Weak GARCH Diffusion

Carol Alexander and Emese Lazar
ICMA Centre, University of Reading

Abstract

Discrete time volatility analysis has focussed almost exclusively on GARCH processes, which are very flexible models for time varying conditional variance. The continuous limit of these processes is therefore of considerable interest for continuous time volatility modelling. Unfortunately, progress in this area has been hampered by conflicting results. The limit of the symmetric normal GARCH model is fundamental for limits of other GARCH processes, yet even this has been the point of much debate amongst econometricians. Nelson (1990) derived the limit of the strong GARCH model as a stochastic volatility process that is uncorrelated with the price process, so this limit has limited applicability. However, since the strong GARCH process is not time aggregating one should question whether it is sensible to derive its continuous limit at all. This paper derives the continuous limit of the weak GARCH process, which is time aggregating. Moreover the limit model is a stochastic volatility model with non-zero price volatility correlation in which both the variance diffusion coefficient and the price-volatility correlation are related to the skewness and kurtosis of the physical returns density. When returns are normally distributed this limit model reduces to Nelson's strong GARCH diffusion, however, more generally it has the flexibility to fit most short term volatility skews without adding jumps in either the price or the volatility process.

JEL Classification Codes: C32, G13.
Keywords: GARCH, stochastic volatility, time aggregation, continuous limit

Author Details:

Professor Carol Alexander
Chair of Risk Management and Director of Research,
ICMA Centre, Business School,
The University of Reading, PO Box 242
Reading RG6 6BA, United Kingdom
Email: c.alexander@icmacentre.rdg.ac.uk
Tel: +44 (0)1183 786431 (ICMA Centre)

Emese Lazar‡
Lecturer in Finance
ICMA Centre, Business School,
The University of Reading, PO Box 242
Reading RG6 6BA, United Kingdom
Email: e.lazar@icmacentre.rdg.ac.uk
Tel: +44 (0)1183 786768 (ICMA Centre)

‡ Corresponding author

I Introduction

The variance of financial returns is not directly observable and has thus been subject to extensive time series analysis based on nonparametric and parametric methods. Whilst non-parametric methods have just started to flourish, parametric methods are more widely spread and in this group models in the class of generalized autoregressive conditional heteroscedasticity (GARCH) processes are considered the most popular discrete time framework to characterize the dynamic behaviour of the variance process. Continuous time parametric modelling has focussed on smile consistent models and uses a very different set of tools, mostly stochastic volatility models with correlated Brownian motions that can incorporate jumps in the price and/or variance processes. The discrete and continuous approaches are quite well differentiated, but recently more and more papers are connecting the two frameworks.

The first study that links GARCH processes with continuous time modelling is the paper of Nelson (1990). In this path breaking work, which also introduces one of the most important approaches for GARCH option pricing, the author derives the continuous limit of strong GARCH processes using a theorem of weak convergence.[1] This limit is a stochastic variance process with independent Brownian motions, i.e. the well-known 'GARCH diffusion' that is commonly applied in practice. However, Corradi (2000) changed Nelson's set of assumptions and arrived at a different limit: a continuous-time model with deterministic variance.[2,3]

A hypothesis has to be made about the behaviour of the GARCH parameters when the step length converges to zero and in GARCH(1,1) there is some freedom to make these assumptions, hence the difference in the limits derived. The debate about the limit of strong GARCH(1,1) can thus be reduced to asking which set of assumptions is correct. The arguments for Nelson's limiting model are the following: first of all, GARCH has a non-zero variance for the variance yet Corradi's limit has a deterministic variance process, making the variance of the variance conditionally zero. Secondly, a simple simulation study performed by the authors suggests that Nelson's assumptions are appropriate.[4]

---

[1] Nelson (1990) also derives the continuous limit of the EGARCH model of Nelson (1991). His results were later generalized by Duan (1997) to a more general family, the so-called augmented GARCH models.

[2] In another paper, Duan, Ritchken and Sun (2006) show that their model converges to a continuous time model with jumps in the both the price and variance processes, but with diffusion in the price process only. If restricted to a normal GARCH, their limit model gives the model derived by Corradi (2000) because they use the same limiting assumptions for the parameters.

[3] See also the paper of Jeantheau (2004) for the convergence of a GARCH-type model. His assumptions are similar to those of Corradi (2000).

[4] The results are available from the authors upon request.

In favour of Corradi's limit it can be argued that discrete time GARCH has only one source of randomness whilst a diffusive variance has two sources.[5] Furthermore, Wang (2002) used the asymptotic non-equivalence of the likelihood functions to demonstrate that the continuous limit of normal GARCH(1,1) must have a deterministic variance, i.e. it cannot be a diffusion model. Brown, Wang and Zhao (2002) consider stronger convergence conditions and again show that there can be no diffusion term in the continuous limit of multiplicative GARCH models. Also, the transition from continuous variance diffusion to discrete time models yields a discrete time stochastic volatility model such as the autoregressive volatility model that was introduced by Taylor (1986), and not a GARCH process. Hence, choosing between Nelson's and Corradi's limiting models is not a straightforward task.

Several papers study the continuous time limit of different GARCH processes. Kallsen and Taqqu's (1998) approach has the advantages that (1) there is only one source of randomness, as in the discrete time model and (2) it keeps the delayed effect of the returns on the variance process present in GARCH. However, this is not the limit of GARCH, but an extension of it, assuming a step function for the variance. Kazmerchuk *et al.* (2002) further developed this model by changing the variance process so that it is no longer a step function, but a continuous function. A critique of this approach is that, when discretized, the model will return the GARCH process for only one given step length and for all other frequencies the process is not GARCH. Also, it is not obvious how the variance should behave between the breakpoints given by this discretization. Mele and Fornari (2000) considered the continuous limit of A-PARCH models where the error term follows a GED distribution (++++++++++++++++++++more). Later Zheng (2005) studied the limit of HARCH type processes proposed by Müller *et al.* (1997). The disadvantage of this model is, besides the lack of time aggregation, that the model is a function of the step length, namely it is a GARCH(1,K)-type process with K increasing as the step length decreases; its limit has deterministic variance as the conditions set are similar to that of Corradi (2000). Trifi (2006) extended the paper of Nelson (1990) to non-normal distributions, case in which a continuous limit similar to ours is obtained, and to Augmented GARCH processes as well as the CEV-GARCH model of Fornari and Mele (2005). However, still it is necessary to assume limiting behaviours for the model parameters.

A continuous time process that features the properties of GARCH and where the residuals follow a Lévy process was introduced by Klüppelberg, Lindner and Maller (2004). This has the advantage that

---

[5] One explanation for this is that, given a normally distributed variable, $x(t)$, a new one can be created (based on the very same process), $x(t)^2$, with Corr($x(t)$, $x(t)^2$) = 0. Hence, with only one source of uncertainty two uncorrelated (but not independent) processes can be created.

it has only one source of randomness. However, it is not the limit of the discrete time GARCH but a continuous time extension.

This paper employs the weak definition of GARCH given by Drost and Nijman (1993) which has the advantage that it is time aggregating: if the weak GARCH(1,1) is the data generating process (DGP) for a given frequency, then the same model will be the DGP for any other frequency. We believe that only under this condition is it legitimate to consider the continuous limit of a model. Drost and Werker (1996) introduce continuous time GARCH processes that exhibit weak GARCH-type behaviour for all discrete frequencies. (+++++++++++++more) Meddahi and Renault (2004) introduce a large class of volatility models that have stochastic volatility models as their continuous time limit. This class is closed under temporal aggregation and it includes GARCH processes as well. However, their definition does not create a closed subgroup for the GARCH processes alone. In other words, taking GARCH(1,1) as the DGP for some frequency, then for any other frequency we have another model in Meddahi and Renault's class, but not a GARCH(1,1) variance model.

The aggregated strong GARCH gives an infinite-leg weak GARCH variance, which has a very different behaviour from a strong GARCH(1,1), as shown by Nijman and Sentana (1996), Komunjer (2001) and Jondeau (2008), although in these papers aggregation is considered contemporaneously (in the context of a portfolio) as opposed to temporally. The behaviour of the temporally aggregated strong GARCH process has been studied by Baillie and Bollerslev (1992) and Breuer and Jandacka (2007). With weak GARCH we find that there is no flexibility to choose assumptions when deriving the limit: the convergence of all the parameters is given by the definition of the process. Here the continuous time limit is proved to be a stochastic volatility model with more general properties than Nelson's GARCH limit and which reduces to Nelson's limit under certain assumptions about the conditional returns densities. Nelson's limit has zero price-volatility correlation, but such stochastic volatility models have poor hedging properties when the volatility smile has a negative skew.[6] By contrast, the limit of weak GARCH derived in this paper has correlated Brownian motions in which both the variance diffusion coefficient and the price-volatility correlation are related to the skewness and kurtosis of the physical returns density.

After deriving the weak GARCH diffusion model we examine its properties. First we show that a discretization of the model does indeed return a weak GARCH process; hence the model does not suffer from this criticism of the strong GARCH process. Then we compare, again favourably,

---

[6] See, for example, Alexander and Nogueira (2007).

simulations from the weak GARCH diffusion with those of the strong GARCH diffusion. The zero price-volatility correlation in the strong GARCH diffusion is one of the serious drawbacks of the model, yet in the weak GARCH diffusion the price-volatility correlation is not only non-zero, but it is related to the skewness and kurtosis of the underlying returns, which is very intuitive. Finally, we discuss the calibration of the model and compare its properties with those of the popular stochastic volatility model introduced by Heston (1993).

The remainder of this paper is organized as follows: Section II presents the weak GARCH process. Section III addresses the issue of parameter convergence and Section IV derives the continuous time limit of the weak GARCH(1,1). Section V examines the discretization of the continuous model and this is followed by a some simulation studies in Section VI. Section VII (to be written) contains calibration results and Section VIII concludes. All proofs are in the Appendix.

II. Weak GARCH process

A GARCH(1,1) process (from now on denoted simply by GARCH), as introduced by Engle (1982) and Bollerslev (1986), is given by an autoregressive conditional variance that also depends on the square of the previous return. We denote the returns by:

$$y_t = \frac{S_t - S_{t-1}}{S_{t-1}} \cong \ln\left(S_t / S_{t-1}\right)$$

and assume that the conditional mean equation is given by $y_t = \mu + \varepsilon_t$ with $E\left(\varepsilon_{t+1} \middle| I_t\right) = 0$ where the 'information set' I_t is the σ-algebra generated by the vector $\left(\varepsilon_t\right)$. The conditional variance h_t is assumed to follow the process:

$$h_t = \omega + \alpha\varepsilon_t^2 + \beta h_{t-1} \tag{1}$$

Now the classical (strong) definition states that:[7]

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

We define the step-length Δ and consider the continuous limit as Δ ↓ 0. Our notation for a time series with step-length Δ indexes time as k Δ, with k = 1, 2, … Hence, for any Δ we can define the Δ-

[7] The subscript *t* here stands for the time that the process becomes known; this means that h_t is the conditional variance for ε_{t+1}^2 and it is revealed at time *t*. See also Zheng(2005).

step process with two components: the residuals and the GARCH process. In the following the pre-subscript in front of the parameters will denote the step-length used.

The first paper that discusses the continuous limit of GARCH is that of Nelson (1990). The main theorem states that, under the conditions:

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

the limit will be a stochastic volatility model:

$$\frac{dS}{S} = \mu\, dt + \sqrt{V} dB_1$$
$$dV = \left(\omega - \theta V\right) dt + \sqrt{2}\alpha V\, dB_2$$

where the two Brownian motions are independent. We have used the notation S and V for the continuous-time limits of S_t and h_t.

On the other hand, Corradi (2000) proves that, if we assume the following convergence rates:

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

then the continuous-time limit is a deterministic variance model:

$$\frac{dS}{S} = \mu\, dt + \sqrt{V} dB$$
$$dV = \left(\omega - \theta V\right) dt$$

The difference between the two assumptions lies with the convergence of alpha. In the first case it is assumed to converge to a constant at rate $\sqrt{\Delta}$, whilst in the second case it is assumed to converge at rate Δ. Which assumption is correct has been the subject of considerable debate. Here we argue that Nelson's assumptions are correct. However, we promote a different continuous limit because it is best to use the time aggregating version of the model. Without time aggregation we have a strong GARCH process for a given frequency, but for any other frequencies the process will not be a strong GARCH process anymore.

When computing the limiting model, we use a definition that guarantees that the process will be the same for any frequency, meaning that the model is aggregating in time; this is the weak definition introduced by Drost and Nijman (1993). The difference is that the weak GARCH specifies that h_t in

(1) is not the conditional variance, but the best linear predictor (BLP) of the squared residuals. In weak GARCH equation (2) is replaced by the conditions:

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

The assumption that 0 and h_t are the BLPs for the residuals and squared residuals at time t + 1, guarantees that the BLP of the squared residuals (not the conditional variance) aggregates in time.

Consider this process using two base step lengths: Δ and δ where δ < Δ. Since we need to compare variances at different time steps, ${}_{\Delta}h_{k\Delta}$ will denote the BLP for ${}_{\Delta}\varepsilon_{k\Delta}^{2}/\Delta$ (note that dividing by the step-length will give us comparable linear predictors). This means that, for an arbitrary step Δ, the weak GARCH process can be written as:

$${}_{\Delta}y_{k\Delta}=\Delta\mu+{}_{\Delta}\varepsilon_{k\Delta} \text{ where } {}_{\Delta}y_{k\Delta}=\frac{S_{k\Delta}-S_{(k-1)\Delta}}{S_{(k-1)\Delta}}\cong\ln\left(S_{k\Delta}/S_{(k-1)\Delta}\right)$$

$${}_{\Delta}h_{k\Delta}={}_{\Delta}\omega+{}_{\Delta}\alpha\,{}_{\Delta}\varepsilon_{k\Delta}^{2}/\Delta+{}_{\Delta}\beta\,{}_{\Delta}h_{(k-1)\Delta}$$

$$E\left({}_{\Delta}\varepsilon_{(k+1)\Delta}\,{}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right)=0 \quad i\geq 0 \quad r=0,1,2$$

$$E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta}^{2}/\Delta-{}_{\Delta}h_{k\Delta}\right){}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right)=0 \quad i\geq 0 \quad r=0,1,2$$

Similarly, ${}_{\delta}h_{k\delta}$ will denote the BLP for ${}_{\delta}\varepsilon_{k\delta}^{2}/\delta$ and a similar set of defining equations can be written for steps of length δ.

III. Parameter convergence

The weak definition of GARCH implies a relationship between the parameters and unconditional kurtosis ${}_{\Delta}\varkappa$ of the Δ-step process ${}_{\Delta}h_{k\Delta}$ and the parameters and unconditional kurtosis of the δ-step process denoted by ${}_{\delta}h_{k\delta}$, where the pre-subscript denotes the step-length. This relationship was derived by Drost and Nijman (1993), and is given by the following set of equations:[8]

$${}_{\Delta}\omega={}_{\delta}\omega\frac{1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}}{1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)};\ {}_{\Delta}\alpha=\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}-{}_{\Delta}\beta$$

[8] Note that we annualise the GARCH processes (dividing ω by the step length) but Drost and Nijman do not.

$$ {}_{\Delta}\varkappa = 3 + \frac{{}_{\delta}\varkappa - 3}{\Delta/\delta} + $$

$$ +6\left({}_{\delta}\varkappa - 1\right)\frac{\left((\Delta/\delta)\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)-\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2}+{}_{\delta}\alpha\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)}{\left(\Delta/\delta\right)^{2}\left(1-{}_{\delta}\alpha-{}_{\delta}\beta\right)^{2}\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2}+{}_{\delta}\alpha^{2}\right)} $$

$$ {}_{\Delta}\beta \text{ is the solution to } \frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^{2}} = \frac{{}_{\Delta}a\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}-{}_{\Delta}b}{{}_{\Delta}a\left(1+\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2\Delta/\delta}\right)-2{}_{\Delta}b} \text{ where} $$

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

$$ +4\frac{\left((\Delta/\delta)\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)-\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1-{}_{\delta}\beta\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)}{1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2}} $$

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

The above formulae give the low frequency parameters in terms of the high frequency parameters for symmetric models. However, to find the continuous limit of this model we are interested in the inverse relationship: assuming that the parameters for low frequency data are given we derive the high frequency parameters (and later on their limit), provided they exist.

We therefore assume that the Δ-step parameters are known and we derive the δ-step parameters for $\delta < \Delta$. Using the above we obtain:

$$ {}_{\delta}\omega = {}_{\Delta}\omega\frac{1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{\delta/\Delta}}{1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)}; $$

$$ {}_{\delta}\alpha = \left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{\delta/\Delta}-{}_{\delta}\beta; $$

$$ {}_{\delta}\varkappa = 1+\frac{{}_{\Delta}\varkappa-3+2\delta/\Delta}{\left(\frac{\delta}{\Delta}\right)+6\frac{\left(\left(1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{\delta/\Delta}\right)/(\delta/\Delta)-\left(1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)\right)\right)}{\left(\left(1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{\delta/\Delta}\right)/(\delta/\Delta)\right)^{2}}\frac{{}_{\delta}\alpha\left(1-{}_{\delta}\beta\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{\delta/\Delta}\right)}{\left(1-\left({}_{\Delta}\alpha+{}_{\Delta}\beta\right)^{2\delta/\Delta}+{}_{\delta}\alpha^{2}\right)}} $$

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

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

$$\times\left(\begin{array}{l}\frac{1}{\delta/\Delta}\left(1-{}_\delta\beta\right)^2+\frac{1}{\delta/\Delta}\left(\frac{1}{\delta/\Delta}-1\right)\left(\frac{2}{{}_\delta k-1}\right)\left(\frac{1-\left({}_\Delta\alpha+{}_\Delta\beta\right)^{\delta/\Delta}}{1+\left({}_\Delta\alpha+{}_\Delta\beta\right)^{\delta/\Delta}}\right)\left(1-\left({}_\Delta\alpha+{}_\Delta\beta\right)^{2\delta/\Delta}+{}_\delta\alpha^2\right)\\ +4\frac{\left(\left(1-\left({}_\Delta\alpha+{}_\Delta\beta\right)^{\delta/\Delta}\right)/\left(\delta/\Delta\right)-\left(1-\left({}_\Delta\alpha+{}_\Delta\beta\right)\right)\right){}_\delta\alpha\left(1-{}_\delta\beta\left({}_\Delta\alpha+{}_\Delta\beta\right)^{\delta/\Delta}\right)}{1-\left({}_\Delta\alpha+{}_\Delta\beta\right)^{2\delta/\Delta}}\end{array}\right)$$

Before deriving the continuous limit of weak GARCH, we need to determine the limits and convergence speeds of the parameters, as the limiting model will depend on these. In contrast to the strong GARCH process, where there is some freedom to choose assumptions about parameter convergence speeds, we now find that it is not possible to make any assumption. Instead the time-aggregation property of weak GARCH implies unique convergence speeds for all parameters, as the following proposition shows:

Proposition 1:

The convergence rates for the parameters implied by the weak GARCH model are as follows:

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

Also, the unconditional kurtosis converges to:

$$\varkappa=\lim_{\Delta\downarrow 0}{}_\Delta\varkappa=\frac{3}{1-\alpha^2/\theta}$$

The proof in the Appendix of this proposition may also be derived from the work of Drost and Nijman (1996), albeit their results serve a different purpose and they use different notation. In our notation, they consider the convergence of:

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

They also show that

$$\frac{{}_\Delta\omega/\Delta}{1-\left({}_\Delta\alpha+{}_\Delta\beta\right)}=\left(\frac{{}_\Delta\omega}{\Delta^2}\right)\left(\frac{1-\left({}_\Delta\alpha+{}_\Delta\beta\right)}{\Delta}\right)^{-1}$$

converges to a constant, which means that they have convergence for $_{\Delta}\omega / \Delta^2$, whilst we have convergence for $_{\Delta}\omega / \Delta$. This apparent inconsistency is caused by the fact that we annualise the GARCH processes to make the processes of different step-lengths comparable.

IV. Continuous limit of weak GARCH

The continuous time limit of a model is undoubtedly very important; however, in general it does not offer equivalence with the discrete time model. Two types of problems might arise. First of all, it can be that the discretization of the continuous model will give a different model from the original one. The second problem arises only for models that do not aggregate in time. If the discretization of the continuous limit will give the same model as the original, then it can be discretized for any frequency, so we have the *same* model for all frequencies. But, if the model is not aggregating in time, then we should have ***different*** model specifications for different frequencies – and this is a contradiction.

The question arises: when is it possible to obtain equivalent discrete time and continuous time models? This is possible only when (1) the original discrete time model is time aggregating, hence the conflict presented above can be avoided. A second requirement is that (2) the model can be discretized in the form of the original model. This gives a very strong relationship between discrete and continuous models as they can be translated with ease into each other. This is a step forward from deriving continuous time limits for discrete time models as it gives equivalence between the two types of models.

Many of the papers mentioned above have a common deficiency: when computing the continuous limit they employ the classical (strong) definition of GARCH that is not aggregating in time. Hence, the continuous time limit is not equivalent with the original model. The reason is that if the strong GARCH(1,1) is a valid description of a data for a given frequency, then for any other frequency GARCH(1,1) will not be the data generating process (DGP).

An argument that decreases the importance of time aggregation is the following: suppose the same discrete time model, GARCH, is not the data generating process for any frequency, and nor is its diffusion limit the continuous time DGP. On the other hand, let's assume that GARCH is a good approximation to the DGP. In this case time aggregation is slightly less important than if GARCH is the true DGP. What is essential, however, is that the continuous time limit of the DGP is the same process as the limit of the GARCH model used. As the frequency of the observations increases, the estimated GARCH variance will offer a good approximation for the variance of the true process (Nelson, 1992). Moreover, under certain assumptions, the misspecified GARCH model may perform well for forecasting as well (Nelson, 1995), even if it is not the DGP. However, the above holds only for high frequencies and only under certain regularity conditions; thus, only as an approximation.

Consider the first two conditional moments and the conditional skewness and kurtosis:

$$ {}_{\Delta}\mu_{k\Delta} = E\left({}_{\Delta}\varepsilon_{(k+1)\Delta} / \Delta \,\middle|\, {}_{\Delta}I_{k\Delta} \right) $$

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

$$ {}_{\Delta}\tau_{k\Delta} = E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta\, {}_{\Delta}\mu_{k\Delta} \right)^3 / \left(\Delta^{3/2}\, {}_{\Delta}\sigma^3_{k\Delta} \right) \,\middle|\, {}_{\Delta}I_{k\Delta} \right) $$

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

where ${}_{\Delta}I_{k\Delta}$ is the σ-algebra generated by the vector $\left({}_{\Delta}\varepsilon_{k\Delta} \right)$. We divide by Δ when computing the conditional mean and variance series because these are additive in time (the mean and variance over a period of length Δ must be comparable with Δ times the 1-step mean and variance). Also, the conditional expectation of the second moment and the kurtosis must be positive.[9]

We assume that the following limits exist:

$$ V(t) := \lim_{\Delta \downarrow 0} {}_{\Delta}h_t \quad \text{where} \quad {}_{\Delta}h_t := {}_{\Delta}h_{k\Delta} \quad \text{for } k\Delta \le t < (k+1)\Delta $$

$$ \mu(t) := \mu + \lim_{\Delta \downarrow 0} {}_{\Delta}\mu_t \quad \text{where} \quad {}_{\Delta}\mu_t := {}_{\Delta}\mu_{k\Delta} \quad \text{for } k\Delta \le t < (k+1)\Delta $$

$$ \tau(t) := \lim_{\Delta \downarrow 0} {}_{\Delta}\tau_t \quad \text{where} \quad {}_{\Delta}\tau_t := {}_{\Delta}\tau_{k\Delta} \quad \text{for } k\Delta \le t < (k+1)\Delta $$

$$ \eta(t) := \lim_{\Delta \downarrow 0} {}_{\Delta}\eta_t \quad \text{where} \quad {}_{\Delta}\eta_t := {}_{\Delta}\eta_{k\Delta} \quad \text{for } k\Delta \le t < (k+1)\Delta $$

It can be seen that:

$$ E\left({}_{\Delta}\varepsilon^2_{(k+1)\Delta} / \Delta \,\middle|\, I_{k\Delta} \right) = {}_{\Delta}\sigma^2_{k\Delta} + \Delta\, {}_{\Delta}\mu^2_{k\Delta} $$

and at least one of the processes ${}_{\Delta}\mu_{k\Delta}$ and ${}_{\Delta}\sigma^2_{k\Delta} + \Delta\, {}_{\Delta}\mu^2_{k\Delta} - {}_{\Delta}h_{k\Delta}$ has to be different from zero, otherwise the process will be a semi-strong GARCH which is not aggregating in time.

[9] Drost and Nijman (1993) showed the time aggregation property for symmetric weak GARCH models. However, for generality, here we consider asymmetric models as well, but it is only symmetric weak models that aggregate in time.

We assume that as the step length Δ converges to zero, the difference between the conditional variance and the BLP of the squared residuals, divided by the square root of time step, converges to a stochastic process, i.e.

$$\lim_{\Delta \downarrow 0} \frac{{}_\Delta\sigma_t^2 - {}_\Delta h_t}{\sqrt{\Delta}} = e(t)$$

In other words, the BLP of the squared residuals is 'close' to the conditional variance process. This intuitive assumption is necessary to prove our results. This is the only assumption we make and we consider that it is non-binding because as the time step decreases, the BLP process becomes more and more informative and so it converges fast to the conditional variance, i.e.

$$V(t) = \lim_{\Delta \downarrow 0} {}_\Delta\sigma_t^2 \text{ where } {}_\Delta\sigma_t^2 := {}_\Delta\sigma_{k\Delta}^2 \text{ for } k\Delta \le t < (k+1)\Delta.$$

Now that we have the convergence speeds and have defined the limits of the parameters and the series we are ready to state the theorem regarding the continuous limit of GARCH:

Theorem 1: *The continuous time limit of the weak GARCH process is the following stochastic volatility model:*

$$\frac{dS(t)}{S(t)} = \mu(t)dt + \sqrt{V(t)}dB_1(t)$$
$$dV(t) = \left(\omega + \alpha e(t) - \theta V(t)\right)dt + \sqrt{\eta(t) - 1}\,\alpha V(t)dB_2(t)$$

where

$$dB_2(t) = \varrho(t)dB_1(t) + \sqrt{1 - \varrho(t)^2}\,dB_3(t) \quad \text{with} \quad \varrho(t) = \frac{\tau(t)}{\sqrt{\eta(t) - 1}},$$

and $\tau(t)$ and $\eta(t)$ are the instantaneous skewness and kurtosis of returns and B_1 and B_3 are independent Brownian motions. The unconditional kurtosis of returns is:

$$\varkappa = \frac{3}{1 - \alpha^2 / \theta}$$

In the GARCH limit model derived in the theorem above, the drift term $\mu(t)$ is time varying and has expectation μ. The variance process has a constant rate of mean-reversion θ and the long-run level of the variance is:

$$\frac{\omega + \alpha e(t)}{\theta}$$

That this is time varying may at first sight appear inconsistent with the limit of the discrete long-term variance, ${}_\Delta\omega / (1 - {}_\Delta\alpha - {}_\Delta\beta)$, but it is not. First, the discrete time long-term variance, denoted by ${}_\Delta\sigma^2$, is not the expression above. It satisfies:

$$\left(1 - {}_\Delta\beta\right)\left({}_\Delta\sigma^2 - \left(e(t) + o(1)\right)\sqrt{\Delta}\right) = {}_\Delta\omega + {}_\Delta\alpha\left({}_\Delta\sigma^2 + \Delta\mu(t)^2\right)$$

which leads to:

$${}_\Delta\sigma^2 = \frac{{}_\Delta\omega + {}_\Delta\alpha\Delta\mu(t)^2 + \left(1 - {}_\Delta\beta\right)\left(e(t) + o(1)\right)\sqrt{\Delta}}{1 - {}_\Delta\alpha - {}_\Delta\beta}$$

Its limit when $\Delta \downarrow 0$ is:

$$\begin{aligned}
\sigma^2 &= \lim_{\Delta\downarrow 0} {}_\Delta\sigma^2 \\
&= \lim_{\Delta\downarrow 0} \frac{{}_\Delta\omega / \Delta + {}_\Delta\alpha\mu(t)^2 + \left(1 - {}_\Delta\alpha - {}_\Delta\beta\right)\left(e(t) + o(1)\right) / \sqrt{\Delta} + {}_\Delta\alpha\left(e(t) + o(1)\right) / \sqrt{\Delta}}{\left(1 - {}_\Delta\alpha - {}_\Delta\beta\right) / \Delta} \\
&= \frac{\omega + \alpha e(t)}{\theta}
\end{aligned}$$

Discrete-time weak GARCH processes are characterized by (1) the existence of a long-term volatility; (2) mean reversion in the variance process; (3) the variance is stochastic, i.e. it has a non-zero variance; and (4) non-zero correlation between the variance and the returns processes, if the returns are asymmetric (have non-zero skewness). As we saw before, all these are properties describe also the continuous limit above.

Corollary 1: *If the conditional mean of the residuals converges to zero and the difference between the BLP and the conditional variance converges to zero at rate* $\sqrt{\Delta}$*, then the continuous time limit of the weak GARCH process is the following stochastic volatility model:*

$$\begin{aligned}
\frac{dS}{S} &= \mu dt + \sqrt{V} dB_1 \\
dV &= \left(\omega - \theta V\right) dt + \sqrt{\eta - 1}\, \alpha V dB_2
\end{aligned}$$

where

$$dB_2 = \varrho dB_1 + \sqrt{1 - \varrho^2}\, dB_3 \quad \textit{with} \quad \varrho = \frac{\tau}{\sqrt{\eta - 1}}$$

and the unconditional kurtosis is:

$$\varkappa = \frac{3}{1 - \alpha^2 / \theta}$$

Hence the limit reduces to the diffusion derived by Nelson (1990) under the further assumptions

$$\tau(t) = 0, \quad \eta(t) = 3.$$

In the above stochastic processes the volatility of the variance is $\sqrt{\eta - 1}\,\alpha V$. For given alpha, the smallest value of the variance diffusion coefficient is $\sqrt{2}\,\alpha V$, as in Nelson's model. More generally we have $\eta > 3$ so that the greater the instantaneous kurtosis, the more volatile is the variance process. Also, the instantaneous correlation between the variance and the returns is directly related to the instantaneous skewness, and inversely related to the instantaneous kurtosis. These properties are intuitive and parallel the observed behaviour of implied volatilities in the risk neutral measure: see for example, Bates (1997, 2000) and Bakshi *et al.* (2003).

V. Discretization of the continuous limit of weak GARCH

We are interested in the discretization such that the original GARCH model is returned. First, assuming a time step of length Δ, the changes in the independent Brownian motions at time $t = k\Delta$ can be expressed as:

$$\Delta B_1(k\Delta) = \sqrt{\Delta}\;{}_{\Delta}\xi_{1,(k+1)\Delta} \quad \text{with} \quad {}_{\Delta}\xi_{1,(k+1)\Delta} \sim N(0,1) \tag{3}$$

$$\Delta B_3(k\Delta) = \sqrt{\Delta}\;{}_{\Delta}\xi_{3,(k+1)\Delta} \quad \text{with} \quad {}_{\Delta}\xi_{3,(k+1)\Delta} \sim N(0,1) \tag{4}$$

These are independent variables. Also, we have:

$${}_{\Delta}\varepsilon_{(k+1)\Delta} = \sqrt{{}_{\Delta}V_{k\Delta}}\,\sqrt{\Delta}\;{}_{\Delta}\xi_{1,(k+1)\Delta} \tag{5}$$

$${}_{\Delta}\varepsilon_{3,(k+1)\Delta} = \sqrt{{}_{\Delta}V_{k\Delta}}\,\sqrt{\Delta}\;{}_{\Delta}\xi_{3,(k+1)\Delta} \tag{6}$$

Proposition 2: *Using approximations* (3) −(6) *and the parameter specifications below, the discretization of the continuous time model returns the original GARCH model and the time aggregation property is preserved:*

$${}_{\Delta}\omega = \omega\,\Delta\frac{\left(1-e^{-\theta\Delta}\right)}{\theta\,\Delta}$$

$${}_{\Delta}\beta = \frac{{}_{\Delta}a^*\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^* + \left(1-e^{-\theta\Delta}\right)\sqrt{{}_{\Delta}a^{*2}\left(1+e^{-\theta\Delta}\right)^2 - 4\,{}_{\Delta}a^*\,{}_{\Delta}b^*}}{2\left({}_{\Delta}a^*\,e^{-\theta\Delta} - {}_{\Delta}b^*\right)} \quad \textit{with}$$

$${}_{\Delta}a^* = \Delta\,\alpha^2 + (\Delta)^2\,\theta\left(\theta-\alpha^2\right) + 2\alpha^2\left(\Delta - \left(1-e^{-\theta\Delta}\right)/\,\theta\right) \quad \textit{and} \quad {}_{\Delta}b^* = \frac{\alpha^2}{2\theta}\left(1-e^{-2\theta\Delta}\right)$$

$${}_{\Delta}\alpha = e^{-\theta\Delta} - {}_{\Delta}\beta$$

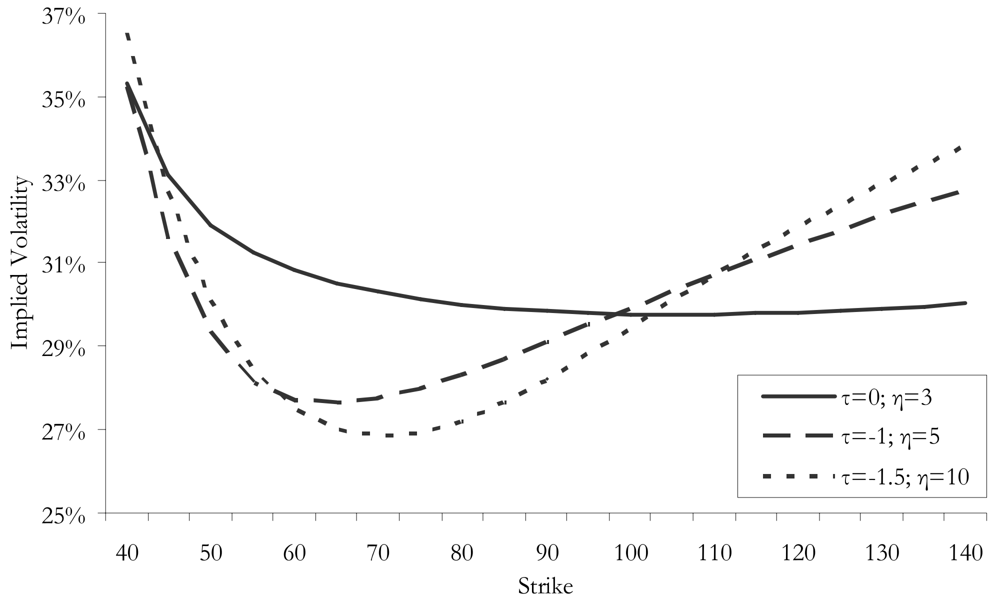
The unconditional kurtosis is discretized as:

$${}_{\Delta}\varkappa = 3 + 6(\varkappa-1)\frac{\alpha^2\left(\theta\Delta - \left(1-e^{-\theta\Delta}\right)\right)}{\theta^2\Delta^2\left(\alpha^2+2\theta\right)}$$

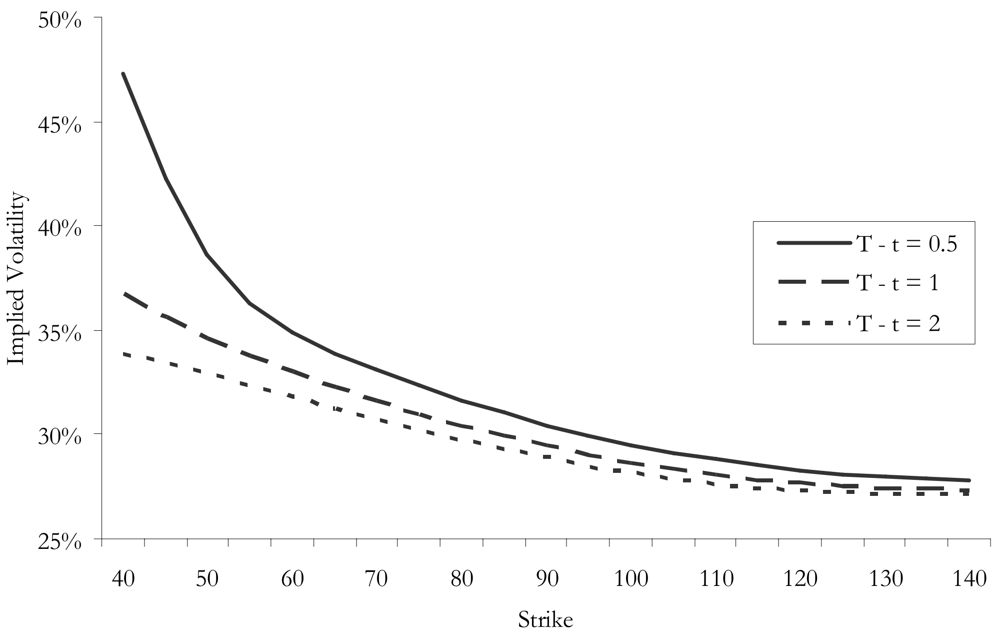
VI. Simulation Results

Figure 1 compares the volatility smile (with zero volatility risk premium) that is generated by Nelson's diffusion with those from the more general model. It shows how different values of instantaneous skewness and kurtosis that give the same instantaneous correlation can influence the shape of the model implied volatility.

Figure 1. Volatility smiles generated by the continuous limit of weak GARCH



(A) $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $T - t = 1$;
$S_0 = 100$; $V_0 = 0.09$; $r = 0\%$; 100 steps and 100,000 runs were used for the simulations
(a) $\tau(t) = 0$; $\eta(t) = 3$, (b) $\tau(t) = -1$; $\eta(t) = 5$ and (c) $\tau(t) = -1.5$; $\eta(t) = 10$



(B) $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $\tau(t) = -1 + t/2$; $\eta(t) = 7 - 2t$;
$S_0 = 100$; $V_0 = 0.09$; $r = 0\%$; 100 steps and 100,000 runs were used for the simulations
(a) $T - t = 0.5$, (b) $T - t = 1$ and (c) $T - t = 2$

The GARCH limit model has considerable flexibility to fit a volatility smile surface though a suitable parameterization of the instantaneous skew and kurtosis functions. Figure 1 (B), for instance, depicts model implied volatility curves when the instantaneous skewness and kurtosis functions are decreasing linearly with time in absolute value.

Figure 2 compares the simulations from four volatility diffusions, with all four simulations are driven by the same realizations for the Brownian motions B_1 and B_2. The simulations are based on the Heston model, the strong GARCH diffusion, the weak GARCH diffusion with a shallow skew, and the weak GARCH diffusion with a deep skew. The model parameters are set as in Table 1, and the Heston model is parameterized as

$$dV = \varphi(m - V)\, dt + \xi\sqrt{V}\left(\varrho dB_1 + \sqrt{1-\varrho^2}\, dB_2\right).$$

Table 1 **Parameters used in simulations**

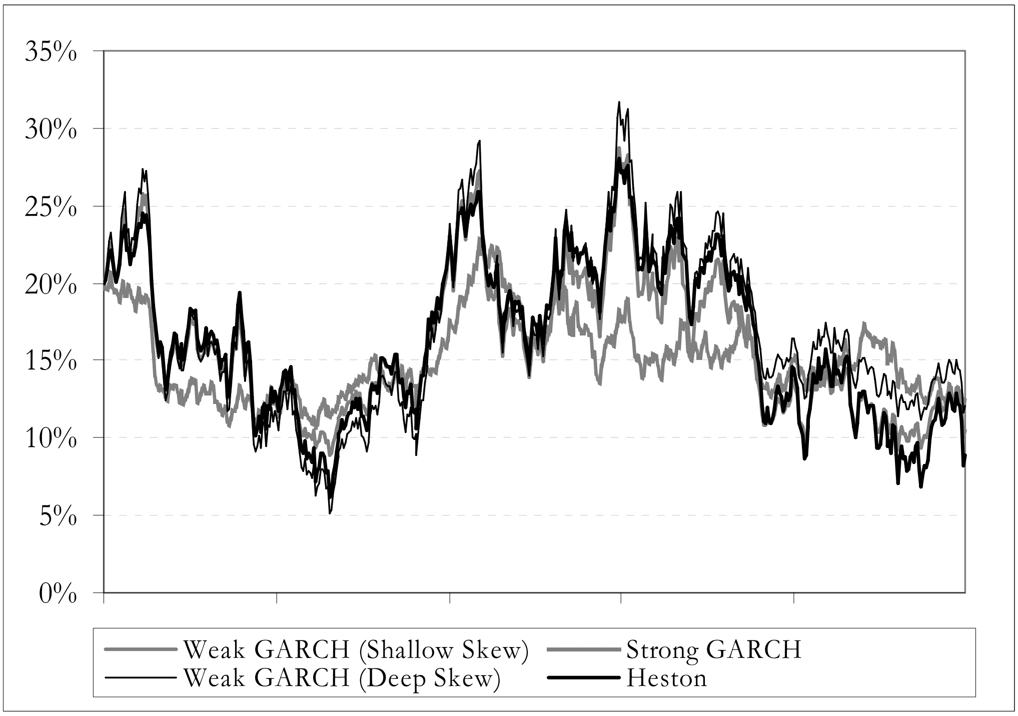
Heston parameters		Value	GARCH parameters	Annual	Daily
Long run volatility	$m^{1/2}$	15%	ω	2.8×10^{-5}	1.5×10^{-6}
Spot volatility	$\sqrt{V_0}$	20%	α	1.2374	0.06477
Vol-of-vol	ξ	35%	β		0.91879
Mean reversion	φ	6	θ		0.01644

In both the weak GARCH models we assume that the price–volatility correlation is –0.75 but in the strong GARCH of course, the price–volatility correlation is zero.[10] We remark that in the weak GARCH diffusion there are infinitely many possible choices of skewness and kurtosis that give a price–volatility correlation of –0.75. For the weak GARCH diffusion with shallow skew we assume the kurtosis is 5 and the skewness is –1.5, and for the weak GARCH diffusion with deep skew we assume the kurtosis is 7 and the skewness is –1.837.

The strong GARCH volatility is very different from the Heston volatility in every simulation, not just the one shown in Figure 2. This is primarily because the price–volatility correlation is zero in the strong GARCH process. The two weak GARCH model simulations are always closer to the Heston model simulations, but which of the two weak GARCH simulations is closest to the Heston model depends on the simulations.

[10] As for Figure III.3.24 we have used a crude method to ensure the discretization always has positive variance.

Figure 2 Comparison of GARCH and Heston volatility simulations



VII. Calibration results

[This section has not yet been written]

VIII. Conclusions

The derivation of the continuous time limit of a discrete time model volatility process is an important to link between the two option pricing schools which have, for the most part, developed quite separate frameworks. This paper derives the continuous limit of the 'vanilla' version of the discrete time volatility process that has become ubiquitous in financial econometrics, but we have shown that for this is it necessary to take the weak rather than the strong version of the model.

In fact, this research was originally tangential to our original agenda, which was to derive the continuous limit of the 'queen' of GARCH processes, i.e. the Markov switching GARCH. In a forthcoming paper we show that this model has a continuous limit where the stochastic volatility process jumps between two volatility states. However, in order to derive this limit we first needed to re-derive the limit of 'vanilla' GARCH, since much controversy surrounded this in the econometrics literature.

Previous work on the continuous limit of GARCH has examined the strong GARCH model. But there are four major problems with the strong GARCH diffusion:

1. The strong GARCH process is not ***time aggregating***. This means that if we generate a GARCH process at one discrete time frequency and then resample the process at another frequency the result is ***not*** another GARCH process. Lack of time aggregation is a real problem because one has to assume time aggregation to derive the continuous limit of any model, and without time aggregation it makes no sense to derive a continuous limit.
2. The limit may only be derived when one is prepared to make a specific assumption about the convergence behaviour of the GARCH parameters and other assumptions lead to different (not necessarily stochastic) limits.
3. Any discretization of the strong GARCH diffusion is not a GARCH model.
4. The fact that the Brownian motions are independent means that the model has very limited applicability, because very few financial markets have zero price–volatility correlation, i.e. symmetric volatility smiles.

This paper has studied the continuous limit of the weak GARCH process. This version of the GARCH is time aggregating and it implies the convergence rates for parameters; there is no uncertainty about these (they follow from the definition) and the limit model derived here is unique. We showed that the continuous time model corresponding to the weak GARCH is a stochastic variance process with correlated Brownian motions in which the variance diffusion coefficient and the price-volatility correlation are related to both the instantaneous kurtosis and the instantaneous skewness. This limit model reduces to Nelson's GARCH diffusion only when the conditional mean, skewness and excess kurtosis converge to zero and the difference between the GARCH BLP process and the conditional variance converges to zero with the square root of the step-length. A discretization of this process that returns the original weak GARCH model is given. When we have a discretization of the continuous model that is the same as the original specification, then the two models are equivalent, i.e. they can be transformed into each other. This has many advantages, offering a bridge between the two different worlds. Finally, by comparing simulations from strong and weak GARCH diffusion processes we show that the non-zero price-volatility correlation is important; and finally calibrations to market prices of liquid options show how the additional skewness and kurtosis parameters enhance the model calibration for short maturity options.

Appendix

Proof of Proposition 1:

We have that:

$$\left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{1/\delta} = \left({}_{\Delta}\alpha + {}_{\Delta}\beta\right)^{1/\Delta}$$

Since this expression is independent of the step-length, it must be a constant between 0 and 1; we denote it by $e^{-\theta}$ with $\theta > 0$. This leads to ${}_{\Delta}\alpha + {}_{\Delta}\beta = e^{-\theta\Delta}$ which gives:

$$\lim_{\Delta\downarrow 0}\left(\frac{1-\left({}_{\Delta}\alpha + {}_{\Delta}\beta\right)}{\Delta}\right) = \lim_{\Delta\downarrow 0}\left(\frac{1-e^{-\theta\Delta}}{\Delta}\right) = \theta$$

Furthermore:

$$\frac{{}_{\delta}\omega}{1-e^{-\theta\delta}} = \frac{{}_{\Delta}\omega}{1-e^{-\theta\Delta}}$$

and this is also independent of the step-length, so it must be a positive constant. We denote it by ω/θ, where $\omega > 0$, so that ${}_{\Delta}\omega = \omega\left(1-e^{-\theta\Delta}\right)/\theta$. As a result, we have the following convergence:

$$\lim_{\Delta\downarrow 0}\left(\frac{{}_{\Delta}\omega}{\Delta}\right) = \omega \lim_{\Delta\downarrow 0}\left(\frac{1-e^{-\theta\Delta}}{\Delta}\right)/\theta = \omega$$

Based on the formula for kurtosis, we can write:

$${}_{\delta}\varkappa = 1 + \frac{{}_{\Delta}\varkappa - 3 + 2\delta/\Delta}{\dfrac{\delta}{\Delta} + 6\dfrac{\left(\Delta\left(1-e^{-\theta\delta}\right)/\delta - \left(1-e^{-\theta\Delta}\right)\right)}{\Delta^2\left(\left(1-e^{-\theta\delta}\right)/\delta\right)^2}\dfrac{{}_{\delta}\alpha\left(1-e^{-2\theta\delta}\right)/\delta + {}_{\delta}\alpha^2/\delta - {}_{\delta}\alpha^2\left(1-e^{-\theta\delta}\right)/\delta}{\left(1-e^{-2\theta\delta}\right)/\delta + {}_{\delta}\alpha^2/\delta}}$$

Taking the limit when $\delta \downarrow 0$, we have (using ${}_{\delta}\alpha \downarrow 0$):

$$\varkappa = \lim_{\delta\downarrow 0} {}_{\delta}\varkappa = 1 + \frac{{}_{\Delta}\varkappa - 3}{6\dfrac{\left(\Delta\theta - \left(1-e^{-\theta\Delta}\right)\right)}{\Delta^2\theta^2}\dfrac{1}{\left(2\theta/\lim_{\delta\downarrow 0}\left({}_{\delta}\alpha^2/\delta\right)+1\right)}}$$

Taking the limit on the RHS as $\Delta \downarrow 0$ gives:

$$3\varkappa = 3 + \left(\varkappa - 3\right)\left(2\theta/\lim_{\delta\downarrow 0}\left({}_{\delta}\alpha^2/\delta\right)+1\right)$$

This can be further expressed as:

$$\varkappa = \frac{3}{1-\lim_{\delta\downarrow 0}\left({}_{\delta}\alpha^2/\delta\right)/\theta}$$

The limit of the unconditional kurtosis needs to be finite and positive. This forces $\lim_{\delta\downarrow 0}\left({}_{\delta}\alpha^2/\delta\right)<\theta$.

As a consequence, $\varkappa$ cannot be equal to 1.

To see the exact speed of convergence for ${}_{\delta}\alpha$ we proceed in the following way: First assume the limit below exists (with w unknown):

$$\alpha := \lim_{\delta\downarrow 0}\left(\frac{{}_{\delta}\alpha}{\delta^{w}}\right) \text{ with } 0<\alpha<\infty$$

Then write

$$\lim_{\delta\downarrow 0}\left(\frac{1-e^{-\theta\delta}+{}_{\delta}\alpha}{\delta^{y}}\right)=\lim_{\delta\downarrow 0}\left(\frac{1-e^{-\theta\delta}}{\delta^{y}}\right)+\lim_{\delta\downarrow 0}\left(\frac{{}_{\delta}\alpha}{\delta^{y}}\right)\in(0,\infty) \text{ for } y = \min(w, 1)$$

$$\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}-{}_{\delta}\alpha\left(1-e^{-\theta\delta}\right)+{}_{\delta}\alpha}{\delta^{y}}\right)=\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}}{\delta^{y}}\right)+\lim_{\delta\downarrow 0}\left(\frac{{}_{\delta}\alpha}{\delta^{y}}\right)\in(0,\infty)$$

$$\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}+{}_{\delta}\alpha^{2}}{\delta^{z}}\right)=\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}}{\delta^{z}}\right)+\lim_{\delta\downarrow 0}\left(\frac{{}_{\delta}\alpha^{2}}{\delta^{z}}\right)\in(0,\infty) \text{ for } z = \min(2w, 1)$$

Also, we have that:

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

$$\times\left(\begin{array}{l}\frac{1}{\delta/\Delta}\left(1-e^{-\theta\delta}+{}_{\delta}\alpha\right)^{2}+\frac{1}{\delta/\Delta}\left(\frac{1}{\delta/\Delta}-1\right)\left(\frac{2}{{}_{\delta}k-1}\right)\left(\frac{1-e^{-\theta\delta}}{1+e^{-\theta\delta}}\right)\left(1-e^{-2\theta\delta}+{}_{\delta}\alpha^{2}\right)\\ +4\frac{\left(\left(1-e^{-\theta\delta}\right)/(\delta/\Delta)-\left(1-e^{-\theta\Delta}\right)\right){}_{\delta}\alpha\left(1-e^{-2\theta\delta}-{}_{\delta}\alpha\left(1-e^{-\theta\delta}\right)+{}_{\delta}\alpha\right)}{1-e^{-2\theta\delta}}\end{array}\right)$$

Multiplying this by δ^{1-w-y} (unless w = ½) and computing the limit as δ tends to zero leads to:

$$\left(2\left(\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^{2}}\right)-1\right)\left(1-e^{-2\theta\Delta}\right)\alpha\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}+{}_{\delta}\alpha}{\delta^{y}}\right)\frac{1}{2\theta}=\left(\left(\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^{2}}\right)\left(1+e^{-2\theta\Delta}\right)-e^{-\theta\Delta}\right)\times$$

$$\times\left(\begin{array}{l}\Delta\lim_{\delta\downarrow 0}\left(\delta^{-x+y}\right)\left(\lim_{\delta\downarrow 0}\left(\frac{1-e^{-\theta\delta}+{}_{\delta}\alpha}{\delta^{y}}\right)\right)^{2}+\Delta^{2}\theta\lim_{\delta\downarrow 0}\left(\delta^{-x-y+z}\right)\left(\frac{1}{k-1}\right)\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}+{}_{\delta}\alpha^{2}}{\delta^{z}}\right)\\ +\frac{2}{\theta}\left(\Delta\theta-\left(1-e^{-\theta\Delta}\right)\right)\alpha\lim_{\delta\downarrow 0}\left(\frac{1-e^{-2\theta\delta}+{}_{\delta}\alpha}{\delta^{y}}\right)\end{array}\right)$$

The LHS is finite; if $w > ½$ then the RHS is infinite, which is a contradiction, so $w \leq ½$. But on the other hand, $\lim_{\delta \downarrow 0}\left({}_{\delta}\alpha^2 / \delta \right) < \theta$, which implies $w \geq ½$. So, the solution is $w = ½$ and this sets the convergence of α to be with the square root of the time-step. □

Proof of Theorem 1: We employ the convergence theorem for stochastic difference equations to stochastic differential equations given by Nelson (1990). The convergence theorem applies if we can show that the conditional first and second moments of the percentage returns process and the changes in the variance process, and their conditional covariance, converge as the step-length decreases to zero. For the returns process we have:

$$E\left(\Delta^{-1}\left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right) \middle| I_{k\Delta} \right) = \mu + E\left(\Delta^{-1}\varepsilon_{(k+1)\Delta} \mid I_{k\Delta} \right) = \mu + \mu_{k\Delta}$$

and[11]

$$E\left(\Delta^{-1}\left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right)^2 \middle| I_{k\Delta} \right) = E\left(\Delta^{-1}\left(\Delta\mu + \varepsilon_{(k+1)\Delta} \right)^2 \middle| I_{k\Delta} \right) =$$
$$= E\left(\Delta\mu^2 + \Delta^{-1}\varepsilon^2_{(k+1)\Delta} + 2\mu\varepsilon_{(k+1)\Delta} \middle| I_{k\Delta} \right) = E\left(\Delta^{-1}\varepsilon^2_{(k+1)\Delta} \middle| I_{k\Delta} \right) + o(1) =$$
$$= \sigma^2_{k\Delta} + \Delta\mu^2_{k\Delta} + o(1) = h_{k\Delta} + \left(\sigma^2_{k\Delta} - h_{k\Delta} \right) + o(1) = h_{k\Delta} + o(1)$$

so as $\Delta \downarrow 0$ the conditional first and second moments per unit time converge to $\mu(t)$ and $V(t)$ respectively. For the variance process we have:

$$E\left(\Delta^{-1}\left(h_{(k+1)\Delta} - h_{k\Delta} \right) \middle| I_{k\Delta} \right) =$$
$$= \frac{{}_{\Delta}\omega}{\Delta} - \frac{\left(1 - {}_{\Delta}\alpha - {}_{\Delta}\beta \right)}{\Delta} h_{k\Delta} + \frac{{}_{\Delta}\alpha}{\sqrt{\Delta}}\left(\sigma^2_{k\Delta} - h_{k\Delta} \right) / \sqrt{\Delta} + \frac{{}_{\Delta}\alpha}{\Delta}\Delta\mu^2_{k\Delta} =$$
$$= \frac{{}_{\Delta}\omega}{\Delta} - \frac{\left(1 - {}_{\Delta}\alpha - {}_{\Delta}\beta \right)}{\Delta} h_{k\Delta} + \frac{{}_{\Delta}\alpha}{\sqrt{\Delta}}\left(\sigma^2_{k\Delta} - h_{k\Delta} \right) / \sqrt{\Delta} + o(1)$$

and this converges to $\omega + \alpha e(t) - \theta V(t)$ by proposition 1. The variance of the variance component is:

---

[11] $o(1)$ denotes a process that converges to zero when $\Delta \downarrow 0$ (*o* is called the Landau symbol).

$$E\left(\Delta^{-1}\left(h_{(k+1)h}-h_{k\Delta}\right)^2\middle|I_{k\Delta}\right)=$$

$$=E\left(\Delta^{-1}\left({}_{\Delta}\omega+{}_{\Delta}\alpha\left(\varepsilon^2_{(k+1)\Delta}/\Delta\right)+\left({}_{\Delta}\beta-1\right)h_{k\Delta}\right)^2\middle|I_{k\Delta}\right)=$$

$$=E\left(\begin{array}{l}\Delta^{-1}{}_{\Delta}\omega^2+\Delta^{-1}{}_{\Delta}\alpha^2\left(\varepsilon^2_{(k+1)\Delta}/\Delta\right)^2+\Delta^{-1}\left({}_{\Delta}\beta-1\right)^2h^2_{k\Delta}+2\Delta^{-1}{}_{\Delta}\omega\,{}_{\Delta}\alpha\left(\varepsilon^2_{(k+1)\Delta}/\Delta\right)+\\+2\Delta^{-1}{}_{\Delta}\omega\left({}_{\Delta}\beta-1\right)h_{k\Delta}+2\Delta^{-1}{}_{\Delta}\alpha\left({}_{\Delta}\beta-1\right)\left(\varepsilon^2_{(k+1)\Delta}/\Delta\right)h_{k\Delta}\end{array}\middle|I_{k\Delta}\right)=$$

$$=E\left(\begin{array}{l}\Delta^{-1}{}_{\Delta}\alpha^2\left(\varepsilon^2_{(k+1)\Delta}/\Delta\right)^2+\Delta^{-1}\left({}_{\Delta}\beta-1\right)^2h^2_{k\Delta}+2\Delta^{-1}{}_{\Delta}\alpha\left({}_{\Delta}\beta-1\right)h^2_{k\Delta}+\\+2\Delta^{-1}{}_{\Delta}\alpha\left({}_{\Delta}\beta-1\right)\left(\sigma^2_{k\Delta}-h_{k\Delta}+\Delta\mu^2_{k\Delta}\right)h_{k\Delta}\end{array}\middle|I_{k\Delta}\right)+o(1)=$$

$$=\Delta^{-1}{}_{\Delta}\alpha^2\left(E\left(\varepsilon^4_{(k+1)\Delta}/\Delta^2\middle|I_{k\Delta}\right)\right)-E\left(\Delta^{-1}\left({}_{\Delta}\beta-1\right)\left(1-{}_{\Delta}\alpha-{}_{\Delta}\beta-{}_{\Delta}\alpha\right)h^2_{k\Delta}\middle|I_{k\Delta}\right)+o(1)=$$

$$=\Delta^{-1}{}_{\Delta}\alpha^2\left(E\left(\varepsilon^4_{(k+1)\Delta}/\Delta^2\middle|I_{k\Delta}\right)\right)-E\left(\Delta^{-1}{}_{\Delta}\alpha^2\,h^2_{k\Delta}\middle|I_{k\Delta}\right)+o(1)=$$

$$=\Delta^{-1}{}_{\Delta}\alpha^2\left(E\left(\left({}_{\Delta}\sigma^4_{k\Delta}\right)\left(\varepsilon^4_{(k+1)\Delta}/\left(\Delta^2\,{}_{\Delta}\sigma^4_{k\Delta}\right)\right)-h^2_{k\Delta}\middle|I_{k\Delta}\right)\right)+o(1)$$

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

$$E\left(\Delta^{-1}\left(\frac{S_{(k+1)\Delta}-S_{k\Delta}}{S_{k\Delta}}\right)\left(h_{(k+1)\Delta}-h_{k\Delta}\right)\middle|I_{k\Delta}\right)=$$

$$=E\left(\Delta^{-1}\left(\Delta\mu+\varepsilon_{(k+1)\Delta}\right)\left({}_{\Delta}\omega+{}_{\Delta}\alpha\varepsilon^2_{(k+1)\Delta}/\Delta+\left({}_{\Delta}\beta-1\right)h_{k\Delta}\right)\middle|I_{k\Delta}\right)=$$

$$=E\left(\left({}_{\Delta}\alpha\left(\varepsilon^3_{(k+1)\Delta}/\Delta^2\right)+\left({}_{\Delta}\beta-1\right)h_{k\Delta}\left({}_{\Delta}\varepsilon_{(k+1)\Delta}/\Delta\right)\right)\middle|I_{k\Delta}\right)+o(1)=$$

$$=E\left(\left({}_{\Delta}\alpha/\sqrt{\Delta}\right)\left({}_{\Delta}\sigma^3_{k\Delta}\right)\left(\varepsilon^3_{(k+1)\Delta}/\left(\Delta^{3/2}\,{}_{\Delta}\sigma^3_{k\Delta}\right)\right)\middle|I_{k\Delta}\right)+o(1)$$

The limits of the expected squared terms and cross-product derived above define the following covariance matrix of the continuous process:

$$A(t)=\begin{pmatrix}V(t) & \alpha V(t)^{3/2}\tau(t)\\ \alpha V(t)^{3/2}\tau(t) & \alpha^2V(t)^2\left(\eta(t)-1\right)\end{pmatrix}$$

The parameters of the variance diffusion are given by the elements of the Cholesky matrix of A. Therefore set

$$C(t)C(t)'=A(t)$$

with $C(t)=\begin{pmatrix} c_{11}(t) & 0 \\ c_{12}(t) & c_{22}(t) \end{pmatrix}$. The solution is:

$$c_{11}(t)=\sqrt{V(t)};\quad c_{12}(t)=\alpha V(t)\tau(t);\quad c_{22}(t)=\alpha V(t)\sqrt{\eta(t)-1-\tau(t)^2}\qquad \square$$

Proof of Proposition 2:

Part A: Parameter discretization

We use the notation:

$${}_{\Delta}\omega=\omega\,\Delta\frac{\left(1-e^{-\theta\,\Delta}\right)}{\theta\,\Delta}$$

$${}_{\Delta}\alpha=e^{-\theta\,\Delta}-{}_{\Delta}\beta$$

$${}_{\Delta}\beta=\frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right)-2\,{}_{\Delta}b^{*}+\left(1-e^{-\theta\Delta}\right)\sqrt{{}_{\Delta}a^{*2}\left(1+e^{-\theta\Delta}\right)^{2}-4\,{}_{\Delta}a^{*}\,{}_{\Delta}b^{*}}}{2\left({}_{\Delta}a^{*}\,e^{-\theta\Delta}-{}_{\Delta}b^{*}\right)}\quad\text{with}$$

$${}_{\Delta}a^{*}=\Delta\,\alpha^{2}+\Delta^{2}\theta\left(\theta-\alpha^{2}\right)+2\alpha^{2}\left(\Delta-\left(1-e^{-\theta\Delta}\right)/\,\theta\right)\quad\text{and}\quad {}_{\Delta}b^{*}=\frac{\alpha^{2}}{2\theta}\left(1-e^{-2\theta\Delta}\right)$$

The discretization of θ and ω is based on the following definitions:

$${}_{\Delta}\alpha+{}_{\Delta}\beta=e^{-\theta\,\Delta}$$

$${}_{\Delta}\omega=\omega\,\Delta\frac{\left(1-e^{-\theta\,\Delta}\right)}{\theta\,\Delta}$$

The discretization of α and β is slightly more complex. We already saw that the unconditional kurtosis for a given frequency Δ can be expressed as a function of the parameters at an arbitrary higher frequency δ as in:

$${}_{\Delta}\varkappa=3+\frac{{}_{\delta}\varkappa-3}{\Delta/\delta}+$$

$$+6\left({}_{\delta}\varkappa-1\right)\frac{\left(\left(\Delta/\delta\right)\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)-\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2}+{}_{\delta}\alpha\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)}{\left(\Delta/\delta\right)^{2}\left(1-{}_{\delta}\alpha-{}_{\delta}\beta\right)^{2}\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2}+{}_{\delta}\alpha^{2}\right)}$$

Denoting the limit of the unconditional kurtosis by $\varkappa := \lim_{\delta \downarrow 0} {}_{\delta}\varkappa$, we get:

$$_{\Delta}\varkappa = 3 + 6(\varkappa - 1)\frac{\alpha^2\left(\theta\Delta - \left(1 - e^{-\theta\Delta}\right)\right)}{\theta^2\Delta^2\left(\alpha^2 + 2\theta\right)}$$

where the limit of the unconditional kurtosis is given by:

$$\varkappa = \frac{3}{1 - \lim_{\delta\downarrow 0}\left({}_{\delta}\alpha^2/\delta\right)/\theta} = \frac{3}{1 - \alpha^2/\theta}$$

Also, we know that for any two time steps $\Delta > \delta$, $_{\Delta}\beta$ is the solution to

$$\frac{_{\Delta}\beta}{1 + {}_{\Delta}\beta^2} = \frac{_{\Delta}a\left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{\Delta/\delta} - {}_{\Delta}b}{_{\Delta}a\left(1 + \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{2\Delta/\delta}\right) - 2\,{}_{\Delta}b} \text{ where}$$

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

$$+4\,\frac{\left((\Delta/\delta)\left(1 - \left({}_{\delta}\alpha + {}_{\delta}\beta\right)\right) - \left(1 - \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1 - {}_{\delta}\beta\left({}_{\delta}\alpha + {}_{\delta}\beta\right)\right)}{1 - \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^2}$$

$$\text{and } {}_{\Delta}b = \left({}_{\delta}\alpha - {}_{\delta}\alpha\,{}_{\delta}\beta\left({}_{\delta}\alpha + {}_{\delta}\beta\right)\right)\frac{1 - \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{2\Delta/\delta}}{1 - \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^2}$$

We want a discretization that ensures that the above relationship will still hold. By taking the limits of the last two equations when δ converges to zero, we obtain that:

$$_{\Delta}a^* := \lim_{\delta\downarrow 0} {}_{\Delta}a = \Delta\,\alpha^2 + (\Delta)^2\,\theta\left(\theta - \alpha^2\right) + 2\alpha^2\left(\Delta - \left(1 - e^{-\theta\Delta}\right)/\theta\right) \quad {}_{\Delta}b^* := \lim_{\delta\downarrow 0} {}_{\Delta}b = \frac{\alpha^2}{2\theta}\left(1 - e^{-2\theta\Delta}\right)$$

This means that we can discretize the continuous model by solving the following equation:

$$\frac{_{\Delta}\beta}{1 + {}_{\Delta}\beta^2} = \frac{_{\Delta}a^*\, e^{-\theta\Delta} - {}_{\Delta}b^*}{_{\Delta}a^*\left(1 + e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^*}$$

First, we have to make sure 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:

$$f(x) = x^2 - \frac{_{\Delta}a^*\left(1 + e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^*}{_{\Delta}a^*\, e^{-\theta\Delta} - {}_{\Delta}b^*}x + 1$$

This has two roots $x_{1,2}$ where $x_1\ x_2 = 1$ and $x_1 + x_2 = \frac{{}_{\Delta}a^*\left(1+e^{-2\theta\Delta}\right)-2\,{}_{\Delta}b^*}{{}_{\Delta}a^*\ e^{-\theta\Delta}-{}_{\Delta}b^*}$. If we show that the sum of the roots is positive, then both roots will be positive, but only one will be smaller than 1. First, for the existence it is enough to show that:

$$\frac{{}_{\Delta}a^*\left(1+e^{-2\theta\Delta}\right)-2\,{}_{\Delta}b^*}{{}_{\Delta}a^*\ e^{-\theta\Delta}-{}_{\Delta}b^*} > 2$$

If ${}_{\Delta}a^*\ e^{-\theta\Delta}-{}_{\Delta}b^*$ is positive, then the above is equivalent to (because ${}_{\Delta}a$ is positive):

$$\left(1-e^{-\theta\Delta}\right)^2 > 0$$

which is obviously true.

Therefore, all we need to show is that ${}_{\Delta}a^*\ e^{-\theta\Delta}-{}_{\Delta}b^* > 0$. This is equivalent to:

$$e^{-\theta\Delta}\left(\Delta+(\Delta)^2\,\theta\left(\theta/\alpha^2-1\right)+2\left(\Delta-\frac{1-e^{-\theta\Delta}}{\theta}\right)\right) > \frac{1-e^{-2\theta\Delta}}{2\theta}$$

$$6\theta\Delta+2(\Delta)^2\,\theta^3/\alpha^2+5e^{-\theta\Delta} > e^{\theta\Delta}+2\theta^2(\Delta)^2+4$$

Both sides converge to 5 when $\Delta \to 0$, but we have to show that the left hand side converges faster. The above is equivalent with the following relationship between the first order derivatives:

$$6\theta+4\Delta\,\theta^3/\alpha^2-5\theta e^{-\theta\Delta} > \theta e^{\theta\Delta}+4\theta^2\Delta$$

Again, the two sides have the same limit when Δ converges to zero. For the above we need the following relationship (between the derivatives of the two sides with respect to Δ) to hold:

$$4\theta^3/\alpha^2+5\theta^2 e^{-\theta\Delta} > \theta^2 e^{\theta\Delta}+4\theta^2$$

This holds for small values of Δ because the relationship between the limits of the two sides is:

$$4\theta^3/\alpha^2+5\theta^2 > 5\theta^2$$

and this is obviously true. Thus, we showed that ${}_{\Delta}a^{*}e^{-\Delta\theta}-{}_{\Delta}b^* > 0$, 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:

$$\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2} = \frac{{}_{\Delta}a^*\ e^{-\theta\Delta}-{}_{\Delta}b^*}{{}_{\Delta}a^*\left(1+e^{-2\theta\Delta}\right)-2\,{}_{\Delta}b^*}$$

The solution is:

$$
{}_{\Delta}\beta = \left(\frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*}}{{}_{\Delta}a^{*}\; e^{-\theta\Delta} - {}_{\Delta}b^{*}} - \sqrt{\left(\frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*}}{{}_{\Delta}a^{*}\; e^{-\theta\Delta} - {}_{\Delta}b^{*}} \right)^{2} - 4} \right) / 2 =
$$

$$
= \frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*} + \sqrt{{}_{\Delta}a^{*2}\left(1-e^{-2\theta\Delta}\right)^{2} - 4a^{*}b^{*}\left(1-e^{-\theta\Delta}\right)^{2}}}{2\left({}_{\Delta}a^{*}\; e^{-\theta\Delta} - {}_{\Delta}b^{*}\right)} =
$$

$$
= \frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*} + \left(1-e^{-\theta\Delta}\right)\sqrt{{}_{\Delta}a^{*2}\left(1+e^{-\theta\Delta}\right)^{2} - 4a^{*}b^{*}}}{2\left({}_{\Delta}a^{*}\; e^{-\theta\Delta} - {}_{\Delta}b^{*}\right)}
$$

<u>Part B: Discretization of the model returns original weak GARCH</u>

The continuous time process is:

$$
\frac{dS}{S} = \mu dt + \sqrt{V}\, dB_1
$$

$$
dV = \left(\omega - \theta V\right) dt + \alpha\tau V dB_1 + \alpha\sqrt{\eta - \tau^2 - 1}\; V\; dB_3
$$

The independent Brownians dB_1 and dB_3 are discretized as:

$$
\Delta B_1\left(k\Delta\right) = \sqrt{\Delta}\;\; {}_{\Delta}\xi_{1,(k+1)\Delta} \quad \text{with} \quad {}_{\Delta}\xi_{1,(k+1)\Delta} \sim N\left(0,1\right) \tag{7}
$$

$$
\Delta B_3\left(k\Delta\right) = \sqrt{\Delta}\;\; {}_{\Delta}\xi_{3,(k+1)\Delta} \quad \text{with} \quad {}_{\Delta}\xi_{3,(k+1)\Delta} \sim N\left(0,1\right) \tag{8}
$$

Also, we have:

$$
{}_{\Delta}\varepsilon_{(k+1)\Delta} = \sqrt{{}_{\Delta}V_{k\Delta}}\;\sqrt{\Delta}\;{}_{\Delta}\xi_{1,(k+1)\Delta} \tag{9}
$$

$$
{}_{\Delta}\varepsilon_{3,(k+1)\Delta} = \sqrt{{}_{\Delta}V_{k\Delta}}\;\sqrt{\Delta}\;{}_{\Delta}\xi_{3,(k+1)\Delta} \tag{10}
$$

Thus, the first equation can be discretized as:

$$
\ln\left(S_{(k+1)\Delta} / S_{k\Delta}\right) \approx \frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} = \mu\,\Delta + {}_{\Delta}\varepsilon_{(k+1)\Delta}
$$

There is no loss of generality assuming the following:

$$
{}_{\Delta}\xi_{3,(k+1)\Delta} = \frac{{}_{\Delta}\xi^{2}_{1,(k+1)\Delta} - 1}{\sqrt{2}}
$$

This will not follow an exact normal distribution (instead, it is derived from the chi-square distribution). However, we show that it satisfies the basic properties of the B_3 Brownian motion: zero mean, variance of 1, and independence from B_1:

$$E\left({}_{\Delta}\xi_{3,(k+1)\Delta} \right) = 0$$

$$\mathrm{Var}\left({}_{\Delta}\xi_{3,(k+1)\Delta} \right) = E\left({}_{\Delta}\xi^2_{3,(k+1)\Delta} \right) = E\left(\frac{ {}_{\Delta}\xi^4_{1,(k+1)\Delta} - 2\,{}_{\Delta}\xi^2_{1,(k+1)\Delta} + 1}{2} \right) = \frac{3-2+1}{2} = 1$$

$$E\left({}_{\Delta}\xi_{1,(k+1)\Delta}\ {}_{\Delta}\xi_{3,(k+1)\Delta} \right) = E\left({}_{\Delta}\xi_{1,(k+1)\Delta}\ \frac{ {}_{\Delta}\xi^2_{1,(k+1)\Delta} - 1}{\sqrt{2}} \right) = 0$$

We obtain that:

$$E\left({}_{\Delta}\varepsilon^2_{(k+1)\Delta} / \Delta \middle| {}_{\Delta}I_{k\Delta} \right) = {}_{\Delta}V_{k\Delta}$$

The parameters are discretized in the following way:

$$\begin{aligned} &\omega dt \rightarrow {}_{\Delta}\omega \\ &\theta dt \rightarrow 1 - \left({}_{\Delta}\alpha + {}_{\Delta}\beta \right) \\ &\alpha\sqrt{dt} \rightarrow {}_{\Delta}\alpha \end{aligned}$$

The first two discretizations are intuitive because of the following:

$$\omega\ \Delta t \approx \omega\ \Delta t \frac{\left(1 - \exp\left(-\theta\ \Delta t\right)\right)}{\theta\ \Delta t} = {}_{\Delta}\omega$$

$$\theta\ \Delta t \approx 1 - \exp\left(-\theta\ \Delta t\right) = 1 - \left({}_{\Delta}\alpha + {}_{\Delta}\beta \right)$$

The processes for the higher moments are discretized as:

$$\begin{aligned} \tau(t) &\rightarrow {}_{\Delta}\tau_{k\Delta} = \tau(k\Delta) \quad \textit{where} \quad k\Delta \le t < (k+1)\Delta \\ \eta(t) &\rightarrow {}_{\Delta}\eta_{k\Delta} = \eta(k\Delta) \quad \textit{where} \quad k\Delta \le t < (k+1)\Delta \end{aligned}$$

The discrete time variance process can be expressed as:

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

$$\begin{aligned} {}_{\Delta}u_{k\Delta} = {}& {}_{\Delta}\alpha\left({}_{\Delta}V_{(k-1)\Delta} - {}_{\Delta}\varepsilon^2_{k\Delta} / \Delta \right) + {}_{\Delta}\alpha\,{}_{\Delta}\tau_{(k-1)\Delta}\sqrt{ {}_{\Delta}V_{(k-1)\Delta}}\ \frac{ {}_{\Delta}\varepsilon_{k\Delta}}{\sqrt{\Delta}} + \\ & + {}_{\Delta}\alpha\sqrt{ {}_{\Delta}\eta_{(k-1)\Delta} - {}_{\Delta}\tau^2_{(k-1)\Delta} - 1}\sqrt{ {}_{\Delta}V_{(k-1)\Delta}}\ \frac{ {}_{\Delta}\varepsilon_{3,k\Delta}}{\sqrt{\Delta}} \end{aligned}$$

So far we considered the conditional variance; however, the value of interest here is not this, but the BLP of the squared residuals. We define the following process:

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

It is easy to see that this follows the GARCH process below:

$$ {}_{\Delta}h_{k\Delta} = {}_{\Delta}\omega + {}_{\Delta}\alpha\, {}_{\Delta}\varepsilon_{k\Delta}^{2} / \Delta + {}_{\Delta}\beta\, {}_{\Delta}h_{(k-1)\Delta} $$

We have to show that ${}_{\Delta}h_{k\Delta}$ is the BLP of ${}_{\Delta}\varepsilon_{(k+1)\Delta}^{2} / \Delta$ (in order to have a weak GARCH specification). This requires showing that:

$$ E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta}^{2} / \Delta - {}_{\Delta}h_{k\Delta}\right) {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i \geq 0 \quad r = 0,1,2 $$

We already know that the following is true:

$$ E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta}^{2} / \Delta - {}_{\Delta}V_{k\Delta}\right) {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i \geq 0 \quad r = 0,1,2 $$

Thus, we need to show that:

$$ E\left(\left(\sum_{j=0}^{k} {}_{\Delta}\beta^{j}\, {}_{\Delta}u_{(k-j)\Delta}\right) {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i \geq 0 \quad r = 0,1,2 $$

Or, equivalently:

$$ E\left({}_{\Delta}u_{(k-j)\Delta}\, {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i,j \geq 0 \quad r = 0,1,2 $$

This can be decomposed into the following:

$$ E\left(\begin{pmatrix} {}_{\Delta}\alpha\left({}_{\Delta}V_{(k-j-1)\Delta} - {}_{\Delta}\varepsilon_{(k-j)\Delta}^{2} / \Delta\right) + \\ + {}_{\Delta}\alpha\, {}_{\Delta}\tau_{(k-j-1)\Delta} \sqrt{{}_{\Delta}V_{(k-j-1)\Delta}}\, \dfrac{{}_{\Delta}\varepsilon_{(k-j)\Delta}}{\sqrt{\Delta}} \end{pmatrix} {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i,j \geq 0 \quad r = 0,1,2 \qquad (11) $$

$$ E\left({}_{\Delta}\alpha \sqrt{{}_{\Delta}\eta_{(k-j-1)\Delta} - {}_{\Delta}\tau_{(k-j-1)\Delta}^{2} - 1} \sqrt{{}_{\Delta}V_{(k-j-1)\Delta}}\, \frac{{}_{\Delta}\varepsilon_{3,(k-j)\Delta}}{\sqrt{\Delta}}\, {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i,j \geq 0 \quad r = 0,1,2 \qquad (12) $$

The last equation is obviously true due to the independence of the two error terms. Equation (11) is satisfied for $r = 0$ always, for $r = 1$ when $i \neq j$, and for $r = 2$ when $i > j$. Thus, we need to prove this for $r = 1$ and $i = j$, and for $r = 2$ and $i \leq j$. We ignore the proof for $r = 2$, and below is the proof that equation (11) is satisfied for $r = 1$ and $i = j$:

We have to show that:

$$ E\left(\begin{pmatrix} {}_{\Delta}\alpha\left({}_{\Delta}V_{(k-i-1)\Delta} - {}_{\Delta}\varepsilon_{(k-i)\Delta}^{2} / \Delta\right) + \\ + {}_{\Delta}\alpha\, {}_{\Delta}\tau_{(k-i-1)\Delta} \sqrt{{}_{\Delta}V_{(k-i-1)\Delta}}\, \dfrac{{}_{\Delta}\varepsilon_{(k-i)\Delta}}{\sqrt{\Delta}} \end{pmatrix} {}_{\Delta}\varepsilon_{(k-i)\Delta}^{r}\right) = 0 \quad i \geq 0 $$

This is equivalent to:

$$E\left(\left(\left({}_{\Delta}V_{(k-i-1)\Delta} - {}_{\Delta}\varepsilon^{2}_{(k-i)\Delta} / \Delta\right) + {}_{\Delta}\tau_{(k-i-1)\Delta}\sqrt{{}_{\Delta}V_{(k-i-1)\Delta}}\ \frac{{}_{\Delta}\varepsilon_{(k-i)\Delta}}{\sqrt{\Delta}}\right){}_{\Delta}\varepsilon_{(k-i)\Delta}\right) = 0 \quad i \geq 0$$

Or

$$E\left(\left(\left({}_{\Delta}V_{(k-i-1)\Delta}\,{}_{\Delta}\varepsilon_{(k-i)\Delta} - {}_{\Delta}\varepsilon^{3}_{(k-i)\Delta} / \Delta\right) + {}_{\Delta}\tau_{(k-i-1)\Delta}\sqrt{{}_{\Delta}V_{(k-i-1)\Delta}}\ \frac{{}_{\Delta}\varepsilon^{2}_{(k-i)\Delta}}{\sqrt{\Delta}}\right)\right) = 0 \quad i \geq 0$$

This reduces to:

$$E\left(E\left(\left({}_{\Delta}V_{(k-i-1)\Delta}\,{}_{\Delta}\varepsilon_{(k-i)\Delta} - {}_{\Delta}\varepsilon^{3}_{(k-i)\Delta} / \Delta\right) + {}_{\Delta}\tau_{(k-i-1)\Delta}\sqrt{{}_{\Delta}V_{(k-i-1)\Delta}}\ \frac{{}_{\Delta}\varepsilon^{2}_{(k-i)\Delta}}{\sqrt{\Delta}}\,\middle|\,{}_{\Delta}I_{(k-i-1)\Delta}\right)\right) = 0 \quad i \geq 0$$

We know that:

$$E\left({}_{\Delta}\varepsilon_{(k+1)\Delta}\,\middle|\,{}_{\Delta}I_{k\Delta}\right) = 0$$

$$E\left({}_{\Delta}\varepsilon^{2}_{(k+1)\Delta} / \Delta\,\middle|\,{}_{\Delta}I_{k\Delta}\right) = {}_{\Delta}V_{k\Delta}$$

$$E\left({}_{\Delta}\varepsilon^{3}_{(k+1)\Delta} / \left(\Delta^{3/2}\,{}_{\Delta}V^{3/2}_{k\Delta}\right)\,\middle|\,{}_{\Delta}I_{k\Delta}\right) = {}_{\Delta}\tau_{k\Delta}$$

$$E\left({}_{\Delta}\varepsilon^{4}_{(k+1)\Delta} / \left(\Delta^{2}\,{}_{\Delta}V^{2}_{k\Delta}\right)\,\middle|\,{}_{\Delta}I_{k\Delta}\right) = {}_{\Delta}\eta_{k\Delta}$$

We get:

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

$$E\left({}_{\Delta}\varepsilon^{3}_{(k-i)\Delta}\,\middle|\,{}_{\Delta}I_{(k-i-1)\Delta}\right) = \Delta^{3/2}\,{}_{\Delta}V^{3/2}_{(k-i-1)\Delta}\,{}_{\Delta}\tau_{(k-i-1)\Delta}$$

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

Thus we need to show that:

$$E\left(-\Delta^{1/2}\,{}_{\Delta}V^{3/2}_{(k-i-1)\Delta}\,{}_{\Delta}\tau_{(k-i-1)\Delta} + \Delta^{1/2}\,{}_{\Delta}V_{(k-i-1)\Delta}\sqrt{{}_{\Delta}V_{(k-i-1)\Delta}}\,{}_{\Delta}\tau_{(k-i-1)\Delta}\right) = 0 \quad i \geq 0$$

Which is obviously true.

Thus, we showed that ${}_{\Delta}h_{k\Delta}$ is the BLP of ${}_{\Delta}\varepsilon^{2}_{(k+1)\Delta} / \Delta$ so we have a weak GARCH specification.

Part C: Time aggregation property is preserved

The time aggregation property is preserved, as we shall see in the following. First, it is easy to see that the two expressions below hold:

$$ {}_{\Delta}\alpha = \left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{\Delta/\delta} - {}_{\Delta}\beta $$

$$ {}_{\Delta}\omega = {}_{\delta}\omega \frac{1-\left({}_{\delta}\alpha + {}_{\delta}\beta\right)^{\Delta/\delta}}{1-\left({}_{\delta}\alpha + {}_{\delta}\beta\right)} $$

Then, we know that for any time step Δ we have:

$$ {}_{\Delta}\beta = \frac{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*} + \left(1-e^{-\theta\Delta}\right)\sqrt{{}_{\Delta}a^{*2}\left(1+e^{-\theta\Delta}\right)^{2} - 4\,{}_{\Delta}a^{*}\,{}_{\Delta}b^{*}}}{2\left({}_{\Delta}a^{*}\,e^{-\theta\Delta t} - {}_{\Delta}b^{*}\right)} $$

with

$$ {}_{\Delta}a^{*} = \Delta\,\alpha^{2} + \Delta^{2}\theta\left(\theta-\alpha^{2}\right) + 2\alpha^{2}\left(\Delta - \left(1-e^{-\theta\Delta}\right)/\theta\right) \quad \text{and} \quad {}_{\Delta}b^{*} = \frac{\alpha^{2}}{2\theta}\left(1-e^{-2\theta\Delta}\right). $$

This means that the following equation holds:

$$ \frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^{2}} = \frac{{}_{\Delta}a^{*}e^{-\theta\Delta} - {}_{\Delta}b^{*}}{{}_{\Delta}a^{*}\left(1+e^{-2\theta\Delta}\right) - 2\,{}_{\Delta}b^{*}} $$

Also, we know that, for any Δ:

$$ {}_{\Delta}\varkappa = 3 + 6\left(\varkappa-1\right)\frac{\alpha^{2}\left(\theta\Delta - \left(1-e^{-\theta\Delta}\right)\right)}{\theta^{2}\Delta^{2}\left(\alpha^{2}+2\theta\right)} \quad \text{with} \quad \varkappa = \frac{3}{1-\alpha^{2}/\theta} $$

We need to prove the following two identities:

$$ {}_{\Delta}\varkappa = 3 + \frac{{}_{\delta}\varkappa - 3}{\Delta/\delta} + $$

$$ +6\left({}_{\delta}\varkappa - 1\right)\frac{\left(\left(\Delta/\delta\right)\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right) - \left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2} + {}_{\delta}\alpha\left({}_{\delta}\alpha+{}_{\delta}\beta\right)\right)}{\left(\Delta/\delta\right)^{2}\left(1-{}_{\delta}\alpha-{}_{\delta}\beta\right)^{2}\left(1-\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2} + {}_{\delta}\alpha^{2}\right)} $$

and

$$ \frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^{2}} = \frac{{}_{\Delta}a\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{\Delta/\delta} - {}_{\Delta}b}{{}_{\Delta}a\left(1+\left({}_{\delta}\alpha+{}_{\delta}\beta\right)^{2\Delta/\delta}\right) - 2\,{}_{\Delta}b} $$

where the following notation is used:

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

$$+4\frac{\left((\Delta/\delta)\left(1-({}_{\delta}\alpha+{}_{\delta}\beta)\right)-\left(1-({}_{\delta}\alpha+{}_{\delta}\beta)^{\Delta/\delta}\right)\right){}_{\delta}\alpha\left(1-{}_{\delta}\beta({}_{\delta}\alpha+{}_{\delta}\beta)\right)}{1-({}_{\delta}\alpha+{}_{\delta}\beta)^2}$$

and

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

The first expression that we need to prove is equivalent to:

$$(\varkappa-1)\alpha^2\left(1-e^{-\theta\delta}\right)^2\left(1-e^{-2\theta\delta}+{}_{\delta}\alpha^2\right)=$$
$$\left(2\delta^2\theta^2\left(\alpha^2+2\theta\right)+6(\varkappa-1)\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right){}_{\delta}\alpha\left(1-e^{-2\theta\delta}+{}_{\delta}\alpha e^{-\theta\delta}\right)$$

This further simplifies to:

$$\left(e^{-\theta\delta}\left(2\delta^2\theta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)-\alpha^2\left(1-e^{-\theta\delta}\right)^2\right)\left(1+{}_{\delta}\beta^2\right)$$
$$\left(\left(2\delta^2\theta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(1+e^{-2\theta\delta}\right)-2\alpha^2\left(1-e^{-\theta\delta}\right)^2e^{-\theta\delta}\right){}_{\delta}\beta$$

which can be reduced to:

$$\frac{{}_{\delta}\beta}{1+{}_{\delta}\beta^2}=\frac{\left(\delta\alpha^2+\delta^2\theta\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)e^{-\theta\delta}-\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta}{\left(\delta\alpha^2+\delta^2\theta\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)\left(1+e^{-2\theta\delta}\right)-2\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta}$$

This expression holds, so we just showed that the kurtosis is time aggregating.

Using the notation:

$$A=\frac{1+{}_{\delta}\beta^2}{{}_{\delta}\beta}$$

the second expression simplifies to showing that:

$$\frac{{}_{\Delta}\beta}{1+{}_{\Delta}\beta^2}=\frac{{}_{\Delta}\tilde{a}e^{-\theta\Delta}-{}_{\Delta}\tilde{b}}{{}_{\Delta}\tilde{a}\left(1+e^{-2\theta\Delta}\right)-2\,{}_{\Delta}\tilde{b}}$$

where

$$
{}_{\Delta}\tilde{a} = (\Delta/\delta)(A-2)\,{}_{\delta}\beta + 2(\Delta/\delta)(\Delta/\delta - 1)\frac{\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right){}_{\delta}\beta}{\left({}_{\delta}\varkappa - 1\right)\left(1-e^{-2\theta\delta}\right)} +
$$

$$
+4\,\frac{\left((\Delta/\delta)\left(1-e^{-\theta\delta}\right)-\left(1-e^{-\theta\Delta}\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right){}_{\delta}\beta}{1-e^{-2\theta\delta}}
$$

$$
{}_{\Delta}\tilde{b} = \left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right){}_{\delta}\beta\,\frac{1-e^{-2\theta\Delta}}{1-e^{-2\theta\delta}}
$$

This can be reduced to showing that:

$$
\frac{{}_{\Delta}\tilde{a}}{{}_{\Delta}\underset{\sim}{b}} = \frac{{}_{\Delta}a^{*}}{{}_{\Delta}b^{*}}
$$

which simplifies to the expression below:

$$
\frac{2\Delta(\Delta-\delta)\theta^2\left(\theta-\alpha^2\right)\dfrac{\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right)}{\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)} + \Delta(A-2)\left(\dfrac{1-e^{-2\theta\delta}}{\delta}\right)+4\left(\Delta\left(\dfrac{1-e^{-\theta\delta}}{\delta}\right)-\left(1-e^{-\theta\Delta}\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}{\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)} =
$$

$$
= \frac{\left(\Delta\alpha^2+\Delta^2\theta\left(\theta-\alpha^2\right)+2\alpha^2\left(\Delta-\left(1-e^{-\theta\Delta}\right)/\theta\right)\right)}{\alpha^2/2\theta}
$$

Furthermore, using simple algebra, this can be expressed as:

$$
\Delta^2\theta\left(\theta-\alpha^2\right)\alpha^2\,\frac{\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right)}{\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)} +
$$

$$
+\Delta\left(\begin{array}{l}
\alpha^2\left(\dfrac{A-2}{\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}\right)\left(\dfrac{1-e^{-2\theta\delta}}{\delta}\right)/2\theta + 2\alpha^2\left(\dfrac{1-e^{-\theta\delta}}{\delta}\right)/\theta - \\
-\delta\theta\left(\theta-\alpha^2\right)\alpha^2\,\dfrac{\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right)}{\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}
\end{array}\right) -
$$

$$
-2\alpha^2\left(1-e^{-\theta\Delta}\right)/\theta =
$$

$$
= \Delta^2\theta\left(\theta-\alpha^2\right)+3\Delta\alpha^2-2\alpha^2\left(1-e^{-\theta\Delta}\right)/\theta
$$

Matching terms on both sides of the equation that contain Δ and Δ^2, this reduces to showing the following two expressions:

$$\frac{\alpha^2\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right)}{\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}=1$$

and

$$\left(\frac{A-2}{\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}\right)\left(\frac{1-e^{-2\theta\delta}}{\delta}\right)/2\theta+2\left(\frac{1-e^{-\theta\delta}}{\delta}\right)/\theta-$$

$$-\delta\theta\left(\theta-\alpha^2\right)\frac{\left(1-e^{-\theta\delta}\right)^2\left(A-2e^{-\theta\delta}\right)}{\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)}=3$$

The first one reduces to:

$$A=\frac{\left(\alpha^2\delta+\theta\delta^2\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)\left(1+e^{-2\theta\delta}\right)-2\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta}{\left(\alpha^2\delta+\theta\delta^2\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)e^{-\theta\delta}-\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta}$$

which is obviously satisfied.

The second expression can be further expressed as:

$$\left(A-2\right)\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)\left(1-e^{-2\theta\delta}\right)/2\theta$$

$$+\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)2\left(1-e^{-\theta\delta}\right)\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)/\theta-$$

$$-\left(A-2e^{-\theta\delta}\right)\delta^2\theta\left(\theta-\alpha^2\right)\left(1-e^{-\theta\delta}\right)^2=$$

$$=\left(e^{-\theta\delta}A-\left(1+e^{-2\theta\delta}\right)\right)3\delta\left(2\theta^2\delta^2\left(\theta-\alpha^2\right)+6\alpha^2\left(\theta\delta-\left(1-e^{-\theta\delta}\right)\right)\right)$$

By expressing *A* in this expression, this reduces to:

$$A=\frac{6\left(e^{-\theta\delta}+\delta\theta-1\right)\left(\left(\alpha^2\delta+\theta\delta^2\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)\left(1+e^{-2\theta\delta}\right)-2\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta\right)}{6\left(e^{-\theta\delta}+\delta\theta-1\right)\left(\left(\alpha^2\delta+\theta\delta^2\left(\theta-\alpha^2\right)+2\alpha^2\left(\delta-\left(1-e^{-\theta\delta}\right)/\theta\right)\right)e^{-\theta\delta}-\alpha^2\left(1-e^{-2\theta\delta}\right)/2\theta\right)}$$

which is obviously true.

References

Alexander, C and L. Nogueira (2007): 'Model-free hedge ratios and scale-invariant models', *Journal of Banking and Finance* Vol. 31 (6), 1839-1861

Baillie, R.T. and T. Bollerslev (1992): 'Prediction in Dynamic Models with Time-dependent Conditional Variances', *Journal of Econometrics* 52, 91-113

Bakshi, G., Kapadia, N. and Madan, D. (2003): 'Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options', *The Review of Financial Studies* Vol. 16 (1), 101-143

Bates, D.S. (1997): 'The Skewness Premium: Option Pricing under Asymmetric Processes', *Advances in Futures and Options Research* Vol. 9, 51-82

Bates, D.S. (2000): 'Post-'87 Crash Fears in the S&P 500 Futures Option Market', *Journal of Econometrics* Vol. 94, 181-238

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

Breuer, T. and M. Jandacka (2007): 'Temporal Aggregation of GARCH Models: Conditional Kurtosis and Optimal Frequency', available at SSRN: http://ssrn.com/abstract=967824

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

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

Duan, J. (1997): 'Augmented GARCH(p,q) Process and its Diffusion Limit', *Journal of Econometrics* Vol. 79, 97-127

Duan, J., Ritchken, P. and Sun, Z. (2006): 'Approximating GARCH-Jump Models, Jump-Diffusion Processes, and Option Pricing', *Mathematical Finance* Vol. 16, 21-52

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

Heston, S (1993): 'A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options', *Review of Financial Studies* Vol. 6 (2), 327-343

Jeantheau, T. (2004): 'A Link Between Complete Models with Stochastic Volatility and ARCH Models', *Finance and Stochastics* Vol. 8, 111-131

Jondeau, E. (2008): 'Contemporaneous Aggregation of GARCH Models and Evaluation of the Aggregation Bias', *Swiss Finance Institute* Research Paper 08-06

Kallsen, J. and Taqqu, M.S. (1998): 'Option Pricing in ARCH-Type Models', *Mathematical Finance* Vol. 8 (1), 13-26

Kazmerchuk, Y.; Swishchuk, A. and Wu, J. (2002): 'A Continuous-time GARCH Model for Stochastic Volatility with Delay', *Department of Mathematics & Statistics, York University Toronto* Discussion Paper

Klüppelberg, 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

Komunjer, I. (2001): 'Consistent Estimation for Aggregated GARCH Processes', *University of California San Diego* DP 2001-08

Meddahi, N. and Renault, E. (200

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

Müller, 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

Nelson, D.B. (1991): 'Conditional Heteroskedasticity in Asset Returns: A New Approach', *Econometrica* Vol. 59 (2), 347-370

Nijman, T.E. and E. Sentana (1996): 'Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes', *Journal of Econometrics* Vol. 71, 71-87

Rossi, P.E. editor (1996): 'Modelling Stock Market Volatility: Bridging the Gap to Continuous Time', *Academic Press Inc.*

Taylor, S.J. (1986): 'Modelling Financial Time Series', *John Wiley and Sons, Chichester, UK*

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

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

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