



THE BUSINESS SCHOOL
FOR FINANCIAL MARKETS



The University of Reading

On The Continuous Limit of GARCH

ICMA Centre Discussion Papers in Finance 2005-13

19 October 2005

Carol Alexander

ICMA Centre, University of Reading

Emese Lazar

ICMA Centre, University of Reading

Copyright 2005 Alexander and Lazar. All rights reserved.

ICMA Centre • The University of Reading
Whiteknights • PO Box 242 • Reading RG6 6BA • UK
Tel: +44 (0)1183 788239 • Fax: +44 (0)1189 314741
Web: www.icmacentre.rdg.ac.uk

Director: Professor John Board, Chair in Finance
The ICMA Centre is supported by the International Capital Market Association



Abstract[§]

GARCH processes constitute the major area of time series variance analysis, hence the limit of these processes is of considerable interest for continuous time volatility modelling. The continuous time limit of the GARCH(1,1) model is fundamental for limits of other GARCH processes, yet it has been the point of much debate between econometricians. The seminal work of Nelson (1990) derived the GARCH(1,1) limit as a stochastic volatility process, uncorrelated with the price process. But then a subsequent paper of Corradi (2000) that derives the limit as a deterministic volatility process and several other contradictory papers followed. We reconsider this continuous limit, arguing that because the strong GARCH model is not aggregating in time it is incorrect to consider its limit. Instead it is legitimate to use the weak definition of GARCH that is aggregating in time. This model differs from strong GARCH by defining the discrete time process on the best linear predictor of the squared errors, rather than the conditional variance itself. We prove that its continuous limit is a stochastic volatility model with 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. Under certain assumptions our limit model reduces to Nelson's GARCH diffusion.

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 786768 (ICMA Centre) Fax: +44 (0)1189 314741

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 786675 (ICMA Centre) Fax: +44 (0)1189 314741

[§] We are grateful for useful comments on an earlier draft from Lars Hansen, University of Chicago, Nour Meddahi of University of Montreal, Xiaohong Chen from New York University, Hyungsok Ahn of Commerz Bank, Jacques Pezier from the ICMA Centre, The University of Reading, Valentina Corradi of Queen Mary, University of London and Roel Oomen from Warwick Business School.

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 nonparametric methods have just started to flourish, parametric methods are more widely spread and in this group GARCH is considered the most popular discrete time framework to characterize the dynamic behaviour of the variance process. On the other hand, continuous time parametric modelling has focussed on smile consistent models and uses a very different set of tools, mostly stochastic volatility models with correlated brownians that can incorporate jumps as well in the price 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, that also introduces one of the most important approaches for GARCH option pricing, the author derives the continuous limit of GARCH 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 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 (2005) 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. This way, choosing between Nelson's and Corradi's limiting models is not a straightforward task.

Other papers that investigate continuous time equivalents for the GARCH process include Kallsen and Taqqu (1998), Kazmerchuk et al. (2002) and Klüppelberg, Lindner and Maller (2004). Kallsen and Taqqu's 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. 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 it has only one source of randomness. However it is not the limit of the discrete time GARCH but a continuous time extension.

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. This means that if GARCH(1,1) is the data generating process (DGP) for a given frequency, then for any other frequency GARCH(1,1) will not be the DGP. The computation of the continuous time limit for such a model is therefore of questionable validity.

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 $\text{Corr}(x(t), x(t)^2) = 0$. This way, with only one source of uncertainty two uncorrelated (but not independent) processes can be created.

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 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.⁶ 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.⁷ By contrast, the limit of 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.

The remainder of this paper is organized as follows: Section II re-examines the continuous time limit of the GARCH(1,1) model and Section III concludes.

II The continuous limit of GARCH processes

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:

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

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

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

⁶ 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) model.

⁷ See, for example, Alexander and Nogueira (2005).

Now the classical (strong) definition states that:⁸

$$E(\epsilon_{t+1}^2 | I_t) = h_t \quad (2)$$

We define the step-length Δ and consider the continuous limit as $\Delta \downarrow 0$. Our notation for a time series with step-length Δ indexes time as $k\Delta$, with $k = 1, 2, \dots$. This way, for any Δ we can define the Δ -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:

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

the limit will be a stochastic volatility model:

$$\begin{aligned} \frac{dS}{S} &= \mu dt + \sqrt{V} dB_1 \\ dV &= (\omega - \theta V) dt + \sqrt{2\alpha V} dB_2 \end{aligned}$$

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:

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

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

$$\begin{aligned} \frac{dS}{S} &= \mu dt + \sqrt{V} dB \\ dV &= (\omega - \theta V) dt \end{aligned}$$

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. But we argue that the limits of both Nelson (1990) and Corradi (2000) are inaccurate, because they have worked with

⁸ 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 .

the strong definition of GARCH. A major disadvantage of the strong definition is that it does not guarantee time aggregation; namely that if we have a strong GARCH process for a given frequency, then for any other frequency the process will not be a strong GARCH process.

Instead it is necessary to use a definition that guarantees that the process will exist and be the same for any frequency, meaning that the model is aggregating in time. This requires the use of the weak definition when computing the limiting model. Drost and Nijman (1993) introduced the definition of the weak GARCH process that, contrary to the strong process, is aggregating in time. The difference is that the weak GARCH specifies that b_t in (1) is not the conditional variance, but the best linear predictor (BLP) of the squared residuals. In weak GARCH (2) is replaced by the conditions:

$$\begin{aligned} E(\varepsilon_t \varepsilon_{t-i}^r) &= 0 \quad i \geq 0 \quad r = 0, 1, 2 \\ E((\varepsilon_{t+1}^2 - b_t) \varepsilon_{t-i}^r) &= 0 \quad i \geq 0 \quad r = 0, 1, 2 \end{aligned}$$

The assumption that 0 and b_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 $\delta < \Delta$. Since we need to compare variances at different time steps, ${}_{\Delta}b_{k\Delta}$ 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:

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

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

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

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

Similarly, ${}_{\delta}b_{k\delta}$ 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 .

The weak definition of GARCH implies a relationship between the parameters and unconditional kurtosis ${}_{\Delta}\kappa$ of the Δ -step process ${}_{\Delta}b_{k\Delta}$ and the parameters and unconditional kurtosis of the δ -step

process denoted by ${}_{{\delta}}b_{{\delta}\delta}$, where the pre-subscript denotes the step-length. This relationship was derived by Drost and Nijman, and is given by the following set of equations:⁹

$$\begin{aligned}\Delta\omega &= {}_{\delta}\omega \frac{1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^{\Delta/\delta}}{1 - ({}_{\delta}\alpha + {}_{\delta}\beta)}; \Delta\alpha = ({}_{\delta}\alpha + {}_{\delta}\beta)^{\Delta/\delta} - \Delta\beta \\ \Delta\kappa &= 3 + \frac{{}_{\delta}\kappa - 3}{\Delta/\delta} + \\ &+ 6({}_{\delta}\kappa - 1) \frac{\left((\Delta/\delta)(1 - ({}_{\delta}\alpha + {}_{\delta}\beta)) - (1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^{\Delta/\delta}) \right) {}_{\delta}\alpha (1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2 + {}_{\delta}\alpha ({}_{\delta}\alpha + {}_{\delta}\beta))}{(\Delta/\delta)^2 (1 - {}_{\delta}\alpha - {}_{\delta}\beta)^2 (1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2 + {}_{\delta}\alpha^2)} \\ \Delta\beta \text{ is the solution to } \frac{\Delta\beta}{1 + \Delta\beta^2} &= \frac{\Delta a ({}_{\delta}\alpha + {}_{\delta}\beta)^{\Delta/\delta} - \Delta b}{\Delta a (1 + ({}_{\delta}\alpha + {}_{\delta}\beta)^{2\Delta/\delta}) - 2\Delta b} \text{ where} \\ \Delta a &= (\Delta/\delta)(1 - {}_{\delta}\beta)^2 + 2(\Delta/\delta)(\Delta/\delta - 1) \frac{(1 - {}_{\delta}\alpha - {}_{\delta}\beta)^2 (1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2 + {}_{\delta}\alpha^2)}{({}_{\delta}\kappa - 1)(1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2)} + \\ &+ 4 \frac{\left((\Delta/\delta)(1 - ({}_{\delta}\alpha + {}_{\delta}\beta)) - (1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^{\Delta/\delta}) \right) {}_{\delta}\alpha (1 - {}_{\delta}\beta ({}_{\delta}\alpha + {}_{\delta}\beta))}{1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2} \\ \Delta b &= ({}_{\delta}\alpha - {}_{\delta}\alpha {}_{\delta}\beta ({}_{\delta}\alpha + {}_{\delta}\beta)) \frac{1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^{2\Delta/\delta}}{1 - ({}_{\delta}\alpha + {}_{\delta}\beta)^2}\end{aligned}$$

The above formulae give the low frequency parameters in terms of the high frequency parameters. 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:

$$\begin{aligned}{}_{\delta}\omega &= \Delta\omega \frac{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}}{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)}; \\ {}_{\delta}\alpha &= ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta} - {}_{\delta}\beta;\end{aligned}$$

⁹ Note that we annualise the GARCH processes (dividing ω by the step length) but Drost and Nijman do not.

$$\begin{aligned}
 {}_{\delta}x &= 1 + \frac{{}_{\Delta}x - 3 + 2\delta/\Delta}{\left(\frac{\delta}{\Delta}\right) + 6 \frac{\left(\left(1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}\right)/(\delta/\Delta) - (1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta))\right) {}_{\delta}\alpha (1 - {}_{\delta}\beta ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta})}{\left(\left(1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}\right)/(\delta/\Delta)\right)^2} \frac{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{2\delta/\Delta} + {}_{\delta}\alpha^2}{\left(1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{2\delta/\Delta} + {}_{\delta}\alpha^2\right)}} \\
 &= \left(2\left(\frac{{}_{\Delta}\beta}{1 + {}_{\Delta}\beta^2}\right) - 1\right) {}_{\delta}\alpha (1 - {}_{\delta}\beta ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}) \frac{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^2}{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{2\delta/\Delta}} = \\
 &= \left(\left(\frac{{}_{\Delta}\beta}{1 + {}_{\Delta}\beta^2}\right) (1 + ({}_{\Delta}\alpha + {}_{\Delta}\beta)^2) - ({}_{\Delta}\alpha + {}_{\Delta}\beta)\right) \times \\
 &\quad \times \left(\frac{1}{\delta/\Delta} (1 - {}_{\delta}\beta)^2 + \frac{1}{\delta/\Delta} \left(\frac{1}{\delta/\Delta} - 1\right) \left(\frac{2}{{}_{\delta}k - 1}\right) \left(\frac{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}}{1 + ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}}\right) (1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{2\delta/\Delta} + {}_{\delta}\alpha^2)\right) \\
 &\quad + 4 \frac{\left(\left(1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta}\right)/(\delta/\Delta) - (1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta))\right) {}_{\delta}\alpha (1 - {}_{\delta}\beta ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{\delta/\Delta})}{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)^{2\delta/\Delta}}
 \end{aligned}$$

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:

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

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

Proof: The proof of this proposition is contained in the work of Drost and Nijman (1996), albeit serving a different purpose and using different notation. In our notation, they consider the convergence of:

$$\frac{1 - ({}_{\Delta}\alpha + {}_{\Delta}\beta)}{\Delta}$$

and

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

They also show that

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

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. Readers interested in the full proof are referred to Appendix A.

Now consider the first two conditional moments and the conditional skewness and kurtosis:

$$\begin{aligned} {}_{\Delta}\mu_{k\Delta} &= E\left({}_{\Delta}\varepsilon_{(k+1)\Delta} / \Delta \middle| {}_{\Delta}I_{k\Delta} \right) \\ {}_{\Delta}\sigma_{k\Delta}^2 &= E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta {}_{\Delta}\mu_{k\Delta} \right)^2 / \Delta \middle| {}_{\Delta}I_{k\Delta} \right) \\ {}_{\Delta}\tau_{k\Delta} &= E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta {}_{\Delta}\mu_{k\Delta} \right)^3 / \left(\Delta^{3/2} {}_{\Delta}\sigma_{k\Delta}^3 \right) \middle| {}_{\Delta}I_{k\Delta} \right) \\ {}_{\Delta}\eta_{k\Delta} &= E\left(\left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta {}_{\Delta}\mu_{k\Delta} \right)^4 / \left(\Delta^2 {}_{\Delta}\sigma_{k\Delta}^4 \right) \middle| {}_{\Delta}I_{k\Delta} \right) \end{aligned}$$

where ${}_{\Delta}I_{k\Delta}$ is the σ -algebra generated by the vector $({}_{\Delta}\varepsilon_{k\Delta})$. 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.

We assume that the following limits exist:

$$\begin{aligned} \varepsilon(t) &:= \lim_{\Delta \downarrow 0} {}_{\Delta}\varepsilon_t \quad \text{where} \quad {}_{\Delta}\varepsilon_t := {}_{\Delta}\varepsilon_{k\Delta} \quad \text{for} \quad k\Delta \leq t < (k+1)\Delta \\ V(t) &:= \lim_{\Delta \downarrow 0} {}_{\Delta}b_t \quad \text{where} \quad {}_{\Delta}b_t := {}_{\Delta}b_{k\Delta} \quad \text{for} \quad k\Delta \leq t < (k+1)\Delta \\ \mu(t) &:= \mu + \lim_{\Delta \downarrow 0} {}_{\Delta}\mu_t \quad \text{where} \quad {}_{\Delta}\mu_t := {}_{\Delta}\mu_{k\Delta} \quad \text{for} \quad k\Delta \leq t < (k+1)\Delta \\ \tau(t) &:= \lim_{\Delta \downarrow 0} {}_{\Delta}\tau_t \quad \text{where} \quad {}_{\Delta}\tau_t := {}_{\Delta}\tau_{k\Delta} \quad \text{for} \quad k\Delta \leq t < (k+1)\Delta \\ \eta(t) &:= \lim_{\Delta \downarrow 0} {}_{\Delta}\eta_t \quad \text{where} \quad {}_{\Delta}\eta_t := {}_{\Delta}\eta_{k\Delta} \quad \text{for} \quad k\Delta \leq t < (k+1)\Delta \end{aligned}$$

It can be seen that:

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

and at least one of the processes $_{\Delta}\mu_{k\Delta}$ and $_{\Delta}\sigma_{k\Delta}^2 + \Delta_{\Delta}\mu_{k\Delta}^2 - _{\Delta}b_{k\Delta}$ has to be different from zero, otherwise the process will be a semi-strong GARCH which is not aggregating 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 converges at rate $\sqrt{\Delta}$, i.e.

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

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.

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

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

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

where

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

and τ and η are the instantaneous skewness and kurtosis of returns and dB_1 and dB_3 are independent Brownian motions.

Proof: 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:

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

and¹⁰

$$\begin{aligned}
 & E \left(\Delta^{-1} \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right)^2 \middle| I_{k\Delta} \right) = E \left(\Delta^{-1} \left(\Delta\mu + \varepsilon_{(k+1)\Delta} \right)^2 \middle| I_{k\Delta} \right) = \\
 & = E \left(\Delta\mu^2 + \Delta^{-1} \varepsilon_{(k+1)\Delta}^2 + 2\mu\varepsilon_{(k+1)\Delta} \middle| I_{k\Delta} \right) = E \left(\Delta^{-1} \varepsilon_{(k+1)\Delta}^2 \middle| I_{k\Delta} \right) + o(1) = \\
 & = \sigma_{k\Delta}^2 + \Delta\mu_{k\Delta}^2 + o(1) = b_{k\Delta} + (\sigma_{k\Delta}^2 - b_{k\Delta}) + o(1) = b_{k\Delta} + o(1)
 \end{aligned}$$

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:

$$\begin{aligned}
 & E \left(\Delta^{-1} (b_{(k+1)\Delta} - b_{k\Delta}) \middle| I_{k\Delta} \right) = \\
 & = \frac{\Delta\omega}{\Delta} - \frac{(1 - \Delta\alpha - \Delta\beta)}{\Delta} b_{k\Delta} + \frac{\Delta\alpha}{\sqrt{\Delta}} (\sigma_{k\Delta}^2 - b_{k\Delta}) / \sqrt{\Delta} + \frac{\Delta\alpha}{\Delta} \Delta\mu_{k\Delta}^2 = \\
 & = \frac{\Delta\omega}{\Delta} - \frac{(1 - \Delta\alpha - \Delta\beta)}{\Delta} b_{k\Delta} + \frac{\Delta\alpha}{\sqrt{\Delta}} (\sigma_{k\Delta}^2 - b_{k\Delta}) / \sqrt{\Delta} + o(1)
 \end{aligned}$$

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

$$\begin{aligned}
 & E \left(\Delta^{-1} (b_{(k+1)\Delta} - b_{k\Delta})^2 \middle| I_{k\Delta} \right) = \\
 & = E \left(\Delta^{-1} \left(\Delta\omega + \Delta\alpha(\varepsilon_{(k+1)\Delta}^2 / \Delta) + (\Delta\beta - 1)b_{k\Delta} \right)^2 \middle| I_{k\Delta} \right) = \\
 & = E \left(\Delta^{-1} \Delta\omega^2 + \Delta^{-1} \Delta\alpha^2 (\varepsilon_{(k+1)\Delta}^2 / \Delta)^2 + \Delta^{-1} (\Delta\beta - 1)^2 b_{k\Delta}^2 + 2\Delta^{-1} \Delta\omega \Delta\alpha (\varepsilon_{(k+1)\Delta}^2 / \Delta) + \right. \\
 & \quad \left. + 2\Delta^{-1} \Delta\omega (\Delta\beta - 1) b_{k\Delta} + 2\Delta^{-1} \Delta\alpha (\Delta\beta - 1) (\varepsilon_{(k+1)\Delta}^2 / \Delta) b_{k\Delta} \middle| I_{k\Delta} \right) = \\
 & = E \left(\Delta^{-1} \Delta\alpha^2 (\varepsilon_{(k+1)\Delta}^2 / \Delta)^2 + \Delta^{-1} (\Delta\beta - 1)^2 b_{k\Delta}^2 + 2\Delta^{-1} \Delta\alpha (\Delta\beta - 1) b_{k\Delta}^2 + \right. \\
 & \quad \left. + 2\Delta^{-1} \Delta\alpha (\Delta\beta - 1) (\sigma_{k\Delta}^2 - b_{k\Delta} + \Delta\mu_{k\Delta}^2) b_{k\Delta} \middle| I_{k\Delta} \right) + o(1) = \\
 & = \Delta^{-1} \Delta\alpha^2 \left(E \left(\varepsilon_{(k+1)\Delta}^4 / \Delta^2 \middle| I_{k\Delta} \right) \right) - E \left(\Delta^{-1} (\Delta\beta - 1) (1 - \Delta\alpha - \Delta\beta - \Delta\alpha) b_{k\Delta}^2 \middle| I_{k\Delta} \right) + o(1) = \\
 & = \Delta^{-1} \Delta\alpha^2 \left(E \left(\varepsilon_{(k+1)\Delta}^4 / \Delta^2 \middle| I_{k\Delta} \right) \right) - E \left(\Delta^{-1} \Delta\alpha^2 b_{k\Delta}^2 \middle| I_{k\Delta} \right) + o(1) = \\
 & = \Delta^{-1} \Delta\alpha^2 \left(E \left((\Delta\sigma_{k\Delta}^4) (\varepsilon_{(k+1)\Delta}^4 / (\Delta^2 \sigma_{k\Delta}^4)) - b_{k\Delta}^2 \middle| I_{k\Delta} \right) \right) + o(1)
 \end{aligned}$$

¹⁰ $o(1)$ denotes a process that converges to zero when $\Delta \downarrow 0$ (o is called the Landau symbol).

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

$$\begin{aligned}
 & E \left(\Delta^{-1} \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right) \left(b_{(k+1)\Delta} - b_{k\Delta} \right) \middle| I_{k\Delta} \right) = \\
 & = E \left(\Delta^{-1} \left(\Delta \mu + \varepsilon_{(k+1)\Delta} \right) \left(\Delta \omega + \Delta \alpha \varepsilon_{(k+1)\Delta}^2 / \Delta + (\Delta \beta - 1) b_{k\Delta} \right) \middle| I_{k\Delta} \right) = \\
 & = E \left(\left(\Delta \alpha \left(\varepsilon_{(k+1)\Delta}^3 / \Delta^2 \right) + (\Delta \beta - 1) b_{k\Delta} \left(\Delta \varepsilon_{(k+1)\Delta} / \Delta \right) \right) \middle| I_{k\Delta} \right) + o(1) = \\
 & = E \left(\left(\Delta \alpha / \sqrt{\Delta} \right) \left(\Delta \sigma_{k\Delta}^3 \right) \left(\varepsilon_{(k+1)\Delta}^3 / \left(\Delta^{3/2} \Delta \sigma_{k\Delta}^3 \right) \right) \middle| I_{k\Delta} \right) + o(1)
 \end{aligned}$$

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

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

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:

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

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:

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

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:

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

which leads to:

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

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

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

Corollary: *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:*

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

where

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

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). 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.

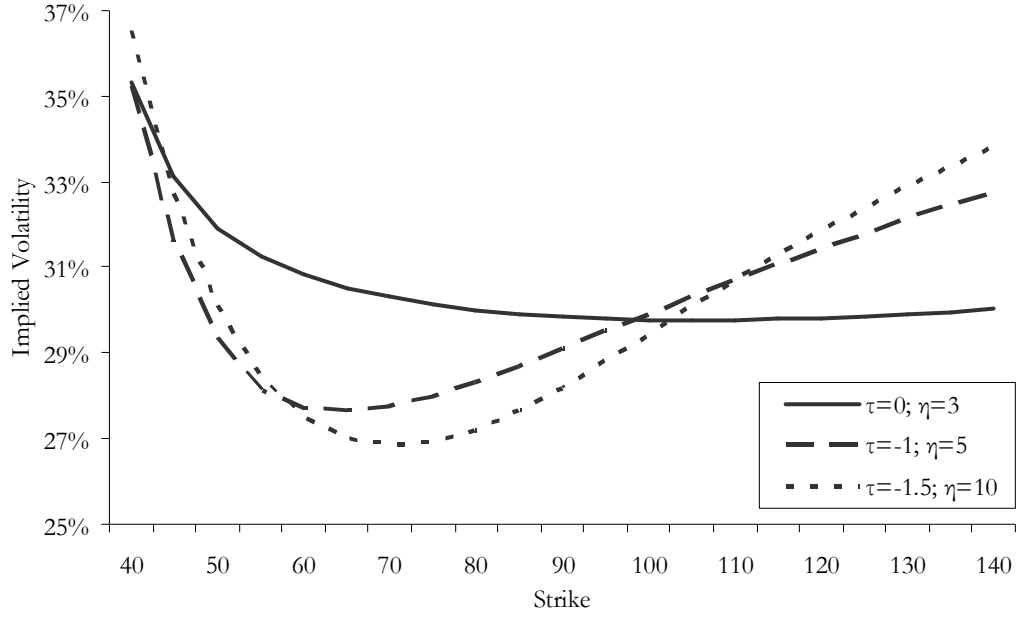


Figure 1: Volatility smiles generated by the continuous limit of weak GARCH, assuming
(a) $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $\tau(t) = 0$; $\eta(t) = 3$
(b) $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $\tau(t) = -1$; $\eta(t) = 5$
(c) $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $\tau(t) = -1.5$; $\eta(t) = 10$
 $S_0 = 100$; $V_0 = 0.09$; $T - t = 1$; $r = 0\%$; 100 steps and 100,000 runs were used for the simulations

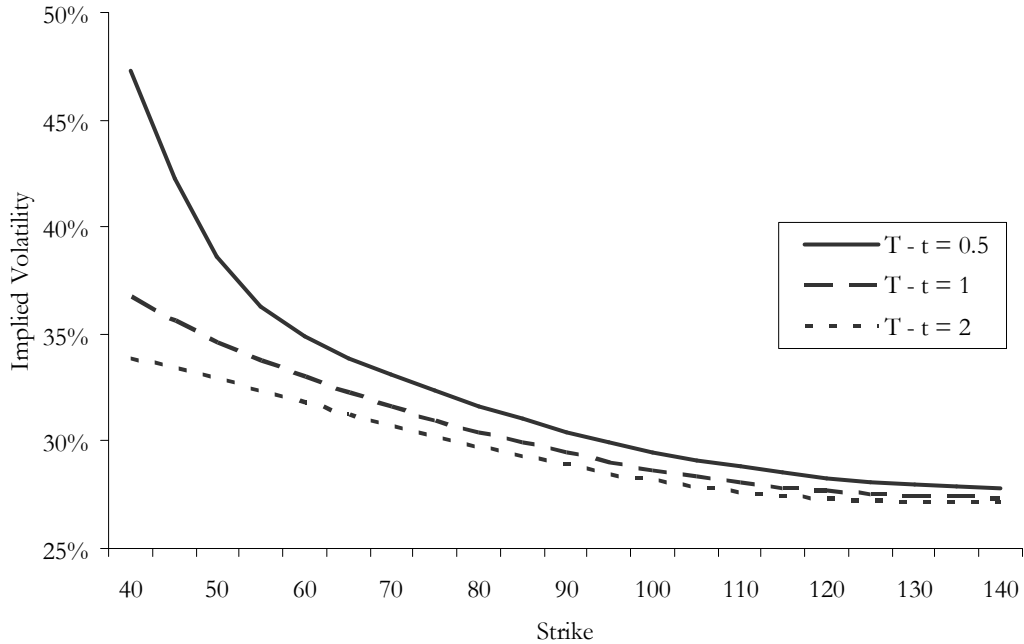


Figure 2: Volatility smiles generated by the continuous limit of weak GARCH, assuming
 $\theta = 0.05$; $\omega = 0.0045$; $\alpha = 0.1$; $\mu(t) = 0$; $e(t) = 0$; $\tau(t) = -1 + t/2$; $\eta(t) = 7 - 2t$;
 $S_0 = 100$; $V_0 = 0.09$; $r = 0\%$; 100 steps and 100,000 runs were used for the simulations
(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 2, for instance, depicts model implied volatility curves when the instantaneous skewness and kurtosis functions are decreasing linearly with time in absolute value.

III Conclusions

To examine the continuous time limit of a discrete time process the process must exist and be the same for any time step. That is, the model must be aggregating in time. This necessitates the use of the weak definition of GARCH. Previous work on the continuous limit of GARCH has examined the strong GARCH model, which is not time aggregating. As a result there was flexibility to choose the rates of convergence of the discrete time GARCH parameters to their continuous limit and different assumptions led to different limit models. By contrast, the weak GARCH model defines the convergence rates for parameters; there is no uncertainty about these and the limit model derived here is unique.

We have shown that the continuous time model corresponding to the weak GARCH process in the physical measure 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. Our limit model can be reduced 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.

Appendix A

We have that:

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

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:

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

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 ${}_{\Delta}\omega = \omega(1 - e^{-\theta\Delta})/\theta$. 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:

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

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

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

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

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

This can be further expressed as:

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

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

As a consequence, κ 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

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

Also, we have that:

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

Multiplying this by δ^{1-w-y} (unless $w = 1/2$) and computing the limit as δ tends to zero leads to:

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

The LHS is finite; if $w > 1/2$ then the RHS is infinite, which is a contradiction, so $w \leq 1/2$. But on the other hand, $\lim_{\delta \downarrow 0} ({}_{\delta}\alpha^2 / \delta) < \theta$, which implies $w \geq 1/2$. So, the solution is $w = 1/2$ and this sets the convergence of α to be with the square root of the time-step. \square

References

- Alexander, C and L. Nogueira (2005) 'A Taxonomy of Options Pricing Models: Scale Invariance and Optimal Hedging' *ICMA Discussion Papers in Finance* 2005-10, available from www.icmacentre.rdg.ac.uk/dp
- Bakshi, G., Kapadia, N. and Madan, D. (2003): 'Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options', *Review of Financial Studies*, Vol. 16 (4), 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
- 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. (2005): 'Approximating GARCH-Jump Models, Jump-Diffusion Processes, and Option Pricing', *Mathematical Finance* (forthcoming)
- Engle, R.F. (1982): 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, Vol. 50, No. 4, 987-1007
- Jeantheau, T. (2004): 'A Link Between Complete Models with Stochastic Volatility and ARCH Models', *Finance and Stochastics* Vol. 8, 111-131
- 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', *Discussion Paper Department of Mathematics & Statistics*, York University Toronto
- 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
- Meddahi, N. and Renault, E. (2004): 'Temporal Aggregation of Volatility Models', *Journal of Econometrics* Vol. 119, 355-379
- 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
- Taylor, S.J. (1986): *Modelling Financial Time Series*, John Wiley and Sons, Chichester, UK
- Wang, Y. (2002) Asymptotic Nonequivalence of GARCH Models and Diffusions. *The Annals of Statistics* Vol. 30, 754-783