The continuous limit of weak GARCH

Article

Accepted Version

The Continuous Limit of Weak GARCH

Carol Alexander* and Emese Lazar†

July 7, 2020

Abstract

GARCH models are called 'strong' or 'weak' depending on the presence of parametric distributional assumptions for the innovations. The symmetric weak GARCH$(1,1)$ is the only model in the GARCH class that has been proved to be closed under the temporal aggregation property (Drost and Nijman, 1993). This property is fundamental in two respects: (a) for a time-series model to be invariant to the data frequency; and (b) for a unique option-pricing model to exist as a continuous-time limit. While the symmetric weak GARCH$(1,1)$ is temporally aggregating precisely because it makes no parametric distributional assumptions, the lack of these also makes it harder to derive theoretical results. Rising to this challenge, we prove that its continuous-time limit is a geometric mean-reverting stochastic volatility process with diffusion coefficient governed by a time-varying kurtosis of log returns. When log returns are normal the limit coincides with Nelson's (1990) strong GARCH$(1,1)$ limit. But unlike strong GARCH models, the weak GARCH$(1,1)$ has a unique limit because it makes no assumptions about the convergence of model parameters. The convergence of each parameter is uniquely determined by the temporal aggregation property. Empirical results show that the additional time-varying kurtosis parameter enhances both term-structure and smile effects in implied volatilities, thereby affording greater flexibility for the weak GARCH limit to fit real-world data from option prices.

Keywords: Stochastic volatility, Time aggregation, Diffusion limit, Option pricing, Implied volatility, Simulations.

JEL Classification Codes C32, G13.

*University of Sussex Business School, c.alexander@sussex.ac.uk

†ICMA Centre, Henley Business School, University of Reading, e.lazar@icmacentre.ac.uk

1 Introduction

The class of 'weak' GARCH processes is characterised by the absence of parametric conditional distributions of the errors, with the familiar autoregressive equation being defined for the best linear predictor of the residuals rather than their conditional variance. The well-known GARCH$(1,1)$ models introduced by Bollerslev(1986) are called 'strong' because they make specific parametric assumptions about the innovations of the process. Drost and Nijman (1993) introduce the class of weak GARCH models and prove that the symmetric weak GARCH$(1,1)$ process satisfies the temporal aggregation property. Thus, doubling or halving the sampling frequency doesn't change the model class, it remains a weak GARCH$(1,1)$ process. There are many other time-series models whose structure is likewise invariant to time partitions of the stochastic process on which it is defined, and there is a voluminous literature on this topic.[1] For example, all models in the autoregressive moving average class are also closed under temporal aggregation – see Tsai and Chan (2005) and many others since.

However, few models for the conditional variance of a stochastic process, rather than the process itself, possess this temporal aggregation property.[2] Drost and Nijman (1993) show that strong GARCH$(1,1)$ models do not have this property. Thus, for instance, consider the classic case of the symmetric normal GARCH$(1,1)$ stochastic process and suppose we simulate observations at one frequency and then re-sample from these at a different frequency. Then the conditional variance of the re-sampled process need not follow a symmetric normal GARCH$(1,1)$ process – indeed it need not follow any type of GARCH(p,q) process at all.

We focus on the temporal aggregation property because it is necessary for the derivation of a unique continuous-time limit. In fact, we believe it is controversial to even consider the derivation of a limit if the model class is not closed under temporal aggregation. A case in point is the classic work of Nelson (1990) which derives a limit of symmetric normal GARCH$(1,1)$ as a two-factor stochastic price and variance process having independent Brownian motions.[3] But this limit is not unique because assumptions about the convergence rate of the parameters have to be made, and different assumptions lead to different limits. For instance, Corradi (2000) chooses different assumptions and derives a strong GARCH limit which has a deterministic variance. Yet another problem arises because a discretized version of either limit need not be a strong GARCH process, because the strong GARCH class is not closed under temporal aggregation.

In this paper we show that the continuous limit of a symmetric weak GARCH$(1,1)$ process is a two-factor stochastic volatility model similar to Nelson's limit but the diffusion coefficient in the variance equation is related to a time-varying kurtosis of the distribution of log returns. It is

[1] In a related strand of the literature, Alexander and Rauch (2020) classify all functions of multivariate stochastic processes having time-series *estimates* that are independent of data frequency, requiring only the estimates to be time aggregating and not the entire model class to be closed under temporal aggregation.

[2] It is not known whether weak GARCH(p,q) processes could be closed under temporal aggregation for $p,q>1$.

[3] We use the term 'price' here and in the following to refer to the price of a financial asset, because the GARCH literature has developed mainly within the context of financial econometrics. Typically, the stochastic process in discrete time is for the log return, defined as the difference in log prices, and in continuous time the stochastic process models the price of the asset.

much more flexible than Nelson's limit which has a constant kurtosis of 3. The additional time-varying kurtosis parameter endows the weak GARCH limit with greater flexibility to fit observed data in the implied risk-neutral measure, i.e. data on option prices or, more specifically, their implied volatilities. In the special case that log returns are normally distributed our limit reduces to Nelson's strong GARCH diffusion. However, log returns almost always have leptokurtic distributions in real-world applications and therefore it comes as no surprise to find that Nelson's limit cannot fit implied volatility smiles as well as the weak GARCH limit. We also specify a practical discretization of the limit model that is a symmetric weak GARCH$(1,1)$ process, and use this to simulate the prices of options on the underlying process. There is no ambiguity about parameter convergence when deriving the weak GARCH limit. It follows uniquely and directly from the definition of the weak GARCH process and this is the reason why the weak GARCH limit is unique. Therefore, knowledge of the discrete-time GARCH(1,1) parameters at only one frequency can determine the coefficients of the GARCH limit process.

The remainder of this paper is organized as follows: Section 2 reviews the literature on continuous limits of other GARCH processes; Section 3 presents the weak GARCH$(1,1)$ process and discusses its properties; Section 4 derives the continuous limit, shows that it is unique and has a time-varying conditional kurtosis as an explicit parameter in the variance process. Then we specify a discretization of the continuous model; Section 5 presents our empirical results. First we compare simulations from the weak GARCH diffusion with those of Nelson's strong GARCH diffusion. Then we apply the limit model to fit implied volatility smiles using data from traded option prices; Section 6 concludes.

2 Continuous Limits of GARCH Processes

The particular form of strong GARCH$(1,1)$ limit derived by Nelson (1990) may be extended to derive limits for other GARCH(p,q) processes, see Lindner (2009) for a brief overview. Also, Hafner (2017) derives a limit of a particular multivariate GARCH model. Trifi (2006) extends the paper of Nelson (1990) to non-normal distributions, and considers some augmented GARCH processes as well as the CEV-GARCH$(1,1)$ model of Fornari and Mele (2005). However, none of these limits may be regarded as unique because it is still necessary to make assumptions about the convergence rates of the model parameters. As stated above, it makes little sense to consider these limits of models that are not closed under temporal aggregation. Nevertheless, there are some other GARCH-type models for which limits could be sensibly stated. In particular, Meddahi and Renault (2004) introduce a new class of square-root stochastic autoregressive volatility models that is closed under temporal aggregation. Their model is a natural extension of the weak GARCH class which allows for skewness and leverage effects, i.e. both the asymmetries that are excluded from weak GARCH models. This class includes strong GARCH processes, but there is no closed sub-group for GARCH processes alone. In other words, taking a GARCH as the process for some frequency, then for any other frequency we have another model in Meddahi and Renault's class, but not necessarily a GARCH process.

Replacing the innovations in a GARCH process by jumps of Lévy processes, Kluppelberg, Lindner and Maller (2004) introduce the COGARCH continuous-time process which features similar properties to GARCH but the residuals follow a Lévy process. Here the variance is not continuous either, but has jumps. Indeed, Kallsen and Vesenmayer (2008) show that any COGARCH process can be represented as the limit in law of a sequence of GARCH$(1,1)$ processes. They also argue, yet only heuristically, that COGARCH and the GARCH(1,1) diffusion limit of Nelson (1990) are probably the only continuous-time limits of strong GARCH processes. In support of this Maller et al. (2008) confirm that COGARCH and Nelson's diffusion limit are the only functional continuous-time limits of GARCH in distribution, showing convergence in probability to COGARCH. They also claim that COGARCH can reproduce more of the stylized facts in financial time series. Buchmann and Müller (2012) show that a GARCH process converges generically to a COGARCH process, provided that the volatility processes are observed. They argue that the COGARCH process can be considered as a continuous-time equivalent of the discrete GARCH$(1,1)$ process, even though they are not actually deriving the limit.

Badescu, Elliott and Ortega (2014) consider minimum-variance and local risk minimizing hedging strategies for diffusion limits of a class of asymmetric, non-Gaussian GARCH option pricing models. Badescu, Elliott and Ortega (2015) obtain weak limits of its discretized version under the same parameter convergence assumptions in both the physical and risk-neutral measures. The continuous limit has a variance diffusion coefficient that depends on both the skewness and kurtosis of the distribution of log returns. Subsequently, Badescu, Cui and Ortega (2017) demonstrate the advantages of extending this class of GARCH processes using a pricing kernel with stochastic equity and variance risk preference parameters, and again they derive the corresponding diffusion limit.

Despite this prolific strand of the literature, the existence and uniqueness of a continuous limit of even the simplest, symmetric GARCH$(1,1)$ process remains obscure. As mentioned in the introduction, there are (at least) two possible limits for the strong GARCH (1,1) process, derived by Nelson (1993) and Corradi (2000) respectively, where each limit employs different assumptions about convergence of model parameters. The COGARCH literature reviewed above supports Nelson's diffusive variance limit. However, in favour of Corradi's limit it can be argued that discrete time GARCH has only one source of randomness whilst Nelson's limit has two sources. Furthermore, Wang (2002) uses the asymptotic non-equivalence of the likelihood functions to demonstrate that the continuous limit of normal GARCH$(1,1)$ cannot be a diffusion. Brown, Wang and Zhao (2002) consider stronger parameter convergence conditions and again show that there can be no diffusion term in the continuous limit of multiplicative GARCH models, as described above. Mele and Fornari (2000) consider the continuous limit of A-PARCH models where the error term follows a GED distribution, and Zheng (2005) study the limit of the HARCH type processes proposed by Muller et al. (1997). Its limit has deterministic variance because the parameter convergence conditions chosen are similar to those of Corradi (2000).

In related work, Drost and Werker (1996) introduce continuous-time symmetric GARCH

diffusion and jump-diffusion processes that exhibit weak GARCH-type behaviour at all discrete frequencies. Their discretization scheme depends on the unconditional kurtosis which is assumed constant and strictly greater than 3.

3 The Weak GARCH$(1,1)$ Process

For a detailed discussion of the classes of weak, semi-strong and strong GARCH processes see Drost and Nijman (1993). In this section we only consider the weak GARCH$(1,1)$ process, because this is the only GARCH process that is known to satisfy the temporal aggregation property. We will show that this property allows us to derive a unique continuous-time limit.

Following Engle (1982) and Bollerslev (1986) the GARCH$(1,1)$ process for a log return y_t can be written as $y_t = \mu + \varepsilon_t \quad$ with $E\left(\varepsilon_{t+1} \middle| I_t\right) = 0$, where I_t is the σ-algebra generated by the residual vector (ε_t). The classical or strong GARCH definition states

$$E\left(\varepsilon_{t+1}^2 \middle| I_t\right) = h_t, \tag{1}$$

where h_t is the conditional variance. In the symmetric version of strong and weak GARCH(1,1) we assume $h_t = \omega + \alpha\varepsilon_t^2 + \beta h_{t-1}$. But in the weak GARCH process h_t is the best linear predictor (BLP) of the squared residuals, and not the conditional variance, replacing (1) with

$$E\left(\varepsilon_{t+1}\varepsilon_{t-i}^r\right) = 0 \quad i \geq 0 \quad r = 0,1,2; \qquad E\left(\left(\varepsilon_{t+1}^2 - h_t\right)\varepsilon_{t-i}^r\right) = 0 \quad i \geq 0 \quad r = 0,1,2.$$

The assumption that 0 and h_t are the BLPs for the residuals and squared residuals respectively, guarantees that the BLP of the squared residuals aggregates in time, but only for *symmetric* processes (Drost and Nijman, 1993). That is, time-aggregating weak GARCH models can have no leverage effect, and since there is no parametric distribution for the innovations, they have no explicitly-modelled asymmetry. By contrast with skewness, the kurtosis plays an important role, as we shall see later.

For a finite step-length Δ (please note that we don't use Δ as an operator, but as a simple notation for the step-length) we consider the Δ-step process for the residuals and the GARCH process. For processes in general, time is indexed as $s\Delta$, for $s = 1, 2, \ldots$. Also, for both parameters (such as ${}_\Delta\omega, {}_\Delta\alpha$ and ${}_\Delta\beta$) and processes (such as ${}_\Delta h_{s\Delta}$) the pre-subscript signifies the time step used for their estimation and, to be able to compare variances for different step-lengths, we standardise the BLP series by dividing the squared returns by the step-length. Thus ${}_\Delta h_{s\Delta}$ denotes the BLP for $\Delta^{-1}{}_\Delta\varepsilon_{s\Delta}^2$.

Using ${}_\Delta\lambda = {}_\Delta\alpha + {}_\Delta\beta$, for $i \geq 0$, and for $r = 0,1,2$ the annualised weak GARCH process may be written

$$\begin{aligned} {}_\Delta y_{s\Delta} &= \Delta\mu + {}_\Delta\varepsilon_{s\Delta}, \qquad & {}_\Delta h_{s\Delta} &= {}_\Delta\omega + {}_\Delta\alpha\Delta^{-1}{}_\Delta\varepsilon_{s\Delta}^2 + {}_\Delta\beta{}_\Delta h_{(s-1)\Delta}, \\ E\left({}_\Delta\varepsilon_{(s+1)\Delta}\,{}_\Delta\varepsilon_{(s-i)\Delta}^r\right) &= 0, \qquad & E\left(\left(\Delta^{-1}{}_\Delta\varepsilon_{(s+1)\Delta}^2 - {}_\Delta h_{s\Delta}\right){}_\Delta\varepsilon_{(s-i)\Delta}^r\right) &= 0. \end{aligned} \tag{2}$$

Only one lag can be used here, otherwise we would obtain a higher-order GARCH process; but it has not been proved that GARCH(p,q) processes satisfy the temporal aggregation property.

The first paper that discusses the continuous limit of GARCH is Nelson (1990). Under the conditions

$$\omega = \lim_{\Delta\downarrow 0}\left(\Delta^{-1}{}_{\Delta}\omega\right); \quad \alpha = \lim_{\Delta\downarrow 0}\left(\Delta^{-1/2}{}_{\Delta}\alpha\right); \quad \theta = \lim_{\Delta\downarrow 0}\left(\Delta^{-1}\left(1-{}_{\Delta}\lambda\right)\right); \quad 0<\omega,\alpha,\theta<\infty$$

the limit will be a stochastic volatility model with independent Brownians, i.e.

$$\begin{aligned} dS_t &= \mu\, S_t dt + \sqrt{V_t}\, S_t dB_{1t}, \\ dV_t &= (\omega - \theta V_t)\, dt + \sqrt{2}\alpha V_t dB_{2t}. \end{aligned}$$

where V_t is the continuous-time limit of h_t. On the other hand, Corradi (2000) proves that, if we assume the following convergence rates

$$\omega = \lim_{\Delta\downarrow 0}\left(\Delta^{-1}{}_{\Delta}\omega\right); \quad \alpha = \lim_{\Delta\downarrow 0}\left(\Delta^{-1}{}_{\Delta}\alpha\right); \quad \theta = \lim_{\Delta\downarrow 0}\left(\Delta^{-1}\left(1-{}_{\Delta}\lambda\right)\right); \quad 0<\omega,\alpha,\theta<\infty$$

then the continuous-time limit is a deterministic variance model with the same price dynamics but with

$$dV_t = (\omega - \theta V_t)\, dt.$$

The difference between the two assumptions lies with the convergence of alpha (at rate $\sqrt{\Delta}$ versus rate Δ). Which assumption is correct has been the subject of considerable debate. Here we argue that the assumptions of Nelson are correct, but we promote a different continuous limit because, as argued above, it is best to use a time aggregating model, viz. weak GARCH(1,1).

For a weak GARCH$(1,1)$ using two different step-lengths Δ and δ, $\delta<\Delta$, Drost and Nijman (1993) proved the following relationship between the parameters

$${}_{\Delta}\omega = {}_{\delta}\omega\left(1-\left({}_{\delta}\lambda\right)^{\delta^{-1}\Delta}\right)\left(1-{}_{\delta}\lambda\right)^{-1} \quad \text{and} \quad {}_{\Delta}\alpha = \left({}_{\delta}\lambda\right)^{\delta^{-1}\Delta} - {}_{\Delta}\beta.$$

The relationship between the unconditional kurtosis coefficients, namely between the kurtosis for the Δ-step process denoted ${}_{\Delta}\kappa$, and the kurtosis for the δ-step process denoted ${}_{\delta}\kappa$, is

$${}_{\Delta}\kappa = 3 + \Delta^{-1}\delta\left({}_{\delta}\kappa - 3\right) + 6\left({}_{\delta}\kappa - 1\right)\frac{\left(\delta^{-1}\Delta\left(1-{}_{\delta}\lambda\right) - \left(1-{}_{\delta}\lambda^{\delta^{-1}\Delta}\right)\right){}_{\delta}\alpha\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha{}_{\delta}\lambda\right)}{\left(\delta^{-1}\Delta\right)^2\left(1-{}_{\delta}\lambda\right)^2\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha^2\right)}. \quad (3)$$

Drost and Nijman (1993) derive the following

$${}_{\Delta}\beta\left(1+{}_{\Delta}\beta^2\right)^{-1} = \left({}_{\Delta,\delta}c_{\delta}\lambda^{\delta^{-1}\Delta}-1\right)\left({}_{\Delta,\delta}c\left(1+{}_{\delta}\lambda^{2\delta^{-1}\Delta}\right)-2\right)^{-1}$$

as the relationship between the low and high frequency parameters for λ and β, where

$$
{}_{\Delta,\delta}c = \left[\begin{array}{l} \delta^{-1}\Delta(1-{}_{\delta}\beta)^2 + 2\delta^{-1}\Delta\left(\delta^{-1}\Delta-1\right)(1-{}_{\delta}\lambda)\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha^2\right)({}_{\delta}\kappa-1)^{-1}(1+{}_{\delta}\lambda)^{-1} \\ +4\left(\delta^{-1}\Delta(1-{}_{\delta}\lambda)-\left(1-{}_{\delta}\lambda^{\delta^{-1}\Delta}\right)\right){}_{\delta}\alpha(1-{}_{\delta}\beta_{\delta}\lambda)\left(1-{}_{\delta}\lambda^2\right)^{-1} \end{array} \right] \times
$$

$$
\left[{}_{\delta}\alpha(1-{}_{\delta}\beta_{\delta}\lambda)\left(1-{}_{\delta}\lambda^{2\delta^{-1}\Delta}\right)\left(1-{}_{\delta}\lambda^2\right)^{-1}\right]^{-1}. \quad (4)
$$

To derive the continuous limit of this model we are interested in the inverse relationship expressing the high frequency (δ-step) parameters and their limit based on the low frequency (Δ-step) parameters, for $\delta < \Delta$. This relationship can be derived as

$$
{}_{\delta}\omega = {}_{\Delta}\omega\left(1-{}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)(1-{}_{\Delta}\lambda)^{-1} \quad \text{and} \quad {}_{\delta}\lambda^{\delta^{-1}} = {}_{\Delta}\lambda^{\Delta^{-1}}
$$

Also

$$
\left(2\left(\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2}\right)-1\right){}_{\delta}\alpha(1-{}_{\delta}\beta_{\delta}\lambda)\left(\frac{1-{}_{\Delta}\lambda^2}{1-{}_{\delta}\lambda^2}\right) =
$$

$$
= \left(\left(\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2}\right)(1+{}_{\Delta}\lambda^2)-{}_{\Delta}\lambda\right)\left(\begin{array}{l} \Delta\delta^{-1}(1-{}_{\delta}\beta)^2 + 2\Delta\delta^{-1}\left(\Delta\delta^{-1}-1\right)({}_{\delta}\kappa-1)^{-1}\left(\frac{1-{}_{\delta}\lambda}{1+{}_{\delta}\lambda}\right)\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha^2\right) \\ +4\left(1-{}_{\delta}\lambda^2\right)^{-1}{}_{\delta}\alpha\left(\Delta\delta^{-1}(1-{}_{\delta}\lambda)-(1-{}_{\Delta}\lambda)\right)(1-{}_{\delta}\beta_{\delta}\lambda) \end{array} \right). \quad (5)
$$

and

$$
{}_{\delta}\kappa = 1+\left(\frac{\left(2+\delta^{-1}\Delta\left({}_{\Delta}\kappa-3\right)\right)\left(1-{}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)\left(1-{}_{\Delta}\lambda^{2\Delta^{-1}\delta}+{}_{\delta}\alpha^2\right)}{\left(\left(1-{}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)\left(1-{}_{\Delta}\lambda^{2\Delta^{-1}\delta}+{}_{\delta}\alpha^2\right)\right)+6{}_{\delta}\alpha\left(1-\Delta^{-1}\delta(1-{}_{\Delta}\lambda)\left(1-{}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)^{-1}\right)\left(1-{}_{\delta}\beta_{\Delta}\lambda^{\Delta^{-1}\delta}\right)}\right). \quad (6)
$$

4 Continuous Limit of Weak GARCH$(1,1)$

The continuous limit may not offer equivalence with the discrete-time model. For equivalence, a discretization of the continuous limit should yield the same class of model as the original. Furthermore, the discretized model must be the same for all frequencies, which cannot happen without the temporal aggregation property. Thus, it is only when (1) the original discrete-time model is time aggregating, and (2) the model can be discretized at any frequency in the form of the original model, that we have an equivalence between discrete and continuous models.

4.1 Parameter Limits

The first step for deriving the continuous limit of symmetric weak GARCH(1,1) is to determine the limits of the parameters and their convergence speeds. In contrast to the strong GARCH(1,1) process, where there is some freedom to choose assumptions about parameter convergence speeds, we now show that assumptions about parameter convergence are unnecessary for weak GARCH(1,1) processes. In fact, the temporal aggregation property of weak GARCH(1,1) implies unique convergence speeds for all parameters. We prove this result in the following:

Proposition 1 The convergence rates for the parameters implied by the weak GARCH(1,1) process are as follows:

$$\omega = \lim_{\Delta\downarrow 0} \Delta^{-1}{}_{\Delta}\omega; \quad \alpha = \lim_{\Delta\downarrow 0} \Delta^{-1/2}{}_{\Delta}\alpha; \quad \theta = \lim_{\Delta\downarrow 0} \Delta^{-1}\left(1 - {}_{\Delta}\lambda\right); \quad 0 < \omega, \alpha, \theta < \infty.$$

Also, the unconditional kurtosis converges to $\kappa = \lim_{\Delta\downarrow 0} {}_{\Delta}\kappa = 3\big(1 - \theta^{-1}\alpha^2\big)^{-1}$.

Proof We have ${}_{\delta}\lambda^{\delta^{-1}} = {}_{\Delta}\lambda^{\Delta^{-1}}$ which, being a constant between 0 and 1, can be denoted $\exp(-\theta)$ with $\theta > 0$. Thus

$${}_{\Delta}\lambda = \exp(-\theta\Delta) \quad \text{and} \quad \lim_{\Delta\downarrow 0} \Delta^{-1}\left(1 - {}_{\Delta}\lambda\right) = \lim_{\Delta\downarrow 0} \Delta^{-1}\left(1 - \exp(-\theta\Delta)\right) = \theta.$$

Also ${}_{\delta}\omega(1 - {}_{\delta}\lambda)^{-1} = {}_{\Delta}\omega(1 - {}_{\Delta}\lambda)^{-1}$ is a positive constant denoted $\omega\theta^{-1}$, $\omega > 0$ and ${}_{\Delta}\omega = \omega\theta^{-1}\left(1 - {}_{\Delta}\lambda\right)$, so

$$\lim_{\Delta\downarrow 0} \left(\Delta^{-1}{}_{\Delta}\omega\right) = \omega\theta^{-1} \lim_{\Delta\downarrow 0} \Delta^{-1}\left(1 - \exp(-\theta\Delta)\right) = \omega.$$

Formula (6) for the kurtosis may now be written

$${}_{\delta}\kappa = 1 + \left({}_{\Delta}\kappa - 3 + 2\Delta^{-1}\delta\right)\left(\Delta^{-1}\delta + 6\frac{\left(\delta^{-1}\left(1 - {}_{\Delta}\lambda^{\Delta^{-1}\delta}\right) - \Delta^{-1}\left(1 - {}_{\Delta}\lambda\right)\right)}{\Delta\left(\delta^{-1}\left(1 - {}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)\right)^2}{}_{\Delta,\delta}A\right)^{-1},$$

with

$${}_{\Delta,\delta}A = \frac{{}_{\delta}\alpha\delta^{-1}\left(1 - {}_{\Delta}\lambda^{2\Delta^{-1}\delta}\right) + \delta^{-1}{}_{\delta}\alpha^2 - \delta^{-1}{}_{\delta}\alpha^2\left(1 - {}_{\Delta}\lambda^{\Delta^{-1}\delta}\right)}{\delta^{-1}\left(1 - {}_{\Delta}\lambda^{2\Delta^{-1}\delta}\right) + \delta^{-1}{}_{\delta}\alpha^2}. \tag{7}$$

But

$$\lim_{\delta\downarrow 0} \delta^{-1}\left(1 - {}_{\Delta}\lambda^{\Delta^{-1}\delta}\right) = \theta \quad \text{and} \quad \lim_{\delta\downarrow 0} \delta^{-1}\left(1 - {}_{\Delta}\lambda^{2\Delta^{-1}\delta}\right) = 2\theta.$$

Thus, using ${}_{\delta}\alpha \downarrow 0$, we have

$$\lim_{\delta\downarrow 0} {}_{\Delta,\delta}A = \left(2\theta\left(\lim_{\delta\downarrow 0}\left(\delta^{-1}{}_{\delta}\alpha^2\right)\right)^{-1} + 1\right)^{-1}.$$

Hence, taking the limit of (7) as $\delta \downarrow 0$ and then letting $\Delta \downarrow 0$ yields the limit of the kurtosis as

$$\kappa = 3\left(1 - \theta^{-1}\lim_{\delta\downarrow 0}\left(\delta^{-1}{}_{\delta}\alpha^2\right)\right)^{-1}. \tag{8}$$

The limit of the unconditional kurtosis must be finite and positive, which forces

$$0 \leq \lim_{\delta\downarrow 0} \delta^{-1}{}_{\delta}\alpha^2 < \theta,$$

so the kurtosis will be greater than 3.

To see the speed of convergence for α we consider the limit $\alpha = \lim_{\delta \downarrow 0} \delta^{-w}{}_{\delta}\alpha$ with $\alpha \in (0, \infty)$ with w unknown. Since $\lim_{\delta \downarrow 0} \delta^{-1}{}_{\delta}\alpha^2 < \theta$, $w \geq 1/2$, for $y = \min(w, 1)$ and $z = \min(2w, 1)$ we write

$$\begin{aligned}
\lim_{\delta \downarrow 0} \delta^{-y} \left(1 - {}_{\delta}\lambda + {}_{\delta}\alpha\right) &= \lim_{\delta \downarrow 0} \delta^{-y} \left(1 - {}_{\delta}\lambda\right) + \lim_{\delta \downarrow 0} \delta^{-y}{}_{\delta}\alpha \in (0, \infty), \\
\lim_{\delta \downarrow 0} \delta^{-y} \left(1 - {}_{\delta}\lambda^2 - {}_{\delta}\alpha \left(1 - {}_{\delta}\lambda\right) + {}_{\delta}\alpha\right) &= \lim_{\delta \downarrow 0} \delta^{-y} \left(1 - {}_{\delta}\lambda^2\right) + \lim_{\delta \downarrow 0} \delta^{-y}{}_{\delta}\alpha \in (0, \infty), \\
\lim_{\delta \downarrow 0} \delta^{-z} \left(1 - {}_{\delta}\lambda^2 + {}_{\delta}\alpha^2\right) &= \lim_{\delta \downarrow 0} \delta^{-z} \left(1 - {}_{\delta}\lambda^2\right) + \lim_{\delta \downarrow 0} \delta^{-z}{}_{\delta}\alpha^2 \in (0, \infty).
\end{aligned}$$

Also, using (5) and noting that ${}_{\delta}\kappa \neq 1$, since ${}_{\delta}\alpha^2 > 0$, we can compute

$$\left(2\left(\frac{{}_{\Delta}\beta}{1 + {}_{\Delta}\beta^2}\right) - 1\right) {}_{\delta}\alpha \left(1 - {}_{\delta}\lambda^2 - {}_{\delta}\alpha \left(1 - {}_{\delta}\lambda\right) + {}_{\delta}\alpha\right) \left(\frac{1 - {}_{\Delta}\lambda^2}{1 - {}_{\delta}\lambda^2}\right).$$

If $w > 1/2$, we can multiply the above expression by δ^{1-w-y} and then computing the limit as δ tends to zero leads to a contradiction in terms of limits. So we must have $w = 1/2$ and this sets the convergence of α. □

4.2 Model Convergence

Now consider the conditional variance and the conditional kurtosis of the residuals where the conditional mean and skewness are equal to zero

$${}_{\Delta}\sigma^2_{s\Delta} = E\left(\Delta^{-1}\left({}_{\Delta}\varepsilon_{(s+1)\Delta}\right)^2 \middle| {}_{\Delta}I_{s\Delta}\right),$$

$${}_{\Delta}\kappa_{s\Delta} = E\left(\Delta^{-2}{}_{\Delta}\sigma^{-4}_{s\Delta}\left({}_{\Delta}\varepsilon_{(s+1)\Delta}\right)^4 \middle| {}_{\Delta}I_{s\Delta}\right),$$

where ${}_{\Delta}I_{s\Delta}$ is the σ-algebra generated by the vector $\left({}_{\Delta}\varepsilon_{s\Delta}\right)$. We divide by Δ when computing the conditional variance series, in order to standardize them, so that the variance over Δ is comparable with Δ times the 1-step variance.

The conditional expectation of the second moment and the kurtosis must be positive, and due to the symmetric nature of the returns we can write

$$E\left(\Delta^{-1}{}_{\Delta}\varepsilon^2_{(s+1)\Delta} \middle| I_{s\Delta}\right) = {}_{\Delta}\sigma^2_{s\Delta}. \tag{9}$$

Note that ${}_{\Delta}\sigma^2_{s\Delta} - {}_{\Delta}h_{s\Delta}$ is non-zero, otherwise the process is a semi-strong GARCH (Drost and Nijman, 1993). Next, for $s\Delta \leq t < (s+1)\Delta$ write $\kappa(t) = \lim_{\Delta \downarrow 0} {}_{\Delta}\kappa_t$ where ${}_{\Delta}\kappa_t = {}_{\Delta}\kappa_{s\Delta}$ and write

${}_{\Delta}h_t = {}_{\Delta}h_{s\Delta}$ and $\Delta\sigma_t^2 = {}_{\Delta}\sigma_{s\Delta}^2$. Because $\lim_{\Delta\downarrow 0}\left({}_{\Delta}\sigma_t^2 - {}_{\Delta}h_t\right) = 0$ we can set

$$V(t) = \lim_{\Delta\downarrow 0} {}_{\Delta}h_t = \lim_{\Delta\downarrow 0} {}_{\Delta}\sigma_t^2.$$

We must now conjecture that the difference between the conditional variance and the BLP of the squared residuals converges to zero at a speed of square root of the time step, i.e.

$$\lim_{\Delta\downarrow 0} \Delta^{-1/2}\left({}_{\Delta}\sigma_t^2 - {}_{\Delta}h_t\right) = 0.$$

Our justification is that the BLP process becomes more and more informative as the time step decreases, so it must converge very fast to the conditional variance.

Theorem 1 The continuous limit of the weak GARCH(1,1) process defined in (2) is the following stochastic volatility model, based on the limiting parameters in Proposition 1 and above:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sqrt{V_t}\, dB_{1t}, \\ dV_t &= (\omega - \theta V_t)\, dt + \alpha\sqrt{(\kappa_t - 1)}\, V_t\, dB_{2t}, \end{aligned}$$

where B_{1t} and B_{2t} are independent Brownian motions and κ_t is the conditional kurtosis, which may be time-varying.

Proof We employ the convergence theorem for stochastic difference equations to stochastic differential equations given by Nelson (1990). For the returns process we have

$$E\left(\Delta^{-1}\,{}_{\Delta}y_{(s+1)\Delta}\middle|\, I_{s\Delta}\right) = \mu + E\left(\Delta^{-1}{}_{\Delta}\varepsilon_{(s+1)\Delta}|I_{s\Delta}\right) = \mu.$$

And, using (9) it can be shown that $E\left(\Delta^{-1}\,{}_{\Delta}y_{(s+1)\Delta}^2\middle|\, I_{s\Delta}\right) = {}_{\Delta}h_{s\Delta} + o(1)$,

$$E\left(\Delta^{-1}\left({}_{\Delta}h_{(s+1)\Delta} - {}_{\Delta}h_{s\Delta}\right)\middle|\, I_{s\Delta}\right) = \Delta^{-1}{}_{\Delta}\omega - \Delta^{-1}(1 - {}_{\Delta}\lambda)\,{}_{\Delta}h_{s\Delta} + \left(\Delta^{-1/2}{}_{\Delta}\alpha\right)\Delta^{-1/2}\left({}_{\Delta}\sigma_{s\Delta}^2 - {}_{\Delta}h_{s\Delta}\right) + o(1)$$

and this converges to $\omega - \theta V_t$ by Proposition 1. The variance of the variance component is

$$E\left(\Delta^{-1}\left({}_{\Delta}h_{(s+1)\Delta} - {}_{\Delta}h_{s\Delta}\right)^2\middle|\, I_{s\Delta}\right) = \Delta^{-1}{}_{\Delta}\alpha^2\left(E\left(\left({}_{\Delta}\sigma_{s\Delta}^4\right)\left(\Delta^2{}_{\Delta}\sigma_{s\Delta}^4\right)^{-1}\left({}_{\Delta}\varepsilon_{(s+1)\Delta}^4\right) - {}_{\Delta}h_{s\Delta}^2\middle|\, I_{s\Delta}\right)\right) + o(1).$$

The covariance between the returns and the changes in the variances converges as follows

$$E\left(\Delta^{-1}\,{}_{\Delta}y_{(s+1)\Delta}\left({}_{\Delta}h_{(s+1)\Delta} - {}_{\Delta}h_{s\Delta}\right)\middle|\, I_{s\Delta}\right) = o(1).$$

Therefore, the limits of the expected squared terms and cross-product between the returns and

variance derived above define the following covariance matrix of the continuous process

$$A_t = \begin{pmatrix} V_t & 0 \\ 0 & \alpha^2 (\kappa_t - 1) V_t^2 \end{pmatrix}. \qquad \square$$

Discussion: Discrete-time weak GARCH(1,1) processes are characterized by (i) the existence of a long-term volatility; (ii) mean reversion in the variance process; (iii) the variance is stochastic, i.e. the variance of variance is non-zero; and (iv) the variance process is uncorrelated with the returns process, which is an implication of the symmetry of the returns' distribution, being a requirement of weak GARCH(1,1) processes. All these properties are present in our limit model which is stated in Theorem 1. In addition, the variance in our model has a higher variance than the variance of the variance process in the limit of Nelson (1990). This is a consequence of the extra kurtosis parameter, which can also be time-varying.

When $\kappa = 3$ the limit process reduces to the diffusion derived by Nelson (1990) and in this case we obtain the smallest value of the volatility of the variance process, i.e. $2^{1/2}\alpha V_t$. Drost and Werker (1996) have conjectured that the conditional kurtosis is independent of t. However, our limit allows the conditional kurtosis to be time-varying.

Finally we note that properties of our limit are not only more flexible than those of Nelson's limit; our model also captures the observed behaviour of implied volatilities in the risk-neutral measure better. See for example, Bates (1997, 2000) and Bakshi et al. (2003). In particular, Section 5 will demonstrate the superior ability of our limit for capturing the smile and term-structure features of implied volatilities.

4.3 Discretization Scheme

Here we find a discretization of the continuous limit when the series of returns and variances are discretized by assuming a time step of length Δ and considering changes at time $t = s\Delta$. Let

$$dt \mapsto \Delta, \quad \frac{dS_t}{S_t} \mapsto {}_\Delta y_{(s+1)\Delta}, \quad V_t \mapsto {}_\Delta V_{s\Delta}, \quad \text{and} \quad dV_t \mapsto {}_\Delta V_{(s+1)\Delta} - {}_\Delta V_{s\Delta}.$$

The conditional kurtosis $\kappa_t \to {}_\Delta\kappa_{s\Delta}$ for $s\Delta \leq t < (s+1)\Delta$ and the parameter μ stays unchanged during discretization. The rest of the parameters are discretized as

$$\omega \mapsto \Delta^{-1}{}_\Delta\omega, \quad \theta \mapsto \Delta^{-1}(1 - {}_\Delta\lambda), \quad \alpha \mapsto \Delta^{-1/2}{}_\Delta\alpha,$$

where $({}_\Delta\omega, {}_\Delta\alpha, {}_\Delta\beta)$ are specified in terms of the parameters (ω, θ, α) of the continuous model and the unconditional kurtosis κ, which is given in Proposition 1, as follows

$$
\begin{aligned}
{}_{\Delta}\omega &= \omega\theta^{-1}\left(1-{}_{\Delta}\lambda\right), \\
{}_{\Delta}\lambda &= \exp\left(-\theta\Delta\right), \\
{}_{\Delta}\alpha &= {}_{\Delta}\lambda - {}_{\Delta}\beta, \\
{}_{\Delta}\kappa &= 3+2\left(\kappa-3\right)\left(\theta\Delta-\left(1-{}_{\Delta}\lambda\right)\right)\theta^{-2}\Delta^{-2},
\end{aligned}
$$

and

$$
{}_{\Delta}\beta = \frac{1}{2}\frac{{}_{\Delta}c\left(1+{}_{\Delta}\lambda^{2}\right)-2+\left(1-{}_{\Delta}\lambda\right)\left({}_{\Delta}c^{2}(1+{}_{\Delta}\lambda)^{2}-4{}_{\Delta}c\right)^{1/2}}{{}_{\Delta}c{}_{\Delta}\lambda-1}, \tag{10}
$$

where

$$
{}_{\Delta}c = 2\,\frac{\left[\Delta\alpha^{2}+2\alpha^{2}\left(\Delta-\theta^{-1}\left(1-{}_{\Delta}\lambda\right)\right)+\Delta^{2}\theta\left(\theta-\alpha^{2}\right)\right]}{\left[\alpha^{2}\theta^{-1}\left(1-{}_{\Delta}\lambda^{2}\right)\right]}. \tag{11}
$$

An alternative but equivalent version of (11) may be derived, after some algebra, viz.

$$
{}_{\Delta}c = 2\,\frac{3\Delta\theta-2\left(1-{}_{\Delta}\lambda\right)+3\Delta^{2}\theta^{2}\left(\kappa-3\right)^{-1}}{1-{}_{\Delta}\lambda^{2}} \tag{12}
$$

where the uncondtional kurtosis κ is defined in Proposition 1. The Brownian motions B_1 and B_2 that drive the price and variance equations are discretized by expressing the changes in the Brownian motions at time $t = s\Delta$ as

$$
\begin{aligned}
B_1\left(\left(s+1\right)\Delta\right)-B_1\left(s\Delta\right) &= \Delta^{1/2}{}_{\Delta}\xi_{(s+1)\Delta}, \\
B_2\left(\left(s+1\right)\Delta\right)-B_2\left(s\Delta\right) &= \Delta^{1/2}{}_{\Delta}\eta_{(s+1)\Delta},
\end{aligned}
$$

where ${}_{\Delta}\xi_{(s+1)\Delta}$is a standard normal variable, ${}_{\Delta}\xi_{(s+1)\Delta}\,|\,{}_{\Delta}I_{s\Delta}\sim N\left(0,1\right)$ and ${}_{\Delta}\eta_{(s+1)\Delta}$is defined as

$$
{}_{\Delta}\eta_{(s+1)\Delta} = 2^{-1/2}\left({}_{\Delta}\xi^{2}_{(s+1)\Delta}-1\right). \tag{13}
$$

Now define the normal variable ${}_{\Delta}\varepsilon_{(s+1)\Delta} = \Delta^{1/2}{}_{\Delta}V^{1/2}_{s\Delta}{}_{\Delta}\xi_{(s+1)\Delta}$ and set ${}_{\Delta}\tilde{\varepsilon}_{s\Delta} = G^{-1}\left[F\left({}_{\Delta}\varepsilon_{s\Delta}\right)\right]$, where F is the normal distribution and G is the distribution for a variable ${}_{\Delta}\tilde{\varepsilon}_{s\Delta}$ that has zero mean and variance $\Delta_{\Delta}V_{s\Delta}$, like ${}_{\Delta}\varepsilon_{s\Delta}$, but kurtosis equal to ${}_{\Delta}\kappa_{s\Delta}$. This way, the errors of the discretized model have non-zero excess kurtosis.

Discussion : The continuous model has two independent sources of randomness yet the discrete model has only one. This is similar to the technique used by Nelson (1990) whereby one source of randomness in discrete time translates to two sources of randomness in continuous time, one for the changes in the variable and the other for changes in the variance process.

Our discretization reduces the number of sources of randomness in the continuous model, via (13). There is no loss of generality using this discretization since the properties of the discretized Brownian motion (mean, variance and correlation) are maintained; ${}_{\Delta}\eta_{(s+1)\Delta}$ is not

exactly normal but it has a zero conditional mean, a unit conditional variance and zero correlation with ${}_{\Delta}\xi_{(s+1)\Delta}$. We are bound to use such a method because the classic discretization does not work in this case, as argued by Lindner (2009, p. 482).[4]

The discretization (12) for β is useful because it makes the dependence on kurtosis explicit at the same time as showing that the scheme does not work when the unconditional kurtosis $\kappa = 3$. We employ (12) in Section 5.1 when we fix parameters and simulate from the discretized model and in Section 5.2 when we calibrate model parameters to real-world data. Finally, we note that there are other discretization schemes to the one above and we tested some of these but they did not return the weak GARCH$(1,1)$ process.

Theorem 2 The above discretization scheme of the weak GARCH limit in Theorem 1 returns a weak GARCH(1,1) process and the time aggregation property is preserved.

Proof The discretization of μdt is obvious, and that for θ and ω will follow from (10) as

$$\omega\Delta \approx \omega\Delta\frac{(1-{}_{\Delta}\lambda)}{\theta\Delta} = {}_{\Delta}\omega \quad \text{and} \quad \theta\Delta \approx 1-{}_{\Delta}\lambda = 1-({}_{\Delta}\alpha+{}_{\Delta}\beta) = 1-{}_{\Delta}\lambda.$$

This gives ${}_{\Delta}\omega = \omega\theta^{-1}(1-{}_{\Delta}\lambda)$ and ${}_{\Delta}\lambda = {}_{\Delta}\lambda$. As is clear from (10) and (11), it is the discretization of β that is most complex. From the aggregation results in Drost and Nijman (1993) we know that the unconditional kurtosis for a given frequency Δ may be expressed as a function of the parameters at an arbitrary higher frequency δ as

$${}_{\Delta}\kappa = 3+\Delta^{-1}\delta\left({}_{\delta}\kappa-3\right)+6\left({}_{\delta}\kappa-1\right)\frac{\left(\delta^{-1}\Delta\left(1-{}_{\delta}\lambda\right)-\left(1-{}_{\delta}\lambda^{\delta^{-1}\Delta}\right)\right){}_{\delta}\alpha\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha{}_{\delta}\lambda\right)}{\left(\delta^{-1}\Delta\right)^2\left(1-{}_{\delta}\lambda\right)^2\left(1-{}_{\delta}\lambda^2+{}_{\delta}\alpha^2\right)}.$$

Denoting the limit of the unconditional kurtosis by $\kappa = \lim_{\delta\downarrow 0}{}_{\delta}\kappa$, we obtain

$${}_{\Delta}\kappa = 3+2\left(\kappa-3\right)\frac{\left(\theta\Delta-\left(1-{}_{\Delta}\lambda\right)\right)}{\theta^2\Delta^2}. \tag{14}$$

From the proof of Proposition 1, we know that for any two time steps $\Delta > \delta$, ${}_{\Delta}\beta$ is the solution to

$$\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2} = \frac{{}_{\Delta,\delta}c_{\delta}\lambda^{\delta^{-1}\Delta}-1}{{}_{\Delta,\delta}c\left(1+{}_{\delta}\lambda^{2\delta^{-1}\Delta}\right)-2}, \tag{15}$$

where ${}_{\Delta,\delta}c$ is given by (4). We want a discretization which ensures that (15) will hold. Taking the limits of (4) when δ goes to 0, we define ${}_{\Delta}c = \lim_{\delta\downarrow 0}{}_{\Delta,\delta}c$ as in either (11) or (12). This means

[4]The direct relationship (13) between the increments of B_{1t} and the squared increments of B_{2t} shows that they have zero correlation – but they are not independent because they have the same source of randomness – just as a standard normal variable is uncorrelated with its square.

that we can discretize the continuous model by solving the following equation

$$\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2}=\frac{{}_{\Delta}c_{\Delta}\lambda-1}{{}_{\Delta}c\left(1+{}_{\Delta}\lambda^2\right)-2}.$$

First, we must ensure that this will have solutions, and then we have to show that there is a unique solution between zero and one. Let's consider the function whose roots we want to find

$$f\left(x\right)=x^2-mx+1,\quad m=\frac{{}_{\Delta}c\left(1+{}_{\Delta}\lambda^2\right)-2}{{}_{\Delta}c_{\Delta}\lambda-1}.$$

This has two roots x_1 and x_2 where $x_1x_2=1$ and $x_1+x_2=m$. If we show that m is positive, then both roots are positive and one will be less than 1. For the existence we need that $m>2$. If ${}_{\Delta}c_{\Delta}\lambda>1$ then $m>2$ is equivalent to $\left(1-{}_{\Delta}\lambda\right)^2>0$. Thus, all we need to show is that ${}_{\Delta}c_{\Delta}\lambda>1$, which is equivalent to $6\theta\Delta+2\alpha^{-2}\Delta^2\theta^3+5{}_{\Delta}\lambda>\exp\left(\theta\Delta\right)+2\Delta^2\theta^2+4$.

Both sides of the above equation converge to 5 when $\Delta\to0$, and it can be shown, using derivatives with respect to Δ, that the left hand side converges faster. Thus ${}_{\Delta}c_{\Delta}\lambda>1$, so for any small step Δ close enough to zero there will always be a unique solution for ${}_{\Delta}\beta$ between zero and one that satisfies the above equation; this solution will be

$${}_{\Delta}\beta=\frac{1}{2}\left(m-\left(m^2-4\right)^{1/2}\right)=\frac{{}_{\Delta}c\left(1+{}_{\Delta}\lambda^2\right)-2+\left(1-{}_{\Delta}\lambda\right)\left({}_{\Delta}c^2(1+{}_{\Delta}\lambda)^2-4{}_{\Delta}c\right)^{1/2}}{2\left({}_{\Delta}c_{\Delta}\lambda-1\right)}.$$

Also, we have that ${}_{\Delta}\alpha={}_{\Delta}\lambda-{}_{\Delta}\beta$. The discretization of the Brownian motions in our scheme is obvious, whilst there is no loss of generality in assuming (13).

Now, we have $y_{(s+1)\Delta}=\mu\Delta+{}_{\Delta}\varepsilon_{(s+1)\Delta}$. Hence $E\left(\Delta^{-1}{}_{\Delta}\varepsilon^2_{(s+1)\Delta}|{}_{\Delta}I_{s\Delta}\right)={}_{\Delta}V_{s\Delta}$ with

$${}_{\Delta}V_{(s+1)\Delta}={}_{\Delta}\omega+{}_{\Delta}\alpha\Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^2_{(s+1)\Delta}+{}_{\Delta}\beta{}_{\Delta}V_{s\Delta}+{}_{\Delta}u_{(s+1)\Delta},$$

$${}_{\Delta}u_{(s+1)\Delta}={}_{\Delta}\alpha{}_{\Delta}V_{s\Delta}\left[1-\left(2^{-1}\left({}_{\Delta}\kappa_{s\Delta}-1\right)\right)^{1/2}+\left(\left(2^{-1}\left({}_{\Delta}\kappa_{s\Delta}-1\right)\right)^{1/2}-1\right){}_{\Delta}\tilde{\xi}^2_{(s+1)\Delta}\right],$$

where ${}_{\Delta}\tilde{\xi}_{(s+1)\Delta}$ has an unconditional kurtosis of ${}_{\Delta}\kappa_{(s+1)\Delta}$, which can be approximated by ${}_{\Delta}\kappa_{s\Delta}$. The above may also be written

$${}_{\Delta}u_{(s+1)\Delta}={}_{\Delta}\alpha{}_{\Delta}V_{s\Delta}\left[\left(\left(2^{-1}\left({}_{\Delta}\kappa_{s\Delta}-1\right)\right)^{1/2}-1\right)\left({}_{\Delta}\tilde{\xi}^2_{(s+1)\Delta}-1\right)\right].$$

So far we have considered the conditional variance. For the BLP of squared residuals we have

$${}_{\Delta}h_{s\Delta}={}_{\Delta}V_{s\Delta}-\sum_{j=0}^{k}{}_{\Delta}\beta^j{}_{\Delta}u_{(k-j)\Delta}.$$

It is easy to see that this follows a GARCH(1,1) process since

$${}_{\Delta}h_{s\Delta}={}_{\Delta}\omega+{}_{\Delta}\alpha\left(\Delta^{-1}{}_{\Delta}\varepsilon^2_{s\Delta}\right)+{}_{\Delta}\beta{}_{\Delta}h_{(s-1)\Delta}.$$

For a weak GARCH(1,1) we must show that ${}_{\Delta}h_{s\Delta}$ is the BLP of $\Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^2_{(s+1)\Delta}$ which requires showing

$$E\left(\left(\Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^2_{(s+1)\Delta} - {}_{\Delta}h_{s\Delta}\right){}_{\Delta}\tilde{\varepsilon}^r_{(s-i)\Delta}\right) = 0 \quad \text{for} \quad i \geq 0, r = 0, 1, 2.$$

Since $E\left(\left({}_{\Delta}V_{(s-j-1)\Delta} - \Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^2_{(s-j)\Delta}\right){}_{\Delta}\tilde{\varepsilon}^r_{(s-i)\Delta}\right) = 0$, this reduces to proving that

$$E\left({}_{\Delta}u_{(s-j)\Delta\Delta}\tilde{\varepsilon}^r_{(s-i)\Delta}\right) = 0.$$

This is clearly satisfied for $i \neq j$ so we now prove it for $r = 1$ and $i = j$. We have to show that

$$E\left(\left({}_{\Delta}V_{(s-i-1)\Delta} - \Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^2_{(s-i)\Delta}\right){}_{\Delta}\tilde{\varepsilon}_{(s-i)\Delta}\right) = 0, \quad i \geq 0$$

or

$$E\left(E\left({}_{\Delta}V_{(s-i-1)\Delta\Delta}\tilde{\varepsilon}_{(s-i)\Delta} - \Delta^{-1}{}_{\Delta}\tilde{\varepsilon}^3_{(s-i)\Delta}\right)|{}_{\Delta}I_{(s-i-1)\Delta}\right) = 0, \quad i \geq 0,$$

which is clearly true. Also

$$E\left({}_{\Delta}\tilde{\varepsilon}^2_{(s-i)\Delta}\middle|\,{}_{\Delta}I_{(s-i-1)\Delta}\right) = \Delta_{\Delta}V_{(s-i-1)\Delta}$$

and

$$E\left({}_{\Delta}\tilde{\varepsilon}^4_{(s-i)\Delta}\middle|\,{}_{\Delta}I_{(s-i-1)\Delta}\right) = \Delta^2{}_{\Delta}V^2_{(s-i-1)\Delta\Delta}\kappa_{(s-i-1)\Delta}.$$

Thus, we have a weak GARCH(1,1) specification; this means that the time aggregation is preserved by our discretization. It is easy to see that ${}_{\Delta}\lambda^{\Delta^{-1}} = {}_{\delta}\lambda^{\delta^{-1}}$ and that ${}_{\Delta}\omega = {}_{\delta}\omega\left(1 - {}_{\delta}\lambda^{\delta^{-1}\Delta}\right)(1 - {}_{\delta}\lambda)$. We also have the relations (14) for the kurtosis and (15) for β. For the kurtosis, we need to prove (3), that is

$$(\kappa - 1)\,\alpha^2(1 - {}_{\delta}\lambda)^2\left(1 - {}_{\delta}\lambda^2 + {}_{\delta}\alpha^2\right) = \left(2\delta^2\theta^2\left(\alpha^2 + 2\theta\right) + 6\,(\kappa - 1)\,\alpha^2\left(\theta\delta - (1 - {}_{\delta}\lambda)\right)\right){}_{\delta}\alpha\left(1 - {}_{\delta}\lambda + {}_{\delta}\alpha_{\delta}\lambda\right).$$

After some algebra, this may be written as

$$\frac{{}_{\delta}\beta}{1 + {}_{\delta}\beta^2} = \frac{\left(\delta\alpha^2 + \delta^2\theta\left(\theta - \alpha^2\right) + 2\alpha^2\left(\delta - (1 - {}_{\delta}\lambda)/\theta\right)\right){}_{\delta}\lambda - {}^1\!/\,{}_2\alpha^2\theta^{-1}\left(1 - {}_{\delta}\lambda^2\right)}{\left(\delta\alpha^2 + \delta^2\theta\left(\theta - \alpha^2\right) + 2\alpha^2\left(\delta - (1 - {}_{\delta}\lambda)/\theta\right)\right)\left(1 + {}_{\delta}\lambda^2\right) - \alpha^2\theta^{-1}\left(1 - {}_{\delta}\lambda^2\right)}.$$

Since the above expression holds, we have shown that the kurtosis is indeed time aggregating.

Theorem 2 has practical importance because it guides how to discretize the weak GARCH limit when simulating returns. In particular, the parameter β which appears in the original discrete version of the model described in Section 3 does not explicitly appear in the limit model of Theorem 1, it only enters implicitly in the sum $\lambda = \alpha + \beta$ with θ as defined in Proposition 1. However, Theorem 2 shows that, although β does not explicitly appear in the limit, it governs the way the continuous model should be discretized.

In the next section we explore two ways in which such simulations can be applied. The first

example in subsection 5.1 takes the limit model parameters as given, discretizes the model and then uses it to generate implied volatility smiles. The second example in subsection 5.2 takes implied volatility smiles as the inputs and uses simulations of the discretized version of the model to calibrate the parameters of the model by matching the simulated smile to the observed smile.

5 Empirical Results

The two most common financial applications of continuous-time stochastic processes are (i) to price non-standard options with complex payoffs that are too illiquid to have a market price and (ii) to hedge any options. Here we consider the second of these applications, for hedging standard European puts and calls, and we do this by simulating and fitting implied volatility 'smiles'.[5] The closer a model fits market implied volatility the better its hedging performance. And since a better hedging performance determines the profits made by option market makers,[6] the ability of a model to fit implied volatility smiles is important in practice.

It is preferable to use a hedging model that is more sophisticated than the basic option-pricing model of Black and Scholes (1973). The Black-Scholes model assumes prices are driven by a simple geometric Brownian motion and its constant volatility assumption yields an implied volatility surface that is flat, i.e. it does not vary with either the option strike or the option's maturity. To fit the 'smile' shape that is ubiquitous when implied volatilities are backed out from real-world option prices requires a stochastic volatility model. But finding one stochastic volatility model that can fit an entire smile surface is very challenging. That is, we seek a single stochastic volatility model which has a parameterization that is flexible enough to calibrating all parameters, in a single optimization, using smiles of different maturities together. Here we demonstrate that the time-varying kurtosis parameter of the weak GARCH limit admits this property.

To this end, we focus on the ability of the weak GARCH limit to capture different shapes of smile features in implied volatility. We consider the model derived in Theorem 1 with $\mu = 0$, viz.

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{V_t}\, dB_{1t}, \\ dV_t &= (\omega - \theta V_t)\, dt + \alpha \sqrt{(\kappa_t - 1)}\, V_t\, dB_{2t}, \end{aligned} \tag{16}$$

since the drift does not affect the fit. Note that (16) encompasses Nelson's limit for strong GARCH$(1,1)$ as the special case that the conditional kurtosis κ_t is constant and equal to 3.

The variance diffusion coefficient in (16), i.e. the product $\alpha\sqrt{(\kappa_t - 1)}$, is colloquially termed the 'vol-of-vol' part of the model because it controls the volatility of the volatility. It has two

[5] The implied volatility smile is obtained by putting the market prices of all options on a given underlying, of different strikes but the same maturity, into the Black-Scholes formula and 'backing out' the diffusion coefficient in the geometric Brown motion that is implied by each observed option price. The smile refers to the shape that is usually seen when implied volatilities are plotted as a function of option strike.

[6] Market makers seek to minimize their spreads because this makes them more competitive, but spreads can only be narrowed when the market maker uses an accurate hedging model. See Alexander et. al. (2012) for further information.

parameters, α and κ_t that do not explicitly appear elsewhere in (16). However, the choice of α affects the discretization scheme derived in Section 4.3 and ${}_{\Delta}\beta$ will depend on the α parameter. Also, the unconditional kurtosis in the limit model is $3/(1-\alpha^2\theta^{-1})$, as in Proposition 1, and so we must have $\alpha^2 < \theta$ in order that this is positive. The parameters ω and θ control the long-term volatility and the speed of mean reversion to it. Specifically the long-term variance is ω/θ and the mean-reversion speed increases with θ.

In subsection 5.1 we explore the effect that different parameters in our limit model (16) have on the shape of the smile, as well as the effect of using different discretization schemes. In subsection 5.2 we demonstrate the gains that can result from using our weak GARCH limit rather than Nelson's strong GARCH limit, by calibrating (16) to implied volatilities backed out from real-world data on option prices.

5.1 Simulations

Weak GARCH processes are a general model class, and since they make no distributional assumptions, the simulation of the discretized limit model is non-trivial because it is based on approximations made to the error term based on normally distributed returns (discretization of Brownian motions). These approximations affect the simulations and it should be noted that the simulated process is in the end an approximation of weak GARCH.

Table 1 defines 12 scenarios that explore the effect of changing the parameters of the model (16) and the expiry date of the options. For the first three sets of simulations we have a constant expiry date of 0.1 years. In the fourth set we explore the term structure effects of weak GARCH implied volatilities and we set the conditional kurtosis $\kappa_t = t^{-1}$, so that $\kappa = 20, 10, 4$ at expiry times $0.05, 0.1$ and 0.25 years.[7]

Next, with each of these 12 different sets of parameters, we employ the discretization scheme of Section 4.3 to simulate prices using the weak GARCH limit. For each parameter set we simulate the process with 100 steps, 50,000 times and thereby simulate a distribution of possible values of the underlying of the options, at the expiry time shown in the table, each time starting at $S_0 = 100$. From this distribution we apply the call option pay-off to simulate an option price for strikes between 65 and 125, at increments of 5 and then for each option price we use the Black-Scholes formula to back-out the implied volatility via a standard search algorithm. The results are displayed in Figures 1 and 2.

The top left graph in Figure 1 is based on scenario set 1 and it displays how the smile effect increases with kurtosis:, which is 3 in the blue smile, 10 in the red smile and 20 in the green smile. This effect is much more pronounced in scenario set 2, in which the value of α changes from 1 to 2. This higher value for α is associated with a greater vol-of-vol, especially when combined with high values of κ_t. In both graphs the blue line (i.e. scenarios 1a and 2a) correspond to Nelson's limit, because $\kappa_t = 3$. The green and the red lines depict the smiles from our weak

[7] It is common to use the term 'instantaneous' in the literature on continuous-time stochastic processes, in place of the term 'conditional' typically used in discrete-time processes. However, we shall continue to use the term conditional kurtosis rather than instantaneous kurtosis in this Section, and for the rest of the paper.

Table 1: Scenarios for Parameters in Model (16)

Parameters above the line are set using scenarios for parameters below the line. The vol-of-vol is $\alpha\sqrt{(\kappa_t - 1)}$, the unconditional kurtosis is $3/(1-\alpha^2\theta^{-1})$, the unconditional volatility is $\sqrt{\omega/\theta}$. The starting values for S_t and V_t are $S_0 = 100$ and $V_0 = 1.5\,\omega/\theta$. This way V_0 is 50% greater than the long-term variance. Note that setting $\kappa_t = 3$ as in scenarios 1a, 2a and 3a corresponds to Nelson's limit for strong GARCH.

	κ_t			α and θ		
Scenario number	1a	1b	1c	2a	2b	2c
V_0	0.375	0.375	0.375	0.375	0.375	0.375
Vol of vol	1.41	3.00	4.36	2.83	6.00	8.72
Unconditional kurtosis	6	6	6	5	5	5
Unconditional vol	50%	50%	50%	50%	50%	50%
Expiry	0.1	0.1	0.1	0.1	0.1	0.1
α	1	1	1	2	2	2
θ	2	2	2	10	10	10
ω	0.5	0.5	0.5	2.5	2.5	2.5
κ_t	3	10	20	3	10	20
	V_0 and ω			Expiry		
Scenario number	3a	3b	3c	4a	4b	4c
V_0	0.75	0.75	0.75	0.375	0.375	0.375
Vol-of-vol	1.41	3.00	4.36	4.36	3.00	1.73
Unconditional kurtosis	6	6	6	6	6	6
Unconditional vol	71%	71%	71%	50%	50%	50%
Expiry	0.1	0.1	0.1	0.05	0.1	0.25
α	1	1	1	1	1	1
θ	2	2	2	2	2	2
ω	1	1	1	0.5	0.5	0.5
κ_t	3	10	20	20	10	4

GARCH limit, which are clearly much more flexible and we observe the smile effect increasing with kurtosis.

The top right graph in Figure 1 illustrates the effect of using a discretization different from the one derived in Section 4.3. The simulations are based on identical parameters to those used for the top left graph, i.e. scenario set 1 in Table 1, but this time we employ a standard discretization scheme for the Brownian motion in the variance process. The kurtosis has less effect on the implied volatility skew now, but we do not find this surprising because the discretized model is no longer a weak GARCH$(1,1)$ under standard discretization. Finally, the bottom right graph depicts the smiles obtained using scenario 3. Here we increase ω which increases the volatility, hence all smiles shift upwards.

Figure 2 explores the volatility term-structure effects in the weak GARCH limit, using the scenario set 4 in Table 1 where the expiry date varies from 0.05 of a year (blue) to 0.1 of a year (red) to 0.25 of a year (green). The conditional kurtosis used in our discretization varies also,

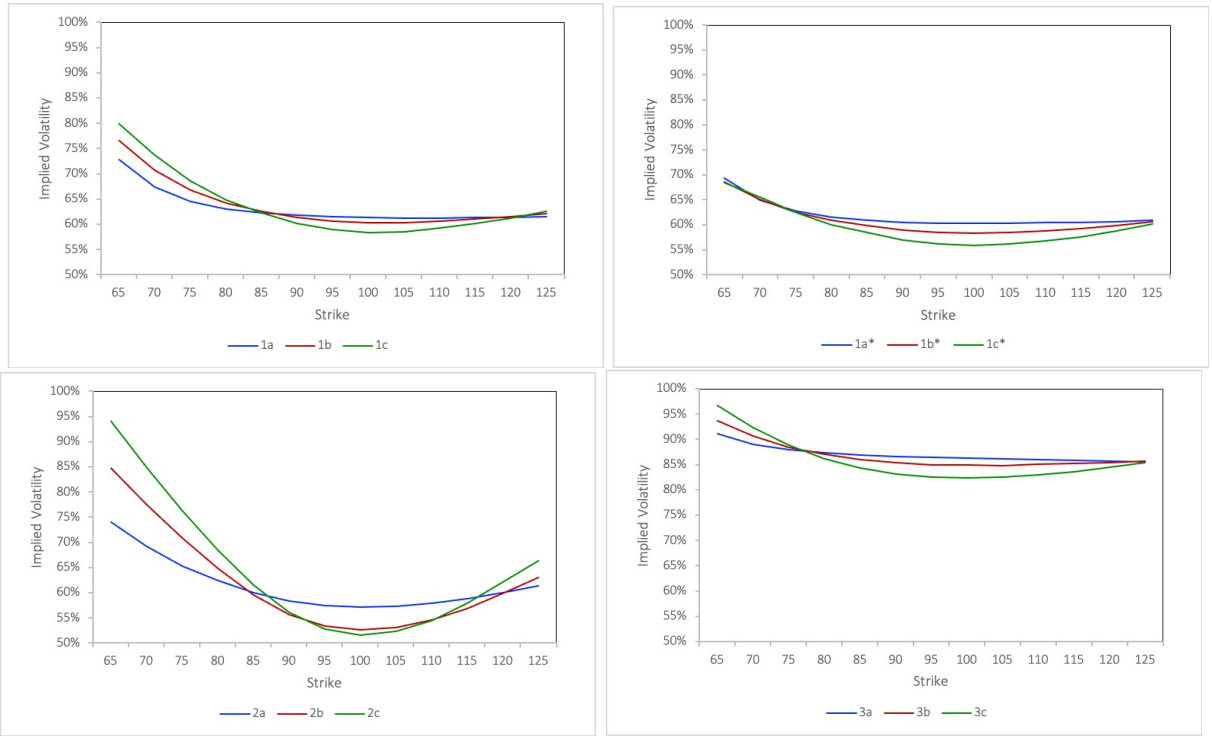

Figure 1: Implied Volatility Smiles Generated by Parameter Values in Table 1
Comparison of volatility smiles of the same maturities generated by the weak GARCH limit model with parameter values given by scenario sets 1, 2 and 3 in Table 1. Results from scenarios labelled a are shown in blue, the b scenarios are shown in red and those labelled c are shown in green. Non-starred curves are derived using simulations based on the discretization scheme in Section 4.3 where kurtosis is constant. The curves 1a*, 1b* and 1c* (shown top right) are derived using the parameter values in scenario set 1, but here we employ the standard discretization for Brownian motions.

as shown in the table. These implied volatility smiles exhibit a 'skew' effect where low strike options have much higher implied volatilities than high-strike options, especially for near-term expiry dates. This is a common feature of equity index options in practice.

The middle graph in Figure 2 employs the standard discretization instead of the one derived in Section 4.3. However, the discretized version of (16) is not a weak GARCH model in this case and again there is a noticeable lessening of the skew effect. This shows that the wrong discretization can produce misleading results. The bottom graph in Figure 2 again depicts three different smiles from the weak GARCH limit, similar to scenarios 4 in Table 1. But this time $\kappa_t = 3 + \eta t^{-1}$ with $\eta = 0.75$, i.e. we parametrize a fully time-varying conditional kurtosis in the discretization instead of assuming it is constant. This time-varying kurtosis allows the smiles to become more symmetric – a feature is often observed in practice, particularly in currency option markets.

5.2 Calibrations

In this section we calibrate the parameters of the weak and strong GARCH limits to some real data on option prices of different strikes and expiry. We expect that the additional time-varying

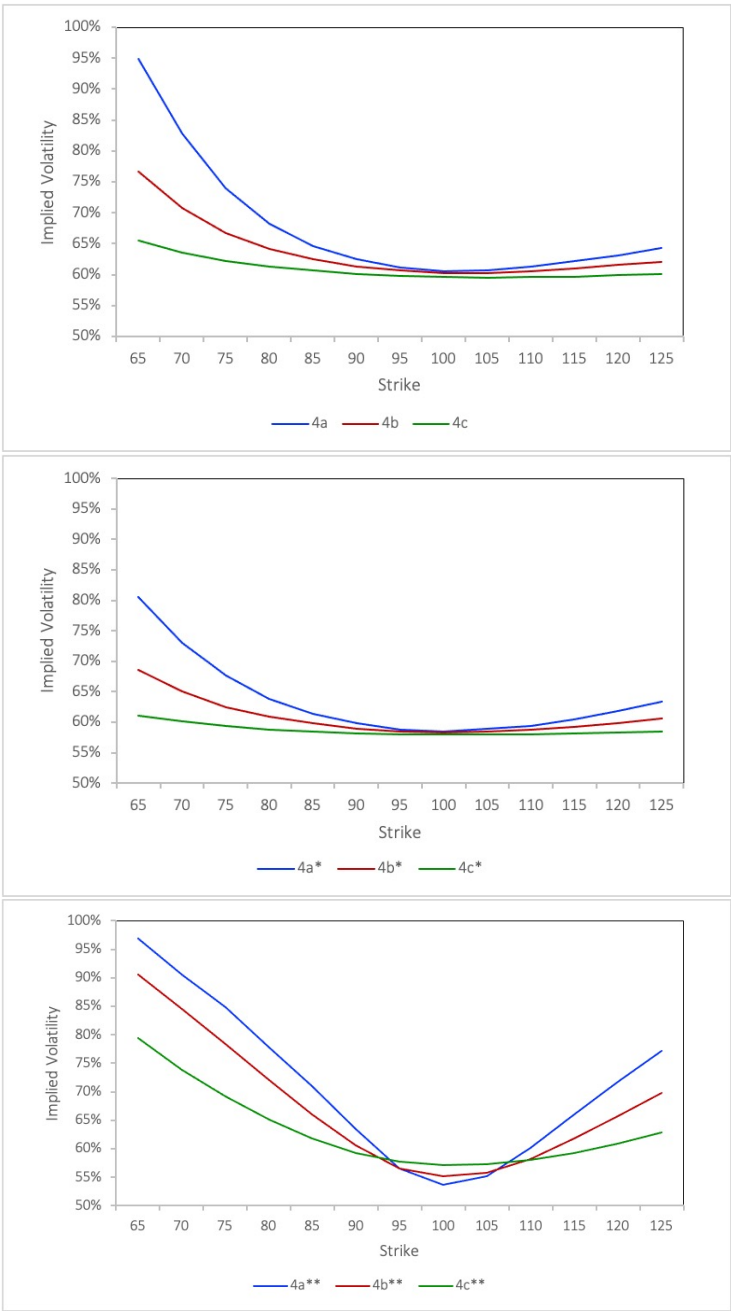

Figure 2: Effect of Discretization on Implied Volatility Smiles

Comparison of volatility smiles of different maturities generated by the weak GARCH limit model with parameter values given by scenario set 4 in Table 1. Results from scenarios labelled a are shown in blue, the b scenarios are shown in red and those labelled c are shown in green. Non-starred curves in the top graph are derived using simulations based on the discretization scheme in Section 4.3 where the conditional kurtosis is constant and κ takes the values shown in scenario 4 of Table 1. The curves 4a*, 4b* and 4c* (middle graph) are based on the same parameters but are derived using the standard discretization for Brownian motions and the curves 4a**, 4b** and 4c** (bottom graph) are based on the discretization scheme in Section 4.3 where kurtosis is time varying of the form $\kappa_t = 3 + \eta t^{-1}$ with $\eta = 0.75$ so that $\kappa = 18, 10.5$ and 6 at expiries $0.5, 0.1$ and 0.25 years respectively. All other parameters are as in scenario set 4 of Table 1.

conditional kurtosis in (16) offers an extra degree of flexibility to match implied volatility smiles more closely. Because currency options typically have deep, symmetric smile effects, and because the cryptocurrency bitcoin is one of the most volatile of all currencies, we have obtained data from the Deribit bitcoin options exchange on the implied volatilities of all bitcoin options that were actively traded on 17 April 2020, when the spot price of bitcoin was almost exactly $7000. There were two actively traded maturities on that day, one expiring on 24 April 2020 (7 days) and the other expiring on 29 May 2020 (42 days).

The two bitcoin option smiles are depicted in the top left graph in Figure 3. The other graphs in Figure 3 repeat the market data on implied volatilities and compare these with the smiles fitted by calibrating parameters of (16). The calibration of model parameters is done my minimizing the root mean square error (RMSE) between the market and model smiles and the resulting fitted smiles are depicted using dashed lines of the same colour as the market smiles. Note that parameters are calibrated using both smiles simultaneously in the RMSE minimization.

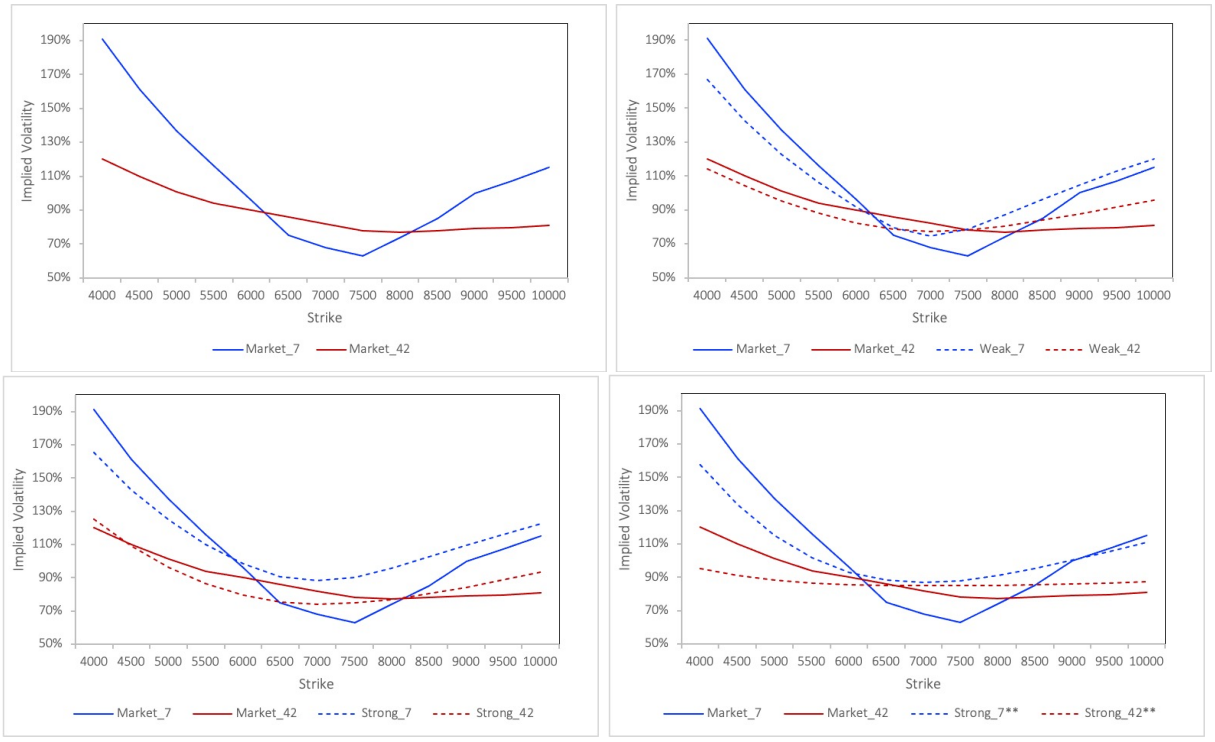

Figure 3: Calibration of GARCH Limit Models to Bitcoin Implied Volatility Smiles
Market smiles are derived from bitcoin option prices on 17 April 2020 at maturities 7 days (blue) and 42 days (red). The top left shows the smiles implied by market prices and the top right depicts the fit obtained using the weak GARCH model (in dotted lines of the same colour as the maturity). The lower two graphs compare the smiles calibrated using strong GARCH when parameters are not constrained (bottom left) and when we constrain the model to have positive unconditional kurtosis (bottom right) . All models are calibrated to both smiles simultaneously.

The top right graph exhibits the fit obtained using the weak GARCH model with time-varying conditional kurtosis and the bottom two graphs depict the smiles obtained by calibrating Nelson's strong GARCH limit with conditional kurtosis fixed at 3. Note that the unconditional kurtosis

is not equal to 3 in Nelson's limit because the volatility is time varying. Indeed, it is equal to $3/(1-\alpha^2\theta^{-1})$ as proved in Proposition 1. Therefore we may need to restrict the values of α and θ so that the unconditional kurtosis is positive.

Table 2: Calibration Results from Different Models
The first column gives the model name, with ** indicating the positive unconditional kurtosis constraint imposed, the second to sixth columns report the fitted model parameters and the last column reports the root mean square error (RMSE) between the calibrated smiles and the fitted smiles. Both smiles were calibrated simultaneously. The parameter η is for the time-varying conditional kurtosis of the weak GARCH limit, where $\kappa_t = 3+\eta t^{-1}$. The fitted unconditional kurtosis is calculated as $\hat{\kappa} = 3/(1-\hat{\alpha}^2\hat{\theta}^{-1})$.

Model	$\hat{\omega}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\kappa}$	RMSE
Weak	2.25	108.71	6.62	0.0248	5.024	0.1023
Strong	7.46	12.89	8.46	0	-0.659	0.1284
Strong**	76.17	104.35	10.21	0	55734	0.1477

Table 2 reports the calibrated values of the parameters of the weak and strong GARCH models corresponding to the fitted smiles exhibited in Figure 3. We used the classic discretization for the strong GARCH limit and the discretization scheme of Section 4.3 for the weak GARCH limit. The parameter η is for the time-varying conditional kurtosis of the weak GARCH limit, where $\kappa_t = 3+\eta t^{-1}$. The ** indicates the condition of positive unconditional kurtosis is imposed.

The results for the weak GARCH model are sensible. The calibrated unconditional kurtosis is $\hat{\kappa} = 5.024$ and the time-varying kurtosis based on $\hat{\eta} = 0.0248$ is $3 + 0.0248\left(\frac{365}{7}\right) = 4.29$ at 7 days and $3+0.0248\left(\frac{365}{42}\right) = 3.22$ at 42 days. Also, this model gives the smallest RMSE of 0.1023, which is in line with the better fit observed for the weak GARCH model in Figure 3. The strong GARCH model, when calibration is unconstrained, returns a negative value for unconditional kurtosis, with $\hat{\kappa} = -0.659$. We therefore impose the restriction that $\hat{\kappa} > 0$ in the last line of results in Table 2. However, even though the fitted smiles appear reasonable to the eye (see the bottom-right graph of Figure 3) the values of the calibrated parameters are not sensible. The strong GARCH model gives such a high value of $\hat{\alpha}$ precisely because it is not flexible enough to fit the given volatility smiles. Also, the RMSE of 0.1477 is almost 50% larger than the RMSE for the weak GARCH calibration.

6 Conclusions

We present four arguments which motivate the use of weak rather than strong GARCH$(1,1)$ for deriving a limit in distribution: (1) Strong GARCH is not time aggregating, so if we generate a GARCH$(1,1)$ variance process and then re-sample at another frequency the result is no longer a GARCH$(1,1)$ process; (2) The limit of strong GARCH may only be derived by making a specific assumption about the convergence of the parameters and different assumptions lead to different limits; (3) Any discretization of the strong GARCH diffusion is not a GARCH model; and (4) the variance of its variance is either zero or too small to fit the implied volatility skew.

By contrast, a weak GARCH process is time aggregating, it implies the convergence rates

for all parameters and our only assumption in deriving its limit is that the difference between the GARCH BLP process and the conditional variance converges to zero with the square root of the step-length. We prove that the weak GARCH diffusion is a stochastic variance process with independent Brownian motions in which the variance diffusion coefficient is related to the conditional kurtosis and the limit reduces to the GARCH diffusion derived by Nelson (1990) when the excess kurtosis is zero. Our limit is unique and we provide a discretization that returns the original weak GARCH model.

An analysis of simulated and real implied volatility skews from option prices demonstrates that the extra conditional kurtosis parameter that is present in the weak GARCH limit – but not in Nelson's strong GARCH limit – adds considerable flexibility for fitting implied volatility surfaces, even those with very steep skews such as those commonly observed in bitcoin options.

Acknowledgements: We are indebted to Shuyuan Qi of the University of Reading for performed the model calibrations in Section 5.2. Bitcoin option price data were kindly provided by CoinAPI.

References

Alexander, C. and J. Rauch (2020) 'A General Property for Time Aggregation' European Journal of Operational Research. (available online: https://doi.org/10.1016/j.ejor.2019.12.045)

Alexander, C., Rubinov, A., Kalepky, M. and S. Leontsinis (2012) 'Regime-dependent smile-adjusted delta hedging' Journal of Futures Markets 32(3), 202-229

Ardia, D., Bluteau, K. and M. Ruedea (2019) 'Regime Changes in Bitcoin GARCH Volatility Dynamics', Finance Research Letters Vol. 29, 266-271.

Badescu A., Elliott R.J., Ortega J-P (2014) 'Quadratic Hedging Schemes for Non-Gaussian GARCH Models', Journal of Economic Dynamics and Control Vol. 42, 13-32.

Badescu A., Elliott R.J., Ortega J-P (2015) 'Non-Gaussian GARCH Option Pricing Models and their Diffusion Limits', European Journal of Operational Research Vol. 247, 820-830.

Badescu A., Cui. Z., Ortega J-P (2017) 'Non-Gaussian GARCH Option Pricing Models, Variance-Dependent Kernels and Diffusion Limits', Journal of Financial Econometrics, Vol. 15, 602–648.

Bollerslev, T. (1986) 'Generalized Autoregressive Conditional Heteroskedasticity', Journal of Econometrics Vol. 31, 309-328.

Brown, L. D., Y. Wang and L.H. Zhao (2002) 'On the Statistical Equivalence at Suitable Frequencies of GARCH and Stochastic Volatility Models with Suitable Diffusion Models', University of Pennsylvania (working paper).

Buchmann, B. and Müller G. (2012) 'Limit Experiments of GARCH', Bernoulli Vol. 18, 64-99.

Corradi, V. (2000) 'Reconsidering the continuous-time limit of the GARCH$(1,1)$ Process', Journal of Econometrics Vol. 96, 145-153.

Drost, F.C. and Nijman, T.E. (1993) 'Temporal Aggregation of GARCH Processes', Econometrica Vol. 61 (4), 909-927.

Drost, F.C. and Werker, B.J.M. (1996) 'Closing the GARCH Gap Continuous Time GARCH Modelling', Journal of Econometrics Vol. 74, 31-57.

Engle, R.F. (1982) 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', Econometrica Vol. 50 (4), 987-1007.

Fornari, F. and A. Mele (2005) 'Approximating Volatility Diffusions with CEV-ARCH Models', Journal of Economic Dynamics and Control Vol. 30 (6), 931-966.

Hafner, C. M., Laurent, S. and F. Violante (2017). 'Weak Diffusion Limits of Dynamic Conditional Correlation Models.' Econometric Theory Vol. 33 (3), 691-716.

Kallsen J. and Vesenmayer B. (2009) 'CO-GARCH as a Continuous-Time Limit of GARCH(1,1)', Stochastic Processes and their Applications Vol. 119, 74-98.

Kluppelberg, C., Lindner, A. and Maller, R. (2004) 'A Continuous Time GARCH Process Driven by Lévy Process Stationarity and Second Order Behaviour', Journal of Applied Probability Vol. 43 (3), 601-622.

Lindner A. (2009) 'Continuous Time Approximations to GARCH and Stochastic Volatility Models', Handbook of Financial Time Series, Edited by Andersen T., Davis R., Kreiss, J-P. and Mikosch, T. (Springer), 481-496.

Maller, R., Miller, G. and Szimayer, A. (2008) 'GARCH Modelling in Continuous Time for Irregularly Spaced Time Series Data'. Bernoulli, Vol. 14, 519-542.

Meddahi, N. and Renault, E. (2004) 'Temporal Aggregation of Volatility Models', Journal of Econometrics Vol. 19, 355 - 379.

Mele, A. and F. Fornari (2000) 'Stochastic Volatility in Financial Markets Crossing the Bridge to Continuous Time', Kluwer Academic Publishers.

Muller, U.A., M.M. Dacorogna, R. Davé, R.B. Olsen, O.V. Pictett and J.E. Von Weizsäcker (1997) 'Volatilities of Different Time Resolutions – Analyzing the Dynamics of Market Components', Journal of Empirical Finance Vol. 4, 213-239.

Nelson, D.B. (1990) 'ARCH Models as Diffusion Approximations', Journal of Econometrics Vol. 45, 7-38.

Trifi, A. (2006) 'Issues of Aggregation Over Time of Conditional Heteroscedastic Volatility Models What Kind of Diffusion Do We Recover?', Studies in Nonlinear Dynamics and Econometrics Vol. 10 (4), 1314-1323.

Tsai, H. and K.S. Chan. (2005) 'Temporal aggregation of stationary and non-stationary discrete-time processes', Journal of Time Series Analysis, 26(4) 613-624.

Wang, Y. (2002) 'Asymptotic Non-Equivalence of GARCH Models and Diffusions', The Annals of Statistics Vol. 30, 754-783.

Zheng, Z. (2005) 'Re-crossing the Bridge from Discrete Time to Continuous Time Towards a Complete Model with Stochastic Volatility I', available at SSRN http//ssrn.com/abstract=694261.