Assignment 3

William Gram

December 5, 2018

1. t-ARCH(1)

Here we consider the ARCH(1) model with t-distributed errors. Consider first the model below:

$$x_t = \sigma_t z_t, \quad z_t \overset{iid}{\sim} t_3\left(0,1\right), \tag{1.1}$$

noting that this implies the z_t's are a scaled t-distribution with 3 degrees of freedom. Consider in Appendix A how this scaling has been achieved. The volatility is given by:

$$\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2. \tag{1.2}$$

1.1 Motivation for using t-distribution

In empirical finance, the normally distributed z_t does not capture the mass in the tails, i.e. it does not manage to explain the frequent extreme returns. As seen in figure 1 below, the student's t-distribution has fatter tails and thus imposes a larger probability of observing these more extreme returns. This motivates the use of a student's t-distributed random variable z_t.

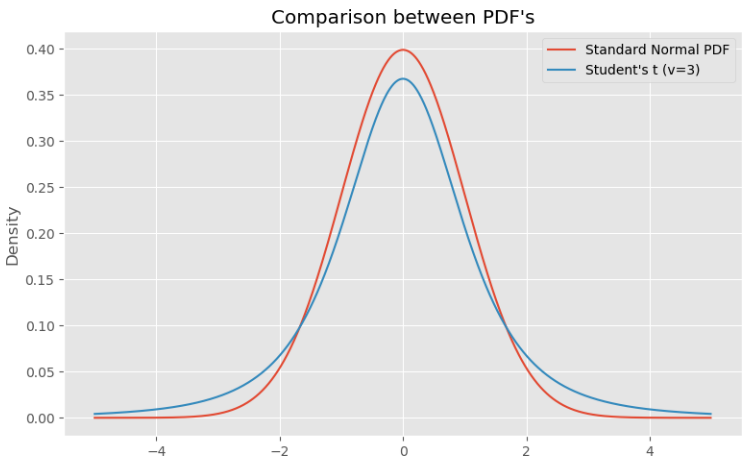
1.2 Existence of moments

As mentioned in the appendix, moments of $\nu - 1$ exist for a student's t-distributed variable z_t with ν degrees of freedom.

1.3 Log-likelihood function

We are asked to find the log-likelihood function for the model given in (1.1)-(1.2). To do this, we first need to find the conditional moments.

Figure 1: Comparison between standard normal PDF and PDF of a t-distributed random variable



Parameters: the t-distributed variable has $\nu = 3$ df.

1.3.1 Moments

Given that x_t is clearly Markov, we have the first two moments given by:

$$\mathbf{E}\left[x_t|x_{t-1}\right] = \mathbf{E}\left[\sqrt{\sigma^2 + \alpha x_{t-1}^2} z_t | x_{t-1}\right] = \sqrt{\sigma^2 + \alpha x_{t-1}^2}\mathbf{E}\left[z_t\right] = 0 \tag{1.3}$$

$$\mathbf{V}\left[x_t|x_{t-1}\right] = \left(\sigma^2 + \alpha x_{t-1}^2\right)\mathbf{E}\left[z_t^2\right] = \sigma^2 + \alpha x_{t-1}^2. \tag{1.4}$$

Notice, that the result above holds because z_t has been scaled, as discussed in Appendix A.

1.3.2 PDF

With the conditional moments defined in equations (1.3) and (1.4), we next need to consider the conditional probability density function. With z_t being t-distributed, we need the PDF for a t-distribution.

We are told that the PDF is given by:

$$p\left(x|\nu,\mu,\sigma\right) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}}\left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \Rightarrow f\left(x|x_{t-1}\right) = \frac{4}{\pi\sigma_t\left(1+z_t^2\right)^2}, \tag{1.5}$$

where both the LHS and the RHS in (1.5) is derived in Appendix B.

1.3.3 The log-likelihood function

Taking the natural logarithm to (1.5) we find:

$$\ell_t\left(x_t|x_{t-1}\right) = \log\left(\frac{4}{\pi}\right) - \frac{1}{2}\log\left(\sigma_t^2\right) - 2\log\left(1+z_t^2\right). \tag{1.6}$$

The result in (1.7) can itself be scaled and the constant removed, to attain:

$$\ell_t(\theta) = -\log\left(\sigma_t^2\right) - 4\log\left(1 + z_t^2\right). \tag{1.7}$$

1.4 Score

Here we are given that the score S_T evaluated at θ_0 is given by:

$$\frac{1}{\sqrt{T}} S_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \ell_t(\theta_0)}{\partial \alpha} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{3z_t^2 - 1}{1 + z_t^2} \frac{x_{t-1}^2}{\sigma^2 + \alpha_0 z_{t-1}^2}, \tag{1.8}$$

where a formal derivation can be found in appendix C.

We are asked to show, that the score in (1.8) is asymptotically Gaussian under appropriate conditions. We can show this, by applying theorem II.1, which is introduced below.

Theorem 1 *Assuming theorem I.2 applies to* $\{X_t\}_{t\geq 0}$*,* X_t *stationary. With* $f(X_t, X_{t-1}, \ldots, X_{t-m}) \in$ $\mathbf{R}$*, assume that:*

1. $\mathbf{E}\left[f(X_t, \ldots, X_{t-m}) \,|X_{t-1}, \ldots, X_{t-m}\right] = 0$
2. $\mathbf{E}\left[f^2(X_t, \ldots, X_{t-m})\right] < \infty$.

In the case where assumptions (1) and (2) hold, the CLT applies as $T \to \infty$*:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(X_t, \ldots, X_{t-m}) \xrightarrow{d} \mathbf{N}\left(0, \mathbf{E}\left[f^2(X_t, \ldots, X_{t-m})\right]\right). \tag{1.9}$$

Thus, for the score $f(X_t, \ldots, X_{t-m}) = S_T(\theta_0)$ to be asymptotically Gaussian, we need to show that assumptions (1) and (2) in Theroem 1 apply to S, taking theorem 1.2 as given.

1.4.1 Assumption 1

With $x_t|x_{t-1}, \ldots, x_0 \overset{d}{=} x_t|x_{t-1}$, we consider the conditional expectation of S_t given x_{t-1}, where each element in the sum of S_T is denoted S_t, $t \in [1, T]$.

First, we can define $y_t \equiv \frac{3z_t^2-1}{1+z_t^2}$ and $f(z_{t-1}) = \xi_{t-1} \equiv \left(\frac{x_{t-1}}{\sigma_t}\right)^2 = \left(\frac{x_{t-1}}{\sigma^2+\alpha x_{t-1}^2}\right)^2$, which depends on the shock z_{t-1} and no other period shocks. That is ξ_{t-1} is independent of z_t for all t. Thus, we need to show that $\mathbf{E}\left[y_t \xi_{t-1} | x_{t-1}\right] = 0$. We are told that $\mathbf{E}\left[y_t\right] = 0$ (see Appendix D for derivation). Thus, applying the tower property we have:

$$\mathbf{E}\left[y_t \xi_{t-1}\right] = \mathbf{E}\left[\mathbf{E}\left[y_t \xi_{t-1} | x_{t-1}\right]\right] = \mathbf{E}\left[\xi_{t-1} \mathbf{E}\left[y_t | x_{t-1}\right]\right] = \mathbf{E}\left[\xi_{t-1} \mathbf{E}\left[y_t\right]\right] = 0, \tag{1.10}$$

where the last equality follows from independence of z_t.

1.4.2 Assumption 2

We need to show that $\mathbf{E}\left[(y_t\xi_{t-1})^2\right] < \infty$. To do this, we first consider ξ_t^2:

$$\xi_t^2 = \frac{\sigma_{t-1}^4 z_{t-1}^4}{\left(\sigma^2 + \alpha\sigma_{t-1}^2 z_{t-1}^2\right)^2} = \frac{\sigma_{t-1}^4 z_{t-1}^4}{\sigma^4 + \alpha^2\sigma_{t-1}^4 z_t^4 + 2\sigma^2\alpha\sigma_{t-1}^2 z_{t-1}^2} < \frac{1}{\alpha}, \quad \alpha^2 > 0. \tag{1.11}$$

Applying (1.12) to the unconditional variance of the score elements, we find that:

$$\mathbf{E}\left[(y_t\xi_{t-1})^2\right] < \mathbf{E}\left[y_t^2\frac{1}{\alpha^2}\right] = \frac{1}{\alpha^2}\mathbf{E}\left[y_t^2\right] < \infty. \tag{1.12}$$

We generally assume $\alpha > 0$ for return series, as this is the ARCH-effect which we are trying to capture and is itself the motivation of considering ARCH-models.

1.4.3 Conclusion

We verified that the conditions in Theorem 1 hold, specifically we showed in section (1.4.1) that condition (1) holds, and in (1.4.2) that condition (2) holds. Consequently the CLT applies to the score, such that it is asymptotically Gaussian.

1.5

In this question, we are asked to simulate 10,000 scores of t-ARCH series of lenght $T = 100$. Subsequently we will plot the QQ plots of the scores their corresponding density estimates. Plotting 10.000 realisations look as found in figure 2. Figure 3 shows the QQ-plot in the left panel and a histogram in the right panel. Remember that the asymptotic distribution is Gaussian, such that the histogram and QQ-plot compares to a Gaussian distribution.

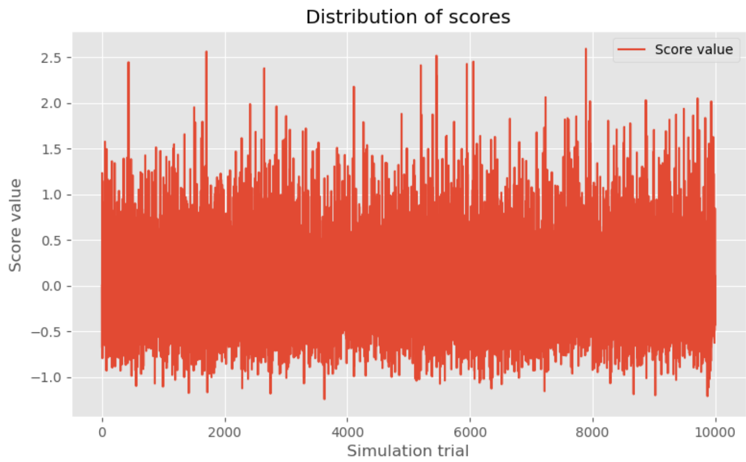
As is clear from the charts in figures 2 and 3, the scores are not normally distributed. Because the theory states that the asymptotic distribution of the scores are Gaussian, I must look for the mistake in the simulation somewhere. The mean comes out seemingly correct, but there is a large right skew.

1.6

We showed in exercise 1.4 that the scores were asymptotically normal, which I did not manage to illustrate in exercise 1.5. Despite this, we accept the theoretical result.

Next, we consider the regularity conditions for the error $\hat{\alpha} - \alpha_0$ to be Gaussian. Here I refer to Theorem III.2 written below:

Figure 2: Realisations of scores



Parameters: the t-distributed variable has been rescaled to have mean zero and unit variance. Df is 3.

Theorem 2 *We first assume that $\ell_T(\theta)$ is three times differentiable in θ with all derivatives continuous. With $\theta_0 \in \Theta$, we assume that:*

1. $\frac{1}{\sqrt{T}} S_T(\theta_0) = \frac{1}{\sqrt{T}} \frac{\partial \ell_T(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}(0, \Omega_S), \quad \Omega_S > 0.$
2. $\frac{1}{T}(\theta_0) = -\frac{1}{T}\frac{\partial^2 \ell_T(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} \Omega_I > 0$
3. $\max_{h,j,j=1,\ldots,k} \sup_{\theta \in N(\theta_0)} \left| \frac{1}{T} \frac{\partial^3 \ell_T(\theta)}{\partial\theta_h \partial\theta_i \partial\theta_j} \right| \leq c_T,$

where $N(\theta_0)$ is a neighbourhood around θ_0, and $0 \leq c_T \xrightarrow{d} c,$, $0 < c < \infty$. Then, in a neighbourhood of θ_0 abd as $T \to \infty$,

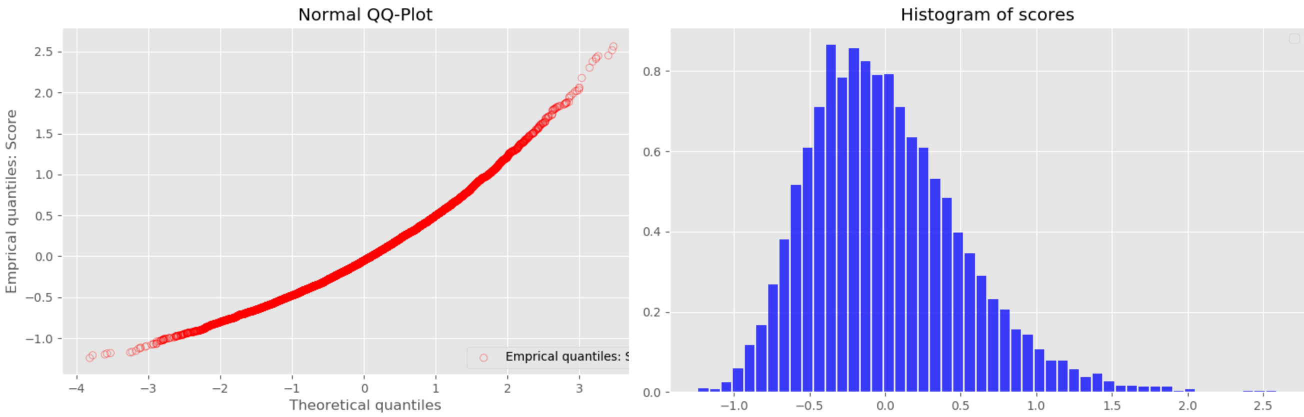
1. *$\hat{\theta}_T$, which solves the estimating equation, $\frac{\partial \ell_T(\hat{\theta}_T)}{\partial \theta} = 0$, is unique (with probability tending to one).*
2. *$\hat{\theta}_T \xrightarrow{P} \theta_0$.*
3. *$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \Omega_I^{-1}\Omega_S\Omega_I^{-1}\right)$.*

Consider now a first-order Taylor-expansion of the score at the point $\theta = \theta_0$:

$$\begin{aligned} S_T\left(\hat{\theta}\right) &\approx S_T(\theta_0) + \left.\frac{\partial S_T(\theta)}{\partial \theta}\right|_{\theta=\theta_0} \left(\hat{\theta} - \theta_0\right) + \ldots \\ &\approx S_T(\theta_0) - \iota_T(\theta_0)\left(\hat{\theta} - \theta_0\right) \\ \Leftrightarrow \sqrt{T}\left(\hat{\theta} - \theta_0\right) &\approx \left(\frac{1}{T}\iota_T(\theta_0)\right)^{-1} \frac{1}{\sqrt{T}} S_T(\theta_0). \end{aligned} \tag{1.13}$$

In practice the third assumption is rarely checked.

Figure 3: Distribution of realised scores



Parameters: the t-distributed variable has been rescaled to have mean zero and unit variance. Df is 3.

Exercise 1.7

We are now trying to fit the T-ARCH(1) model to the S&P 500-index. To do this, we want to maximise the conditional log-likelihood function, which is the sum over all the contributions given in (1.7), i.e.:

$$\ell_T(\theta) = \sum_{t=1}^{T} \ell_t(\theta) = -\sum_{t=1}^{T} \left(\log\left(\sigma_t^2\right) + 4\log\left(1 + z_t^2\right)\right). \tag{1.14}$$

We recall that maximising (1.14) is equivalent to minimising its negative, which is what we do in practice. Next, we choose $\theta = \left(\sigma^2 = 0.05, \alpha = 0.5\right)$ as starting values, leading to estimates and results as given in table 1 below.

Table 1: t-studen'ts ARCH(1) Maximum Likelihood results

	ML Estimate	SE	SE Robust	ML t-value	Robust t-value
σ^2	1.046543	0.04987	0.07121	20.98403	14.69684
α	0.288111	0.05388	0.07751	5.34679	3.716861

The maximisation returned a ML-value of 2,514.74316236.

We not taht the estimated parameters are statistically significant regardless of which standard errors we use for computation. Treating the residuals as normal implies that we are running quasi maximum likelihood (QML) which is consistent, and is the reason why we usually consider standard normal residuals, rather than t-distributed noise.

Exercise 2

We consider here the GJR-ARCH(1) model given below:

$$x_t = \sigma_t z_t, \quad z_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{1.15}$$

$$\sigma_t^2 = \sigma^2 + \alpha x_{t-1}^2 + \gamma \mathbf{1}_{\{x_{t-1}<0\}} x_{t-1}^2, \quad \sigma^2 > 0, \quad \alpha, \gamma \geq 0. \tag{1.16}$$

2.1 Comparison to A-ARCH and discussion

This model is a reformulation of the asymmetric ARCH(1) model considered in week 4. To see this, write the following:

$$\sigma_{t,G}^2 = \begin{cases} \sigma^2 + (\alpha+\gamma)\, x_{t-1}^2, & x_{t-1} < 0 \\ \sigma^2 + \alpha x_{t-1}^2, & x_{t-1} \geq 0 \end{cases} \tag{1.17}$$

$$\sigma_{t,A}^2 = \begin{cases} \sigma^2 + \alpha_n x_{t-1}^2, & x_{t-1} < 0 \\ \sigma^2 + \alpha_p x_{t-1}^2, & x_{t-1} \geq 0 \end{cases}, \tag{1.18}$$

where the subscripts G and A refer to the GJR-ARCH and A-ARCH models, respectively. Thus, $\alpha_n \equiv \alpha + \gamma$ and $\alpha_p \equiv \alpha$.

In consequence, the interpretation is the same as for the A-ARCH model. The model allows for asymmetries in conditional variance, e.g. a negative shock may have a larger ARCH-effect on σ_t^2 than a positive shock. This effect is coined the "leverage effect".

2.2 Log-likelihood function

For this model, we note that there are three parameters to estimate, namely $\theta = \left(\sigma^2, \alpha, \gamma\right)$. The conditional moments of degrees one and two are given by:

$$\mathbf{E}\left[x_t | x_{t-1}\right] = 0, \quad \mathbf{E}\left[x_t^2 | x_{t-1}\right] = \sigma_t^2, \tag{1.19}$$

with the process $\{x_t\}_{t\geq 1}$ being conditionally Gaussian distributed. Thus, the log-likelihood contributions are given by:

$$\ell_t(\theta) = \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_t}{\sigma_t}\right)^2\right\}\right) = -\frac{1}{2}\left(\log\left(2\pi\sigma_t^2\right) + \left(\frac{x_t}{\sigma_t}\right)^2\right), \tag{1.20}$$

with the conditional log-likelihood function given by:

$$\ell_T(\theta) = \sum_{t=1}^{T} \ell_t. \tag{1.21}$$

2.3 The score (not the album by the Fugees, unfortunately)

We are asked to derive the score with respect to α. This is given by the partial derivative below:

$$S_T^{(\alpha)}(\theta) = \frac{\partial \ell_T(\theta)}{\partial \alpha} = -\frac{1}{2}\sum_{t=1}^{T}\left(\frac{x_{t-1}^2}{\sigma_t^2} - \frac{x_t^2}{\sigma_t^4}x_{t-1}^2\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{x_t^2}{\sigma_t^2}\right)\frac{x_{t-1}^2}{\sigma_t^2}. \tag{1.22}$$

2.4 Asymptotic distribution of the score

We are asked to discuss which assumptions are needed in order to have the score in (1.22) being asymptotically Gaussian distributed. We refer to theorem 1 above to emphasise the required assumptions. For the following derivations, we define by f the elements in the score, i.e.:

$$f(x_t, x_{t-1}) = \left(1 - z_t^2\right)\frac{x_{t-1}^2}{\sigma_t^2}. \tag{1.23}$$

2.4.1 Assumption (i)

The first assumption of a conditionally zero mean, is shown by:

$$\mathbf{E}\left[f(x_t, x_{t-1})\,|x_{t-1}\right] = \mathbf{E}\left[\left(1-z_t^2\right)\frac{x_{t-1}^2}{\sigma_t^2}\middle| x_{t-1}\right] = \left(\mathbf{E}\left[z_t^2\right]-1\right)\frac{x_{t-1}^2}{\sigma_t^2} = 0,$$

where the second equality follows from z_t being i.i.d. such that $\mathbf{E}\left[z_t|x_{t-1}\right] = \mathbf{E}\left[z_t\right]$.

2.4.2 Assumption (ii)

The second assumption is that of finite variance, which is shown as follows:

$$\begin{aligned}
\mathbf{E}\left[f^2(x_t, x_{-1})\right] &= \mathbf{E}\left[\left(\left(1-z_t^2\right)\frac{x_{t-1}^2}{\sigma_t^2}\right)^2\right] \\
&= \mathbf{E}\left[\left(z_t^4 + 1 - 2z_t^2\right)\left(\frac{x_{t-1}}{\sigma_t}\right)^4\right] \\
&= \left(\mathbf{E}\left[z_t^4\right] + 1 - 2\mathbf{E}\left[z_t^2\right]\right)\mathbf{E}\left[\left(\frac{x_{t-1}}{\sigma_t}\right)^4\right] \\
&= 2\mathbf{E}\left[\left(\frac{x_{t-1}}{\sigma_t}\right)^4\right] < \infty, \quad \alpha_0 > 0 \vee \mathbf{E}\left[x_t^4\right] < 0,
\end{aligned} \tag{1.24}$$

where again the third equality follows from z_t being i.i.d.

As these assumptions are shown to hold, we have that:

$$\frac{1}{\sqrt{T}} S_T(\theta_0) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{E}\left[f^2(x_t, x_{t-1})\right]\right). \tag{1.25}$$

Table 2: GJR Arch Maximum Likelihood results

	ML Estimate	SE	SE Robust	ML t-value	Robust t-value
σ^2	0.827574	0.05103	0.13355	16.21879	6.196656
α	0.360091	0.25919	0.81889	1.38931	0.439732
γ	0.252586	0.82697	0.73821	0.88020	0.342158

The maximisation returned a ML-value of 2,029.17384396.

2.5 S&P 500 estimation in the GJR-ARCH(1) framework

Repeating the estimation steps in exercise 1.7 leads to the estimates in table 2 below.

Note that the estimates are highly unstable. In class we produced the estimates in table 3 below. Making an initial guess of $\theta = (0.8, 0.05, 0.2)$ leads to the very same estimates to at least four decimals. Deviating too much from this initial guess will lead to other very different estimates.

Table 3: GJR Arch Maximum Likelihood results

	ML Estimate	SE	SE Robust	ML t-value	Robust t-value
σ^2	0.833698	0.03782	0.07550	22.04257	11.041992
α	0.062095	0.02829	0.04098	2.19150	1.515284
γ	0.226405	0.07332	0.09670	3.08793	2.341301

The maximisation returned a ML-value of 2,005.915907.

Appendix A

Consider an unscaled t-distributed random variable z_t^0 with $\nu > 3$ degrees of freedom. This has the following moments:

$$\mathbf{E}\left[z_t^0\right] = 0, \quad \mathbf{V}\left(z_t^0\right) = \frac{\nu}{\nu - 2}. \tag{A.1}$$

In general, $\nu - 1$ moments exist for a t-distributed random variable with ν degrees of freedom. To achieve the normalised t-distributed random variable z_t in exercise 1, we have that:

$$z_t = z_t^0 \sqrt{\frac{\nu - 2}{\nu}} \tag{A.2}$$

This scaling implies that:

$$\mathbf{V}\left[z_t\right] = \mathbf{V}\left[z_t^0\right] \frac{\nu - 2}{\nu} = \frac{\nu}{\nu - 2} \frac{\nu - 2}{\nu} = 1. \tag{A.3}$$

Appendix B

Here we are deriving the PDF given in exercise 1.3. We start by considering the PDF to z_t, which is t-distributed. In exercise 1 we consider a scaled z_t, but we first consider the unscaled case. Thereafter we adapt it to match the scaling of z_t.

The PDF is in general given by:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{B.1}$$

Implementing a location parameter μ and a scale-parameter σ such that a random variable X is defined as:

$$X_t = \mu + \sigma T,$$

where T is a t-distributed random variable, changes the PDF in (B.1) to:

$$p(x|\nu,\mu,\sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}. \tag{B.2}$$

It is (B.2) which is provided in the hint in exercise 1.3. Yet, we can do better still.

Finally, we must consider that T has been rescaled, so that $\sigma^2 = \tilde{\sigma}^2\frac{\nu-2}{\nu}$, with $\mu = 0$ being unchanged. With these changes, we have that:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)\sigma_t^2}}\left(1+\frac{1}{\nu-2}\frac{x_t^2}{\sigma_t^2}\right)^{-\frac{\nu+1}{2}}.$$

With the Γ-function i.a. defined through:

$$\Gamma(n) = (n-1)!, \quad \Gamma\left(\frac{1}{2}+n\right) = \frac{(2n)!}{4^n n!}\sqrt{\pi},$$

we have that for $\nu = 3$:

$$\Gamma\left(\frac{\nu+1}{2}\right) = \Gamma(2) = 2, \quad \Gamma\left(\frac{\nu}{2}\right) = \Gamma\left(\frac{1}{2}+1\right) = \frac{2}{4}\sqrt{\pi} = \frac{1}{2}\sqrt{\pi}. \tag{B.3}$$

Thus, applying (B.4) to (B.3) we find:

$$f(x|x_{t-1}) = \frac{2}{\frac{1}{2}\sqrt{\pi}\sqrt{\pi}\sigma_t}\left(1+z_t^2\right)^{-2} = \frac{4}{\pi\sigma_t\left(1+z_t^2\right)^2}. \tag{B.4}$$

Appendix C

This section shows a short derivation of the score provided in exercise 1.4. Here we were given that the score S_T evaluated at θ_0 is given by:

$$\frac{1}{\sqrt{T}} S_T\left(\theta_0\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \ell_t\left(\theta_0\right)}{\partial \alpha} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{3z_t^2 - 1}{1 + z_t^2} \frac{x_{t-1}^2}{\sigma^2 + \alpha_0 z_{t-1}^2}.$$

Recall that we found that the (transformed) log likelihood contributions were given by:

$$\ell_t\left(\theta\right) = -\log\left(\sigma_t^2\right) - 4\log\left(1 + z_t^2\right). \tag{C.1}$$

Differentiating ℓ_t in (C.1) with respect to α at $\alpha = \alpha_0$ gives:

$$\begin{aligned}
\left.\frac{\partial \ell_t\left(\theta\right)}{\partial \alpha}\right|_{\alpha=\alpha_0} &= -\left(\frac{1}{\sigma_t^2} x_{t-1}^2 + 4\frac{1}{1+z_t^2}\left(-\frac{x_t^2}{\sigma_t^4}\right) x_{t-1}^2\right) \\
&= 4\left(\frac{x_{t-1}}{\sigma_t}\right)^2 z_t^2 \frac{1}{1+z_t^2} - \left(\frac{x_{t-1}}{\sigma_t}\right)^2 \\
&= \left(\frac{x_{t-1}}{\sigma_t}\right)^2 \left(\frac{4z_t^2}{1+z_t^2} - 1\right) \\
&= \left(\frac{x_{t-1}}{\sigma_t}\right)^2 \frac{4z_t^2 - \left(1+z_t^2\right)}{1+z_t^2} \\
&= \left(\frac{x_{t-1}}{\sigma_t}\right)^2 \frac{3z_t^2 - 1}{1+z_t^2}. \qquad \text{(C.2)}
\end{aligned}$$

It is the result in (C.2) that yields the score in (1.8).

Appendix D

In question 1.4 we are told that $\mathbf{E}\left[y_t\right]=0$ and that $\mathbf{E}\left[y_t^4\right]<\infty$. We first show the first moment, with a referral to [2] as the source for the proof.

D.1 Zero expectation

In a general setting, we define:

$$y_t=(\nu+1)\frac{z_t^2}{(\nu-2)+z_t^2}-1, \tag{D.1}$$

and we note that for $\nu=3$ we have that $y_t=\frac{4z_t^2}{1+z_t^2}-1=\frac{4z_t^2-\left(1+z_t^2\right)}{1+z_t^2}=\frac{3z_t^2-1}{1+z_t^2}$ as given in problem 1.4. Next, define:

$$\eta_t=\frac{z_t^2}{(\nu-2)+z_t^2}\in[0,1[. \tag{D.2}$$

By Johnson, Kempt and Kotz (1995) [1], we have that:

$$\eta\sim\text{Beta}\left(\frac{1}{2},\frac{\nu}{2}\right), \tag{D.3}$$

where a Beta-distributed random variable $X\sim\text{Beta}\left(p,q\right)$ has expectation $\mathbf{E}\left[X\right]=\frac{p}{p+q}$. Thus, noting that y_t in (D.1) can be written as a function of η_t as defined in (D.2) rather than z_t:

$$y_t=(\nu+1)\frac{z_t^2}{\nu-2+z_t^2}-1=(\nu+1)\,\eta_t-1, \tag{D.4}$$

we have that $\mathbf{E}\left[y_t\right]$ is given by:

$$\mathbf{E}\left[y_t\right]=(\nu+1)\,\mathbf{E}\left[\eta_t\right]-1=(\nu+1)\frac{\frac{1}{2}}{\frac{1+\nu}{2}}-1=\frac{\nu+1}{\nu+1}-1=0,\quad\forall\nu>2. \tag{D.5}$$

The result in (D.5) shows the desired result.

D.2 Finite variance

Next, we want to show that the second moment is finite. More specifically, Pedersen and Rahbek (2016) show that $\mathbf{E}\left[y_t^2\right]=\frac{2\nu}{\nu+3}<\infty$, for $\nu\geq 3$. We will apply the fact that $\mathbf{E}\left[X^2\right]=p\left(p+q\right)^{-1}\left(p+1\right)\left(p+q+1\right)^{-1}$ for $X\sim\text{Beta}\left(p,q\right)$.

Recall the distribution stated in (D.3) and consider the variance of y_t as defined in (D.4):

$$\begin{aligned}
\mathbf{V}\left(y_t\right) = \mathbf{E}\left[y_t^2\right] &= \mathbf{E}\left[\left(\left(1+\nu\right)\eta_t - 1\right)^2\right] \\
&= \mathbf{E}\left[\left(1+\nu\right)^2 \eta_t^2 + 1 - 2\left(1+\nu\right)\eta_t\right] \\
&= \left(1+\nu\right)^2 \mathbf{E}\left[\eta_t^2\right] + 1 - 2\left(1+\nu\right)\mathbf{E}\left[\eta_t\right] \\
&= \left(1+\nu\right)^2 \left(\frac{1}{2}\frac{2}{1+\nu}\frac{1+2}{2}\frac{2}{1+\nu+2}\right) + 1 - 2\left(1+\nu\right)\frac{1}{1+\nu} \\
&= \frac{3\left(1+\nu\right)}{3+\nu} - 1 \\
&= \frac{3+3\nu-\left(3+\nu\right)}{3+\nu} \\
&= \frac{2\nu}{3+\nu} < \infty.
\end{aligned}$$

The result in (D.6) is what we wanted to show.

Bibliography

[1] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*. Wiley series in probability and mathematical statistics: Applied probability and statistics vb. 2. Wiley & Sons, 1995.

[2] Rs Pedersen and A Rahbek. “Nonstationary GARCH with t-distributed innovations”. In: *Economics Letters* 138 (2016), pp. 19–21.