Ruprecht-Karls-Universität Heidelberg

Fakultät für Mathematik und Informatik

Bachelorarbeit

zur Erlangung des akademischen Grades

Bachelor of Science (B. Sc.)

GARCH(1,1) models

vorgelegt von

Brandon Williams

15. Juli 2011

Betreuung: Prof. Dr. Rainer Dahlhaus

Abstrakt

In dieser Bachelorarbeit werden GARCH(1,1)-Modelle zur Analyse finanzieller Zeitreihen untersucht. Dabei werden zuerst hinreichende und notwendige Bedingungen dafür gegeben, dass solche Prozesse überhaupt stationär werden können. Danach werden asymptotische Ergebnisse über relevante Schätzer hergeleitet und parametrische Tests entwickelt. Die Methoden werden am Ende durch ein Datenbeispiel illustriert.

Abstract

In this thesis, GARCH(1,1)-models for the analysis of financial time series are investigated. First, sufficient and necessary conditions will be given for the process to have a stationary solution. Then, asymptotic results for relevant estimators will be derived and used to develop parametric tests. Finally, the methods will be illustrated with an empirical example.

Contents

1 Introduction

Modelling financial time series is a major application and area of research in probability theory and statistics. One of the challenges particular to this field is the presence of heteroskedastic effects, meaning that the volatility of the considered process is generally not constant. Here the volatility is the square root of the conditional variance of the log return process given its previous values. That is, if P_t is the time series evaluated at time t, one defines the log returns

$$X_t = \log P_{t+1} - \log P_t$$

and the volatility σ_t, where

$$\sigma_t^2 = \mathrm{Var}[X_t^2 \mid \mathcal{F}_{t-1}]$$

and $\mathcal{F}_{t-1}$ is the σ-algebra generated by $X_0, ..., X_{t-1}$. Heuristically, it makes sense that the volatility of such processes should change over time, due to any number of economic and political factors, and this is one of the well known "stylized facts" of mathematical finance.

The presence of heteroskedasticity is ignored in some financial models such as the Black-Scholes model, which is widely used to determine the fair pricing of European-style options. While this leads to an elegent closed-form formula, it makes assumptions about the distribution and stationarity of the underlying process which are unrealistic in general. Another commonly used homoskedastic model is the Ornstein-Uhlenbeck process, which is used in finance to model interest rates and credit markets. This application is known as the Vasicek model and suffers from the homoskedastic assumption as well.

ARCH (autoregressive conditional heteroskedasticity) models were introduced by Robert Engle in a 1982 paper to account for this behavior. Here the conditional variance process is given an autoregressive structure and the log returns are modelled as a white noise multiplied by the volatility:

$$X_t = e_t \sigma_t$$

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + ... + \alpha_p X_{t-p}^2,$$

where e_t (the 'innovations') are i.i.d. with expectation 0 and variance 1 and are assumed independent from σ_k for all $k \leq t$. The lag length $p \geq 0$ is part of the model specification and may be determined using the Box-Pierce or similar tests for autocorrelation significance, where the case $p = 0$ corresponds to a white noise process. To ensure that σ_t^2 remains positive, $\omega, \alpha_i \geq 0 \quad \forall i$ is required.

Tim Bollerslev (1986) extended the ARCH model to allow σ_t^2 to have an additional autoregressive structure within itself. The GARCH(p,q) (generalized ARCH) model is given by

$$X_t = e_t \sigma_t$$

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + ... + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2.$$

This model, in particular the simpler GARCH(1,1) model, has become widely used in financial time series modelling and is implemented in most statistics and econometric software packages. GARCH(1,1) models are favored over other stochastic volatility models by many economists due to their relatively simple implementation: since they are given by stochastic difference equations in discrete time, the likelihood function is easier to handle than continuous-time models, and since financial data is generally gathered at discrete intervals.

However, there are also improvements to be made on the standard GARCH model. A notable problem is the inability to react differently to positive and negative innovations, where in reality, volatility tends to increase more after a large negative shock than an equally large positive shock. This is known as the leverage effect and possible solutions to this problem are discussed further in section 6.

Without loss of generality, the time t will be assumed in the following sections to take values in either $\mathbb{N}_0$ or in $\mathbb{Z}$.

2 Stationarity

The first task is to determine suitable parameter sets for the model. In the introduction, we considered that $\omega, \alpha, \beta \geq 0$ is necessary to ensure that the conditional variance σ_t^2 remains non-negative at all times t. It is also important to find parameters ω, α, β which ensure that σ_t^2 has finite expected value or higher moments. Another consideration which will be important when studying the asymptotic properties of GARCH models is whether σ_t^2 converges to a stationary distribution. Unfortunately, we will see that these conditions translate to rather severe restrictions on the choice of parameters.

Definition 1. : A process X_t is called stationary (strictly stationary), if for all times $t_1, ..., t_n, h \in \mathbb{Z}$:

$$F_X(x_{t_1+h}, ..., x_{t_n+h}) = F_X(x_{t_1}, ...x_{t_n})$$

where $F_X(x_{t_1}, ..., x_{t_n})$ is the joint cumulative distribution function of $X_{t_1}, ..., X_{t_n}$.

Theorem 2. *Let $\omega > 0$ and $\alpha, \beta \geq 0$. Then the GARCH(1,1) equations have a stationary solution if and only if $\mathbb{E}[log(\alpha e_t^2 + \beta)] < 0$. In this case the solution is uniquely given by*

$$\sigma_t^2 = \omega\Big(1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)\Big).$$

Proof. With the equation $\sigma_t^2 = \omega + (\alpha e_{t-1}^2 + \beta)\sigma_{t-1}^2$, by repeated use on σ_{t-1}, etc. we arrive at the equation

$$\sigma_t^2 = \omega(1 + \sum_{j=1}^{k} \prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)) + (\prod_{i=1}^{k+1} (\alpha e_{t-i}^2 + \beta))\sigma_{t-k-1}^2,$$

which is valid for all $k \in \mathbb{N}$. In particular,

$$\sigma_t^2 \geq \omega(1 + \sum_{j=1}^{k} \prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)),$$

since $\alpha, \beta \geq 0$. Assume that σ_t^2 is a stationary solution and that $\mathbb{E}[\log(\alpha e_t^2 + \beta)] \geq 0$. We have

$$\log \mathbb{E}[\prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)] \geq \mathbb{E}[\log \prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)] = \sum_{i=1}^{j} \mathbb{E}[\log(\alpha e_{t-i}^2 + \beta)]$$

and therefore, if $\mathbb{E}[\log(\alpha e_t^2 + \beta)] > 0$, then the product $\prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta)$ diverges a.s. by the strong law of large numbers. In the case that $\mathbb{E}[\log(\alpha e_t^2 + \beta)] = 0$, then $\sum_{i=1}^{j} \log(\alpha e_{t-i}^2 + \beta)$ is a random walk process so that

$$\limsup_{j \to \infty} \sum_{i=1}^{j} \log(\alpha e_{t-i}^2 + \beta) = \infty \text{ a.s.}$$

so that in both cases we have

$$\limsup_{j \to \infty} \prod_{i=1}^{j} (\alpha e_{t-i}^2 + \beta) = \infty \text{ a.s.}$$

Since all terms are negative we then have

$$\sigma_t^2 \geq \limsup_{j\to\infty} \omega \prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta) = \infty \text{ a.s.}$$

which is impossible; therefore, $\mathbb{E}[\log(\alpha e_t^2 + \beta)] < 0$ is necessary for the existence of a stationary solution. On the other hand, let $\mathbb{E}[\alpha e_t^2 + \beta] < 0$. Then there exists a $\xi > 1$ with $\log \xi + \mathbb{E}[\log(\alpha e_t^2 + \beta)] < 0$. For this ξ we have by the strong law of large numbers:

$$\log \xi + \frac{1}{n}\sum_{i=1}^{n} \log(\alpha e_{t-i}^2 + \beta) \xrightarrow{\text{a.s.}} \log \xi + \mathbb{E}[\log(\alpha e_t^2 + \beta)] < 0,$$

so

$$\log(\xi^n \prod_{i=1}^{n}(\alpha e_{t-i}^2 + \beta)) = n(\log \xi + \frac{1}{n}\sum_{i=1}^{n} \log(\alpha e_{t-i}^2 + \beta)) \xrightarrow{\text{a.s.}} -\infty,$$

and

$$\xi^n \prod_{i=1}^{n}(\alpha e_{t-i}^2 + \beta) \xrightarrow{\text{a.s.}} 0.$$

Therefore, the series

$$\omega\Big(1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)\Big)$$

converges a.s. To show uniqueness, assume that σ_t and $\hat{\sigma}_t$ are stationary: then

$$|\sigma_t - \hat{\sigma}_t| = (\alpha e_{t-1}^2 + \beta)|\sigma_{t-1}^2 - \hat{\sigma}_{t-1}^2| = (\xi^n \prod_{i=1}^{n}(\alpha e_{t-i}^2 + \beta))\xi^{-n}|\sigma_{t-n}^2 - \hat{\sigma}_{t-n}^2| \xrightarrow{\mathbb{P}} 0.$$

This means that $\mathbb{P}(|\sigma_t - \hat{\sigma}_t| > \epsilon) = 0 \; \forall \epsilon > 0$, so $\sigma_t = \hat{\sigma}_t$ a.s.

□

Corollary 3. *The GARCH(1,1) equations with $\omega > 0$ and $\alpha, \beta \geq 0$,have a stationary solution with finite expected value if and only if $\alpha + \beta < 1$, and in this case: $\mathbb{E}[\sigma_t^2] = \frac{\omega}{1-\alpha-\beta}$.*

Proof. : Since $\mathbb{E}[\log(\alpha e_t^2 + \beta)] \leq \log(\mathbb{E}[\alpha e_t^2 + \beta]) = \log(\alpha + \beta) < 0$, the conditions of Theorem 1 are fulfilled. We have

$$\begin{aligned}
\mathbb{E}[\sigma_t^2] &= \mathbb{E}[\omega\Big(1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)\Big)] \\
&= \omega(1 + \sum_{j=1}^{\infty}\mathbb{E}[\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)]) \\
&= \omega(1 + \sum_{j=1}^{\infty}(\alpha + \beta)^j) \\
&= \frac{\omega}{(1-\alpha-\beta)}
\end{aligned}$$

if this series converges, that is, if $\alpha + \beta < 1$, and ∞ otherwise. □

Remark 4. This theorem shows that strictly stationary IGARCH(1,1) processes (those where $\alpha + \beta = 1$) exist. For example, if e_t is normally distributed, and $\alpha = 1, \beta = 0$, then

$$\mathbb{E}[\log(\alpha e_t^2 + b)] = \mathbb{E}[\log e_t^2] = -(\gamma + \log 2) < 0,$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Therefore, the equations $X_t = e_t\sigma_t$; $\sigma_t^2 = X_{t-1}^2$, or equivalently $X_t = e_t X_{t-1}$ define a stationary process which has infinite variance at every t. On the other hand, $\sigma_t^2 = \sigma_{t-1}^2$ has no stationary solution.

In some applications, we may require that the GARCH process have finite higher-order moments; for example, when studying its tail behavior it is useful to study its excess kurtosis, which requires the fourth moment to exist and be finite. This leads to further restrictions on the coefficients α and β.

For a stationary GARCH process,

$$\begin{aligned}
\mathbb{E}[X_t^4] &= \mathbb{E}[e_t^4]\mathbb{E}[\sigma_t^4] \\
&= \mathbb{E}[e_t^4]\mathbb{E}[\omega^2\Big(1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)\Big)^2] \\
&= \omega^2\mathbb{E}[e_t^4]\mathbb{E}[1 + 2\sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)^2 + \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\prod_{i=1}^{k}\prod_{j=1}^{l}(\alpha e_{t-i}^2 + \beta)(\alpha e_{t-j}^2 + \beta)] \\
&= \omega^2\mathbb{E}[e_t^4]\Big(1 + 2\sum_{j=1}^{\infty}(\alpha + \beta)^j + \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mathbb{E}[(\alpha e_t^2 + \beta)^2]^{k\wedge l}(\alpha + \beta)^{k\vee l - k\wedge l}\Big),
\end{aligned}$$

which is finite if and only if $\rho^2 := \mathbb{E}[(\alpha e_t^2 + \beta)^2] < 1$. In this case, using the recursion $\sigma_t^2 = \omega + \sigma_{t-1}^2(\alpha e_{t-1}^2 + \beta)$,

$$\begin{aligned}
\mathbb{E}[\sigma_t^4] &= \omega^2 + 2\omega\mathbb{E}[\alpha e_{t-1}^2 + \beta]\mathbb{E}[\sigma_{t-1}^2] + \mathbb{E}[(\alpha e_{t-1}^2 + \beta)^2]\mathbb{E}[\sigma_{t-1}^4] \\
&= \omega^2 + 2\omega^2\frac{\alpha + \beta}{1 - \alpha - \beta} + \rho^2\mathbb{E}[\sigma_t^4],
\end{aligned}$$

so

$$\mathbb{E}[X_t^4] = \mathbb{E}[\sigma_t^4]\mathbb{E}[e_t^4] = \omega^2\mathbb{E}[e_t^4]\frac{1 + \alpha + \beta}{(1 - \rho^2)(1 - \alpha - \beta)}.$$

In the case of normally distributed innovations (e_t), the condition $\rho^2 < 1$ means

$$\mathbb{E}[(\alpha e_t^2 + \beta)^2] = \alpha^2\mathbb{E}[e_t^4] + 2\alpha\beta\mathbb{E}[e_t^2] + \beta^2 = 3\alpha^2 + 2\alpha\beta + \beta^2 = (\alpha + \beta)^2 + 2\alpha^2 < 1.$$

The excess kurtosis of X_t with normally distributed innovations is then

$$\begin{aligned}
\frac{\mathbb{E}[X_t^4]}{\mathrm{Var}[X_t]^2} - 3 &= \frac{3(1 + \alpha + \beta)\omega^2}{(1 - \alpha - \beta)(1 - 2\alpha^2 - (\alpha + \beta)^2)(\frac{\omega}{1 - \alpha - \beta})^2} - 3 \\
&= 3\frac{1 - (\alpha + \beta)^2}{1 - 2\alpha^2 - (\alpha + \beta)^2} - 3 \\
&= \frac{2\alpha^2}{1 - 2\alpha^2 - (\alpha + \beta)^2} > 0,
\end{aligned}$$

which means that X_t is leptokurtic, or heavy-tailed. This implies that outliers in the GARCH model should occur more frequently than they would with a process of i.i.d. normally distributed variables, which is consistent with empirical studies of financial processes.
More generally, for X_t to have a finite $2n$-th moment ($n \in \mathbb{N}$) a necessary and sufficient condition is that $\mathbb{E}[(\alpha e_t^2 + \beta)^n] < 1$.

Another interesting feature of GARCH processes is the extent to which innovations e_t at time t persist in the conditional variance at a later time σ^2_{t+h}. To consider this mathematically we will use the following definition. For the GARCH(1,1)-process $X = (X_t)$, define

$$\pi_X(t,n) = \sigma_0^2 \prod_{i=1}^{t+n} (\alpha e^2_{t+n-i} + \beta) + \omega(\sum_{k=n}^{t+n-1} \prod_{j=1}^{k} (\alpha e^2_{t+n-j} + \beta)).$$

Definition 5. The innovation e_t does **not** persist in X in L^1 iff

$$\mathbb{E}[\pi_X(t,n)] \to 0 \; (n \to \infty),$$

and almost surely (a.s.) iff

$$\pi_X(t,n) \overset{a.s.}{\to} 0 \; (n \to \infty).$$

If every innovation e_t persists in X, then we call X persistent.

To see how this definition reflects the heuristic meaning of a shock innovation persisting in the conditional variance, consider that for a GARCH time series with finite variance,

$$\begin{aligned}\mathbb{E}[\pi_X(t,n)] &= \mathbb{E}[(\mathbb{E}[\sigma^2_{t+n}] - \mathbb{E}[\sigma^2_{t+n} \mid e_t]) \frac{1-\alpha-\beta}{\omega(\alpha+\beta)^{n-1}} \prod_{i=1}^{n-1} (\alpha e^2_{t+n-i} + \beta)] \\ &= (\mathbb{E}[\sigma^2_{t+n}] - \mathbb{E}[\sigma^2_{t+n} \mid e_t]) \frac{1-\alpha-\beta}{\omega},\end{aligned}$$

which tends to zero if and only if $\mathbb{E}[\sigma^2_{t+n}] - \mathbb{E}[\sigma^2_{t+n} \mid e_t]$ tends to zero as well.

Theorem 6. *(i) If e_t persists in X in L^1 for any t, then e_t persists in X in L^1 for all t. This is the case if and only if $\alpha + \beta \geq 1$.*
(ii) If e_t persists in X a.s. for any t, then e_t persists in X a.s. for all t. This is the case if and only if $\mathbb{E}[\log(\alpha e_t^2 + \beta)] \geq 0$.

Proof. (i) First,

$$\mathbb{E}[\pi_X(t,n)] = \sigma_0^2 \prod_{i=1}^{t+n} \mathbb{E}[\alpha e^2_{t+n-i} + \beta] + \omega \sum_{k=n}^{t+n-1} \prod_{j=1}^{k} \mathbb{E}[\alpha e^2_{t+n-j} + \beta].$$

For this value to be converge to zero (that is, for e_t to not persist), we need $\mathbb{E}[\sigma_0^2]$ to be finite, which means $\alpha + \beta < 1$. On the other hand, let $\alpha + \beta < 1$. Then we have

$$\begin{aligned}\mathbb{E}[\pi_X(t,n)] &= \frac{\omega}{1-\alpha-\beta} (\alpha+\beta)^{t+n} + \omega \sum_{k=n}^{t+n-1} (\alpha+\beta)^k \\ &= \frac{\omega(\alpha+\beta)^n}{1-\alpha-\beta} \to 0 \; (n \to \infty),\end{aligned}$$

so e_t does not persist.
(ii) Let $\mathbb{E}[\log(\alpha e_t^2 + \beta)] < 0$. By the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\log(\alpha e_{t-i}^2 + \beta) \overset{a.s.}{\to} \mathbb{E}[\log(\alpha e_t^2 + \beta)] < 0,$$

so

$$\log\prod_{i=1}^{n}(\alpha e_{t-i}^2 + \beta) = n\Big(\frac{1}{n}\sum_{i=1}^{n}\log(\alpha e_{t-i}^2 + \beta)\Big) \overset{a.s.}{\to} -\infty$$

and therefore

$$\prod_{i=1}^{n}(\alpha e_{t-i}^2 + \beta) \overset{a.s.}{\to} 0.$$

This means that we have

$$\pi_X(t,n) = \underbrace{\sigma_0^2\prod_{i=1}^{t+n}(\alpha e_{t+n-i}^2 + \beta)}_{\to 0} + \omega(\underbrace{\sum_{k=n}^{t+n-1}\prod_{j=1}^{k}(\alpha e_{t+n-j}^2 + \beta))}_{\to 0} \overset{a.s.}{\to} 0.$$

On the other hand, let $\mathbb{E}[\log(\alpha e_t^2 + \beta)] \geq 0$. Then by the argument in the proof of Theorem 1, we have

$$\limsup_{j\to\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta) = \infty \text{ a.s.}$$

so that $\pi_X(t,n)$ cannot converge to zero.

□

It is a peculiar property of GARCH(1,1) models that the concept of persistence depends strongly on the type of convergence used in the definition. Persistence in L^1 is the more intuitive sense, since it excludes pathological volatility processes such as $\sigma_t^2 = 3\sigma_{t-1}^2$, which is strongly stationary since $\mathbb{E}[\log(3e_t^2 + 0)] = -(\gamma + \log 2) + \log 3 < 0$.

Definition 7. We call $\pi := \alpha + \beta$ the persistence of a GARCH(1,1) model with parameters ω, α, β.

As we have seen earlier, the persistence of the model limits the kurtosis the process can take. Since the estimated best-fit GARCH process to a time series often has persistence close to 1, this severely limits the value of α to ensure the existence of the fourth moment. From the representation of σ_t^2 in theorem 2.1, we immediately have

Theorem 8. *The autocorrelation function (ACF) of $\{X_n^2\}$ decays exponentially to zero with rate π if $\pi < 1$.*

3 A central limit theorem

Having derived the admissible parameter space, we consider the task of estimating the parameters and predicting the values of X_t at future times t. Since X_t is centered at every t, a natural estimator for its variance is the average of the squares $\frac{1}{n}\sum_{t=1}^{n} X_t^2$. The following theorem will show that, under a stationarity and moment assumption, this is a consistent and asymptotically normal estimator.

Theorem 9. *For a wide-sense stationary GARCH(1,1)-process X_t with $Var[X_t^2] < \infty$, $\mathbb{E}[e_t^4] < \infty$ and parameters ω, α, β, the following theorem holds:*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(X_t^2 - \mathbb{E}[X_t^2]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega^2 \frac{1+\alpha+\beta}{(1-\alpha-\beta)^2}\Big(\frac{\mathbb{E}[e_t^4](1+\alpha-\beta)+2\beta}{1-\rho^2} - \frac{1}{1-\alpha-\beta}\Big),$$

where $\rho^2 := \mathbb{E}[(\alpha e_t^2 + \beta)^2]$ as in section 2.

Proof. By Corollary 2.2 the condition $\mathbb{E}[X_t^2] < \infty$ implies that $\alpha + \beta < 1$ and we have

$$\mathbb{E}[X_t^2] = \mathbb{E}[\sigma_t^2] = \frac{\omega}{1-\alpha-\beta}.$$

Similarly, we have seen that $\mathbb{E}[X_t^4]$ is finite if and only if $\rho^2 < 1$ and in this case one has

$$\mathbb{E}[X_t^4] = \mathbb{E}[\sigma_t^4]\mathbb{E}[e_t^4] = \omega^2\mathbb{E}[e_t^4]\frac{1+\alpha+\beta}{(1-\rho^2)(1-\alpha-\beta)}.$$

Define $Y_t := X_t^2 - \mathbb{E}[X_t^2]$. Then the variables $\{Y_1, Y_2, ...\}$ are weakly dependent in the following sense:

Lemma 10. *Let $s_1 < s_2 < ... < s_u < s_u + r = t$ and let $f : \mathbb{R}^u \to \mathbb{R}$ be quadratically integrable and measurable. Then*

$$|\text{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t]| \leq C\sqrt{\mathbb{E}[f^2(Y_{s_1}, ..., Y_{s_u})]}\rho^r$$

for a constant C which is independent of $s_1, ..., s_u, r$.

Proof. Let $w = \omega - \mathbb{E}[X_t^2]$. Then we have

$$Y_t = we_t^2(1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)).$$

We define the helper variable

$$\tilde{Y}_t = we_t^2(1 + \sum_{j=1}^{r-1}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta)).$$

Then $\tilde{Y}_t$ is independent of $(Y_{s_1}, ..., Y_{s_u})$ and by the Cauchy-Schwarz inequality:

$$\begin{aligned}|\text{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t]| &= |\text{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t - \tilde{Y}_t]| \\ &\leq \sqrt{\mathbb{E}[f^2(Y_{s_1}, ..., Y_{s_u})]}\sqrt{\mathbb{E}[(Y_t - \tilde{Y}_t)^2]}.\end{aligned}$$

However, we have

$$\begin{aligned}
\mathbb{E}[(Y_t - \tilde{Y}_t)^2] &= \mathbb{E}[w^2 e_t^4 (\sum_{j=1}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta) - \sum_{j=1}^{r-1}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta))^2] \\
&= w^2 \mathbb{E}[e_t^4]\mathbb{E}[(\sum_{j=r}^{\infty}\prod_{i=1}^{j}(\alpha e_{t-i}^2 + \beta))^2] \\
&= w^2 \mathbb{E}[e_t^4]\mathbb{E}[(\prod_{k=1}^{r}(\alpha e_{t-k}^2 + \beta))^2]\mathbb{E}[(1 + \sum_{j=r+1}^{\infty}\prod_{i=r+1}^{j}(\alpha e_{t-i}^2 + \beta))^2] \\
&= \rho^{2r}\frac{w^2}{\omega^2}\mathbb{E}[e_t^4]\mathbb{E}[\sigma_t^4] \\
&= \rho^{2r}\mathbb{E}[Y_t^2],
\end{aligned}$$

and therefore

$$|\mathrm{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t]| \leq \underbrace{\sqrt{\mathbb{E}[Y_t^2]}}_{C}\sqrt{\mathbb{E}[f^2(Y_{s_1}, ..., Y_{s_u})]}\rho^r.$$

□

Similarly, we will need an additional inequality:

Lemma 11. *Let $s_1 < s_2 < ... < s_u < s_u + r = t$ and let $f : \mathbb{R}^u \to \mathbb{R}$ be bounded, measurable and integrable. Then*

$$|\mathrm{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t Y_{t+h}]| \leq C\|f\|_\infty \rho^r$$

for any $h > 0$.

Proof. Define $\tilde{Y}_t$ as earlier. Then by Hölder's inequality,

$$\begin{aligned}
|\mathrm{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t Y_{t+h}]| &= |\mathrm{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t Y_{t+h} - \tilde{Y}_t\tilde{Y}_{t+h}]| \\
&\leq 2\|f\|_\infty \mathbb{E}[|Y_t Y_{t+h} - \tilde{Y}_t\tilde{Y}_{t+h}|].
\end{aligned}$$

Using the triangle and Cauchy-Schwarz inequalities, we have

$$\begin{aligned}
\mathbb{E}[|Y_t Y_{t+h} - \tilde{Y}_t\tilde{Y}_{t+h}|] &\leq \mathbb{E}[|Y_t - \tilde{Y}_t|Y_{t+h}] + erw|Y_{t+h} - \tilde{Y}_{t+h}|\tilde{Y}_t \\
&\leq \sqrt{\mathbb{E}[(Y_t - \tilde{Y}_t)]}\sqrt{\mathbb{E}[Y_{t+h}^2]} + \sqrt{\mathbb{E}[(Y_{t+h} - \tilde{Y}_{t+h})^2]}\sqrt{\mathbb{E}[\tilde{Y}_t^2]} \\
&\leq \rho^r\sqrt{\mathbb{E}[Y_t^2]}\sqrt{\mathbb{E}[Y_{t+h}^2]} + \rho^r\sqrt{\mathbb{E}[Y_{t+h}^2]}\sqrt{\mathbb{E}[\tilde{Y}_t^2]},
\end{aligned}$$

so that

$$|\mathrm{Cov}[f(Y_{s_1}, ..., Y_{s_u}), Y_t Y_{t+h}]| \leq \underbrace{2(\mathbb{E}[Y_t^2] + \sqrt{\mathbb{E}[Y_t^2]}\sqrt{\mathbb{E}[\tilde{Y}_t^2]})}_{C}\|f\|_\infty \rho^r.$$

□

The theorem to be proved is now

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega^2 \frac{1+\alpha+\beta}{(1-\alpha-\beta)^2}(\frac{\mathbb{E}[e_t^4](1+\alpha-\beta)+2\beta}{1-\rho^2} - \frac{1}{1-\alpha-\beta})).$$

Define

$$\sigma^2 := \sum_{k=-\infty}^{\infty} \mathbb{E}[Y_0 Y_k] = \mathbb{E}[Y_0^2] + 2\sum_{k=1}^{\infty} \mathbb{E}[Y_0 Y_k]$$

and

$$\sigma_n^2 := \mathrm{Var}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t].$$

Then we have

$$\sigma^2 = \mathbb{E}[X_0^4] - \mathbb{E}[X_0^2]^2 + 2\sum_{k=1}^{\infty}(\mathbb{E}[X_0^2 X_k^2] - \mathbb{E}[X_0^2]\mathbb{E}[X_k^2]).$$

However,

$$\begin{aligned}
\mathbb{E}[X_0^2 X_k^2] &= \mathbb{E}[\sigma_0^2 \sigma_k^2] \\
&= \mathbb{E}[\sigma_0^2 (a_0 \sum_{j=1}^{k-1}\prod_{i=1}^{j}(\alpha e_{k-i}^2 + \beta) + \sigma_0^2 \prod_{i=1}^{k}(\alpha e_{k-i}^2 + \beta))] \\
&= (\alpha+\beta)^k \mathbb{E}[\sigma_0^4] + \omega \frac{1-(\alpha+\beta)^k}{1-\alpha-\beta}\mathbb{E}[\sigma_0^2] \\
&= \omega^2 (\frac{(1+\alpha+\beta)(\alpha+\beta)^k}{(1-\rho^2)(1-\alpha-\beta)} + \frac{1-(\alpha+\beta)^k}{(1-\alpha-\beta)^2}),
\end{aligned}$$

so

$$\begin{aligned}
\sigma^2 &= \omega^2 \mathbb{E}[e_t^4]\frac{1+\alpha+\beta}{(1-\rho^2)(1-\alpha-\beta)} - \omega^2 \frac{1}{(1-\alpha-\beta)^2} + \\
&\quad + 2\sum_{k=1}^{\infty} \omega^2 (\frac{(1+\alpha+\beta)(\alpha+\beta)^k}{(1-\rho^2)(1-\alpha-\beta)} - \frac{(\alpha+\beta)^k}{(1-\alpha-\beta)^2}) \\
&= \frac{\omega^2}{1-\alpha-\beta}(\mathbb{E}[e_t^4]\frac{1+\alpha+\beta}{1-\rho^2} - \frac{1}{1-\alpha-\beta} + 2(\frac{1+\alpha+\beta}{1-\rho^2} - \frac{1}{1-\alpha-\beta})\sum_{k=1}^{\infty}(\alpha+\beta)^k) \\
&= \frac{\omega^2}{(1-\alpha-\beta)^2}(\mathbb{E}[e_t^4]\frac{1-(\alpha+\beta)^2}{1-\rho^2} - 1 + \frac{2+2\alpha+2\beta}{1-\rho^2} - \frac{2}{1-\alpha-\beta}) \\
&= \omega^2 \frac{1+\alpha+\beta}{(1-\alpha-\beta)^2}(\frac{\mathbb{E}[e_t^4](1+\alpha-\beta)+2\beta}{1-\rho^2} - \frac{1}{1-\alpha-\beta}).
\end{aligned}$$

Since Y_t is centered for every t, we have

$$\begin{aligned}
|\tilde{\sigma}_n^2 - \sigma^2| &= |(\frac{1}{n}\sum_{s,t=1}^{n}\mathbb{E}[Y_sY_t]) - \sigma^2| \\
&= |\mathbb{E}[Y_0^2] + 2\sum_{k=1}^{n-1}(1-\frac{k}{n})\mathbb{E}[Y_0Y_k] - \sigma^2| \\
&\leq 2\sum_{k=n}^{\infty}|\mathbb{E}[Y_0Y_k]| + 2\sum_{k=1}^{n-1}\frac{k}{n}|\mathbb{E}[Y_0Y_k]| \\
&\leq \underbrace{2C\sqrt{\mathbb{E}[Y_0^2]}\sum_{k=n}^{\infty}\rho^k}_{\to 0} + \underbrace{2C\sqrt{\mathbb{E}[Y_0^2]}\sum_{k=1}^{n-1}\frac{k\rho^k}{n}}_{\to 0} \to 0 \ (n\to\infty),
\end{aligned}$$

using the weak independence of the variables $\{Y_k\}$, where the limit of the second sum is due to Kronecker's lemma since $\sum_{k=1}^{\infty}\rho^k < \infty$. This means that $\tilde{\sigma}_n^2 \to \sigma^2$ for $n \to \infty$.
We now define

$$W_{n,t} := \frac{Y_t}{\tilde{\sigma}_n\sqrt{n}} = \frac{X_t^2 - \mathbb{E}[X_t^2]}{\tilde{\sigma}_n\sqrt{n}} \ \ (t \leq n).$$

By Slutsky's theorem, it is enough to show that $\sum_{t=1}^{n} W_{n,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. For $k \leq n$ define

$$v_{n,k} := \mathrm{Var}[\sum_{t=1}^{k} W_{n,t}] - \mathrm{Var}[\sum_{t=1}^{k-1} W_{n,t}].$$

Then we have

$$\begin{aligned}
v_{n,k} &= \mathrm{Var}[W_{n,k}] + 2\sum_{t=1}^{k-1}\mathrm{Cov}[W_{n,t}, W_{n,k}] \\
&= \frac{1}{n\tilde{\sigma}_n^2}(\mathbb{E}[Y_0^2] + 2\sum_{t=1}^{k-1}\mathbb{E}[Y_0Y_t]) \\
&= \frac{1}{n\tilde{\sigma}_n^2}(\sigma^2 - \underbrace{2\sum_{t=k}^{\infty}\mathbb{E}[Y_0Y_t]}_{\to 0}),
\end{aligned}$$

so $nv_{n,k}$ tends to 1 as $n \to \infty$ for all k. In particular we have $\max_{k\leq n} v_{n,k} \to 0$ $(n \to \infty)$ and that $v_{n,k}$ is positive for all $n \geq k > N_0$ for a large enough $N_0 \in \mathbb{N}$. Without loss of generality, we may assume that $v_{n,k} > 0$ is true for all n, k.
For every $n \in \mathbb{N}$ let $Z_{n,1}, Z_{n,2}, ..., Z_{n,n}$ be independent random variables, also independent from $W_{n,k}$ for every k, and such that $Z_{n,k} \sim \mathcal{N}(0, v_{n,k})$. Since

$$\underbrace{Z_{n,1} + ... + Z_{n,n}}_{\sim\mathcal{N}(0,\sum_k v_{n,k})} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

it is enough to show that

$$\mathbb{E}[f(W_{n,1}+...+W_{n,n})-f(Z_{n,1}+...+Z_{n,n})]\to 0\ \ (n\to\infty)$$

for any function $f\in C^3(\mathbb{R})$. Let f be any such function. For $1\le k\le n$ we define the partial sums

$$S_{n,k}:=W_{n,1}+...+W_{n,k-1},\qquad T_{n,k}:=Z_{n,k+1}+...+Z_{n,n}$$

(where $S_{n,1}=T_{n,n}=0$), as well as

$$\Delta_{n,k}:=\mathbb{E}[f(S_{n,k}+W_{n,k}+T_{n,k})-f(S_{n,k}+Z_{n,k}+T_{n,k})].$$

Clearly,

$$\mathbb{E}[f(W_{n,1}+...+W_{n,n})-f(Z_{n,1}+...+Z_{n,n})]=\sum_{k=1}^{n}\Delta_{n,k}.$$

We now split $\Delta_{n,k}=\Delta_{n,k}^{(1)}-\Delta_{n,k}^{(2)}$ by defining

$$\Delta_{n,k}^{(1)}:=\mathbb{E}[f(S_{n,k}+W_{n,k}+T_{n,k})]-\mathbb{E}[f(S_{n,k}+T_{n,k})]-\frac{v_{n,k}}{2}\mathbb{E}[f''(S_{n,k}+T_{n,k})]$$

$$\Delta_{n,k}^{(2)}:=\mathbb{E}[f(S_{n,k}+Z_{n,k}+T_{n,k})]-\mathbb{E}[f(S_{n,k}+T_{n,k})]-\frac{v_{n,k}}{2}\mathbb{E}[f''(S_{n,k}+T_{n,k})]$$

and consider each term separately. By applying Taylor's theorem to f as a function in $Z_{n,k}$ around 0, there exists a random variable $\xi_{n,k}\in(0,1)$ with

$$\Delta_{n,k}^{(2)}=\underbrace{\mathbb{E}[Z_{n,k}f'(S_{n,k}+T_{n,k})]}_{=0}+\mathbb{E}[\frac{Z_{n,k}^2}{2}(f''(S_{n,k}+\xi_{n,k}Z_{n,k}+T_{n,k})]-\frac{v_{n,k}}{2}\mathbb{E}[f''(S_{n,k}+T_{n,k}))]$$

where the first term is zero because $Z_{n,k}$ is independent of $S_{n,k},T_{n,k}$. By the mean value theorem,

$$|\sum_{k=1}^{n}\Delta_{n,k}^{(2)}|\le C\sum_{k=1}^{n}\mathbb{E}[|Z_{n,k}^3|]=C\sum_{k=1}^{n}v_{n,k}^{3/2}\le C\max_{1\le k\le n}v_{n,k}^{1/2}\underbrace{\sum_{k=1}^{n}v_{n,k}}_{=1}\to 0(n\to\infty).$$

Showing this for $\Delta_{n,k}^{(1)}$ is somewhat more difficult. Similarly to above, we find by Taylor's theorem a random variable $\tau_{n,k}\in(0,1)$ so that

$$\Delta_{n,k}^{(1)}=\mathbb{E}[W_{n,k}f'(S_{n,k}+T_{n,k})]+\mathbb{E}[\frac{W_{n,k}^2}{2}f''(S_{n,k}+\tau_{n,k}W_{n,k}+T_{n,k})]-\frac{v_{n,k}}{2}\mathbb{E}[f''(S_{n,k}+T_{n,k})]$$

where

$$\begin{aligned}\mathbb{E}[W_{n,k}f'(S_{n,k}+T_{n,k})]&=\sum_{j=1}^{k-1}\mathbb{E}[W_{n,k}(f'(S_{n,j+1}+T_{n,k})-f'(S_{n,j}+T_{n,k}))]\\&=\sum_{j=1}^{k-1}\mathbb{E}[W_{n,k}W_{n,j}f''(S_{n,j}+\xi_{n,k,j}W_{n,j}+T_{n,k})]\end{aligned}$$

for a random variable $\xi_{n,k,j} \in (0,1)$, again by Taylor's theorem. Since $v_{n,k} = \mathbb{E}[W_{n,k}^2] + 2\sum_{j=1}^{k-1}\mathbb{E}[W_{n,k}W_{n,j}]$, we then have

$$\Delta_{n,k}^{(1)} = \underbrace{\sum_{j=1}^{k-1}\mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j} + \xi_{n,k,j}W_{n,j} + T_{n,k}) - \mathbb{E}[f''(S_{n,k} + T_{n,k})])]}_{=\Delta_{n,k}^{(1,1)}}$$

$$+ \underbrace{\frac{1}{2}\mathbb{E}[W_{n,k}^2(f''(S_{n,k} + \tau_{n,k}W_{n,k} + T_{n,k}) - \mathbb{E}[f''(S_{n,k} + T_{n,k})])]}_{=\Delta_{n,k}^{(1,2)}}$$

For every $d \in \mathbb{N}$ we split $\Delta_{n,k}^{(1,2)}$ as follows:

$$\Delta_{n,k}^{(1,2)} = \Delta_{n,k,d}^{(1,2,1)} + \Delta_{n,k,d}^{(1,2,2)} + \Delta_{n,k,d}^{(1,2,3)}$$

where

$$\Delta_{n,k,d}^{(1,2,1)} = \mathbb{E}[W_{n,k}^2(f''(S_{n,k} + \tau_{n,k}W_{n,k} + T_{n,k}) - f''(S_{n,k-d} + T_{n,k}))]$$
$$\Delta_{n,k,d}^{(1,2,2)} = \mathbb{E}[W_{n,k}^2(f''(S_{n,k-d} + T_{n,k}) - \mathbb{E}[f''(S_{n,k-d} + T_{n,k})])]$$

and

$$\Delta_{n,k,d}^{(1,2,3)} = \mathbb{E}[W_{n,k}^2(\mathbb{E}[f''(S_{n,k-d} + T_{n,k})] = \mathbb{E}[f''(S_{n,k} + T_{n,k})])]$$

From the second weak dependence lemma (Lemma 11), it follows that

$$\begin{aligned}
|\Delta_{n,k,d}^{(1,2,2)}| &= \frac{1}{n\tilde{\sigma}_n^2}|\mathbb{E}[Y_k^2(f''(S_{n,k-d} + T_{n,k}) - \mathbb{E}[f''(S_{n,k-d} + T_{n,k})])]| \\
&= \frac{1}{n\tilde{\sigma}_n^2}|\mathrm{Cov}[f''(S_{n,k-d} + T_{n,k}), Y_k^2]| \\
&\le \frac{1}{n\tilde{\sigma}_n^2}\|f''\|_\infty \rho^d,
\end{aligned}$$

and therefore for every $\delta > 0$ there exists a $D_0 \in \mathbb{N}$ so that $\forall d > D_0$:

$$\sum_{k=1}^{n}\Delta_{n,k,d}^{(1,2,2)} \le \frac{1}{\tilde{\sigma}_n^2}\|f''\|_\infty \rho^d \le \delta.$$

For any such d, using the Cauchy-Schwarz inequality and the fact that f'' is bounded, for any $\epsilon > 0$:

$$\begin{aligned}
|\sum_{k=1}^{n}\Delta_{n,k,d}^{(1,2,1)}| &\le C(\sum_{k=1}^{n}\mathbb{E}[W_{n,k}^2\mathbb{I}_{\{|Y_k|\ge\epsilon\sqrt{n}\}}] + \sum_{k=1}^{n}\mathbb{E}[W_{n,k}^2\mathbb{I}_{\{|Y_k|\le\epsilon\sqrt{n}\}}\sum_{j=k-d}^{k}|W_{n,j}|]) \\
&= \frac{C}{\tilde{\sigma}_n^2}\mathbb{E}[Y_1^2\mathbb{I}_{\{|Y_1|\ge\epsilon\sqrt{n}\}}] + \frac{C\epsilon}{\tilde{\sigma}_n}\sum_{k=1}^{n}\mathbb{E}[|W_{n,k}|\sum_{j=k-d}^{k}|W_{n,j}|]
\end{aligned}$$

for an appropriate bound $C > 0$. Using the Cauchy-Schwarz inequality on every term in the above sum, we have

$$
\begin{aligned}
|\sum_{k=1}^{n} \Delta_{n,k,d}^{(1,2,1)} &\leq \frac{C}{\tilde{\sigma}_n^2}\mathbb{E}[Y_1^2 \mathbb{I}_{\{|Y_1|\geq \epsilon\sqrt{n}\}}] + \frac{C\epsilon}{\tilde{\sigma}_n}\sum_{k=1}^{n}\sum_{j=k-d}^{k}\sqrt{\mathbb{E}[W_{n,k}^2]\mathbb{E}[W_{n,j}^2]} \\
&\leq \frac{C}{\tilde{\sigma}_n^2}\mathbb{E}[Y_1^2 \mathbb{I}_{\{|Y_1|\geq \epsilon\sqrt{n}\}}] + \frac{C(d+1)\epsilon}{\tilde{\sigma}_n^3}\sum_{j\in\mathbb{N}}\mathbb{E}[Y_j^2] \leq \delta
\end{aligned}
$$

for large enough n and small enough ϵ. In addition, we have

$$
\begin{aligned}
|\Delta_{n,k,d}^{(1,2,3)}| &= |\mathbb{E}[W_{n,k}^2]||\mathbb{E}[f''(S_{n,k-d} + T_{n,k})] - \mathbb{E}[f''(S_{n,k} + T_{n,k})]| \\
&\leq C|\mathbb{E}[W_{n,k}^2]|\sum_{j=k-d}^{k-1}\mathbb{E}[|W_{n,j}|] \\
&= \frac{Cd}{n^{-3/2}\tilde{\sigma}_n^3}\mathbb{E}[Y_1^2]\mathbb{E}[Y_1],
\end{aligned}
$$

which is summable over n. With $\delta \to 0$ we have

$$
\sum_{k=1}^{n}\Delta_{n,k}^{(1,2)} \to 0 \ \ (n\to\infty).
$$

For the term $\Delta_{n,k}^{(1,1)}$ we use the triangle inequality and have for any d between 1 and k:

$$
\begin{aligned}
|\Delta_{n,k}^{(1,1)}| \leq |&\sum_{j=1}^{k-d}\mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j} + \xi_{n,k,j}W_{n,j} + T_{n,k}) - \mathbb{E}[f''(S_{n,k} + T_{n,k})])]| + \\
&+ |\sum_{j=k-d+1}^{k-1}\mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j} + \xi_{n,k,j}W_{n,j} + T_{n,k}) - \mathbb{E}[f''(S_{n,k} + T_{n,k})])]|.
\end{aligned}
$$

The first term tends to zero: for any $\delta > 0$,

$$
\begin{aligned}
&|\sum_{j=1}^{k-d}\mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j} + \xi_{n,k,j}W_{n,j} + T_{n,k}) - \mathbb{E}[f''(S_{n,k} + T_{n,k})])]| \\
&\leq \frac{\|f\|_\infty}{n\tilde{\sigma}_n^2}\sum_{j=1}^{k-d}\rho^{k-j} \\
&\leq \frac{\|f\|_\infty(1-\rho^d)}{n(1-\rho)\tilde{\sigma}_n^2} \leq \frac{\delta}{n}
\end{aligned}
$$

for large enough d (and $n \geq d$). The other term is split into three parts:

$$
\Delta_{n,k,d}^{(1,1,1)} := \sum_{j=k-d+1}^{k-1}\mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j} + \xi_{n,k,j}W_{n,j} + T_{n,k}) - f''(S_{n,j-d} + T_{n,k}))],
$$

$$\Delta_{n,k,d}^{(1,1,2)} := \sum_{j=k-d+1}^{k-1} \mathbb{E}[W_{n,k}W_{n,j}(f''(S_{n,j-d}+T_{n,k}) - \mathbb{E}[f''(S_{n,j-d}+T_{n,k})])],$$

$$\Delta_{n,k,d}^{(1,1,3)} := \sum_{j=k-d+1}^{k-1} \mathbb{E}[W_{n,k}W_{n,j}(\mathbb{E}[f''(S_{n,j-d}+T_{n,k})] - \mathbb{E}[f''(S_{n,j}+T_{n,k})])].$$

Using the weak dependence shown in Lemma 11, for large enough d:

$$\begin{aligned}
|\Delta_{n,k,d}^{(1,1,2)}| &\leq \frac{1}{n\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} |\mathrm{Cov}[f''(S_{n,j-d}+T_{n,j}), Y_jY_k]| \\
&\leq \frac{d\rho^d}{n\tilde{\sigma}_n^2}\|f\|_\infty \leq \delta.
\end{aligned}$$

Additionally, for every $\epsilon > 0$,

$$\begin{aligned}
|\sum_{k=1}^{n} \Delta_{n,k,d}^{(1,1,1)}| \leq C &\sum_{k=1}^{n} \sum_{j=k-d+1}^{k-1} \mathbb{E}[W_{n,k}W_{n,j}\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}] + \\
&+ C\sum_{k=1}^{n} \sum_{j=k-d+1}^{k-1} \mathbb{E}[W_{n,k}W_{n,j}\mathbb{I}_{\{|Y_kY_j| < n\epsilon^2\}} \sum_{i=j-d}^{k} |W_{n,i}|] \\
&= \frac{C}{\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \mathbb{E}[Y_kY_j\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}] + \frac{C\epsilon}{\tilde{\sigma}_n} \sum_{k=1}^{n} \sum_{j=k-d}^{k-1} \sum_{i=j-d}^{j} \mathbb{E}[\sqrt{|W_{n,k}W_{n,j}|}|W_{n,i}|] \\
&\leq \frac{C}{\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \mathbb{E}[Y_kY_j\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}] + \frac{C\epsilon}{\tilde{\sigma}_n} \sum_{k=1}^{n} \sum_{j=k-d+1}^{k-1} \sum_{i=j-d}^{j} \sqrt{\mathbb{E}[|W_{n,k}W_{n,j}|]\mathbb{E}[W_{n,i}^2]} \\
&\leq \frac{C}{\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \mathbb{E}[Y_kY_j\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}] + \frac{C\epsilon}{n\tilde{\sigma}_n^3} \sum_{k=1}^{n} \sum_{j=k-d+1}^{k-1} \sum_{i=j-d}^{j} \sqrt{\sqrt{\mathbb{E}[Y_j^2]}\rho^{k-j}\mathbb{E}[Y_i^2]} \\
&\leq \frac{C}{\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \mathbb{E}[Y_kY_j\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}] + \frac{Cd\epsilon}{\tilde{\sigma}_n^3} \sum_{j=k-d+1}^{k-1} \rho^{\frac{k-j}{2}}\mathbb{E}[Y_j^2]^{\frac{3}{4}} \\
&= \underbrace{\frac{C}{\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \mathbb{E}[Y_kY_j\mathbb{I}_{\{|Y_kY_j| \geq n\epsilon^2\}}]}_{\to 0} + \underbrace{\frac{Cd\epsilon\mathbb{E}[Y_1^2]^{\frac{3}{4}}(1-\rho^{\frac{d}{2}})}{\tilde{\sigma}_n^3(1-\sqrt{\rho})}}_{\to 0} \leq \delta
\end{aligned}$$

for large enough n with an appropriate $\epsilon = \epsilon(n)$. Finally,

$$\begin{aligned}
|\Delta^{(1,1,3)}_{n,k,d}| &\leq \sum_{j=k-d+1}^{k-1} |\mathbb{E}[W_{n,k}W_{n,j}]|(\mathbb{E}[f''(S_{n,j-d}+T_{n,k}) - f''(S_{n,j}+T_{n,k})]) \\
&\leq Cn\tilde{\sigma}_n^2 \sum_{j=k-d+1}^{k-1} (\sqrt{\mathbb{E}[Y_j^2]}\rho^{k-j}\mathbb{E}[\sum_{i=j-d}^{j-1} |W_{n,i}|]) \\
&= \frac{Cd\sqrt{\mathbb{E}[Y_1^2]}}{n^{\frac{3}{2}}\tilde{\sigma}_n^2} \sum_{j=k-d+1}^{k-1} \rho^{k-j} \\
&= \frac{Cd\sqrt{\mathbb{E}[Y_1^2]}(1-\rho^d)}{n^{\frac{3}{2}}\tilde{\sigma}_n^2(1-\rho)}.
\end{aligned}$$

Altogether, we have

$$\begin{aligned}
\sum_{k=1}^{n} |\Delta^{(1,1)}_{n,k,d}| &\leq \delta + \frac{d\rho^d\|f''\|_\infty}{\tilde{\sigma}_n^2} + \frac{Cd\epsilon\mathbb{E}[Y_j^2]^{\frac{3}{4}}(1-\rho^{\frac{d}{2}})}{\tilde{\sigma}_n^3(1-\sqrt{\rho})} + \frac{Cd\sqrt{\mathbb{E}[Y_1^2]}(1-\rho^d)}{\sqrt{n}\tilde{\sigma}_n^2(1-\rho)} \\
&\leq 3\delta + \frac{Cd\sqrt{\mathbb{E}[Y_1^2]}(1-\rho^d)}{\sqrt{n}\tilde{\sigma}_n^2(1-\rho)} \to 0,
\end{aligned}$$

letting n tend to ∞ and δ to zero. Altogether, we now have

$$|\sum_{k=1}^{n} \Delta^{(1)}_{n,k}| \to 0 \ (n \to \infty)$$

and the theorem is proved. $\square$

4 Parameter estimation

Before attempting to estimate the parameters ω, α, β of a GARCH(1,1) process, we first have to show that they are unique - that they are the only parameters capable of defining the process.

Lemma 12. *Let X_t be a strictly stationary GARCH(1,1) model with $\alpha + \beta < 1$. Then*

$$\sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{k=1}^{\infty} \alpha\beta^{k-1} X_{t-k}^2.$$

Proof. Since $\beta < 1$ it is clear that the series converges with probability 1. Let $Z_t := \frac{\omega}{1-\beta} + \sum_{k=1}^{\infty} \alpha\beta^{k-1} X_{t-k}^2$. Then one has

$$\begin{aligned} \omega + \alpha X_{t-1}^2 + \beta Z_{t-1} &= \omega + \alpha X_{t-1}^2 + (-\omega) + \frac{\omega}{1-\beta} + \sum_{k=1}^{\infty} \alpha\beta^k X_{t-(k+1)}^2 \\ &= \frac{\omega}{1-\beta} + \sum_{k=0}^{\infty} \alpha\beta^k X_{t-(k+1)}^2 = Z_t, \end{aligned}$$

so that Z_t fulfills the same recursive equation as σ_t^2. However, by Theorem 1 the strictly stationary solution is unique, so we must have $Z_t = \sigma_t^2 \; \forall t$. □

Theorem 13. *: Let X_t be a GARCH(1,1) model with $\alpha + \beta < 1$, $\omega > 0$ and where e_t^2 is not a.s. constant. Then the parameters ω, α, β are unique.*

Proof. : Assume that X_t has two representations as a GARCH(1,1) process with parameters ω, α, β and $\hat{\omega}, \hat{\alpha}, \hat{\beta}$. By the above lemma, we can write

$$\sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{k=1}^{\infty} \alpha\beta^{k-1} X_{t-k}^2 = \frac{\hat{\omega}}{1-\hat{\beta}} + \sum_{k=1}^{\infty} \hat{\alpha}\hat{\beta}^{k-1} X_{t-k}^2.$$

This means that $\alpha\beta^{k-1} = \hat{\alpha}\hat{\beta}^{k-1} \; \forall k$; assuming otherwise, let k_0 be the smallest k with $\alpha\beta^{k_0-1} \neq \hat{\alpha}\hat{\beta}^{k_0-1}$. Then

$$e_{t-k_0}^2 = \frac{\frac{\omega}{1-\beta} - \frac{\hat{\omega}}{1-\hat{\beta}} + \sum_{j=k_0+1}^{\infty} (\alpha\beta^{j-1} - \hat{\alpha}\hat{\beta}^{j-1}) X_{t-j}^2}{\sigma_{t-k_0}^2 (\alpha\beta^{k_0-1} - \hat{\alpha}\hat{\beta}^{k_0-1})}.$$

However, since $e_{t-k_0}^2$ is also stochastically independent from the right side of this equation, it must be a.s. constant, which contradicts our assumption. Since $\alpha\beta^{k-1} = \hat{\alpha}\hat{\beta}^{k-1}$ for all k, we have

$$\frac{\alpha z}{1-\beta z} = \sum_{k=1}^{\infty} \alpha\beta^{k-1} z^k = \sum_{k=1}^{\infty} \hat{\alpha}\hat{\beta}^{k-1} z^k = \frac{\hat{\alpha} z}{1-\hat{\beta} z} \;\; \forall |z| \leq 1,$$

and by analytic extension we have

$$\alpha(1-\hat{\beta}z) = \hat{\alpha}(1-\beta z) \;\; \forall z.$$

Letting $z = 0$ we have $\alpha = \hat{\alpha}$, which immediately leads to $\beta = \hat{\beta}$ and $\omega = \hat{\omega}$. □

Due to the complexity of the maximum likelihood function for GARCH(1,1) models, one generally uses an approximating function. Let $\theta = (\omega, \alpha, \beta)$ and let $(\mathcal{F}_t)$ be the filtration on Ω generated by $\{X_s \colon s \leq t\}$. Then one has

$$\mathcal{L}(\theta \mid x_0, ..., x_n) = f(x_0, ..., x_n \mid \theta) = f(x_n \mid \mathcal{F}_{n-1}) f(x_{n-1} \mid \mathcal{F}_{n-2}) ... f(x_1 \mid \mathcal{F}_0) f(x_0 \mid \theta),$$

so

$$-\log \mathcal{L}(\theta \mid x_0, ..., x_n) = -\log f(x_0 \mid \theta) - \sum_{k=1}^{n} \log f(x_k \mid \mathcal{F}_{k-1}),$$

where $f(x_0, ..., x_n \mid \theta)$ is the joint probability distribution of $\{X_0, ..., X_n\}$ in a GARCH model with parameters θ. With a large enough sample size, the contribution from $f(x_0 \mid \theta)$ is commonly assumed to be relatively small and is dropped, resulting in the quasi maximum likelihood function

$$\mathcal{QL}(\theta \mid x_0, ..., x_n) = -\sum_{k=1}^{n-1} \log f(x_k \mid \mathcal{F}_{k-1})$$

which is to be minimized. In the case of normally distributed innovations e_t, we have

$$f(x_{k+1} \mid \mathcal{F}_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp(-\frac{x_k^2}{2\sigma_k}),$$

so

$$\mathcal{QL}(\theta \mid x_0, ..., x_n) = \frac{n-1}{2} \log 2\pi + \frac{1}{2} \sum_{k=1}^{n} \Big(\log \sigma_k^2(\theta) + \frac{x_k^2}{\sigma_k^2(\theta)} \Big).$$

The parameter $\hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)$ which minimizes this function given observations $X_0 = x_0, ..., X_n = x_n$, or equivalently which minimizes $l_n(\theta) = \frac{1}{n} \sum_{k=1}^{n} (\log \sigma_k^2 + \frac{X_k^2}{\sigma_k^2})$, is called the Quasi-Maximum-Likelihood estimator.

Theorem 14. *(QMLE - strong consistency)*
Let X_t be a GARCH-time series with parameters $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. Under the conditions
(i) θ_0 lies in a compact set $K \subset (0, \infty) \times [0, \infty) \times [0, 1)$
(ii) e_t^2 is not a.s. constant
(iii) $\beta_0 < 1$, $\alpha_0 \neq \beta_0$, $\alpha_0 + \beta_0 < 1$, $\omega_0 > 0$
the QML estimator is consistent: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$ for $n \to \infty$

Proof. By Theorem 4.2, θ is identifiable.
For each parameter θ, we define recursively $\sigma_0^2(\theta) = \frac{\omega}{1-\alpha-\beta}$ and

$$\sigma_k^2(\theta) = \omega + \alpha X_{k-1}^2 + \beta \sigma_{k-1}^2(\theta)$$

for $k = 1, ..., n$. Clearly, $\sigma_k^2(\theta_0)$ coincides with the true volatility. Since $\alpha_0 + \beta_0 < 1$, we know that $\mathbb{E}_{\theta_0}[X_k^2] = \mathbb{E}_{\theta_0}[\sigma_k^2(\theta_0)]$ are finite for all k and therefore

$$\mathbb{E}_{\theta_0}[|l_n(\theta_0)|] = 1 + \sum_{k=1}^{n} \mathbb{E}_{\theta_0}[|\log \sigma_k^2|]$$

is also finite. Using the inequality $\log x \leq x - 1$ (for $x > 0$),

$$\begin{aligned}
\mathbb{E}_{\theta_0}[|l_n(\theta)|] - \mathbb{E}_{\theta_0}[|l_n(\theta_0)|] &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{\theta_0}[\log(\sigma_k^2(\theta)) - \log(\sigma_k^2(\theta_0)) + \frac{\sigma_k^2(\theta_0)e_k^2}{\sigma_k^2(\theta)} - e_k^2] \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_{\theta_0}[-\log(\frac{\sigma_k^2(\theta_0)}{\sigma_k^2(\theta)}) + \frac{\sigma_k^2(\theta_0)}{\sigma_k^2(\theta)} - 1] \geq 0,
\end{aligned}$$

with equality only when $\frac{\sigma_k^2(\theta_0)}{\sigma_k^2(\theta)} = 1$ $\mathbb{P}_{\theta_0}$-a.s. for all k. Since θ_0 is identifiable, this happens only when $\theta = \theta_0$. Since the parameter set K is compact, we have $\sup_{\theta\in K}\beta < 1$. Define

$$\tilde{l}_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}(\log\sigma_k^2(\theta) + \frac{X_k^2}{\sigma_k^2(\theta)}) = \frac{1}{n}\sum_{k=1}^{n}(\log\sigma_k^2(\theta) + \frac{e_k^2\sigma_k^2(\theta_0)}{\sigma_k^2(\theta)}).$$

Then

$$\sup_{\theta\in K}|\sigma_t^2(\theta_0) - \sigma_t^2(\theta)| = \sup\sigma_0^2|\prod_i(\alpha_0 e_{t-i}^2 + \beta_0) - \prod_i(\alpha e_{t-i}^2 + \beta)| \leq C\sup_{\theta\in K}\beta^t.$$

Using the inequality $|\log(\frac{x}{y})| \leq \frac{|x-y|}{\min(x,y)}$ for $x, y > 0$, we have

$$\begin{aligned}
\sup_{\theta\in K}|l_n(\theta_0) - \tilde{l}_n(\theta)| &\leq \frac{1}{n}\sum_{k=1}^{n}\sup_{\theta\in K}\frac{|\sigma_k^2(\theta_0) - \sigma_k^2(\theta)|}{\sigma_k^2(\theta_0)\sigma_k^2(\theta)}X_k^2 + |\log(\frac{\sigma_k^2(\theta_0)}{\sigma_k^2(\theta)})| \\
&\leq C\sup_{\theta\in K}\frac{1}{\omega^2}\frac{1}{n}\sum_{k=1}^{n}\beta^k X_k^2 + C\sup_{\theta\in K}\frac{1}{\omega}\frac{1}{n}\sum_{k=1}^{n}\beta^k,
\end{aligned}$$

where $\sup_{\theta\in K}\beta^k X_k^2 \xrightarrow{\text{a.s.}} 0$ since $\sup_{\theta\in K}\beta < 1$ and X_k^2 has a finite moment of order greater than 0. Using Kronecker's lemma, we see that

$$\sup_{\theta\in K}|l_n(\theta) - \tilde{l}_n(\theta)| \xrightarrow{\text{a.s.}} 0.$$

Finally, for every $\theta \in K$ and $r > 0$ let $B_r(\theta)$ be the open sphere with center θ and radius r. Then we have

$$\begin{aligned}
\liminf_{n\to\infty}\inf_{\hat{\theta}\in B_r(\theta)\cap K}\tilde{l}_n(\hat{\theta}) &\geq \liminf_{n\to\infty}\inf_{\hat{\theta}\in B_r(\theta)\cap K} l_n(\hat{\theta}) - \limsup_{n\to\infty}\sup_{\hat{\theta}\in K}|l_n(\hat{\theta}) - \tilde{l}_n(\hat{\theta})| \\
&\geq \liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\inf_{\hat{\theta}\in B_r(\theta)\cap K}\Big(\log\sigma_k^2(\hat{\theta}) + \frac{e_k^2\sigma_k^2(\theta_0)}{\sigma_k^2(\hat{\theta})}\Big) \\
&= \mathbb{E}_{\theta_0}\Big[\inf_{\hat{\theta}\in B_r(\theta)\cap K}\Big(\log\sigma_k^2(\hat{\theta}) + \frac{e_k^2\sigma_k^2(\theta_0)}{\sigma_k^2(\hat{\theta})}\Big)\Big]
\end{aligned}$$

by Birkhoff's ergodic theorem, using the fact that X_t is ergodic by condition (iii) . By Beppo-Levi's theorem (monotone convergence), we have

$$\mathbb{E}_{\theta_0}\Big[\inf_{\hat{\theta}\in B_r(\theta)\cap K}\Big(\log\sigma_1^2(\hat{\theta}) + \frac{e_k^2\sigma_1^2(\theta_0)}{\sigma_1^2(\hat{\theta})}\Big)\Big] \to \mathbb{E}_{\theta_0}\Big[\log\sigma_1(\theta_0) + \frac{e_1^2\sigma_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\Big]$$

as $r \to 0$. This means that for every $\theta \neq \theta_0 \in K$, there is an open neighborhood $U(\theta)$ of θ in K such that

$$\liminf_{n\to\infty} \inf_{\hat{\theta}\in U(\theta)} \tilde{l}_n(\hat{\theta}) > \mathbb{E}_{\theta_0}\Big[\log\sigma_1(\theta_0) + \frac{e_1^2\sigma_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\Big].$$

Since K is compact, by the Heine-Borel theorem K is covered by a finite set of these open neighborhoods $U(\theta_1), ..., U(\theta_k)$ and $B_r(\theta)$ for any $r > 0$. Since

$$\limsup_{n\to\infty} \inf_{\hat{\theta}\in B_r(\theta)} \tilde{l}_n(\hat{\theta}) \leq \lim_{n\to\infty} \tilde{l}_n(\theta_0) = \lim_{n\to\infty} l_n(\theta_0) = \mathbb{E}_{\theta_0}[\log\sigma_1(\theta_0) + \frac{e_1^2\sigma_1^2(\theta_0)}{\sigma_1^2(\theta_0)}],$$

for any $r > 0$, $\hat{\theta}_n$ must lie in $B_r(\theta_0)$ for large enough n, and the theorem is proved. □

Theorem 15. *(QMLE - asymptotic distribution)*
Let X_t be a GARCH-time series with parameters $\theta_0 = (\omega_0, \alpha_0, \beta_0)$. Define $M = \mathbb{R}\times M_\alpha \times M_\beta \subset \mathbb{R}^3$, where $M_\alpha = [0,\infty)$ if $\alpha_0 = 0$ and $\mathbb{R}$ otherwise, and M_β analogously. Under the conditions
(i) θ_0 lies in a compact set $K \subset (0,\infty)\times[0,\infty)\times[0,1)$;
(ii) e_t is not a.s. constant;
(iii) $\beta_0 < 1$, $\alpha_0 \neq \beta_0$, $\alpha_0 + \beta_0 < 1$, $\omega_0 > 0$;
(iv) $\kappa := \mathbb{E}[e_t^4] < \infty$;
(v) $\mathbb{E}[X_t^6] < \infty$;
we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \xi$, where

$$\xi = \arg\inf_{t\in M}(t - Z)^T J(\theta_0)(t - Z), \quad Z \sim \mathcal{N}(0, (\eta - 1)J(\theta_0)^{-1})$$

and

$$J(\theta_0) = \mathbb{E}_{\theta_0}\Big[\frac{1}{\sigma_1^4(\theta_0)}\frac{\partial^2}{\partial\theta^2}\sigma_1^2(\theta_0)\Big]$$

is a positive definite matrix.

The proof of this theorem can be found in [4], Chapter 8. However, it is useful to see the following equality. Let $f_t(\theta) = \log\sigma_t^2(\theta) + \frac{X_t^2}{\sigma_t^2(\theta)}$, so that $l_n(\theta) = \frac{1}{n}\sum_{k=1}^n f_k(\theta)$. Since $e_t^2 = \frac{X_t^2}{\sigma_t^2(\theta_0)}$ is independent of $\sigma_t^2(\theta_0)$ and all derivatives of σ_t^2 at θ_0, we have

$$\mathbb{E}_{\theta_0}\Big[\frac{\partial f_t(\theta_0)}{\partial\theta}\Big] = \mathbb{E}_{\theta_0}[1 - e_t^2]\cdot\mathbb{E}_{\theta_0}\Big[\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\Big] = 0,$$

where the derivatives are understood as one-sided in case θ_0 lies on the boundary of $(0,\infty)\times[0,\infty)\times[0,1)$. Additionally,

$$\begin{aligned}\mathbb{E}_{\theta_0}\Big[\frac{\partial^2}{\partial\theta^2}{}_\theta f_t(\theta_0)\Big] &= \mathbb{E}_{\theta_0}[1 - e_t^2]\cdot\mathbb{E}_{\theta_0}\Big[\frac{1}{\sigma_t^2}\frac{\partial^2}{\partial\theta^2}\sigma_t^2\Big] + \mathbb{E}_{\theta_0}[2e_t^2 - 1]\cdot\mathbb{E}_{\theta_0}\Big[\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial^2}{\partial\theta^2}{}_\theta\sigma_t^2(\theta_0)\Big]\\ &= J(\theta_0),\end{aligned}$$

so we may equivalently write

$$J(\theta_0) = \mathbb{E}_{\theta_0}\Big[\frac{\partial^2}{\partial\theta^2}{}_\theta f_t(\theta_0)\Big].$$

Note that when θ_0 lies in the interior of K, $M = \mathbb{R}^3$ and $\xi = Z$ is normally distributed.

5 Tests

We can use the QML estimator from section 4 to construct tests for the significance of the coefficients in the model; for example, to test between the hypothesis and alternative

$$H_0 : \alpha_0 = 0 \qquad H_1 : \alpha_0 > 0.$$

However, for technical reasons, it is difficult to test $\alpha_0 = \beta_0 = 0$ simultaneously.

The LM test

Let $l_n(\theta) := \sum_{k=1}^n \log \sigma_k^2(\theta) + \frac{X_k^2}{\sigma_k^2(\theta)}$ be the QMLE as in section 4 (removing the factor $\frac{1}{n}$ for convenience in the below equations). We construct $\tilde{\theta}_n$ as the QML estimator under the constraint that $\alpha = 0$. First, we assume that the innovations e_t actually are standard normally distributed, so that the QML estimator l_n is the true maximum likelihood estimator. Under the assumptions that θ_0 is identifiable and X_t has finite moments of up to 6th order, one has the convergence

$$\frac{1}{\sqrt{n}} \frac{\partial l_n(\theta_0)}{\partial \theta} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I)$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1})$$

where $I = \lim_{n \to \infty} \frac{1}{n} \frac{\partial^2 l_n(\theta_0)}{\partial \theta^2}$ converges almost surely. The constrained estimator is then derived through the method of Lagrangian multipliers: define $\Lambda(\theta, \lambda) = l_n(\theta) - \lambda\alpha$. Then $(\tilde{\theta}_n, \tilde{\lambda}) = \arg\sup \Lambda(\theta, \lambda)$. We have $\nabla\Lambda(\tilde{\theta}_n, \tilde{\lambda}) = 0$ and therefore

$$\tilde{\alpha} = 0, \ \ \frac{\partial}{\partial \theta} l_n(\tilde{\theta}) = \tilde{\lambda} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Let $R = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$. Then we have

$$\sqrt{n} R(\hat{\theta}_n - \tilde{\theta}_n) = \sqrt{n} R(\hat{\theta}_n - \tilde{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, RI^{-1}R^T).$$

We have

$$0 = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l_n^(\hat{\theta}_n) = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l_n(\theta_0) - I\sqrt{n}(\hat{\theta}_n - \theta_0) + o(1)$$

and

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l_n(\tilde{\theta}_n) = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l_n(\theta_0) - I\sqrt{n}(\tilde{\theta}_n - \theta_0) + o(1),$$

so that

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = I^{-1} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l_n^(\tilde{\theta}_n) = I^{-1} \frac{1}{\sqrt{n}} R^T \tilde{\lambda}$$

and therefore

$$\frac{\tilde{\lambda}}{\sqrt{n}} = (RI^{-1}R^T)^{-1} \sqrt{n} R(\tilde{\theta}_n - \theta_0) + o(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, (RI^{-1}R^T)^{-1}).$$

This means that
$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}l_n(\tilde{\theta}_n) = R^T\frac{\tilde{\lambda}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, R^T(RI^{-1}R^T)^{-1}R).$$
Let $\hat{I}_n = \frac{1}{n}\frac{\partial^2}{\partial\theta^2}l_n(\tilde{\theta}_n)$, so that $\hat{I}_n$ is a strongly consistent estimator for I. Then the LM statistic
$$LM(n) = \frac{1}{n}(\frac{\partial}{\partial\theta}l_n(\tilde{\theta}_n))^T\hat{I}_n^{-1}\frac{\partial}{\partial\theta}l_n(\tilde{\theta}_n)$$
is asymptotically χ^2-distributed with Rank$(R) = 1$ degree of freedom.
In the case where e_t are not normally distributed, one has in general
$$J := \lim_{n\to\infty}\mathrm{Var}_{\theta_0}[\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}l_n(\theta)] \neq I$$
and it can be shown that the following limits hold:
$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}l_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, J)$$
and
$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}JI^{-1}),$$
and therefore the LM statistic
$$LM_n = \frac{1}{n}(\frac{\partial l_n(\tilde{\theta}_n)}{\partial\theta})^T\hat{I}_n^{-1}R^T(R\hat{I}_n^{-1}\hat{J}_n\hat{I}_n^{-1}R^T)^{-1}R\hat{I}_n^{-1}\frac{\partial l_n(\tilde{\theta}_n)}{\partial\theta}.$$
It can be shown that this statistic is also asymptotically χ^2-distributed with Rank$(R) = 1$ degree of freedom.

First, we show the problem that appears when attempting to test both $\alpha_0 = \beta_0 = 0$. Applying the previous results to this case, we have $\theta_0 = (\omega_0, 0, 0)$ under hypothesis H_0 and therefore
$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial l_n(\theta_0)}{\partial\theta} &= \frac{1}{2\sqrt{n}}\sum_{k=1}^{n}\frac{X_k^2 - \sigma_k^2(\theta_0)}{\sigma_k^4(\theta_0)}\frac{\partial\sigma_k^2(\theta_0)}{\partial\theta} \\ &= \frac{1}{2\sqrt{n}}\sum_{k=1}^{n}\frac{e_k^2 - 1}{\omega_0}\begin{pmatrix}1\\ X_{k-1}^2\\ \omega_0\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, J),\end{aligned}$$
where, letting $\kappa = \mathbb{E}[e_k^4]$,
$$J = \frac{1}{4\omega_0^2}\mathbb{E}\Big[(e_k^2 - 1)^2\begin{pmatrix}1\\ X_{k-1}^2\\ \omega_0\end{pmatrix}\begin{pmatrix}1 & X_{k-1}^2 & \omega_0\end{pmatrix}\Big] = \frac{\kappa - 1}{4\omega_0^2}\begin{pmatrix}1 & \omega_0 & \omega_0\\ \omega_0 & \kappa\omega_0^2 & \omega_0^2\\ \omega_0 & \omega_0^2 & \omega_0^2\end{pmatrix}.$$
Since J is singular, no LM test can immediately be constructed.

Testing $\alpha_0 = 0$, we have $\theta_0 = (\omega_0, 0, \beta_0)$ under H_0 and therefore $\mathbb{E}[\sigma_k^2(\theta_0)] = \frac{\omega_0}{1-\beta_0}$. We then have

$$\frac{1}{\sqrt{n}}\frac{\partial l_n(\theta_0)}{\partial\theta} = \frac{1}{2\sqrt{n}}\sum_{k=1}^{n}\frac{(e_k^2-1)}{\sigma_k^2(\theta_0)}\begin{pmatrix}1\\ X_{k-1}^2\\ \sigma_{k-1}^2(\theta_0)\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,J),$$

where $\sigma_k^2(\theta_0) = \omega_0\frac{1-\beta_0^k}{1-\beta_0}$, so $\frac{\sigma_{k-1}^2(\theta_0)}{\sigma_k^2(\theta_0)} = \frac{1-\beta_0^{k-1}}{1-\beta_0^k} =: \gamma_k$ and therefore

$$J = \lim_{n\to\infty}\mathrm{Var}[\frac{1}{\sqrt{n}}\frac{\partial l_n(\theta_0)}{\partial\theta}] = \lim_{n\to\infty}\frac{\kappa-1}{4n}\sum_{k=1}^{n}\begin{pmatrix}1 & \gamma_k & \gamma_k\\ \gamma_k & \kappa\gamma_k\sigma_{k-1}^2(\theta_0) & \gamma_k\sigma_{k-1}^2(\theta_0)\\ \gamma_k & \gamma_k\sigma_{k-1}^2(\theta_0) & \gamma_k\sigma_{k-1}^2(\theta_0)\end{pmatrix},$$

and we have with $g_n = \sum_{k=1}^n \gamma_k$, $s_n = \sum_{k=1}^n \gamma_k\sigma_{k-1}^2(\theta)$,

$$J^{-1} = \lim_{n\to\infty}\frac{4n}{\kappa-1}\begin{pmatrix}\frac{s_n}{s_n-g_n^2} & 0 & \frac{-g_n}{s_n-g_n^2}\\ 0 & \frac{1}{s_n(\kappa-1)} & \frac{-1}{s_n(\kappa-1)}\\ \frac{-g_n}{s_n-g_n^2} & \frac{-1}{s_n(\kappa-1)} & \frac{s_n\kappa-g_n^2}{s_n(s_n-g_n^2)(\kappa-1)}\end{pmatrix}$$

where the limit in each matrix entry exists. To see this, consider for example that with $0<\beta_0<1$,

$$g_n = \sum_{k=1}^{n}\frac{1-\beta_0^{k-1}}{1-\beta_0^k} = \frac{1-\beta_0}{\beta_0\log\beta_0}\psi_{\beta_0}(n) + n + O(1),$$

where $\psi_{\beta_0}(n)$ is the β_0-digamma function and

$$\lim_{z\to\infty}\frac{\psi_{\beta_0}(z)}{z} = \log\beta_0\sum_{n=0}^{\infty}\lim_{z\to\infty}\frac{\beta_0^{n+z}}{z(1-\beta_0^{n+z})} = 0$$

by uniform convergence, so that $\lim_{n\to\infty}\frac{n}{g_n} = 1$ Additionally, we have

$$\begin{aligned} s_n &= \sum_{k=1}^{n}\frac{\omega_0(1-\beta_0^{k-1})^2}{(1-\beta_0)(1-\beta_0^k)}\\ &= \frac{\beta_0^n-\beta_0^2(n-1)+\beta_0(n-2)}{(\beta_0-1)^2\beta_0} + \frac{(\beta-1)\psi_{\beta_0}(n+1)}{\beta_0^2\log\beta_0} + O(1), \end{aligned}$$

and $\frac{n}{s_n}\to 1-\beta_0$ follows. Thus the limit J is indeed invertible, with inverse

$$J^{-1} = \frac{1}{\kappa-1}\begin{pmatrix}\frac{1}{1-\beta_0} & 0 & -1\\ 0 & \frac{1-\beta_0}{\kappa-1} & -\frac{1-\beta_0}{\kappa-1}\\ -1 & -\frac{1-\beta_0}{\kappa-1} & \frac{1-\beta_0}{\kappa-1}\end{pmatrix}.$$

On the other hand,

$$\begin{aligned}
I_n &:= \frac{1}{n}\frac{\partial^2}{\partial\theta^2} l_n(\theta_0) \\
&= \frac{1}{n}\sum_{k=1}^{n} \frac{\partial}{\partial\theta}\Big[\frac{X_k^2-\sigma_k^2(\theta_0)}{2\sigma_k^4(\theta_0)}\frac{\partial\sigma_k^2(\theta_0)}{\partial\theta}\Big] \\
&= \frac{1}{n}\sum_{k=1}^{n} \frac{2X_k^2-\sigma_k^2(\theta_0)}{2\sigma_k^6(\theta_0)}\frac{\partial\sigma_k^2(\theta_0)}{\partial\theta}(\frac{\partial\sigma_k^2(\theta_0)}{\partial\theta})^T \\
&= \frac{1}{n}\sum_{k=1}^{n} \frac{2e_k^2-1}{2\sigma_k^4(\theta_0)}\begin{pmatrix} 1 & X_{k-1}^2 & \sigma_{k-1}^2(\theta_0) \\ X_{k-1}^2 & X_{k-1}^4 & X_{k-1}^2\sigma_{k-1}^2(\theta_0) \\ \sigma_{k-1}^2(\theta_0) & X_{k-1}^2\sigma_{k-1}^2(\theta_0) & \sigma_{k-1}^4(\theta_0)\end{pmatrix},
\end{aligned}$$

which converges almost surely to $\frac{2}{\kappa-1}J$ due to the factor of $\frac{1}{n}$. Defining $\hat{J}_n = \frac{\kappa-1}{2}\hat{I}_n$ where

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n}\frac{2X_k^2-\sigma_k^2(\tilde{\theta}_n)}{2\sigma_k^6(\tilde{\theta}_n)}\begin{pmatrix}1 \\ X_{k-1}^2 \\ \sigma_{k-1}^2(\tilde{\theta}_n)\end{pmatrix}\begin{pmatrix}1 & X_{k-1}^2 & \sigma_{k-1}^2(\tilde{\theta}_n)\end{pmatrix},$$

we have

$$R\hat{I}_n^{-1}\hat{J}_n\hat{I}_n^{-1}R^T = \frac{\kappa-1}{2}R\hat{I}_n^{-1}R^T = \frac{2n}{s_n(\kappa-1)},$$

and considering that $\frac{\partial}{\partial\theta}l_n(\theta_0) = \frac{1}{2}\sum_{k=1}^{n}\frac{X_k^2-\sigma_k^2(\tilde{\theta}_n)}{\sigma_k^4(\tilde{\theta}_n)}\begin{pmatrix}1 \\ X_{k-1}^2 \\ \sigma_{k-1}^2(\tilde{\theta}_n)\end{pmatrix}$, the LM statistic can be calculated.

Testing for covariance stationarity

The GARCH(1,1)-process is covariance stationary if and only if $\alpha_0+\beta_0<1$. The problem of determining whether the process has a stationary solution or not may therefore be addressed with a parametric test. Let

$$H_0: \alpha_0+\beta_0 \geq 1 \qquad H_1: \alpha_0+\beta_0 < 1.$$

Here, θ lies in the interior of a compact parameter space K, so that the QML estimator is asymptotically normal:

$$\sqrt{n}(\theta_0-\hat{\theta}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,(\kappa-1)J^{-1})$$

and with $R = \begin{pmatrix}0 & 1 & 1\end{pmatrix}$, if $R\theta_0 = 1$, we have

$$\sqrt{n}(R\hat{\theta}_n - 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0,(\kappa-1)RJ^{-1}R^T).$$

This leads to the asymptotic normality of the Wald statistic:

$$W_n = \frac{\sqrt{n}(\hat{\alpha}_n+\hat{\beta}_n-1)}{\sqrt{(\kappa-1)R\hat{J}_n^{-1}R^T}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

and H_0 is rejected if $W_n < u_\alpha$, where u_α is the α-quantile of the standard normal distribution.

6 Variants of the GARCH(1,1) model

While the standard GARCH(1,1) and related GARCH(p,q) models are useful tools in econometrics, they are unable to describe certain aspects often found in financial data. An important weakness is their inability to react differently to positive and negative innovations - the conditional variance considers only the squares of the innovations. However, many datasets display a leverage effect, where negative returns correspond to higher increases in volatility than positive returns. Another problem is the lack of clarity with regard to stationarity and persistence, where shocks may persist in one norm but not in another in the GARCH model. The existence of almost-surely stationary GARCH(1,1)-processes with infinite variance at every time t is inconvenient.

A notable variant of the GARCH model which addresses these problems is the Exponential ARCH (EARCH) model due to Nelson (1989). This has the additional advantage of greater flexibility in the parameters by imposing the autoregressive relationship on $\log \sigma_t^2$, which can take negative values. The general form of the EARCH(1) model is

$$\log \sigma_t^2 = \omega + \beta(|e_{t-1}| - \mathbb{E}[|e_{t-1}|]) + \gamma e_{t-1}.$$

It can also be shown that the conditions for stationarity, unlike the GARCH(1,1) model, are the same for both wide-sense (almost sure) and covariance stationarity. A necessary and sufficient condition for this is $\beta < 1$. However, the asymptotic properties of QML estimation for EARCH models are not as well known as the GARCH case.
Another possible extension of the GARCH(1,1) model to allow for asymmetry is the QGARCH(1,1) model of Sentana (1995):

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma e_{t-1},$$

where appropriate restrictions are necessary to ensure that σ_t remains positive. These are given by $\omega, \alpha, \beta > 0$ and $|\gamma| \leq 2\sqrt{\alpha\omega}$. Many of the properties of the GARCH(1,1) model carry over immediately to QGARCH. For example, the QGARCH model has unconditional variance $\frac{\omega}{1-\alpha-\beta}$ if $\alpha + \beta < 1$ and undefined otherwise, and $\alpha + \beta < 1$ is also necessary for covariance stationarity.

The AGARCH(1,1) (asymmetric GARCH) model developed by Engle and Ng (1993) is another approach to allowing the GARCH model to react asymmetrically. It is defined by

$$X_t = e_t \sigma_t, \quad \sigma_t^2 = \omega + \alpha(X_{t-1} + \gamma)^2 + \beta \sigma_{t-1}^2$$

where γ is the noncentrality parameter.
Another interesting model is the TGARCH(1,1) (threshold GARCH) model developed by Jean-Michel Zakoian. Here, the autoregressive specification is given for the conditional standard deviation instead of the variance:

$$X_t = e_t \sigma_t$$

$$\sigma_t = \omega + \alpha^+ X_{t-1}^+ + \alpha^- X_{t-1}^- + \beta \sigma_{t-1}$$

where $X_t^+ = \max\{0, X_t\}$ is the positive part of X_t and $X_t^- = \min\{0, X_t\}$ the negative. This is another model developed to account for asymmetric reactions to shocks, which has the advantage of being closer to the original GARCH formulation but also requires some non-negativity assumptions for the parameters.

7 GARCH(1,1) in continuous time

The diffusion limit

Heuristically, we will consider in this section whether increasingly frequent observations of the time series will lead to a better model. Although this may seem intuitive, it is easy to see that this should not be the case in general: even in a non-stochastic setting, increasingly fine interpolation can lead to large errors without the assumption of differentiability. However, we will show that highly frequent observations will lead to a more accurate model in certain cases, where the precise meaning of "leading to a better model" will be the weak convergence of processes.

Nelson (1990) has studied the relationship between GARCH(1,1) and similar models, and stochastic differential equations in continuous time. His main result is that in a sequence of GARCH(1,1) models to increasingly small time intervals, under certain assumptions on the parameters, the conditional variance process converges in distribution to a stochastic differential equation with an inverse-gamma stationary distribution. This means that for sufficiently short time intervals the GARCH log returns can be approximately modelled with a Student's t distribution. Since a GARCH(1,1) process is Markovian, it is enough to consider convergence of Markov chains.
The following theorem gives conditions under which a sequence of Markov processes $X_{hk}^{(h)}$ $(k \in \mathbb{N})$ converges weakly to an Ito process as h tends to zero. Let D be the Skorokhod space of càdlàg mappings from $[0, \infty)$ into $\mathbb{R}^n$, endowed with the Skorokhod metric. For every $h > 0$ let $\mathcal{F}_{kh}^{(h)}$ be the σ-algebra generated by $X_0^{(h)}, ..., X_{kh}^{(h)}$, $\mathbb{P}^{(h)}$ be a probability measure on $\mathcal{B}^n$, the Borel σ-algebra over $\mathbb{R}^n$, and for every $h > 0$ and $k \in \mathbb{N}_0$ let $K_{kh}^{(h)}$ be a transition function on $\mathbb{R}^n$; that means:
(i) for every $x \in \mathbb{R}^n$, $K_{kh}^{(h)}(x, .)$ is a probability measure on $\mathcal{B}^n$, and
(ii) for every $A \in \mathcal{B}^n$, $K_{kh}^{(h)}(., A)$ is a measurable function.
and such that we have $\mathbb{P}^{(h)}(X_{(k+1)h}^{(h)} \in A \mid \mathcal{F}_{kh}^{(h)}) = K_{kh}^{(h)}(X_{kh}^{(h)}, A)$ $\mathbb{P}^{(h)}$-a.s. $\forall A \in \mathcal{B}^n$, as well as $\mathbb{P}^{(h)}(X_t^{(h)} = X_{kh}^{(h)}, kh \leq t < (k+1)h) = 1$.
We define $X_t^{(h)}$ as the extension of $X_{kh}^{(h)}$ into continuous time; that is, $X_t^{(h)}$ is a step function with discontinuities at kh for all k. For $h > 0$ and $\epsilon > 0$ define

$$a_h(x,t) = \frac{1}{h} \int\limits_{\mathbb{R}^n} (y-x)(y-x)^T K_{h,\lfloor t/h \rfloor}^{(h)}(x, dy)$$

$$b_h(x,t) = \frac{1}{h} \int\limits_{\mathbb{R}^n} (y-x) K_{h,\lfloor t/h \rfloor}^{(h)}(x, dy)$$

and for each $i = 1, ..., n$ define

$$c_{h,i,\epsilon}(x,t) = \frac{1}{h} \int\limits_{\mathbb{R}^n} |(y-x)_i|^{2+\epsilon} K_{h,\lfloor t/h \rfloor}(x, dy)$$

where a_h and b_h are finite if $c_{h,i,\epsilon}$ is finite for all i with some $\epsilon > 0$.

Theorem 16. *Let the following assumptions be fulfilled:*
(i): There is an $\epsilon > 0$ so that for every $R > 0$, $T > 0$ and $i = 1, ..., n$:

$$\lim_{h \to 0} \sup_{\|x\| \leq R,\ 0 \leq t \leq T} c_{h,i,\epsilon}(x,t) = 0$$

and continuous functions $a : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^{n\times n}$, $b : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n$ *so that*

$$\lim_{h\to 0} \sup_{\|x\|\le R,\, 0\le t\le T} \|a_h(x,t) - a(x,t)\|_F = 0$$

$$\lim_{h\to 0} \sup_{\|x\|\le R,\, 0\le t\le T} \|b_h(x,t) - b(x,t)\| = 0$$

where $\|-\|_F$ *denotes the Frobenius norm.*
(ii): There exists a continuous function $\sigma : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^{n\times n}$ *so that*

$$a(x,t) = \sigma(x,t)\sigma(x,t)^T \;\; \forall x \in \mathbb{R}^n,\, t \ge 0.$$

(iii): There exists a random variable X_0 *with distribution* $\mathbb{P}_0$ *on* $(\mathbb{R}^n, \mathcal{B}^n)$ *so that* $X_0^{(h)} \xrightarrow{\mathcal{D}} X_0$.
(iv): a, b, *and* $\mathbb{P}_0$ *uniquely determine the distribution of a diffusion process* X_t *with starting distribution* $\mathbb{P}_0$, *diffusion matrix* $a(x,t)$ *and drift* $b(x,t)$.
Then $X_t^{(h)} \Rightarrow X_t$ *where* X_t *is given by the stochastic differential equation*

$$X_t = X_0 + \int_0^t b(X_s, s)ds + \int_0^t \sigma(X_s, s)dB_{n,s}$$

where $B_{n,t}$ *is an* n*-dimensional Brownian motion independent of* X_0. X_t *does not depend on the choice of matrix square root* σ. *In addition, for every* $T > 0$, *we have*

$$\mathbb{P}(\sup_{0\le t\le T} \|X_t\| < \infty) = 1.$$

This theorem can be applied to the GARCH(1,1) model with normally distributed innovations, allowing the parameters α, β, ω to depend on h and making the innovations proportional to h. Let $Y_t = \sum_{s<t} X_t$. Then one has the difference equations

$$Y_{hk}^{(h)} = Y_{h(k-1)}^{(h)} + \sigma_{hk}^{(h)} e_{hk}^{(h)}$$

$$(\sigma_{hk}^{(h)})^2 = \omega_h + (\sigma_{h(k-1)}^{(h)})^2(\beta_h + \frac{1}{h}\alpha_h(e_{h(k-1)}^{(h)})^2)$$

with i.i.d. random variables $e_{hk}^{(h)} \sim \mathcal{N}(0,h)$ $(k = 0,1,...)$. For $B \in \mathcal{B}^2$ let $\nu_h(B) = \mathbb{P}((Y_0^{(h)}, \sigma_0^2) \in B)$ where we may assume that ν_h satisfies condition (iii) of the theorem.
Let $\mathcal{F}_{hk}$ be the σ-algebra generated by $Y_0^{(h)}, ..., Y_{h(k-1)}^{(h)}, \sigma_0^{(h)}, ..., \sigma_{hk}^{(h)}$. Then

$$\mathbb{E}[h^{-1}(Y_{hk}^{(h)} - Y_{h(k-1)}^{(h)}) \mid \mathcal{F}_{hk}] = \sigma_{hk}^{(h)}\mathbb{E}[e_{hk}^{(h)}] = 0$$

and

$$\mathbb{E}[h^{-1}((\sigma_{h(k+1)}^{(h)})^2 - (\sigma_{hk}^{(h)})^2) \mid \mathcal{F}_{hk}] = \frac{1}{h}(\omega_h + (\beta_h + \alpha_h - 1)(\sigma_{hk}^{(h)})^2),$$

so the limit condition in (i) can be fulfilled only if the limits

$$\lim_{h\to 0} \frac{\omega_h}{h} = \omega \ge 0 \qquad \lim_{h\to 0} \frac{1-\alpha_h-\beta_h}{h} = \theta$$

exist. In addition, assuming $\lim_{h\to 0} \frac{\alpha_h^2}{h} = \frac{\alpha^2}{2}$ exists and is finite,

$$
\begin{aligned}
&\mathbb{E}[h^{-1}\{(\sigma_{h(k+1)}^{(h)})^2 - (\sigma_{hk}^{(h)})^2\}^2 \mid \mathcal{F}_{hk}] \\
&= h^{-1}\Big(\omega_h^2 - 2(1-\alpha_h-\beta_h)(\sigma_{hk}^{(h)})^2 + (1-\alpha_h-\beta_h)^2(\sigma_{hk}^{(h)})^4 + 2\alpha_h^2(\sigma_{hk}^{(h)})^4\Big) \\
&= \alpha^2(\sigma_{hk}^{(h)})^4 + o(1),
\end{aligned}
$$

and

$$\mathbb{E}[h^{-1}(Y_{hk}^{(h)} - Y_{h(k-1)}^{(h)})^2 \mid \mathcal{F}_{hk}] = (\sigma_{hk}^{(h)})^2,$$

and

$$\mathbb{E}[h^{-1}(Y_{hk}^{(h)} - Y_{h(k-1)}^{(h)})((\sigma_{h(k+1)}^{(h)})^2 - (\sigma_{hk}^{(h)})^2) \mid \mathcal{F}_{hk}] = 0,$$

for a term $o(1)$ which tends to zero uniformly on compacta. Since

$$\mathbb{E}[h^{-1}(Y_{hk}^{(h)} - Y_{h(k-1)}^{(h)})^4 \mid \mathcal{F}_{hk}] = 3h^{-1}(\sigma_{hk}^{(h)})^4 \to 0$$

and

$$\mathbb{E}[h^{-1}\{(\sigma_{h(k+1)}^{(h)})^2 - (\sigma_{hk}^{(h)})^2\}^4 \mid \mathcal{F}_{hk}] \to 0,$$

we have with $\epsilon = 2$,

$$c_{h,i,\epsilon}(x,t) = \mathbb{E}[h^{-1}(Y_{hk}^{(h)} - Y_{h(k-1)}^{(h)})^4 \mid \mathcal{F}_{hk}] \to 0;$$

and setting

$$
\begin{aligned}
a(Y,\sigma) &= \begin{pmatrix} \sigma^2 & 0 \\ 0 & \alpha^2\sigma^4 \end{pmatrix} \\
b(Y,\sigma) &= \begin{pmatrix} 0 \\ \omega - \theta\sigma^2 \end{pmatrix},
\end{aligned}
$$

assuming the limit conditions on $\alpha_h, \beta_h, \omega_h$, condition (i) is satisfied. Letting $\tau = \begin{pmatrix} \sigma & 0 \\ 0 & \alpha\sigma^2 \end{pmatrix}$ fulfills condition (ii). Under the assumption of distributional uniqueness, by Theorem 7.1 we have the diffusion limit

$$
\begin{aligned}
dY_t &= \sigma_t dB_t \\
d\sigma_t^2 &= (\omega - \theta\sigma_t^2)dt + \alpha\sigma_t^2 dW_t,
\end{aligned}
$$

for independent Brownian motions B_t, W_t. Since the drift and variation terms are Lipschitz-continuous, there exists a unique strong solution and therefore a unique distributional solution. The corresponding Fokker-Planck equation for σ_t^2 is

$$\frac{\partial}{\partial t} f(x,t) = -\frac{\partial}{\partial x}[(\omega - \theta x)f] + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\alpha^2 x^2 f)$$

where $f(x,t)$ is the probability density of σ_t^2 given σ_0^2. A stationary distribution with p.d.f. g for σ_t^2 must satisfy $g(x) = \lim_{t\to\infty} f(x,t)$ and therefore

$$\frac{d}{dx}\alpha^2 x^2 g(x) = 2(\omega - \theta x)g(x),$$

so

$$g'(x) = \frac{2}{\alpha^2 x^2}(\omega - \theta x - \alpha^2 x) g(x)$$

with solution

$$g(x) = C x^{-2\theta/\alpha^2 - 2} \exp(\frac{-2\omega}{\alpha^2 x})$$

for a normalizing constant C. This is the p.d.f. of an inverse gamma distribution up to the constant factor and therefore

$$C = \frac{(2\omega/\alpha^2)^{(2\theta/\alpha^2+1)}}{\Gamma(2\theta/\alpha^2+1)},$$

so under the assumption that $2\theta/\alpha^2 + 1 > 0$,

$$\sigma_t^2 \xrightarrow{\mathcal{D}} \Gamma^{-1}(2\theta/\alpha^2+1,\ 2\omega/\alpha^2).$$

Theorem 17. *Assume that $2\theta/\alpha^2 + 1 > 0$ with θ, α defined as above, and that*

$$(\sigma_0^{(h)})^2 \xrightarrow{\mathcal{D}} \Gamma^{-1}(2\theta/\alpha^2+1,\ 2\omega/\alpha^2)\ \ (h \to 0).$$

Then

$$(\sigma_{hk}^{(h)})^2 \xrightarrow{\mathcal{D}} \Gamma^{-1}(2\theta/\alpha^2+1,\ 2\omega/\alpha^2)$$

and

$$\sqrt{\frac{(2\theta+\alpha^2)/2\omega}{h}}\, e_{hk}^{(h)} \sigma_{hk}^{(h)} \xrightarrow{\mathcal{D}} t(2+4\theta/\alpha^2)$$

as $(h \to 0)$ and kh remains constant.
Here, $t(2+4\theta/\alpha^2)$ is the Student's t-distribution with $2+4\theta/\alpha^2$ degrees of freedom.

Proof. The first statement follows immediately from the above considerations since $\sigma_{hk}^{(h)}$ converges in distribution to σ_t for $h \to 0$ and $k = \frac{t}{h}$. For the second statement, consider that $\frac{1}{h} e_{hk}^{(h)}$ is standard normally distributed and independent of $\sigma_{hk}^{(h)}$ for every h, k. Therefore we assume without loss of generality that the process is stationary, that is,

$$(\sigma_{hk}^{(h)})^2 \sim \Gamma^{-1}(\lambda, \mu)$$

for every h, k, defining $\lambda = 2\theta/\alpha^2 + 1,\ \mu = 2\omega/\alpha^2$. Since the density of $(\sigma_{hk}^{(h)})^2$ is

$$g(x) = C x^{-\lambda-1} \exp(\frac{-\mu}{x})$$

with C defined as earlier, the density of $\sigma_{hk}^{(h)}$ is

$$f_\sigma(x) = \frac{g(x^2)}{|\frac{d\sqrt{x}}{dx}|} = 2C x^{-2\lambda-1} \exp(\frac{-\mu}{x^2}),$$

so that the density of the product $\sqrt{\frac{1}{h}}e^{(h)}_{hk}\sigma^{(h)}_{hk}$ is

$$
\begin{aligned}
f(y) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{1}{x}e^{\frac{-x^2}{2}} f_\sigma(\frac{y}{x})dx \\
&= \frac{2C}{\sqrt{2\pi}}y^{-2\lambda-1}\int_0^\infty x^{2\lambda}e^{-x^2(\frac{1}{2}+\frac{\mu}{y^2})}dx \\
&= \frac{2C}{\sqrt{2\pi}}(y^2)^{-\lambda-\frac{1}{2}}\frac{1}{2}(\frac{\mu}{y^2}+\frac{1}{2})^{-\lambda-\frac{1}{2}}\Gamma(\lambda+\frac{1}{2}) \\
&= \mu^\lambda(\mu+\frac{y^2}{2})^{-\lambda-\frac{1}{2}}\frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{2\pi}\Gamma(\lambda)},
\end{aligned}
$$

so that the distribution of the product $\sqrt{\frac{(2\theta+\alpha^2)/2\omega}{h}}e^{(h)}_{hk}\sigma^{(h)}_{hk} = \sqrt{\frac{\lambda}{\mu h}}e^{(h)}_{hk}\sigma^{(h)}_{hk}$ is

$$
\begin{aligned}
f(y) &= \sqrt{\frac{\mu}{\lambda}}\mu^\lambda(\mu+\frac{y^2\mu}{2\lambda})^{-\lambda-\frac{1}{2}}\frac{\Gamma(\lambda+\frac{1}{2})}{\sqrt{2\pi}\Gamma(\lambda)} \\
&= (1+\frac{y^2}{2\lambda})^{\frac{-2\lambda-1}{2}}\frac{\Gamma(\frac{2\lambda+1}{2})}{\sqrt{\pi\cdot 2\lambda}\Gamma(\frac{2\lambda}{2})},
\end{aligned}
$$

which is the density of a t-distributed random variable with $2\lambda = 2 + 4\theta/\alpha^2$ degrees of freedom. The theorem is proved.

$\square$

The COGARCH(1,1) model

Recall that in the GARCH(1,1)-process with $\omega > 0, \alpha, \beta \geq 0$, (see Theorem 3.1), we have the representation

$$
\begin{aligned}
\sigma_t^2 &= \omega\big(\sum_{j=0}^{t-1}\prod_{i=1}^{j}(\alpha e_{t-i}^2+\beta)\big) + \sigma_0^2\prod_{k=0}^{t-1}(\alpha e_{t-k}^2+\beta) \\
&= \omega\sum_{j=0}^{n-1}\exp(\sum_{i=1}^{j}\log(\alpha e_{t-i}^2+\beta)) + \sigma_0^2\exp(\sum_{k=0}^{n-1}(\alpha e_k^2+\beta)).
\end{aligned}
$$

This motivates the following continuous-time GARCH(1,1) variant due to Klüppelberg, Lindner and Maller:

Definition 18. Let $(L_t)_{t\geq 0}$ be a Lévy process adapted to a filtrated probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ satisfying the usual conditions: $\mathcal{F}_t$ contains every $\mathbb{P}$-null set for any t, and $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$. Define the cadlag process

$$X_t = -t\log\beta - \sum_{s\leq t}\log\Big(1+\frac{\alpha(\Delta L_s)^2}{\beta}\Big) \quad (t \geq 0)$$

and with a random variable σ_0 independent of $(L_t)_{t\geq 0}$, the caglad volatility process

$$\sigma_t^2 = (\omega \int_0^t e^{X_s} ds + \sigma_0^2) e^{-X_{t-}} \quad (t \geq 0)$$

and the cadlag (integrated) COGARCH(1,1) process

$$G_t = \int_0^s \sigma_s dL_s \quad (t > 0), \; G_0 = 0.$$

G_t plays the same role as the cumulative log-returns $Y_t = X_t + ... + X_0$ from the GARCH model described in section 1. The simplest example of a COGARCH(1,1) process is driven by a Brownian motion: $L_t = B_t$. Since B_t is almost surely continuous, we have $X_t = -t \log \beta$, so $\sigma_t^2 = (\omega \int_0^t \beta^{-s} ds + \sigma_0^2)\beta^t = \frac{\omega(\beta^t - 1)}{\log \beta} + \sigma_0^2 \beta^t$. In this case, σ_t^2 is deterministic. This is not surprising: in general, the jumps ΔL_s are the analogons to the innovations e_n in discrete time.

Lemma 19. *Let X_t and σ_t be given as above. Then σ_t^2 solves the stochastic differential equation*

$$d\sigma_{t+}^2 = \omega dt + \sigma_t^2 e^{X_{t-}} d(e^{-X_t})$$

and therefore

$$\sigma_t^2 = \omega t + \log \beta \int_0^t \sigma_s^2 ds + \frac{\alpha}{\beta} \sum_{0<s<t} \sigma_s^2 (\Delta L_s)^2 + \sigma_0^2.$$

Proof. : Let $Y_t = \prod_{s\leq t}(1 + \frac{\alpha(\Delta L_s)^2}{\beta})$. Then

$$e^{-X_t} = \beta^t \prod_{s\leq t}(1 + \frac{\alpha(\Delta L_s)^2}{\beta}) = \beta^t Y_t,$$

and defining $f(t,x) = \beta^t x$, we have by Ito's lemma:

$$df(t, Y_t) = (\beta^t Y_t \log \beta)dt + \beta^t dY_t,$$

and therefore

$$e^{-X_t} = 1 + \log \beta \int_0^t e^{-X_s} ds + \frac{\alpha}{\beta} \sum_{s\leq t} e^{-X_{s-}} (\Delta L_s)^2$$

using the fact that Y_t has bounded variation. Since

$$\begin{aligned} e^{-X_t} \int_0^t e^{X_s} ds &= \int_0^t e^{-X_s} d(\int_0^s e^{X_r} dr) + \int_0^t \int_0^s e^{X_r} dr d(e^{-X_s}) + \left[e^{-X_t}, \int_0^t e^{X_s} ds\right]_t \\ &= \int_0^t e^{-X_s} e^{X_s} ds + \int_0^t \int_0^s e^{X_r} dr d(e^{-X_s}) + \left[e^{-X_t}, \int_0^t e^{X_s} ds\right]_t \end{aligned}$$

by integration by parts and the chain rule, and since

$$\left[e^{-X_t}, \int_0^t e^{X_s} ds\right]_t = \left[\log\beta \int_0^t e^{-X_s}, \int_0^t e^{X_s}\right]_t = \int_0^t d[s\log\beta, s]_s = 0$$

we have

$$e^{-X_t} \int_0^t e^{X_s} ds = t + \int_0^t \int_0^r e^{X_r} dr d(e^{-X_s}),$$

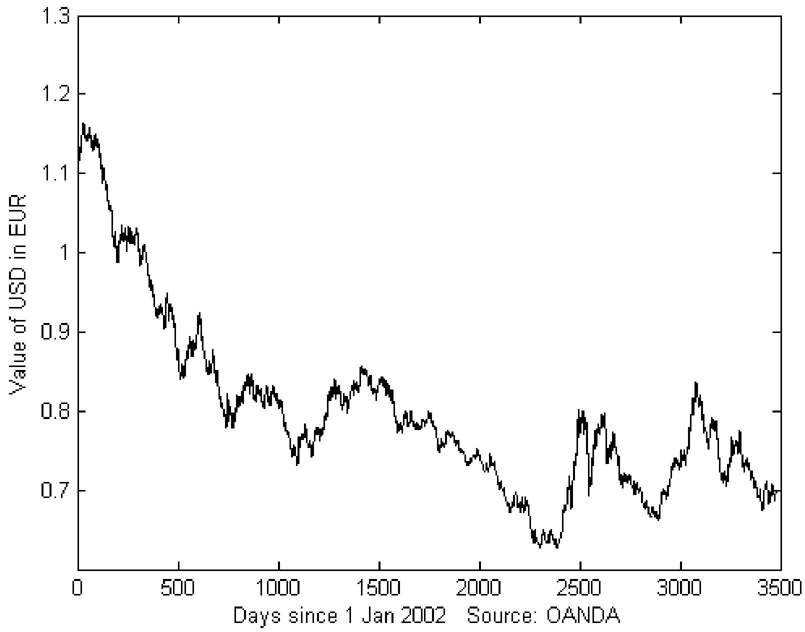
so

$$\begin{aligned} d\sigma_t^2 &= \omega d(e^{-X_t} \int_0^t e^{X_s} ds) + \sigma_0^2 d(e^{-X_t}) \\ &= \omega dt + \omega \int_0^t e^{X_s} ds d(e^{-X_t}) + \sigma_0^2 d(e^{-X_t}) = \omega dt + \sigma_t^2 e^{X_{t-}} d(e^{-X_t}). \end{aligned}$$

□

8 Example with MATLAB

In this section, we use the GARCH methodology to analyze the exchange rate between the U.S. dollar and the euro since its full introduction in 2002.

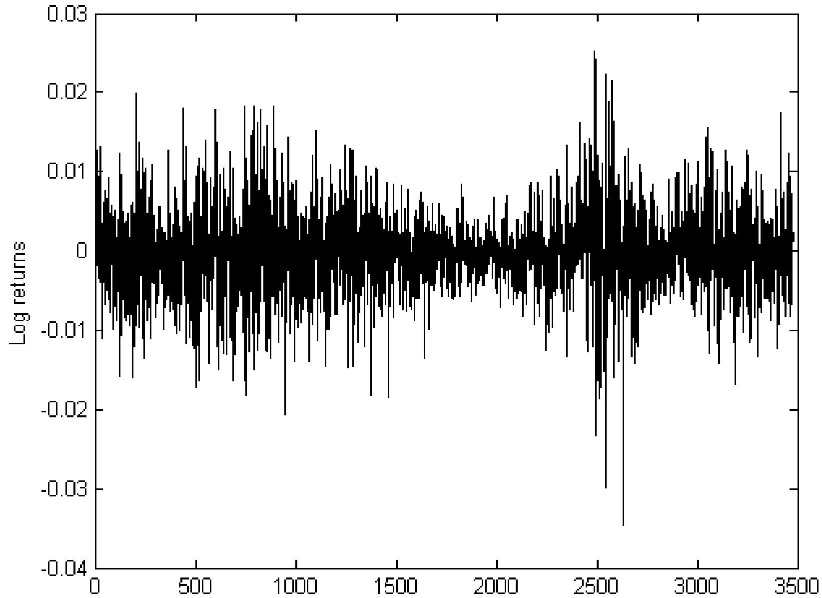


The above graph shows the daily average value of USD in euros each day from January 1st, 2002 until July 9, 2011. First, the data is transformed into its log returns. With the original time series named "data", the MATLAB code is simple:

```
for i = 1:length(data)-1
   temp(i) = log(data(i+1) / data(i));
end
global log_returns = temp;
plot(1:length(log_returns),log_returns);
```

where the log returns are saved as a global variable so that they can be accessed easily in other functions later on.

This returns the graph below:



Alternating periods of volatility and relative quiet are visible, as well as a period of intense volatility in late 2008 (around 2500 days from the start) which likely corresponds to the subprime mortgage crisis. At first glance, the data appear to have heteroskedastic effects. We now estimate the parameters using MATLAB's optimization toolbox. The Gaussian quasi-maximum likelihood function is implemented and stored separately in a file called *QMLE.m*:

```
function y = QMLE(param)
global log_returns;
sigma2(1) = param(1)/(1 - (param(2) + param(3)));
y = log(sigma2(1)) + (log_returns(1)^2)/sigma2(1);
for i=2:length(log_returns)
    sigma2(i) = param(1) + param(2)*log_returns(i-1)^2...
       + param(3)*sigma2(i-1);
    assert(sigma2(i) > 0);
    y = y + log(sigma2(i)) + (log_returns(i)^2)/sigma2(i);
end
```

where *param* is a vector (ω, α, β), *sigma2(i)* is σ_i^2 and *log_returns(i)* is X_i.

The function QMLE is now minimized over the compact set

$$[1 * 10^{-15}, 5] \times [1 * 10^{-15}, 1] \times [1 * 10^{-15}, 1]$$

under the additional constraint that $\alpha + \beta \leq 1 - 10^{-15}$. The reason for this is that the constraints must be given in the form $Ax \leq b$ instead of $Ax < b$. This optimization is done with

```
[param, fval] = fmincon(@QMLE,[0.0000002,0.03,0.96],...
                      [0,1,1],1-10^-15,[],[],lb,ub,[],options)
```

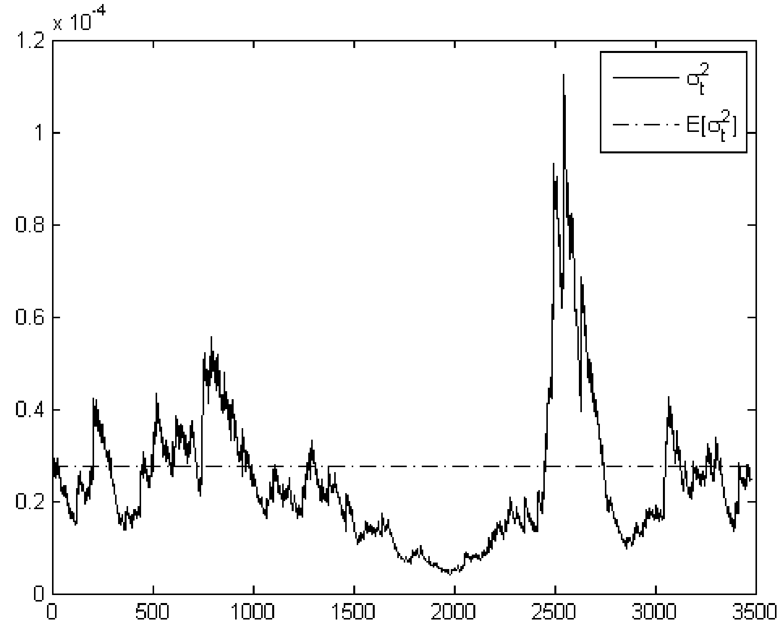
where $lb = 10^{-5}[1; 1; 1]$ is the lower bound and $ub = [5; 1; 1]$ the upper bound for the parameters. MATLAB generates the output

```
param =
    0.0000    0.0248    0.9744
fval =
 -3.4017e+004
```

and we find (after increasing the precision) our QML estimator

$$\hat{\omega} = 2.27 * 10^{-8}, \quad \hat{\alpha} = 0.0248, \quad \hat{\beta} = 0.9744$$

The function also generates the reconstructed volatility process $\sigma_t^2(\hat{\theta}_n)$ described in section 4:

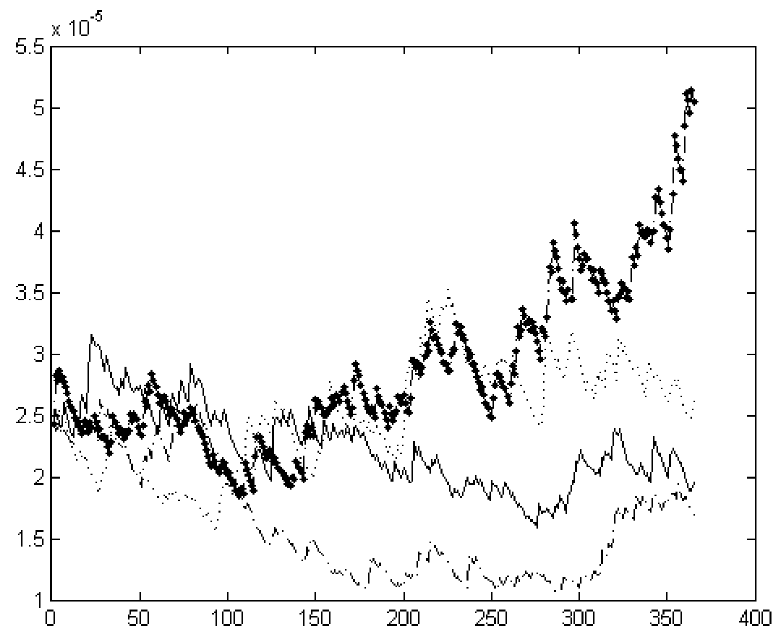


where the dotted line represents the stationary volatility.

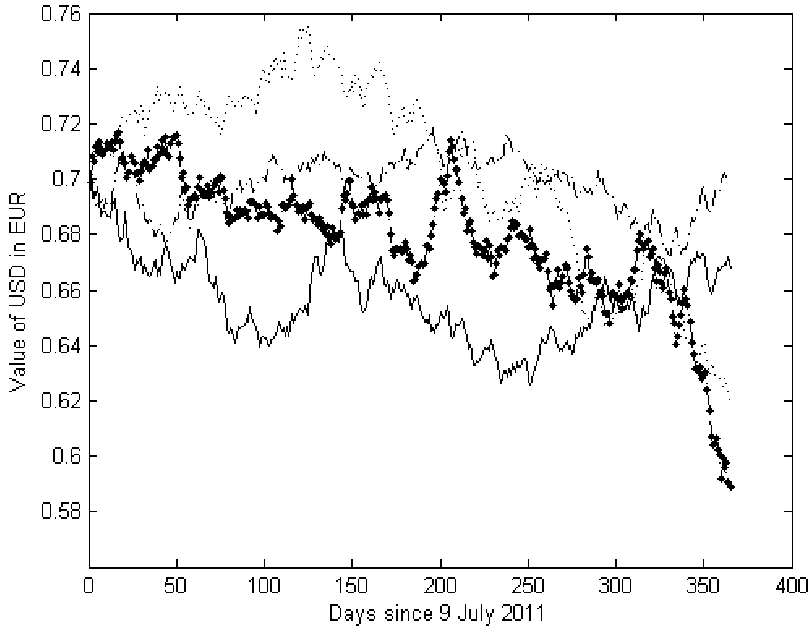
We can now simulate the process for the following year. Since the parameters were estimated using Gaussian quasi-maximum likelihood, it is appropriate that normally distributed innovations should be used for the simulation. Normally distributed pseudorandom numbers are generated in MATLAB with *randn*. We define the function

```
function [sim,sigma2] = simul(param,last_sigma2,last_observ,t)
assert(length(param) == 3);
assert(last_sigma2 > 0);
assert(param(1)*param(2)*param(3) > 0);
innov = randn([1,t]);
sigma2(1) = param(1) + param(2)*last_observ^2...
                   + param(3)*last_sigma2;
sim(1) = innov(1)*sqrt(sigma2(1));
for i=2:t
    sigma2(i) = param(1) + param(2)*sim(i-1)^2...
                      + param(3)*sigma2(i-1);
    sim(i) = innov(i)*sqrt(sigma2(i));
end
```

Four simulated volatilities are shown below:



which correspond to the following predicted exchange rates:



9 Discussion

In this thesis, we have considered the strengths and weaknesses of the GARCH(1,1) model in mathematical finance, as well as the practical questions of parameter estimation and implementation. GARCH models and variants have become ubiquitous in the theory of economic time series since their introduction only 25 years prior. This is due to the relative simplicity of the model and the wide range of processes it can approximate. Related models such as the EARCH and EGARCH models of Nelson (1989) provide the ability to account for emperical phenomena such as the leverage effect at the cost of a more complex asymptotic theory for the typical estimators.

The investigation of multivariate GARCH models remains an active area of research. The need for such models arises when one considers a set of time series with significant interdependence; an example of this is the stock price of a manufacturing firm and the commodity prices for the resources it requires. The most general model replaces the GARCH specification with matrix-valued coefficients as well as a log-returns vector X_t and a vectorized volatility matrix σ_t (that is, such that σ_t^2 is the conditional covariance of X_t). This is known as the Vec model. However, this can be very difficult to work with, as necessary and sufficient conditions to ensure that σ_t^2 is positive definite are difficult or impossible to derive. Therefore, the Vec model is often restricted to models such as the BEKK model. The theory of these models is beyond the scope of this paper.

Another area of further research is the connection between GARCH models and stochastic volatility models. Two examples of continuous-time processes which are related to the discrete GARCH equations are mentioned in section 7; these are the only such classes of processes known at this time. The stationary distribution of the diffusion limit may also contain information about the stationary distribution of the GARCH model and thus its long-term behavior. In addition, the convergence to the diffusion limit is weak and does not necessarily imply that the GARCH model and the diffusion limit must be asymptotically equivalent.

References

[1] Bollerslev, T., 1986, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics 31, 307-327 Journal of Economics, 31, 307-327

[2] Engle, R.F., 1982, *Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation*, Econometrica, 50, 987-1008

[3] Engle, R.F. and Ng, V.K., 1993, *Measuring and testing the impact of news on volatility*, Journal of Finance, 48, 1749-1778

[4] Francq, C. and Zakoïan, J-M. *Modèles GARCH: structure, inférence statistique et applications financières*, Economica: Paris, 2009

[5] Kreiss, J-P. and Neuhaus, G., *Einführung in die Zeitreihenanalyse*, Berlin, Heidelberg: Springer-Verlag, 2006

[6] Klüppelberg, C., Lindner, A. and Maller, R., 2004, *A continuous time GARCH(1,1) process driven by a LÂ ´evy process: stationarity and second order behaviour*, Journal of Applied Probability, 41, 1-22

[7] MATLAB version 7.11.0. Natick, Massachusetts: The MathWorks Inc., 2010.

[8] Nelson, D.B., 1988, *Stationarity and persistence in the GARCH(1,1) model*, Econometric Theory, 6, 318-334

[9] Nelson, D.B., 1989, *Conditional heteroskedasticity in asset returns: A new approach*, Econometrica, 59, 347-370

[10] Nelson, D.B., 1990, *ARCH models as diffusion approximations*, Journal of Economics, 45, 7-28

[11] Sentana, E., 1995, *Quadratic ARCH models*, Review of Economic Studies, 62, 639-661

[12] Stroock, D.W. and Varadhan, S.R.S., *Multidimensional Diffusion Processes*, Berlin: Springer-Verlag, 2006

[13] Zakoïan, J-M., 1994, *Threshold heteroskedastic models*, Journal of Economic Dynamics and Control, 18, 931-955