\sigma_t^2\right) = 2\sigma_t^4$. Even in this case, the scaling quantity $s\left(\sigma_t^2\right)$ is related to the inverse of the Fisher information. Note also that the two fourth moments $\xi_{kh}^{(4)}$ and

$\zeta^{(4)}$ exist and are finite. By virtue of equations (2.26), (2.27), the limiting SDE of the Gamma-GED is:

$$dx_t = \sqrt{\zeta^{(2)}}\sigma_t dW_t^{(1)} \tag{3.24}$$
$$d\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right) dt + \alpha\sqrt{\nu}\sigma_t^2 dW_t^{(2)} \tag{3.25}$$

If $\nu = 2$, the GED(ν) reduces to the normal and we recover Nelson's limit. If $\nu < 2$, the GED(ν) has heavier tails than the normal. As in the Beta-t, the dynamics of volatility are less sensitive to large returns, and the volatility of volatility is smaller compared to the GARCH diffusion. In contrast, if $\nu > 2$, the GED(ν) has lighter tails than the normal. In this case, the score overreacts to large returns and the volatility of volatility is larger compared to the GARCH diffusion.

As done with the Beta-t, we compare the limiting SDE of the Gamma-GED with that of a GARCH model with a GED conditional density. We refer to the latter as GED-GARCH. The GED-GARCH reads:

$$y_t = \sigma_t \epsilon_t \tag{3.26}$$
$$\sigma_{t+1}^2 = \omega + \beta\,\sigma_t^2 + \alpha\sigma_t^2 u_t \tag{3.27}$$

where $\epsilon_t \overset{d}{\sim} \text{GED}\,(\nu)$, $\omega > 0$, $\beta \geq 0$, $\alpha \geq 0$. In Appendix D.2, we prove the following:

Theorem 3.2. *Let ω_h, β_h, α_h be as in Theorem 3.1. Then the continuous-time process $\left(x_t^{(h)}, \sigma_t^{(h),2}\right)$ constructed from a* GED-GARCH *converges in distribution to the following Itô process as h goes to zero:*

$$dx_t = \sqrt{\zeta^{(2)}}\sigma_t dW_t^{(1)} \tag{3.28}$$
$$d\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right) dt + \alpha \left[\frac{\Gamma\left(\frac{1}{\nu}\right) + 4^{\frac{1}{\nu}}\left(4^{\frac{1}{\nu}}\Gamma\left(\frac{5}{\nu}\right) - 2\,\Gamma\left(\frac{3}{\nu}\right)\right)}{\Gamma\left(\frac{1}{\nu}\right)}\right]^{1/2} dW_t^{(2)} \tag{3.29}$$

where $W_t^{(1)}$, $W_t^{(2)}$ are independent standard Brownian motions, independent of the initial values (x_0, σ_0^2), and $\Gamma(\cdot)$ denotes the Gamma function.

As in the Gamma-GED, if $\nu = 2$, we recover Nelson's limit. Figure (3.1.2) compares the diffusion coefficient $g_1(\nu) = \sqrt{\nu}$ of the Gamma-GED to the diffusion coefficient $g_2(\nu)$ of the GED-GARCH in Eq. (3.29). If $\nu < 2$, the latter is larger than $g_1(\nu)$ and diverges as $\nu \downarrow 0$. This circumstance is similar to that in Section 3.1.1, as the GED(ν) has heavier tails than the normal. In contrast, if $\nu > 2$, the GED(ν) has lighter tails and $g_2(\nu)$ is smaller than $g_1(\nu)$. In this case, the score overreacts to large returns and the volatility of volatility of the Gamma-GED diffusion turns out to be larger than that of the GED-GARCH diffusion.

Note also that $g_2(\nu)$ converges to a finite value as ν goes to infinity, in particular we obtain $\lim_{\nu\to+\infty} g_2(\nu) = \frac{8}{15}$.

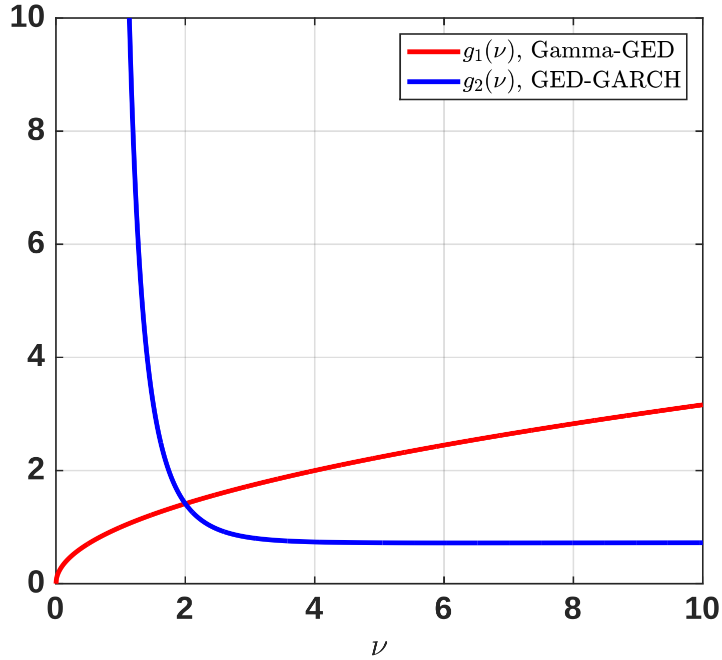

Figure 2: We report the diffusion coefficient $g_1(\nu) = \sqrt{\nu}$ of the Gamma-GED and the diffusion coefficient $g_2(\nu) = \left[\frac{\Gamma(\frac{1}{\nu})+4^{\frac{1}{\nu}}\left(4^{\frac{1}{\nu}}\Gamma(\frac{5}{\nu})-2\Gamma(\frac{3}{\nu})\right)}{\Gamma(\frac{1}{\nu})}\right]^{1/2}$ of the GED-GARCH as a function of ν.

3.2 Case 2: $\Lambda(\lambda_t) = e^{2\lambda_t}$

In this class of models the link function is the exponential, namely $\Lambda(\lambda_t) = \exp(2\lambda_t)$. As done in Section 3.1, we compute the scaling function $s(\lambda_t)$, the conditional second moment $\chi(\lambda_t)$ of the score, and the second moment $\zeta^{(2)}$ of the standardized variable z. Theorem 2.4 ensures the weak existence, the uniqueness and the finiteness of the solution of the limiting SDE on compact sets.

3.2.1 Beta-t-E

Consider the model:

$$y_t = \exp(\lambda_t)\,\epsilon_t \tag{3.30}$$

$$\lambda_{t+1} = \omega + \beta\,\lambda_t + \alpha\,u_t \tag{3.31}$$

where $\epsilon_t = ((\nu - 2)/\nu)^{1/2} \tilde{\epsilon}_t$, $\tilde{\epsilon}_t \overset{d}{\sim} t_\nu$, $\nu > 4$, $\omega, \beta, \alpha \in \mathbb{R}$, and u_t is given by:

$$u_t = \frac{(\nu+1)\, y_t^2}{(\nu-2)\exp(2\lambda_t) + y_t^2} - 1 \tag{3.32}$$

The conditional density function $p(y_t|\mathcal{G}_{t-1}; \lambda_t, \nu)$ coincides with the one in Eq. (3.4), with $\exp(2\lambda_t)$ in place of σ_t^2. Thus:

$$\nabla_t = \frac{(\nu+1)\, y_t^2}{(\nu-2)\exp(2\lambda_t) + y_t^2} - 1 \tag{3.33}$$

$$\chi(\lambda_t) = 2\frac{\nu}{\nu+3} \tag{3.34}$$

$$\zeta^{(2)} = 1 \tag{3.35}$$

In particular, $s(\lambda_t) = 1$. Note also that the two fourth moments $\xi_{kh}^{(4)}$ and $\zeta^{(4)}$ exist and are finite under the assumption $\nu > 4$. Hence, the Beta-t-E diffusion limit can be written as:

$$dx_t = \exp(\lambda_t)\, dW_t^{(1)} \tag{3.36}$$

$$d\lambda_t = (\omega - \theta\lambda_t)\, dt + \alpha\sqrt{2\left(\frac{\nu}{\nu+3}\right)} dW_t^{(2)} \tag{3.37}$$

The main difference with respect to the Beta-t limit is that the diffusive term in the log-volatility process does not depend on λ_t. The diffusion coefficient $f_1(\nu) = \sqrt{2\left(\frac{\nu}{\nu+3}\right)}$ is the same as in the Beta-t. In the Gaussian limit, it reduces to $\sqrt{2}$, which is larger than $f_1(\nu)$. As found for the Beta-t, this is due to the particular form of the score, which takes into account the shape of the Student-t density.

3.2.2 Gamma-GED-E

Consider the model:

$$y_t = \exp(\lambda_t)\, \epsilon_t \tag{3.38}$$

$$\lambda_{t+1} = \omega + \beta\lambda_t + \alpha u_t \tag{3.39}$$

where $\epsilon_t \overset{d}{\sim} \text{GED}(\nu)$, $\omega, \beta, \alpha \in \mathbb{R}$ and u_t is given by:

$$u_t = \frac{\nu}{2}\left|\frac{y_t}{\exp(\lambda_t)}\right|^\nu - 1 \tag{3.40}$$

The conditional density $p\left(y_t|\mathcal{G}_{t-1};\lambda_t,\nu\right)$ is obtained from the expression in Eq. (3.20) by replacing σ_t^2 with $\exp\left(2\lambda_t\right)$. We have:

$$\nabla_t = \frac{\nu}{2}\left|\frac{y_t}{\exp\left(\lambda_t\right)}\right|^{\nu} - 1 \tag{3.41}$$

$$\chi\left(\lambda_t\right) = \nu \tag{3.42}$$

$$\zeta^{(2)} = \frac{4^{1/\nu}\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} \tag{3.43}$$

from which $s(\lambda_t) = 1$. Note that the two fourth moments $\xi_{kh}^{(4)}$ and $\zeta^{(4)}$ exist and are finite. Therefore, the Gamma-GED-E diffusion limit is:

$$dx_t = \sqrt{\zeta^{(2)}}\exp\left(\lambda_t\right)dW_t^{(1)} \tag{3.44}$$

$$d\lambda_t = \left(\omega - \theta\lambda_t\right)dt + \alpha\sqrt{\nu}dW_t^{(2)} \tag{3.45}$$

As in the case of the Student-t, the diffusive term in the log-volatility process does not depend on λ_t. The coefficient $f_2(\nu) = \sqrt{\nu}$ is the same as in the Gamma-GED. The Gaussian limit is obtained for $\nu = 2$. For $\nu < 2$, the distribution is fat-tailed and the volatility of volatility is smaller than the Gaussian. Similarly, for $\nu > 2$ the distribution is light-tailed and the volatility of volatility is larger than the Gaussian.

In order to compare with a different update for the log-volatility, we consider the EGARCH model of Nelson (1991). The EGARCH reads[3]:

$$y_t = \exp(\lambda_t/2)\epsilon_t \tag{3.46}$$

$$\lambda_{t+1} = \omega + \beta\lambda_t + \alpha_*(|\epsilon_t| - \mathbb{E}[|\epsilon_t|]) + \alpha\epsilon_t \tag{3.47}$$

where $\epsilon_t \sim \text{GED}\left(\nu\right)$. The GED is generally employed in place of the Student-t, since the latter does not guarantee the existence of the unconditional mean and variance of the λ_t process. The term $\alpha(|\epsilon_t| - \mathbb{E}[|\epsilon_t|]) + \alpha\epsilon_t$ responds asymmetrically to positive and negative shocks, thus capturing the leverage effect. In Appendix D.3 we prove the following:

Theorem 3.3. *Under the assumption that:*

$$\lim_{h\to 0} h^{-1}\omega_h = \omega \tag{3.48}$$

$$\lim_{h\to 0} h^{-1}\left(1-\beta_h\right) = \theta \tag{3.49}$$

and α, α_* are constant, the continuous-time process $\left(x_t^{(h)},\lambda_t^{(h)}\right)$ constructed from an EGARCH

[3]Note that in the EGARCH the log-variance is modeled instead of the log-volatility.

converges in distribution to the following Itô process as h goes to zero:

$$dx_t = \sqrt{\zeta^{(2)}}\exp(\lambda_t/2)dW_t^{(1)} \tag{3.50}$$

$$d\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right)dt + \left[\alpha_*^2\frac{4^{1/\nu}\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} + \alpha^2\left(\frac{4^{1/\nu}\Gamma(\frac{3}{\nu})}{\Gamma(\frac{1}{\nu})} - \frac{2^{\frac{6}{\nu}}\Gamma(\frac{1}{2}+\frac{1}{\nu})^2}{4\pi}\right)\right]^{1/2} dW_t^{(2)} \tag{3.51}$$

where $W_t^{(1)}$, $W_t^{(2)}$ *are independent standard Brownian motions, independent of the initial values* (x_0, σ_0^2).

It is immediate to see that the diffusion coefficient in the log-variance process diverges as ν goes to zero. Even for the exponential link function, we thus find that the volatility of volatility of the diffusion limit of score-driven models has a different behavior compared to that of other popular update rules.

4 Monte-Carlo analysis: score-driven models as diffusion approximations

Theorem A.1, which establishes the conditions for the weak convergence of a discrete-time Markov process to a diffusion, suggests that the discrete-time model can be regarded as an approximation of the continuous-time model. This has lead to the use of GARCH-type models for approximate estimation and filtering of diffusion processes. "Quasi Approximate Maximum Likelihood" (QAML), advocated among others by Engle and Lee (1996), Barone-Adesi et al. (2005), Fornari and Mele (2006), Stentoft (2011), consists in first estimating the simple, approximating discrete-time model, and then recovering the continuous-time parameters using a set of moment conditions as those in Eq. (2.11)-(2.13). Similarly, Nelson (1992), Nelson and Foster (1994) and Nelson (1996) study the asymptotic optimality of misspecified GARCH filters. Using continuous record asymptotics, they show that the volatility generated by a class of stochastic differential equations is consistently estimated by GARCH-type models.

In this Monte-Carlo analysis, we study the behavior of score-driven models when employed as diffusion approximations. To this end, we examine their performance in recovering: (i) QAML parameter estimates of their limiting SDE, and (ii) approximate filtered estimates of the volatility of the underlying diffusion. As a discrete-time approximation, we consider the Beta-t model described in Section 3.1.1. We also consider two alternative discrete-time specifications. The first is the GARCH, which differs from the Beta-t in the probability density function and in the update of volatility. The second is the t-GARCH, which differs from the Beta-t only in the update of volatility. This comparison allows us to characterize the properties of the discrete-time model that are more relevant in the approximation to the continuous-time model.

The general form of the diffusion limit of GARCH, t-GARCH and Beta-t is:

$$dx_t = \sigma_t dW_t^{(1)} \tag{4.1}$$

$$d\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right) dt + \kappa\sigma_t^2 dW_t^{(2)} \tag{4.2}$$

We simulate $N = 1000$ time-series of n_s observations of the above SDE using the EM scheme. The latter provides a first-order Gaussian approximation to the true transition density. We thus sample the n_s observations on intervals of length $s \in \mathbb{N}$, i.e. we keep every s-th observation. Without loss of generality, we assume that n_s is a multiple of s and set $n_s = ns$, so that the same number n of observations is available for each sampling frequency. This allows us to compare the results obtained for each s by avoiding the potential distortion due to the use of different sample sizes. Furthermore, n is chosen to be large in order to avoid potential finite-sample effects that may hide the error due to the diffusion approximation. We set $n = 20000$ and $s = \{50, 100, 200, 400, 800, 1600\}$. In the Euler discretization, we set $dt = 1/n$ and thus the time between successive observations is given by $h_s = s/n$. The parameters of the SDE are chosen as $\omega = 0.01$, $\theta = 0.2$, $\kappa = 2.5$.

The QAML estimates of the SDE parameters are recovered from the discrete-time parameters using the moment conditions in Eq. (2.11)-(2.13). In particular, the diffusion coefficient is computed as $\hat{\kappa} = \hat{\alpha}\sqrt{2}$ in the case of GARCH, and $\hat{\kappa} = \hat{\alpha}\sqrt{2\frac{\hat{\nu}-1}{\hat{\nu}-4}}$, $\hat{\kappa} = \hat{\alpha}\sqrt{2\frac{\hat{\nu}}{\hat{\nu}+3}}$ in the case of t-GARCH and Beta-t, respectively. Here, the estimated parameter $\hat{\alpha}$ is rescaled as indicated in Eq. (2.13). Table (1) reports the average of the estimated QAML parameters and the mean-square-error (MSE) of the filtered volatility estimates. For each of the three discrete-time models, the MSE is computed as $\frac{1}{nN}\sum_{i=1}^{N}\sum_{t=1}^{n}(\sigma_{t,i}^2 - \hat{\sigma}_{t,i}^2)^2$, where $\sigma_{t,i}$ is the volatility sampled from the continuous-time process in Eq. (4.1), (4.2) in the i-th simulation, and $\hat{\sigma}_{t,i}$ is the volatility filtered by the discrete-time model.

The likelihood of discrete-time observations sampled from a continuous-time diffusion is generally non-normal; see among others Ait-Sahalia (2002), Ait-Sahalia and Yu (2006). The non-normality increases with the aggregation, as can be seen in Table (1) from the estimated t-GARCH and Beta-t degrees of freedom parameter. For high sampling frequencies (small s), the volatility can be regarded as approximately constant in the short time interval $h_s = s/n$, and thus the diffusion looks locally as a Wiener process. In this case, the three models provide close QAML estimates and similar filtered volatilities. As the sampling frequency decreases (large s), the stochastic nature of the volatility emerges more clearly, and the conditional density estimated by the t-GARCH and Beta-t becomes fat-tailed. They thus lead to better volatility estimates and lower average MSE's compared to the GARCH. It is interesting to note that, while the degrees of freedom parameter of the two fat-tailed models is similar for each s, the MSE of the Beta-t is substantially lower at large s. This is due to

		$s = 50$	$s = 100$	$s = 200$	$s = 400$	$s = 800$	$s = 1600$
ω	GARCH	0.0113	0.0119	0.0123	0.0133	0.0140	0.0131
		(0.0031)	(0.0027)	(0.0029)	(0.0066)	(0.0055)	(0.0035)
	t-GARCH	0.0123	0.0128	0.0126	0.0143	0.0146	0.0124
		(0.0013)	(0.0013)	(0.0030)	(0.0029)	(0.0054)	(0.0083)
	Beta-t	0.0119	0.0122	0.0129	0.0135	0.0142	0.0143
		(0.0013)	(0.0012)	(0.0009)	(0.0008)	(0.0007)	(0.0006)
θ	GARCH	0.5050	0.4797	0.4790	0.4803	0.4972	0.4828
		(0.4022)	(0.3248)	(0.2868)	(0.1959)	(0.2410)	(0.2029)
	t-GARCH	0.5007	0.5322	0.6223	0.5949	0.6527	0.7012
		(0.4098)	(0.3409)	(0.2874)	(0.2755)	(0.2907)	(0.6639)
	Beta-t	0.5340	0.5328	0.6731	0.6110	0.7017	0.7361
		(0.4231)	(0.3450)	(0.2893)	(0.2140)	(0.1577)	(0.1199)
κ	GARCH	2.2883	2.2766	2.2201	1.7819	1.7543	1.7428
		(0.3902)	(0.3878)	(0.4219)	(0.5681)	(0.7637)	(0.4645)
	t-GARCH	2.7352	2.8288	2.7775	3.3486	4.0970	12.1784
		(0.1211)	(0.2325)	(0.6075)	(0.5099)	(0.9492)	(5.6790)
	Beta-t	2.4873	2.4354	2.3537	2.2045	1.9854	1.8026
		(0.0904)	(0.0799)	(0.0716)	(0.0651)	(0.0587)	(0.0538)
ν	t-GARCH	11.4397	8.5891	6.1133	4.8426	4.2762	4.0006
		(0.6019)	(0.3456)	(0.1891)	(0.1273)	(0.1046)	(0.0087)
	Beta-t	12.5682	9.5384	6.8371	5.4044	4.7474	4.0223
		(0.6806)	(0.4057)	(0.2251)	(0.1493)	(0.1245)	(0.0193)
MSE	GARCH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	t-GARCH	0.9727	0.9707	0.9486	0.9069	0.9860	0.9912
	Beta-t	0.9742	0.9716	0.9508	0.8821	0.8735	0.9374

Table 1: For each $s = \{50, 100, 200, 400, 800, 1600\}$, we show the average, over $N = 1000$ simulations of the DGP in Eq. (4.1), (4.2), of the QAML parameters ω, θ, κ and the MSE of GARCH, t-GARCH and Beta-t. The latter is computed as $\text{MSE} = \frac{1}{nN}\sum_{i=1}^{N}\sum_{t=1}^{n}(\sigma_{t,i}^2 - \hat{\sigma}_{t,i}^2)^2$, where $\sigma_{t,i}$ is the SDE volatility generated in the i-th simulation, and $\hat{\sigma}_{t,i}$ is the volatility estimated by the discrete-time model. Each MSE is normalized by the MSE of the GARCH. We also show the degrees of freedom parameter estimated by the t-GARCH and Beta-t. The standard deviations of the estimated parameters are reported in parenthesis.

the robust score update in the Beta-t, which better captures the dynamics of volatility in the presence of fat-tails. Koopman et al. (2016) obtain similar simulation results when analyzing score-driven models as approximations to discrete-time stochastic volatility models.

The QAML estimates of the parameter θ confirm the result of Wang (2002) that QAML is not necessarily consistent; see also Hafner et al. (2017) for further simulation evidences. All the three models provide significantly upward biased estimates of θ for each sampling frequency. The parameter ω has instead a small positive bias, which slightly increases with s. The diffusion coefficient κ is correctly estimated at high frequencies. As s increases, the estimate provided by the t-GARCH rapidly increases and significantly overestimates the true diffusion coefficient. This is due to the explosive behavior, for small ν, of the diffusion limit of the t-GARCH model; see Section 3.1. The estimates of κ of the GARCH and Beta-t are downward biased for large s, however the bias is considerably smaller compared to the t-GARCH. Note also that the standard deviations of the Beta-t estimates are significantly smaller compared to both GARCH and t-GARCH estimates.

We thus find that, when the likelihood of the discrete-time observations is non-normal, the use of the Beta-t for approximate estimation and filtering of the underlying diffusion is preferable to that of models with a GARCH update. The non-normality of the likelihood depends, among others, on the sampling frequency. At large sampling frequencies, the diffusion is locally well approximated by a Wiener process. In this case, the t-GARCH and Beta-t reduce to the GARCH, and thus the three models behave similarly. As the sampling frequency decreases, the likelihood becomes highly non-normal, and we see the benefits of adopting the robust score-driven model. In reality, the last circumstance is of interest because data is rarely available at very large sampling frequencies, for instance due to liquidity constraints. Our results indicate that the use of score-driven models as diffusion approximations is valuable in this real-world scenario. We obtain similar results in the case of models with an exponential link function, given the analogous explosive nature of the diffusion coefficient of the EGARCH model of Nelson (1991); see Section 3.2.2.

5 Generalizations

The results of Section 2 can be generalized into several directions. In this section, we examine volatility models with a time-varying conditional mean in log-returns, and multivariate volatility models based on elliptical distributions. The former are relevant in markets with a time-varying risk-premium, whereas the latter are of interest in financial risk management.

5.1 Dynamic location-scale family models

Let us assume that, conditionally on the past, financial log-returns $\{y_t\}_{t=1}^n$ are generated by a location-scale family density:

$$y_t|\mathcal{G}_{t-1} \overset{d}{\sim} p\left(y_t|\mathcal{G}_{t-1}; c_t, \Theta\right) \tag{5.1}$$

$$p\left(y_t|\mathcal{G}_{t-1}; \mu_t, c_t, \Theta\right) = \frac{1}{\sqrt{c_t}}\Psi\left(\frac{y_t-\mu_t}{\sqrt{c_t}}, \Theta\right) \tag{5.2}$$

where $\mu_t \in \mathbb{R}$ is a $\mathcal{G}_{t-1}-$measurable location parameter. It is immediate to see that $\frac{y_t-\mu_t}{\sqrt{c_t}}|\mathcal{G}_{t-1} \overset{d}{\sim} \Psi\left(\cdot, \Theta\right)$. Similarly to the scale parameter λ_t, the location μ_t is updated based on observations available up to time $t-1$. More precisely, μ_t and λ_t obey the following laws of motion:

$$\mu_t = d + b\,\mu_{t-1} + a\,v_{t-1} \tag{5.3}$$

$$\lambda_t = \omega + \beta\,\lambda_{t-1} + \alpha\,u_{t-1} \tag{5.4}$$

where v_t and u_t are martingale difference sequences given by:

$$v_t = s_\mu(\mu_t, \lambda_t)\nabla_{t,\mu}, \qquad \nabla_{t,\mu} = \frac{\partial \log p(y_t|\mathcal{G}_{t-1}; \mu_t, \Lambda\left(\lambda_t\right), \Theta)}{\partial \mu_t} \tag{5.5}$$

$$u_t = s_\lambda(\mu_t, \lambda_t)\nabla_{t,\lambda}, \qquad \nabla_{t,\lambda} = \frac{\partial \log p(y_t|\mathcal{G}_{t-1}; \mu_t, \Lambda\left(\lambda_t\right), \Theta)}{\partial \lambda_t} \tag{5.6}$$

with $s_\mu(\mu_t, \lambda_t)$ and $s_\lambda(\mu_t, \lambda_t)$ continuous and measurable functions of μ_t and λ_t. In particular, we set $s_\mu(\mu_t, \lambda_t) = \mathbb{E}[\nabla_{t,\mu}^2]^{-1}$. We declare since the beginning the choice of $s_\mu(\mu_t, \lambda_t)$ because, differently from $\nabla_{t,\lambda}$, the moments of $\nabla_{t,\mu}$ depend on the length h of the discretization, as it will be shown in the following. Thus, the form of the limiting SDE depends on the power of the Fisher information that is used to normalize the score. The limiting SDE of models with a different choice of $s_\mu(\mu_t, \lambda_t)$ can be obtained following a similar method.

Our goal is to derive the weak diffusion limit of the class of score-driven location-scale family models described by Eq. (5.1)-(5.4). Without loss of generality, we set $y_t = \mu_t + \sqrt{c_t}\epsilon_t$, where ϵ_t is conditionally distributed according to a scale family density with scale equal to one. Writing $x_t = \sum_{i=1}^t y_t$, we have:

$$x_t = x_{t-1} + \mu_t + \sqrt{\Lambda\left(\lambda_t\right)}\epsilon_t \tag{5.7}$$

$$\mu_{t+1} = d + b\,\mu_t + a s_\mu(\mu_t, \lambda_t)\nabla_{t,\mu} \tag{5.8}$$

$$\lambda_{t+1} = \omega + \beta\lambda_t + \alpha s_\lambda(\mu_t, \lambda_t)\,\nabla_{t,\lambda} \tag{5.9}$$

At this point, we assume a timestamp of length h and allow the static parameters in the previous system to depend on h. Formally, for $k \in \mathbb{N}$, we have:

$$\begin{aligned} x_{kh}^{(h)} &= x_{(k-1)h}^{(h)} + \mu_{kh}^{(h)} h + \sqrt{\Lambda(\lambda_{kh}^{(h)})}\epsilon_{kh}^{(h)} && (5.10)\\ \mu_{(k+1)h}^{(h)} &= d_h + b_h\, \mu_{kh}^{(h)} + a_h s_\mu(\mu_{kh}^{(h)}, \lambda_{kh}^{(h)})\nabla_{kh,\mu}^{(h)} && (5.11)\\ \lambda_{(k+1)h}^{(h)} &= \omega_h + \beta_h\, \lambda_{kh}^{(h)} + \alpha_h s_\lambda(\mu_{kh}^{(h)}, \lambda_{kh}^{(h)})\, \nabla_{kh,\lambda}^{(h)} && (5.12) \end{aligned}$$

and:

$$\mathbb{P}\left[\left(x_0^{(h)}, \mu_0^{(h)}, \lambda_0^{(h)}\right) \in \Gamma\right] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma \in \mathcal{B}\left(\mathbb{R}^3\right) \tag{5.13}$$

where $\left\{\epsilon_{kh}^{(h)}\right\}$ has scale $\sqrt{h}$ and ν_h is a sequence of probability measures on $(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3))$ satisfying Assumption 3. As done in Section 2, we construct the continuous-time processes $\left(x_t^{(h)}, \mu_t^{(h)}, \lambda_t^{(h)}\right)$ in the following way: $x_t^{(h)} = x_{kh}^{(h)}$, $\mu_t^{(h)} = \mu_{kh}^{(h)}$ and $\lambda_t^{(h)} = \lambda_{kh}^{(h)}$ for $kh \le t \le (k+1)h$.

Let us consider again the filtration $\mathcal{F}_{kh}^{(h)} = \sigma\left(x_0^{(h)}, \ldots, x_{(k-1)h}^{(h)}\right)$, the score $\nabla_{kh,\lambda}^{(h)}$ defined as in Eq. (2.10), and the score $\nabla_{kh,\mu}^{(h)}$ computed with respect to the time-varying mean $\mu_{kh}^{(h)}$:

$$\nabla_{kh,\mu}^{(h)} = \frac{\partial \log p\left(y_{kh}^{(h)} | \mathcal{F}_{kh}^{(h)}; \mu_{kh}^{(h)}, \Lambda(\lambda_{kh})\right)}{\partial \mu_{kh}^{(h)}} \tag{5.14}$$

Let also define $\nabla_z = \frac{\partial \log \Psi(z,\Theta)}{\partial z}$, $\chi_z = \mathbb{E}\left[\nabla_z^2\right]$ and $\xi_{kh,\mu}^{(\ell)} = \mathbb{E}\left[\left(\nabla_{kh,\mu}^{(h)}\right)^{\ell} | \mathcal{F}_{kh}^{(h)}\right]$. Finally, let $\zeta^{(\ell)}$, $\chi(\lambda_t)$ being defined as in Section 2. We state now the main result of this section, which is the analogue of Theorem 2.1:

Theorem 5.1. *Let $\Psi(\cdot)$ be symmetric and such that $\zeta^{(2)} \ge \chi_z^{-1}$. Under the following assumptions on the parameters:*

$$\begin{aligned} &\lim_{h\to 0} h^{-1}\omega_h = \omega && \lim_{h\to 0} h^{-1} d_h = d && (5.15)\\ &\lim_{h\to 0} h^{-1}(1-\beta_h) = \theta && \lim_{h\to 0} h^{-1}(1-b_h) = \vartheta && (5.16)\\ &\lim_{h\to 0} h^{-1}\alpha_h^2 = \alpha^2 && \lim_{h\to 0} h^{-2} a_h^2 = a^2 && (5.17) \end{aligned}$$

where ω, d, θ, ϑ, α, $a \in \mathbb{R}$, and under the assumption that the moments $\xi_{kh,\mu}^{(\ell)}$, $\xi_{kh,\lambda}^{(\ell)}$, $\zeta^{(\ell)}$ exist and are finite for $\ell \le 4$, the continuous-time process $q_t = (x_t^{(h)}, \mu_t^{(h)}, \lambda_t^{(h)})'$ converges in distribution to the following Itô process as h goes to zero:

$$dq_t = b(q_t, t)\, dt + \sigma(q_t, t) dW_t \tag{5.18}$$

where W_t is a 3-dimensional vector of standard Brownian motions, independent from q_0.

The drift, $b(q_t, t)$, is given by:

$$b(q_t,t) = \begin{bmatrix} \mu_t \\ d - \vartheta\,\mu_t \\ \omega - \theta\,\lambda_t \end{bmatrix}, \tag{5.19}$$

while $\sigma(q_t, t)$ is a mapping such that, for all $q_t \in \mathbb{R}^3$ and $t \geq 0$, $a(q_t,t) = \sigma(q_t,t)\sigma(q_t,t)'$, where $a(q_t,t)$ is defined as:

$$a(q_t,t) = \begin{bmatrix} \Lambda(\lambda_t)\zeta^{(2)} & a\Lambda(\lambda_t)\chi_z^{-1} & 0 \\ a\Lambda(\lambda_t)\chi_z^{-1} & a^2\Lambda\left(\lambda_t\right)\chi_z^{-1} & 0 \\ 0 & 0 & \alpha^2 s(\lambda_t)^2\chi\left(\lambda_t\right) \end{bmatrix} \tag{5.20}$$

The assumption $\zeta^{(2)} \geq \chi_z^{-1}$ guarantees that the matrix $a(q_t,t)$ is positive semi-definite. The assumption that Ψ is symmetric can be relaxed at the expense of additional nonzero entries $a_{23}(q_t,t)$, $a_{32}(q_t,t)$, which would imply more involved conditions for positive semi-definiteness. Note that, if $\zeta^{(2)} = \chi_z^{-1}$, the matrix $a(q_t,t)$ becomes singular and x_t, μ_t are driven by two perfectly correlated Brownian motions. This degeneracy is due to the nonzero correlation between the innovations $\epsilon_{kh}^{(h)}$ and the score $\nabla_{kh,\mu}^{(h)}$, see Appendix E.2. In contrast, the correlation between the innovations $\epsilon_{kh}^{(h)}$ and the score $\nabla_{kh,\lambda}^{(h)}$ is zero, as shown in Appendix C.1, implying that λ_t is always driven by an independent stochastic component.

The same methodology used for proving Theorem 2.1 can be applied here. More precisely, the determination of the dependence of the conditional moments $\xi_{kh,\mu}^{(\ell)}$ and $\xi_{kh,\lambda}^{(\ell)}$ on h enables us to compute the limits in (**A1.1**), (**A1.2**) and (**A1.3**). In particular, it is not difficult to see that, for $\xi_{kh,\lambda}^{(\ell)}$, the same result of Theorem 2.2 holds:

$$\xi_{kh,\lambda}^{(\ell)} = (-1)^\ell \left[\frac{1}{2}\frac{\Lambda'(\lambda_{kh}^{(h)})}{\Lambda(\lambda_{kh}^{(h)})}\right]^\ell \int_{-\infty}^{\infty} \left(1 + \frac{\Psi'\left(z,\Theta\right)}{\Psi\left(z,\Theta\right)} z\right)^\ell \Psi\left(z,\Theta\right)\,dz$$

i.e. $\xi_{kh,\lambda}^{(\ell)}$ is independent of h. Concerning $\xi_{kh,\mu}^{(\ell)}$, instead, it results that it scales as $h^{\ell/2}$, as the following theorem shows.

Theorem 5.2. *For the class of conditional location-scale family densities*

$$p\left(y_{kh}^{(h)}|\mathcal{F}_{(k-1)h}^{(h)};\mu_{kh}^{(h)},\Lambda\left(\lambda_{kh}^{(h)}\right),\Theta\right) = \frac{1}{\sqrt{\Lambda\left(\lambda_{kh}^{(h)}\right)h}}\Psi\left(\frac{y_{kh}^{(h)} - \mu_{kh}^{(h)}h}{\sqrt{\Lambda(\lambda_{kh}^{(h)})h}},\Theta\right)$$

the moments $\xi^{(\ell)}_{kh,\mu} = \mathbb{E}\left[\left(\nabla^{(h)}_{kh,\mu}\right)^{\ell} |\mathcal{F}^{(h)}_{kh}\right]$ *are given by:*

$$\xi^{(\ell)}_{kh,\mu} = (-1)^{\ell}\left[\frac{h}{\Lambda(\lambda^{(h)}_{kh})}\right]^{\ell/2}\int_{-\infty}^{\infty}\left(\frac{\Psi'(z,\Theta)}{\Psi(z,\Theta)}\right)^{\ell}\Psi(z,\Theta)\,dz$$

Proof. See Appendix E.1. □

In particular, for location-scale family models, the asymptotic behavior of the expressions in (**A1.1**), (**A1.2**) and (**A1.3**) depends also on $\xi^{(\ell)}_{kh,\mu}$, in addition to the parameters $\omega_h, d_h, \beta_h, b_h, \alpha_h$ and a_h. In the following, we report explicit expressions for the limit of the these expectations, computed under the assumptions of Theorem 5.1. We refer the reader to Appendix E.2 for details on their derivations.

The drift per unit of time (Condition **A1.1**) is given by:

$$h^{-1}\mathbb{E}\left[\left(x^{(h)}_{kh} - x^{(h)}_{(k-1)h}\right)|\mathcal{F}^{(h)}_{kh}\right] = \mu^{(h)}_{kh} \tag{5.21}$$

$$h^{-1}\mathbb{E}\left[\left(\mu^{(h)}_{(k+1)h} - \mu^{(h)}_{kh}\right)|\mathcal{F}^{(h)}_{kh}\right] = d - \vartheta\,\mu^{(h)}_{kh} \tag{5.22}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda^{(h)}_{(k+1)} - \lambda^{(h)}_{kh}\right)|\mathcal{F}^{(h)}_{kh}\right] = \omega - \theta\,\lambda^{(h)}_{kh} \tag{5.23}$$

Again, constraints on the parameters d, ϑ, ω and θ depend upon the link function and will be discussed in specific cases. The second moments per unit of time (Condition **A1.2**) are given by:

$$h^{-1}\mathbb{E}\left[\left(x^{(h)}_{kh} - x^{(h)}_{(k-1)h}\right)^2|\mathcal{F}^{(h)}_{kh}\right] = \Lambda\left(\lambda^{(h)}_{kh}\right)\zeta^{(2)} + o(1) \tag{5.24}$$

$$h^{-1}\mathbb{E}\left[\left(\mu^{(h)}_{(k+1)h} - \mu^{(h)}_{kh}\right)^2|\mathcal{F}^{(h)}_{kh}\right] = a^2\Lambda\left(\lambda^{(h)}_{kh}\right)\chi_z^{-1} + o(1) \tag{5.25}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda^{(h)}_{(k+1)h} - \lambda^{(h)}_{kh}\right)^2|\mathcal{F}^{(h)}_{kh}\right] = \alpha^2 s_\lambda\left(\mu^{(h)}_{kh},\lambda^{(h)}_{kh}\right)^2\xi^{(2)}_{kh} + o(1) \tag{5.26}$$

$$h^{-1}\mathbb{E}\left[\left(x^{(h)}_{kh} - x^{(h)}_{(k-1)h}\right)\left(\mu^{(h)}_{(k+1)h} - \mu^{(h)}_{kh}\right)|\mathcal{F}^{(h)}_{kh}\right] = a\Lambda(\lambda^{(h)}_{kh})\chi_z^{-1} + o(1) \tag{5.27}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda^{(h)}_{kh} - \lambda^{(h)}_{(k-1)}\right)\left(\mu^{(h)}_{(k+1)h} - \mu^{(h)}_{kh}\right)|\mathcal{F}^{(h)}_{kh}\right] = o(1) \tag{5.28}$$

$$h^{-1}\mathbb{E}\left[\left(x^{(h)}_{kh} - x^{(h)}_{(k-1)h}\right)\left(\lambda^{(h)}_{(k+1)} - \lambda^{(h)}_{kh}\right)|\mathcal{F}^{(h)}_{kh}\right] = o(1) \tag{5.29}$$

The considerations of the previous section apply also here. Concerning Condition (**A1.3**), we choose, as in Nelson (1991), $\delta = 2$, and then set $d_h = hd$, $\omega_h = h\omega$, $\beta_h = 1 - h\theta$,

$b_h = 1 - h\vartheta$, $\alpha_h = h^{1/2}\alpha$ and $a_h = ha$. These choices lead to the following results:

$$h^{-1}\mathbb{E}\left[\left(x_{kh}^{(h)} - x_{(k-1)h}^{(h)}\right)^4 |\mathcal{F}_{kh}^{(h)}\right] = h\,\Lambda^2\left(\lambda_{kh}^{(h)}\right)\zeta^{(4)} \tag{5.30}$$

$$h^{-1}\mathbb{E}\left[\left(\mu_{kh}^{(h)} - \mu_{(k-1)h}^{(h)}\right)^4 |\mathcal{F}_{kh}^{(h)}\right] = O(h^{\varsigma}), \quad \varsigma \geq 4 \tag{5.31}$$

$$h^{-1}\mathbb{E}\left[\left(\lambda_{kh}^{(h)} - \lambda_{(k-1)h}^{(h)}\right)^4 |\mathcal{F}_{kh}^{(h)}\right] = h\,\alpha^4 s_\lambda\left(\mu_{kh}^{(h)}, \lambda_{kh}^{(h)}\right)^4 \xi_{kh,\lambda}^{(4)} + O\left(h^{\varrho}\right) \tag{5.32}$$

with $\varrho \geq 3/2$. In particular, since $\zeta^{(4)}$ and $\xi_{kh,\lambda}^{(4)}$ are finite by assumption, all the expressions above converge to zero as h goes to zero. As a consequence, Assumption 1 is satisfied and the two coefficients $b(q_t, t)$, $a(q_t, t)$ are given by Eq. (5.19), (5.20).

In order to conclude the proof of the Theorem 5.1, we need to discuss the finiteness of the process in finite intervals and uniqueness of the diffusion limit. This is related to the growth and regularity of the coefficients $b(x_t, \mu_t, \lambda_t, t)$ and $a(x_t, \mu_t, \lambda_t, t)$, and on the link function $\Lambda(\lambda_t)$. Again, we distinguish between models for which the link function is the identity ($\Lambda(\lambda_t) = \lambda_t$), and models for which the link function is the exponential ($\Lambda(\lambda_t) = e^{\lambda_t}$).

In the first class of models, the equation describing the evolution of σ_t^2 is decoupled from the other two equations, and it has a unique solution since the coefficients are continuous and globally Lipschitz (see proof of Theorem 2.3). The Cholesky decomposition of $a(x_t, \mu_t, \sigma_t^2)$ gives us the matrix $\sigma\left(x_t, \mu_t, \sigma_t^2\right)$, which is such that $\sigma(x_t, \mu_t, \sigma_t^2)\sigma(x_t, \mu_t, \sigma_t^2)' = a(x_t, \mu_t, \sigma_t^2)$:

$$\sigma\left(x_t, \mu_t, \sigma_t^2\right) = \begin{bmatrix} \sigma_t\sqrt{\zeta^{(2)}} & \frac{a\sigma_t}{\sqrt{\zeta^{(2)}\chi_z}} & 0 \\ 0 & \frac{\sigma_t a\sqrt{\zeta^{(2)}\chi_z - 1}}{\sqrt{\zeta^{(2)}\chi_z}} & 0 \\ 0 & 0 & \alpha C(\Theta)\sigma_t^2 \end{bmatrix}$$

where $a > 0$ and $C(\Theta)$ is a positive constant function (possibly) dependent on the vector of static parameter. Under the assumptions of Theorem 5.1 both $b(x_t, \mu_t, \sigma_t^2)$ and $\sigma\left(x_t, \mu_t, \sigma_t^2\right)$ are globally Lipschitz and, therefore, Assumption 4 is satisfied. In models with an exponential link function, the matrix $\sigma(x_t, \mu_t, \lambda_t)$ of the Cholesky decomposition of $a\left(x_t, \mu_t, \lambda_t\right)$ is given by:

$$\sigma\left(x_t, \mu_t, \lambda_t\right) = \begin{bmatrix} \exp\left(\lambda_t\right)\sqrt{\zeta^{(2)}} & \frac{a\exp(\lambda_t)}{\sqrt{\zeta^{(2)}\chi_z}} & 0 \\ 0 & \frac{\exp(\lambda_t)a\sqrt{\chi_z\,\zeta^{(2)} - 1}}{\sqrt{\zeta^{(2)}\chi_z}} & 0 \\ 0 & 0 & C(\Theta) \end{bmatrix}$$

where $a > 0$ and $C(\Theta)$ is a positive constant functions (possibly) dependent on a vector of static parameters Θ. Under the assumptions of Theorem 5.1, $b(x_t, \mu_t, \lambda_t)$ and $\sigma(x_t, \mu_t, \lambda_t)$ are continuous and globally Lipschitz and, therefore, Assumption 4 is satisfied.

5.1.1 A Student-t location-scale family model

In this section, we specialize the general limit result of Theorem 5.1 to the case of a Student-t density with time-varying location and scale. Specifically, we assume that the conditional density in (5.1)-(5.2) is:

$$\begin{aligned} y_t|\mathcal{G}_{t-1} &\overset{d}{\sim} p(y_t|\mathcal{G}_{t-1};\sigma_t^2,\Theta) \\ p(y_t|\mathcal{G}_{t-1};\mu_t,\sigma_t^2,\Theta) &= \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi\sigma_t^2}}\left[1+\frac{(y_t-\mu_t)^2}{(\nu-2)\sigma_t^2}\right]^{-\frac{(\nu+1)}{2}}. \end{aligned}$$

Note that the link function is the identity and $\Psi(z,\Theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\left[1+\frac{z^2}{(\nu-2)}\right]^{-\frac{(\nu+1)}{2}}$. In order to derive the drift vector $b(x_t,\nu_t,\sigma_t^2)$ and the matrix $a(x_t,\nu_t,\sigma_t^2)$ in Eq. (5.18), it is sufficient to compute the following quantities (note that $s(\sigma_t^2)$ and $\chi(\sigma_t^2)$ are the same as in Section 3.1.1):

$$\zeta^{(2)} = 1 \tag{5.33}$$

$$\chi_z = \frac{\nu(1+\nu)}{\nu^2+\nu-6} \tag{5.34}$$

$$s(\sigma_t^2) = 2\sigma_t^4 \tag{5.35}$$

$$\chi(\sigma_t^2) = \frac{1}{2\sigma_t^4}\frac{\nu}{\nu+3} \tag{5.36}$$

In particular, it results that $\zeta^{(2)} \geq \chi_z^{-1}$ for all $\nu > 2$, whereas $b(x_t,\mu_t,\sigma_t^2)$ and $a(x_t,\mu_t,\sigma_t^2)$ are given by:

$$b(x_t,\mu_t,\sigma_t^2) = \begin{bmatrix} \mu_t \\ d-\vartheta\,\mu_t \\ \omega-\theta\,\sigma_t^2 \end{bmatrix} \qquad a(x_t,\mu_t,\sigma_t^2) = \sigma_t^2\begin{bmatrix} 1 & a\chi_z^{-1} & 0 \\ a\chi_z^{-1} & a^2\chi_z^{-1} & 0 \\ 0 & 0 & \alpha^2 4\sigma_t^6\chi(\sigma_t^2) \end{bmatrix} \tag{5.37}$$

where $\omega > 0$, and d, θ and $\vartheta \in \mathbb{R}$. If ν is finite, $\chi_z^{-1} < 1$, implying that the matrix $a(x_t,\mu_t,\lambda_t)$ is non-singular and that q_t is driven by three independent Brownian motions. In the Gaussian limit, $\chi_z = 1$ and x_t, μ_t are driven by two perfectly correlated Brownian motions.

5.2 Dynamic multivariate elliptical models

To develop the results in the present section, we first need some matrix notations and definitions. Let $\mathrm{M}_{m,n}(\mathbb{R})$ be the set of $m\times n$ matrices with entries in $\mathbb{R}$. We denote the (i,j) entry of an $m\times n$ matrix A by $(A)_{ij}$ or a_{ij}. In addition, we set $\mathrm{M}_n(\mathbb{R}) = \mathrm{M}_{n,n}(\mathbb{R})$ and denote by $\mathrm{S}_n^+(\mathbb{R})$ the set of $n\times n$ non-negative definite symmetric matrix. Now, let

$A \in \mathrm{M}_{m,n}(\mathbb{R})$ and $B \in \mathrm{M}_{m',n'}$ be two matrices. The Kronecker product is denoted by $A \otimes B$, with $A \otimes B \in \mathrm{M}_{mm',nn'}(\mathbb{R})$. When $A = B$ we write $A_{\otimes}$. Moreover, the operator $\mathrm{vec}(A)$ vectorizes the matrix A into a column vector, whereas $\mathrm{vech}(A)$ vectorizes the lower-triangular part of a matrix A into a column vector. The operator $\oplus$ for two matrices A and B is given by: $A \oplus B = (A \otimes B) + (B \otimes A)$. Finally, we denote by $\mathcal{D}_n$ the duplication matrix, i.e. the unique matrix $\mathcal{D}_n$ such that $\mathcal{D}_n \mathrm{vech}(A) = \mathrm{vec}(A)$ for any matrix $A \in \mathrm{S}_n^+(\mathbb{R})$, by $\mathcal{B}_n$ the elimination matrix, i.e. the matrix $\mathcal{B}_n$ such that $\mathcal{B}_n \mathrm{vec}(A) = \mathrm{vech}(A)$, and by $\mathcal{C}_n$ the commutation matrix, i.e. the matrix $\mathcal{C}_n$ such that $\mathcal{C}_n \mathrm{vec}(A) = \mathrm{vec}\left(A'\right)$. For further details on matrix operations, we refer to Abadir and Magnus (2005).

In the present section, we deal with the class of dynamic covariance models based on elliptical conditional densities. More precisely, let $\{y_t\}_{t=1}^T$ be a multivariate time-series where each $y_t \in \mathbb{R}^N$ is a vector of log-returns, and denote by $\mathcal{G}_t = \sigma(y_1, \ldots, y_t)$ the σ-algebra generated by the observables up to time t. In the following, we assume that $\{y_t\}_{t=1}^T$ are sampled from the following conditional density:

$$y_t | \mathcal{G}_{t-1} \overset{d}{\sim} p(y_t | \mathcal{G}_{t-1}; \Sigma_t, \Theta) \tag{5.38}$$

$$p(y_t | \mathcal{G}_{t-1}; \Sigma_t, \Theta) = \frac{g}{|\Sigma_t|^{1/2}} \Psi\left(y_t' \Sigma_t^{-1} y_t, \Theta\right) \tag{5.39}$$

where $g \in \mathbb{R}^+$ is a normalization factor, Ψ is a probability density function, Θ denotes a set of static parameters and $|\cdot|$ is the determinant. Note that this class of conditional densities include the multivariate Student-t distribution, and thus the t-GAS model of Creal et al. (2011). In Eq. (5.38), $\Sigma_t \in \mathrm{S}_N^+(\mathbb{R})$ is $\mathcal{G}_{t-1}$-measurable and depends upon a vector of time-varying parameters $f_t \in \mathbb{R}^K$. More precisely, $\Sigma_t = \Sigma(f_t)$ with $\Sigma : \mathbb{R}^K \to \mathrm{S}_N^+(\mathbb{R})$ being a differentiable function of the $\mathcal{G}_{t-1}$-measurable vector f_t. The time-varying vector f_t obeys the following law of motion:

$$f_{t+1} = c + B f_t + A s_t \tag{5.40}$$

where $c \in \mathbb{R}^K$ is a vector of constants, $A, B \in \mathrm{M}_K(\mathbb{R})$ and $s_t = s(y_t, f_t) \in \mathbb{R}^K$ is a measurable function depending on y_t and f_t. We assume that $s_t = S(f_t)\nabla_t$, where $S(f_{t-1}) \in \mathrm{M}_K(\mathbb{R})$ is a scaling matrix that may depend on f_t in a continuous way, whereas, under the assumption that $p(y_t | \mathcal{G}_{t-1}; \Sigma_t, \Theta)$ is differentiable with respect to f_t, the score ∇_t is given by:

$$\nabla_t = \frac{\partial \log p(y_t | \mathcal{G}_{t-1}; \Sigma(f_t), \Theta)}{\partial f_t}$$

We write the cumulative log-return process $x_t = \sum_{i=1}^{t} y_i$ as:

$$x_t = x_{t-1} + \eta_t \tag{5.41}$$

$$f_{t+1} = c + Bf_t + As_t \tag{5.42}$$

where $\eta_t \in \mathbb{R}^N$ has an elliptical conditional density with zero mean and normalized such that $\mathbb{E}[\eta_t \eta_t' | \mathcal{G}_{t-1}] = \Sigma_t$. By assuming a timestamp of length h, and allowing the static parameters to depend on h, we write, for $k \in \mathbb{N}$:

$$x_{kh}^{(h)} = x_{(k-1)}^{(h)} + \eta_{kh}^{(h)} \tag{5.43}$$

$$f_{(k+1)h}^{(h)} = c_h + B_h f_{kh}^{(h)} + A_h S(f_{kh}^{(h)}) \nabla_{kh}^{(h)} \tag{5.44}$$

and

$$\mathbb{P}\left[\left(x_0^{(h)}, f_0^{(h)}\right) \in \Gamma\right] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma \in \mathbb{B}\left(\mathbb{R}^{N+K}\right), \tag{5.45}$$

where $\eta_{kh}^{(h)}$ has scale $h\Sigma_{kh}^{(h)}$ and the sequence $\{\nu_h\}$ satisfies Assumption 3. Let $\mathcal{F}_{kh}^{(h)} = \sigma\left(x_0^{(h)}, \ldots, x_{(k-1)h}^{(h)}\right)$. The score $\nabla_{kh}^{(h)}$ is computed as:

$$\nabla_{kh}^{(h)} = \left[\frac{\partial \log p\left(y_{kh}^{(h)} | \mathcal{F}_{(k-1)h}^{(h)}; f_{kh}^{(h)}, \Theta\right)}{\partial f_{kh}^{(h)\prime}}\right]' \tag{5.46}$$

where $y_{kh} = x_{kh} - x_{(k-1)h}$ are log-returns. Note that, using the chain rule, it is possible to write:

$$\nabla_{kh}^{(h)} = \Upsilon_{kh}' \tilde{\nabla}_{kh}^{(h)} \tag{5.47}$$

where $\Upsilon_{kh} = \frac{\partial \text{vech}(\Sigma_{kh})}{\partial f_{kh}'}$ and $\tilde{\nabla}_{kh}^{(h)}$ denotes the score computed with respect to the lower-triangular elements of Σ_{kh}:

$$\tilde{\nabla}_{kh}^{(h)} = \left[\frac{\partial \log p\left(y_{kh}^{(h)} | \mathcal{F}_{(k-1)h}^{(h)}; f_{kh}^{(h)}, \Theta\right)}{\partial \text{vech}(\Sigma_{kh})'}\right]' \tag{5.48}$$

The matrix Υ_{kh} depends explicitly on the parameterization of Σ_t.

As done in Section 2, we consider the continuous-time process $\left(x_t^{(h)\prime}, f_t^{(h)\prime}\right)' \in \mathbb{R}^{N+K}$ constructed in the following way: $x_t^{(h)} = x_{kh}^{(h)}$, $f_t^{(h)} = f_{kh}^{(h)}$ for $kh \le t \le (k+1)h$. We first determine the converges rates of the parameters c_h, B_h, A_h which ensure that Assumptions (**A1.1**), (**A1.2**), (**A1.3**) are satisfied as $h \downarrow 0$. As in previous cases, it is important to determine the dependence on h of the moments of the score. Let us denote by $\xi_{kh}^{(2)} =$

$\mathbb{E}[\nabla_{kh}\nabla'_{kh}|\mathcal{F}^{(h)}_{kh}]$ the conditional Fisher information matrix and by $\xi^{(4)}_{kh,i} = \mathbb{E}[\nabla^4_{kh,i}|\mathcal{F}^{(h)}_{kh}]$, $i = 1,\ldots,K$, the conditional fourth moments of the i-th component of the score. In Appendix F, we show that, for the class of elliptical densities, $\xi^{(2)}_{kh}$ and $\xi^{(4)}_{kh,i}$ are independent of h. Thanks to this, we have the following result for the diffusion limit of the process $\left(x^{(h)\prime}_{kh}, f^{(h)\prime}_{kh}\right)'$:

Theorem 5.3. *Under the following assumptions on the parameters:*

$$\begin{aligned}
&\lim_{h\to 0} h^{-1}c_h = c\\
&\lim_{h\to 0} h^{-1}\left(I_K - B_h\right) = \Lambda\\
&\lim_{h\to 0} h^{-1}A^2_h = A^2
\end{aligned} \tag{5.49}$$

where $c \in \mathbb{R}^K$, Λ, $A \in \mathrm{M}_K(\mathbb{R})$, and under the assumption that $\xi^{(2)}_{kh}$ and $\xi^{(4)}_{kh,i}$, $i = 1,\ldots,K$ exist and are finite, the continuous-time process $\left(x^{(h)\prime}_t, f^{(h)\prime}_t\right)'$ weakly converges as $h \to 0$ to the diffusion process $m_t = (x'_t, f'_t)'$ which is the solution to the system of stochastic differential equations:

$$dm_t = b(m_t)dt + \sigma(m_t)dW_t$$

where $W_t \in \mathbb{R}^{N+K}$ is a vector of mutually independent Brownian motions, independent from m_0. The drift, $b(m_t)$, is given by:

$$b(m_t) = \begin{bmatrix} 0_N \\ c - \Lambda f_t \end{bmatrix} \tag{5.50}$$

while $\sigma(m_t)$ is a continuous mapping such that, for all $m_t \in \mathbb{R}^{N+K}$ and $t \geq 0$, $a(m_t) = \sigma(m_t)\sigma(m_t)'$, where $a(m_t)$ is given by:

$$a(m_t) = \begin{bmatrix} \Sigma(f_t) & 0_{(N\times K)} \\ 0_{(K\times N)} & AS(f_t)\,\mathcal{I}(f_t)\,S(f_t)'A' \end{bmatrix} \tag{5.51}$$

with $\mathcal{I}(f_t) = \lim_{h\to 0}\xi^{(2)}_{kh}$.

Proof. See Appendix F.1. □

Using the chain rule, we can re-write the conditional Fisher information matrix as:

$$\xi^{(2)}_{kh} = \Upsilon'_{kh}\tilde{\xi}^{(2)}_{kh}\Upsilon_{kh} \tag{5.52}$$

where $\tilde{\xi}^{(2)}_{kh} = \mathbb{E}[\tilde{\nabla}_{kh}\tilde{\nabla}'_{kh}|\mathcal{F}^{(h)}_{kh}]$. As previously stated, the expression of the matrix Υ_{kh} depends on the parameterization used for Σ_t. As in Engle (2002) and Creal et al. (2011), we decompose the covariance matrix as $\Sigma_t = D_tR_tD_t$, where D_t is a diagonal matrix of standard deviations and R_t is a correlation matrix. The latter is written as $R_t = \Delta_t^{-1}Q_t\Delta_t^{-1}$, where

$Q_t \in S_N^+(\mathbb{R})$ and $\Delta_t = \text{diag}[Q_t]^{1/2}$. We specify the vector of time-varying parameters as

$$f_t = \begin{pmatrix} \log\left(\text{diag}\left(D_t^2\right)\right) \\ \text{vech}\left(Q_t\right) \end{pmatrix} \tag{5.53}$$

which contains $K = N(N+3)/2$ elements. Under this parameterization, Creal et al. (2011) obtain the following expression for the matrix Υ_t:

$$\Upsilon_t = \mathcal{B}_N\left(\mathrm{I}_N \oplus D_t R_t\right) W_{D_t} S_D + \mathcal{B}_N D_{t\otimes}\Delta_{t\otimes}^{-1}\left[\mathcal{D}_N - \left(\Delta_t \oplus Q_t\right)\Delta_{t\otimes}^{-1} W_{\Delta_t} S_\Delta\right] S_Q \tag{5.54}$$

where i) W_{Δ_t} is constructed by having a $N^2 \times N^2$ diagonal matrix with diagonal elements $0.5\text{vec}\left(\Delta_t^{-1}\right)$ and then dropping the columns containing only $0s$; ii) W_{D_t} is constructed by having a $N^2 \times N^2$ diagonal matrix with diagonal elements $0.5\text{vec}\left(D_t\right)$ and then dropping the columns containing only $0s$; iii) the matrices S_Δ, S_D and S_Q are selection matrices containing only 1s and 0s such that $\text{diag}\left(\Delta_t^2\right) = S_\Delta \text{vech}\left(Q_t\right)$, $\log(\text{diag}\left(D_t^2\right)) = S_D f_t$, and $\text{vech}\left(Q_t\right) = S_Q f_t$.

We illustrate the continuous-time limit in Theorem 5.3 for a multivariate Student-t density. In this case, the discrete-time process reduces to the t-GAS model of Creal et al. (2011). For simplicity, we set $N = 2$ and assume that the variances are constant over time, i.e. we set $D_t = I_{2\times 2}$. The conditional density is given by:

$$p(y_t|\mathcal{G}_{t-1}; R_t, \Theta) = \frac{\Gamma\left((\nu+2)/2\right)}{\Gamma(\nu/2)\left[(\nu-2)\pi\right]\left|R_t\right|^{1/2}} \times \left[1 + \frac{y_t' R_t^{-1} y_t}{(\nu-2)}\right]^{-(\nu+2)/2} \tag{5.55}$$

where $R_t = [1, \rho_t; \rho_t, 1]$. We employ the DCC parameterization, with $f_t = (q_{11,t}, q_{12,t}, q_{22,t})'$ and $\rho_t = q_{12,t}/\sqrt{q_{11,t}q_{22,t}}$. The matrix Υ_t becomes:

$$\Upsilon_t = \frac{1}{2\sqrt{q_{11,t}q_{22,t}}}(-q_{12,t}/q_{11,t}, 2, -q_{12,t}/q_{22,t})'(1, 0, 1) \tag{5.56}$$

By standard linear algebra computations, we obtain that the matrix $\mathcal{I}\left(f_t\right)$ is given by:

$$\mathcal{I}(f_t) = \Pi\left(\rho_t, \nu\right) \begin{bmatrix} \frac{1}{q_{11t}^2} & -\frac{2}{q_{11t}q_{12t}} & \frac{1}{q_{11t}q_{22t}} \\ -\frac{2}{q_{11t}q_{12t}} & \frac{4}{q_{12t}^2} & -\frac{2}{q_{12t}q_{22t}} \\ \frac{1}{q_{11t}q_{22t}} & -\frac{2}{q_{12t}q_{22t}} & \frac{1}{q_{22t}^2} \end{bmatrix}, \tag{5.57}$$

where

$$\Pi\left(\rho_t, \nu\right) = \frac{\rho_t^2\left(2 + \nu\left(1 + \rho_t^2\right)\right)}{4\left(4+\nu\right)\left(1-\rho_t^2\right)^2} \tag{5.58}$$

In particular, $\mathcal{I}(f_t)$ is of rank one, due to the redundancy of the DCC parameterization (see e.g. discussions in Creal et al., 2011). For this reason, we set $S(f_t) = I_{3\times 3}$ and thus the 3×3 block in matrix $a(m_t)$ is simply $A\mathcal{I}(f_t)A'$.

It is interesting to examine the Gaussian limit. We have:

$$\lim_{\nu\to\infty} \mathcal{I}(f_t) = \Pi(\rho_t) \begin{bmatrix} \frac{1}{q_{11t}^2} & -\frac{2}{q_{11t}q_{12t}} & \frac{1}{q_{11t}q_{22t}} \\ -\frac{2}{q_{11t}q_{12t}} & \frac{4}{q_{12t}^2} & -\frac{2}{q_{12t}q_{22t}} \\ \frac{1}{q_{11t}q_{22t}} & -\frac{2}{q_{12t}q_{22t}} & \frac{1}{q_{22t}^2} \end{bmatrix} \tag{5.59}$$

where $\Pi(\rho_t) = \frac{\rho_t^2(1+\rho_t^2)}{4(1-\rho_t^2)^2}$. This is different from the continuous-time limit of the DCC model (Hafner et al., 2017). In discrete-time, it is known that the t-GAS does not reduce to the DCC as ν goes to infinity. We thus recover in continuous-time a result analogous to the one known in discrete-time.

6 Conclusions

The relation between score-driven stochastic difference equations and continuous-time stochastic differential equations usually employed in the theoretical finance literature is unknown. The present paper sheds light on this topic by providing the weak diffusion limit of a class of score-driven volatility models based on scale family densities, thus generalizing the popular diffusion limit of Nelson (1990). We determine the continuous-time limit of well-known score-driven models, namely Beta-t, Gamma-GED, Beta-t-E and Gamma-GED-E (see Harvey, 2013). Two interesting properties have emerged: a) the form of the coefficients characterizing the diffusion limit depends only on the link function and on the conditional second moment of the score; b) the properties of the diffusion limit are strictly related with those of the corresponding discrete-time process; the volatility of volatility of the diffusion limit of models with a GARCH update can diverge as the density becomes more and more fat-tailed, whereas in score-driven models this quantity remains finite.

We explore in an extensive Monte-Carlo study the implications of such results on approximate estimation and filtering of diffusion models. It is found that, as a consequence of the non-normal nature of the likelihood of the underlying diffusion, the use of fat-tailed score-driven models improves significantly in parameter estimation and volatility filtering. We finally provide a generalization of the previous results to time-varying conditional mean and to conditional variance models.

7 Acknowledgements

We thank Andrew Harvey for useful suggestions. Giulia Livieri acknowledges the support by Unicredit S.P.A. under the project "Dynamics and Information Research Institute-Quantum

Information (Teoria dell'Informazione), Quantum Technologies".

References

Abadir, K.M., Magnus, J.R., 2005. Matrix algebra. volume 1. Cambridge University Press.

Ait-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. Econometrica 70, 223–262.

Ait-Sahalia, Y., Yu, J., 2006. Saddlepoint approximations for continuous-time markov processes. Journal of Econometrics 134, 507 – 551.

Alexander, C., Lazar, E., 2005. On the continuous limit of GARCH. Working paper.

Barone-Adesi, G., Rasmussen, H., Ravanelli, C., 2005. An option pricing formula for the garch diffusion model. Computational Statistics & Data Analysis 49, 287 – 310.

Barunik, J., Krehlik, T., Vacha, L., 2016. Modeling and forecasting exchange rate volatility in time-frequency domain. European Journal of Operational Research 251, 329–340.

Blasques, F., Koopman, S.J., Lucas, A., 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. Biometrika 102, 325–343.

Blasques, F., Lucas, A., Silde, E., 2018. A stochastic recurrence equations approach for score driven correlation models. Econometric Reviews 37, 166–181.

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

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69, 542–547.

Brown, L.D., Wang, Y., Zhao, L.H., 2003. On the statistical equivalence at suitable frequencies of garch and stochastic volatility models with the corresponding diffusion model. Statistica Sinica 13, 993–1013.

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

Cox, D., 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. Scandinavian Journal of Statistics 8, 93–115.

Creal, D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Journal of Business & Economic Statistics 29, 552–563.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of empirical finance 1, 83–106.

Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of garch processes. Econometrica 61, 909–927.

Drost, F.C., Werker, B.J., 1996. Closing the garch gap: Continuous time garch modeling. Journal of Econometrics 74, 31 – 57.

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

Duffie, D., Glynn, P., 2004. Estimation of continuous-time markov processes sampled at random time intervals. Econometrica 72, 1773–1808.

Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20, 339–350.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007.

Engle, R.F., Lee, G.G., 1996. 11 - Estimating diffusion models of stochastic volatility, in: Rossi, P.E. (Ed.), Modelling Stock Market Volatility. Academic Press, San Diego, pp. 333 – 355.

Ethier, S.N., Kurtz, T.G., 1986. Markov processes: characterization and convergence. John Wiley & Sons.

Fornari, F., Mele, A., 1996. Modeling the changing asymmetry of conditional variances. Economics Letters 50, 197–203.

Fornari, F., Mele, A., 2006. Approximating volatility diffusions with cev-arch models. Journal of Economic Dynamics and Control 30, 931 – 966.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

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

Hansen, L., Scheinkman, J., 1995. Back to the future: Generating moment implications for continuous-time markov processes. Econometrica 63, 767–804.

Harvey, A., Lange, R.J., 2017. Volatility modeling with a generalized t distribution. Journal of Time Series Analysis 38, 175–190.

Harvey, A., Lange, R.J., 2018. Modeling the interactions between volatility and returns using egarch-m. Journal of Time Series Analysis 39, 909–919.

Harvey, A.C., 2013. Dynamic models for volatility and heavy tails: with applications to financial and economic time series. volume 52. Cambridge University Press.

Hobson, D.G., Rogers, L.C., 1998. Complete models with stochastic volatility. Mathematical Finance 8, 27–48.

Jeantheau, T., 2004. A link between complete models with stochastic volatility and arch models. Finance and Stochastics 8, 111–131.

Kallsen, J., Taqqu, M.S., 1998. Option pricing in arch-type models. Mathematical Finance 8, 13–26.

Karatzas, I., Shreve, S.E., 1991. Brownian motion and stochastic calculus. 2 ed., Springer.

Kitagawa, G., 1987. Non-gaussian state-space modeling of nonstationary time series. Journal of the American Statistical Association 83, 1032–1041.

Klüppelberg, C., Lindner, A., Maller, R., 2004. A continuous-time garch process driven by a lévy process: stationarity and second-order behaviour. Journal of Applied Probability 41, 601–622.

Koopman, S.J., Lucas, A., Scharth, M., 2016. Predicting time-varying parameters with parameter-driven and observation-driven models. The Review of Economics and Statistics 98, 97–110.

Kushner, H.J., 1984. Approximation and weak convergence methods for random processes, with applications to stochastic systems theory. volume 6. MIT press.

Meddahi, N., Renault, E., 2004. Temporal aggregation of volatility models. Journal of Econometrics 119, 355 – 379. Dynamic factor models.

Nelson, D.B., 1990. Arch models as diffusion approximations. Journal of Econometrics 45, 7–38.

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

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

Nelson, D.B., 1996. Asymptotically optimal smoothing with ARCH models. Econometrica 64, 561–573.

Nelson, D.B., Foster, D.P., 1994. Asymptotic filtering theory for univariate ARCH models. Econometrica 62, 1–41.

Protter, P., 1992. Stochastic Integration and Differential Equation. Second ed., Springer-Verlag, Berlin, Heidelberg.

Stentoft, L., 2011. American option pricing with discrete and continuous time models: An empirical comparison. Journal of Empirical Finance 18, 880 – 902.

Stroock, D.W., Varadhan, S.S., 1979. Multidimensional diffusion processes. Springer.

Trifi, A., 2006. Issues of aggregation over time of conditional heteroscedastic volatility models: What kind of diffusion do we recover? Studies in Nonlinear Dynamics & Econometrics 10, 1–24.

Wang, Y., 2002. Asymptotic nonequivalence of garch models and diffusions. Ann. Statist. 30, 754–783.