The QGARCH(1,1) model

The QGARCH(1,1) model is a well known discrete-time model defined as

$$\begin{aligned} R_t &= \sqrt{V_t}\, Z_t \\ V_t &= \omega + \alpha R_{t-1}^2 + \beta V_{t-1} + \gamma R_{t-1} \end{aligned} \tag{1}$$

for $t = 1, 2...$ (e.g. days) where $R_t = (S_t - S_{t-1})/S_{t-1}$ is the t'th **stock price return** (note $R_t \geq -1$ since $S_t \geq 0$) and $\omega, \alpha, \beta > 0$, and Z_t is a sequence of i.i.d random variables with zero mean and variance σ^2, e.g. $N(0,1)$ or a student t-distribution with ν degrees of freedom if we want fatter tails for which $\sigma^2 = \frac{\nu}{\nu-2}$, so we need $\nu > 2$. Since we can re-write the model as

$$V_t = V_{t-1} + (1-\beta)(\bar{\omega} - V_{t-1}) + \alpha R_{t-1}^2 + \gamma R_{t-1} \tag{2}$$

where $\bar{\omega} = \frac{\omega}{1-\beta}$, we see that $1-\beta$ controls the **mean reversion** speed for V, and $\bar{\omega}$ is level around which V mean reverts. α controls the extent of **volatility clustering**, i.e. past large volatility giving rise to large future volatility and vice versa, and γ is a **skew term** which captures that squared volatility V_t tends to increase if $R_{t-1} < 0$ since usually $\gamma < 0$ as well so $\gamma R_{t-1} > 0$ (the so-called **leverage** effect). $\gamma < 0$ also allows the model to produce negatively skewed non-symmetric implied volatility smiles for European options which are seen in practice, particularly for Index and Equity options. The original Engle&Bollerslev **GARCH** model from 1986 has $\gamma = 0$, so the model above is sometimes known as the **asymmetric GARCH** model.

If we now instead say that V_{t+1} is V_t, then we can re-write the model in the **Euler-scheme** type form

$$\begin{aligned} S_t &= S_{t-1} + S_{t-1}\sqrt{V_{t-1}}\, Z_t \\ V_t &= V_{t-1} + (1-\beta)(\bar{\omega} - V_{t-1}) + \alpha R_t^2 + \gamma R_t \\ &= V_{t-1} + (1-\beta)(\bar{\omega} - V_{t-1}) + \alpha V_{t-1} Z_t^2 + \gamma\sqrt{V_{t-1}}\, Z_t \end{aligned} \tag{3}$$

then we see that (S_t, V_t) is **discrete-time Markov process**, since the distribution of S_t, V_t at time $t-1$ depends only on (S_{t-1}, V_{t-1}) and does not require any further history of these two processes (note our original V_t is now V_{t-1} here).

Taking expectations in (1), we see that

$$\mathbb{E}(V_t) = \omega + \alpha\mathbb{E}(R_{t-1}^2) + \beta\mathbb{E}(V_{t-1}) + \gamma\mathbb{E}(R_{t-1}).$$

Using the **tower property** of conditional expectations, we can further re-write this as

$$\begin{aligned} \mathbb{E}(V_t) &= \omega + \alpha\mathbb{E}(\mathbb{E}(R_{t-1}^2)|V_{t-1}) + \beta\mathbb{E}(V_{t-1}) + \gamma\mathbb{E}(\mathbb{E}(R_{t-1}|V_{t-1})) \\ &= \omega + \alpha\mathbb{E}(\sigma^2 V_{t-1}) + \beta\mathbb{E}(V_{t-1}) + 0 \end{aligned} \tag{4}$$

where we have also used that $\mathbb{E}(R_{t-1}^2|V_{t-1}) = \mathbb{E}(V_{t-1}Z_{t-1}^2|V_{t-1}) = V_{t-1}\mathbb{E}(Z_{t-1}^2|V_{t-1}) = V_{t-1}\mathbb{E}(Z_{t-1}^2) = V_{t-1}\sigma^2$. For V_t to have a **stationary distribution**, i.e. for V_t to have the same distribution for all t, this clearly requires that $\mathbb{E}(V_t) = \mathbb{E}(V_{t-1})$, so we can further re-write (4) as

$$\mathbb{E}(V_t) = \omega + \alpha\sigma^2\mathbb{E}(V_t) + \beta\mathbb{E}(V_t).$$

and

$$\mathbb{E}(R_t^2) = \mathbb{E}(\mathbb{E}(R_t^2|V_t)) = \mathbb{E}(V_t).$$

Re-arranging, we see that

$$\mathbb{E}(V_t) = \frac{\omega}{1 - \alpha\sigma^2 - \beta}.$$

Since V_t cannot be negative, we must have that $\alpha\sigma^2 + \beta < 1$, which we call the **stationarity condition**. If V starts at time zero, then

$$\begin{aligned} \mathbb{E}(V_t) &= \frac{1}{1-\alpha\sigma^2}(\omega + \beta\mathbb{E}(V_{t-1})) \\ \Rightarrow \quad \mathbb{E}(V_t) - \bar{V} &= \frac{1}{1-\alpha\sigma^2}(\omega + \beta\mathbb{E}(V_{t-1})) - \bar{V} = \frac{\beta}{1-\alpha\sigma^2}(\mathbb{E}(V_{t-1}) - \bar{V}) \end{aligned}$$

i.e. a **linear recurrence** relation of the form $r_t = ar_{t-1}$, with solution $r_t = \mathbb{E}(V_t) - \bar{V} = (\frac{\beta}{1-\alpha\sigma^2})^t(V_0 - \bar{V})$.

Moreover

$$V_t = \omega + \alpha R_{t-1}^2 + \beta V_{t-1} + \gamma R_{t-1} \geq \omega + \alpha R_{t-1}^2 + \gamma R_{t-1}$$

and (using basic calculus) the right-hand side is ≥ 0 for all R_{t-1} if $\omega \geq \frac{\gamma^2}{4\alpha}$. This is known as the **positivity condition**.

Let

$$\mathbb{E}(R_t^4) \;=\; \mathbb{E}(\mathbb{E}(R_t^4|\mathcal{F}_{t-1}) \;=\; \mathbb{E}(V_t^2\mathbb{E}(Z_t^4|\mathcal{F}_{t-1})) \;=\; \mathbb{E}(Z_t^4)\mathbb{E}(V_t^2)\,. \tag{5}$$

For $\gamma = 0$ and $\sigma = 1$, we have

$$\begin{aligned}\mathbb{E}(V_t^2) &= (3+K_\varepsilon)\mathbb{E}(V_t^2)\alpha^2 + 2\mathbb{E}(R_{t-1}^2V_{t-1})\alpha\beta + \mathbb{E}(V_t^2)\beta^2 + 2\mathbb{E}(V_t)\alpha\omega + 2\mathbb{E}(V_t)\beta\omega + \omega^2\\ &= (...) + 2\alpha\beta\mathbb{E}(V_{t-1}\mathbb{E}_{t-2}(R_{t-1}^2))\\ &= (...) + 2\alpha\beta\mathbb{E}(V_t^2)\end{aligned}$$

Re-arranging the final expression, we see that

$$\mathbb{E}(V_t^2) \;=\; \frac{\omega(2\mathbb{E}(V_t)(\beta+\alpha)+\omega}{1-((3+K_\varepsilon)\alpha^2+\beta^2+2\alpha\beta)}\,.$$

if the denominator is positive.

Maximum likelihood estimates for the GARCH parameters and asymptotic normality

The joint density of $R_1, ..., R_n$ can be easily expressed as a product of conditional densities of the returns:

$$L \;=\; p(R_1)\,p(R_2|R_1)\,p(R_3|R_1,R_2)... \;=\; p(R_1)\,p(R_2|V_2)...\,p(R_n|V_n) \;=\; \prod_{j=1}^{n} f(\frac{R_j}{\sqrt{V_j}})\frac{1}{\sqrt{V_j}} \;=\; p(R_1)\,p(R_2|V_2)...\,p(R_n|V_n)$$

where f is the density of each Z_t in (1). This is true because

$$\mathbb{P}(R_j \leq x|V_j) \;=\; \mathbb{P}(Z_j \leq \frac{x}{\sqrt{V_j}}|V_j) \;=\; F(\frac{x}{\sqrt{V_j}})$$

where F is the distribution function of Z_t. Using *observed* values for $R_1, ..., R_n$, and given parameter values for the model, the values of $Z_j = \frac{R_j}{\sqrt{V_j}}$ are known as the **residuals** and L is the likelihood function of $R_1, ..., R_n$. We can then maximize L over all admissible parameter combinations to compute MLEs for the model parameters $\omega, \alpha, \beta, \gamma$, and the parameter(s) for the distribution of each Z_t (this is conceptually similar to Part 2). Then LL is

$$\ell(\theta)_n \;=\; \sum_{i=1}^{n} \log f(\frac{R_j}{\sqrt{V_j}}) - \frac{1}{2}\log V_j$$

and recall that V_j actually depends on $R_1, ..., R_{j-1}$ and the model parameters which we collectively denote by θ. Then the Fisher information matrix when the residuals are i.i.d. $N(0,1)$ is

$$\begin{aligned}I(\theta) \;=\; -\mathbb{E}(\frac{\partial^2}{\partial\theta^2}\ell(\theta)_n)^2) &= \sum_{j=1}^{n}\mathbb{E}(-\frac{-2R_j^2+V_j(\theta)}{2V_j(\theta)^3}\frac{\partial V_j(\theta)}{\partial\theta_i}\frac{\partial V_j(\theta)}{\partial\theta_j} + (R_j^2 - V_j(\theta))V_j(\theta)\frac{\partial^2 V_j(\theta)}{\partial\theta_i\partial\theta_j})\\ &= \sum_{j=1}^{n}\mathbb{E}(\frac{1}{2V_j(\theta)^2}\frac{\partial V_j(\theta)}{\partial\theta_i}\frac{\partial V_j(\theta)}{\partial\theta_j})\\ &= n\mathbb{E}(\frac{1}{2V_j(\theta)^2}\frac{\partial V_j(\theta)}{\partial\theta_i}\frac{\partial V_j(\theta)}{\partial\theta_j})\end{aligned} \tag{6}$$

using the stationarity of V, where we have also used the tower property in the final line. Hence $\sqrt{n}(\hat{\theta}-\theta)$ tends to a multivariate $N(0, I(\theta)^{-1})$ random variable as $n \to \infty$.

Goodness-of-fit tests for the residuals

If e.g. we assume $Z_t \sim N(0,1)$, we can then perform standard normality tests like **Kolmogorov Smirnov**, **Shapiro-Wilk**, **Jarque-Bera** or **Andersen-Darling** to test whether the Z_t values are indeed i.i.d. Normals. Otherwise, if we use a different distribution for Z_t (e.g. a ***t*-distribution** with ν degrees of freedom which will give the returns fatter tails), we have to transform these back Z values to Normal RVs before applying these normality tests, using inverse cdfs.

Estimating V_0 from the stock price history

If we assume $\gamma = 0$ for simplicity, then iterating the definition of V_t we see that

$$\begin{aligned} V_t &= \omega + \beta V_{t-1} + \alpha R_{t-1}^2 \\ &= \omega + \beta(\omega + \beta V_{t-2} + \alpha R_{t-2}^2) + \alpha R_{t-1}^2 \\ &= \omega + \beta(\omega + \beta(\omega + \beta V_{t-3} + \alpha R_{t-3}^2) + \alpha R_{t-2}^2) + \alpha R_{t-1}^2 \\ &= \omega(1+\beta+\beta^2+...) + \frac{\alpha}{\beta}\sum_{\tau=1}\beta^{\tau} R_{t-\tau}^2 \;=\; \bar{\omega} + \frac{\alpha}{\beta}\sum_{\tau=1}^{\infty} e^{-b\tau} R_{t-\tau}^2 \end{aligned} \tag{7}$$

where b is defined by $\beta = e^{-b}$ and $\bar{\omega}$ is defined above, and note the first term on the right-hand side is the mean reversion level from above. So we see that the effect of past returns on volatility decays exponentially, and re-doing this computation with $\gamma \neq 0$, we find that the last line just changes to

$$V_t = \frac{\omega}{1-\beta} + \frac{\alpha}{\beta}\sum_{\tau=1}^{\infty} e^{-b\tau} R_{t-\tau}^2 + \frac{\gamma}{\beta}\sum_{\tau=1}^{\infty} e^{-b\tau} R_{t-\tau}\,.$$

In particular, we also see that

$$V_0 = \bar{\omega} + \frac{\alpha}{\beta}\sum_{\tau=1}^{\infty} e^{-b\tau} R_{-\tau}^2 + \frac{\gamma}{\beta}\sum_{\tau=1}^{\infty} e^{-b\tau} R_{-\tau}$$

so we can estimate V_0 by truncating this sum in practice rather than fitting V_0 as an additional free parameter for the MLE maximization computation described above, since V_0 is already fixed by the history of the returns.

Stochastic volatility as the diffusive limit of QGARCH

Consider the following variant of the model above:

$$\begin{aligned} S_t &= S_{t-\Delta t} + S_{t-\Delta t}\sqrt{V_{t-\Delta t}}\, Z_t \\ V_t &= V_{t-\Delta t} + \kappa\theta\Delta t + \frac{\eta}{\sqrt{\Delta t}}(R_t^2 - V_{t-\Delta t}\Delta t) - \kappa V_{t-\Delta t}\Delta t + \gamma R_t \\ &= V_{t-\Delta t} + \kappa(\theta - V_{t-\Delta t})\Delta t + \frac{\eta}{\sqrt{\Delta t}} V_{t-\Delta t}(Z_t^2 - \Delta t) + \gamma\sqrt{V_{t-\Delta t}} Z_t \\ &= V_{t-\Delta t} + \bar{\kappa}(\bar{\theta} - V_{t-\Delta t})\Delta t + \frac{\eta}{\sqrt{\Delta t}} R_t^2 + \gamma R_t \end{aligned}$$

for some $\bar{\kappa}$, $\bar{\theta}$, with $Z_1, Z_2, ...$ i.i.d. as above and V_{t-1} here is our old V_t, and now assume $\mathrm{Var}(Z_t) = \Delta t$ and $\eta = O(1)$, and impose that $\nu > 4$ so $\mathbb{E}(Z_i^4) < \infty$, and from the final line we see that V_t is still of the QGARCH(1,1) form in (3). Then as $\Delta t \to 0$, the model tends to the mean-reverting **Markov stochastic volatility** model:

$$\begin{aligned} dS_t &= S_t\sqrt{V_t}dW_t \\ dV_t &= \kappa(\theta - V_t)dt + 2\eta V_t dB_t + \gamma\sqrt{V_t}dW_t \end{aligned} \tag{8}$$

where W and B are standard independent Brownian motions, so we see that the specific form of the distribution of the Z_t's does not show up in the $\Delta t \to 0$ limit and the independent Brownian motion B appears almost by magic. When η is larger, the implied volatility smile will be more U-shaped as a function of strike K, and will be symmetric as a function of $x = \log\frac{K}{S_0}$ if $\gamma = 0$. If ν is smaller, the smile may just be monotonically decreasing as a function of K over relevant strike ranges.

The limiting model in (7) is hybrid of the well known **Hull-White** and **Heston** models (the well known Heston model has a $\sqrt{V_t}$ term in it). To see why this is true, we first note that

$$\frac{1}{\sqrt{\Delta t}}\sum_{i=1}^{[nt]}(Z_i^2 - \Delta t) = \sqrt{n}\sum_{i=1}^{[nt]}(\Delta t\tilde{Z}_i^2 - \Delta t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(\tilde{Z}_i^2 - 1) \tag{9}$$

where $\tilde{Z}_i = Z_t/\sqrt{\Delta t} \sim N(0,1)$, and that $\mathrm{Var}(\tilde{Z}_i^2 - 1) = \mathbb{E}((\tilde{Z}_i^2-1)^2) = 3 - 2 + 1 = 2$.

We now recall **Donsker's theorem**. Let X_i be a sequence of i.i.d. random variables with $\mathbb{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = 1$, and let $S_n = \sum_{i=1}^n X_i$. Now consider the **random function**:

$$W_t^n = \frac{S_{[nt]}}{\sqrt{n}} \qquad (t \in [0,1])$$

where $[nt]$ denotes the largest integer less than or equal to nt. Then by the **Central Limit Theorem**, $W_1^n = \frac{S_n}{\sqrt{n}}$ tends to an $N(0,1)$ random variable as $n \to \infty$. More precisely, $\lim_{n\to\infty} \mathbb{E}(F(W_1^n)) = \mathbb{E}(F(Z))$ for any bounded continuous function F (this is known as **weak convergence**). Donsker's theorem, states that the random function W_t^n tends weakly to a random function which is a Brownian motion as $n \to \infty$. This shows that we can numerically approximate Brownian motion using X_i's with any distribution with finite variance. Thus (8) falls exactly under the framework of Donsker's theorem, aside from $\tilde{Z}_i^2 - 1$ having a variance of 2 not 1, which is why there is a **factor of** 2 in (7).

Changing from $\mathbb{P}$ to $\mathbb{Q}$ measure

If the Z_t's have a non-zero density under $\mathbb{P}$, then the Z_t's can have any non-zero density under $\mathbb{Q}$ (does not have to be equal to the original density), so long as $\mathbb{E}^{\mathbb{Q}}(Z_t) = 0$, then S will still be a martingale under $\mathbb{Q}$, which is equivalent to $\mathbb{P}$ since both densities are non-zero by assumption.

Intraday dynamics consistent with the QGARCH model

The t-distribution is infinitely divisible which means a random variable Z with this distribution can be written as a sum of n i.i.d random variables Z_i^n, for any n. The characteristic function $\mathbb{E}(e^{iuZ_i^n})$ of Z_i^n is then $\phi(u)^{1/n}$ where $\phi(u) = \mathbb{E}(e^{iuZ})$. This gives us a way to extend the model from modelling daily returns to intraday returns with n i.i.d residuals per day, keeping V constant within any given day.

Bayesian analysis

If we set $X = (R_1, ..., R_n)$ and $\theta = (\alpha, \beta, \gamma, \nu)$, then from Bayes formula, we know that

$$p(\theta|X) \quad = \quad \frac{p(X|\theta)\, p(\theta)}{p(X)}$$

where the p's refer to densities or conditional densities here. $p(X)$ does not depend on θ, and if assume a uniform prior $p(\theta) = const.$ for θ on some finite hypercube in $\mathbb{R}^4$ (and zero elsewhere), then

$$p(\theta|X) \quad = \quad const. \times p(X|\theta)$$

so the conditional density of θ given X is proportional to the likelihood function $p(X|\theta)$, and by integrating in the other 3 parameters we can compute e.g. the marginal density of α, β, γ or ν given X. This is easier if e.g. we fix $\gamma = 0$ and fix $1 - \beta$ to its lower bound, so we only have two free parameters.

Power kernel model

We can modify the model as follows:

$$\begin{aligned} R_t &= \sqrt{V_t}\, Z_t \\ V_t &= \omega + c\sum_{\tau=1}^{\infty} \tau^{-\alpha} R_{t-\tau}^2 + \gamma \sum_{\tau=1}^{\infty} \tau^{-\alpha_2} R_{t-\tau} \end{aligned}$$

for $\alpha, \alpha_2 > 2$ (add mean reversion?) which corresponds to **power decay**, and again we have to take care to ensure positivity and stationarity. In this case, using the same tower law argument as above

$$\mathbb{E}(V_t) \;=\; \omega + c\sum_{\tau=1}^{\infty} \tau^{-\alpha}\mathbb{E}(R_{t-\tau}^2) \;=\; \omega + c\sum_{\tau=1}^{\infty} \tau^{-\alpha}\mathbb{E}(\mathbb{E}(R_{t-\tau}^2|V_{t-\tau}) \;=\; \omega + c\sigma^2 \sum_{\tau=1}^{\infty} \tau^{-\alpha}\mathbb{E}(V_{t-\tau})\,.$$

If V is stationary, then

$$\mathbb{E}(V_t) \;=\; \omega + c\sigma^2 \sum_{\tau=1}^{\infty} \tau^{-\alpha}\mathbb{E}(V_t) \;=\; \omega + c\sigma^2 \mathbb{E}(V_t)\zeta(\alpha)$$

which we can re-arrange as $\mathbb{E}(V_t) = \frac{\omega}{1 - c\sigma^2\zeta(\alpha)}$, where $\zeta(\alpha) = \sum_{n=1}^{\infty} n^{-\alpha}$ denotes the **zeta function**, so clearly a necessary condition for stationarity is that $c\sigma^2\zeta(\alpha) < 1$.

If $\alpha = \alpha_2$, then can re-write as

$$V_t \;=\; \sum_{\tau=1}^{\infty} \tau^{-\alpha}(\bar{\omega} + cR_{t-\tau}^2 + \gamma R_{t-\tau})$$

where $\bar{\omega} = \frac{\omega}{\zeta(a)}$, so we have essentially the same **positivity condition** as before $\bar{\omega} \geq \frac{\gamma^2}{4c}$. This is a discrete-time version of the **rough Heston model**.

Quadratic Rough Heston-type model

We can also generalize to a quadratic rough Heston-type model:

$$V_t \;=\; \omega + c\sum_{\tau=1}^{\infty}\tau^{-\alpha}R_{t-\tau}^2 + b(\sum_{\tau=1}^{\infty}\tau^{-\alpha}R_{t-\tau}-a)^2 + \gamma\sum_{\tau=1}^{\infty}\tau^{-\alpha}R_{t-\tau}\,.$$

Then again assuming stationarity, we now see that

$$\begin{aligned}\mathbb{E}(V_t) &= \omega + c\sigma^2\zeta(\alpha)\mathbb{E}(V_t) + b\,\mathbb{E}(\sum_{\tau=1}^{\infty}\tau^{-\alpha}R_{t-\tau}-a)^2 \\ &= \omega + c\sigma^2\zeta(\alpha)\mathbb{E}(V_t) + b\mathbb{E}(\sum_{\tau=1}^{\infty}\tau^{-\alpha}R_{t-\tau})^2+a^2) \\ &= \omega + c\sigma^2\zeta(\alpha)\mathbb{E}(V_t) + b(\zeta(2\alpha)\mathbb{E}(V_t)+a^2)\end{aligned}$$

using that $\mathbb{E}(R_iR_j) = \mathbb{E}(R_i\mathbb{E}(R_j|R_i,V_j)) = 0$ for $i<j$, so the stationarity condition now reads as $c\sigma^2\zeta(\alpha) + b(\zeta(2\alpha) < 1$.

Numerical results

Below we compute MLEs and apply the Kolmogorov Smirnov, Shapiro-Wilk and Jarque-Bera normality tests on the (transformed) residuals implied by the MLEs for the model in (1) using daily prices, with a 1yr/3yr/1yr test window (the initial 1yr window is used to compute the V_0 for the middle window from the initial 1yr history of returns; the middle 3yr period is used for in-sample (i/s) testing, and final year used for out-of-sample testing, all three periods are consecutive with no gaps/overlap), ending 11/08/2023. Although the fits are very good, the sample variance of the MLEs using synthetic paths with the fitted parameters are much higher than we would ideally like.

MLEs/p-vals	α	β	γ	ν	KS i/s	SW i/s	JB i/s	KS o/s	SW o/s	JB o/s
EUR/USD	0.0293	0.962	-5.405e-05	8.684	0.835	0.870	0.706	0.912	0.714	0.643
GBP/USD	0.0303	0.932	-0.000252	6.192	0.966	0.836	0.712	0.119	0.224	0.279
USD/JPY	0.0830	0.875	-0.000299	5.9611	0.292	0.476	0.352	0.0603	0.0907	0.229
AMZN	0.03482	0.9420	-0.000505	5.008	0.401	0.811	0.951	0.560	0.607	0.570
BRK-B	0.103	0.868	-0.00103	8.929	0.168	0.921	0.950	0.611	0.676	0.984
INTC	0.0280	0.943	-5.940e-05	3.914	0.375	0.0634	0.0404	0.229	0.262	0.236
AZN	0.0496	0.904	-0.000897	4.153	0.247	0.587	0.428	0.103	0.206	0.195
N225	0.0982	0.856	-0.00129	6.271	0.281	0.443	0.349	0.0713	0.236	0.354
HSI	0.06222	0.898	-0.000834	5.108	0.491	0.226	0.358	0.530	0.121	0.161

To fix SPX historical prices well, we need a skewed t-distribution for the residuals