Financial Econometrics Exam

January 17-22, 2019

1 Theoretical part

1.1 Derive the GAS updating step for φ_t

In general the updating step in the GAS model with no scaling is

$$\begin{aligned}\psi_t &= \psi(\mathbf{y}_{1:t-1}) \\ &= \omega + \alpha u_{t-1} + \beta \psi_{t-1},\end{aligned}$$

where ψ_t is the variable of interest, i.e. the variable that we want to filter out, and $u_t = \nabla_t = \frac{\partial \log p(y_t|\mathbf{y}_{1:t-1};\psi)}{\partial \psi}$, which is the unscaled score of the conditional distribution. The general idea is to use this score of the conditional distribution to give the direction of the update step.
In some applications it is convenient to introduce a link function, such that we restrict our variable in some way. In this particular exercise for the GAS-GED model, we want to have that the scale parameter, φ_t, which is the variable of interest, is positive. We impose this restriction by introducing a exponential link equation such that

$$\begin{aligned}\varphi_t &= \exp(\tilde{\varphi}_t) \\ \tilde{\varphi}_t &= \omega + \alpha \tilde{u}_{t-1} + \beta \tilde{\varphi}_{t-1},\end{aligned}$$

with $\tilde{u}_t \equiv s_t$ as stated in the exercise, where

$$s_t = \frac{\partial \log p(y_t \mid \mathbf{y}_{1:t-1}; \tilde{\varphi}, \nu, \mu)}{\partial \tilde{\varphi}_t} = \frac{\partial \log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu, \mu)}{\partial \varphi_t} \frac{\partial \varphi_t}{\partial \tilde{\varphi}_t},$$

is the score of of $Y_t \mid \mathcal{F}_{t-1}$ with respect to $\tilde{\varphi}_t$. This means that in this exercise we need to find the logarithm to the PDF of the conditional distribution of Y_t given information up through time $t-1$, $\mathcal{F}_{t-1}$, which is the generalized error distribution, i.e. $Y_t \mid \mathcal{F}_{t-1} \sim GED(0, \varphi, \nu)$ (note that $\mu = 0$). With $\mu = 0$, the PDF is

$$p(y_t \mid \mathbf{y}_{1:t-1}; \varphi_t, \nu) = \left[2^{(1+1/\nu)} \varphi_t \Gamma(1 + 1/\nu)\right]^{-1} \exp\left(-\frac{|y_t/\varphi_t|^\nu}{2}\right)$$

and log-transformed

$$\log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi_t, \nu) = -\left[(1 + 1/\nu)\log(2) + \log(\varphi_t) + \log(\Gamma(1 + 1/\nu))\right] - \frac{|y_t/\varphi_t|^\nu}{2}.$$

Let us now find s_t by using the chain rule and then insert $\varphi_t = \exp(\tilde{\varphi}_t)$:

$$\begin{aligned}s_t &= \frac{\partial \log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)}{\partial \varphi_t} \frac{\partial \varphi_t}{\partial \tilde{\varphi}_t} \\ &= \left[-\frac{1}{\varphi_t} + \nu \frac{|y_t|^\nu / |\varphi_t|^{\nu+1}}{2}\right] \exp(\tilde{\varphi}_t) \\ &= \left[-\frac{1}{\exp(\tilde{\varphi}_t)} + \nu \frac{|y_t|^\nu / \exp(\tilde{\varphi})^{\nu+1}}{2}\right] \exp(\tilde{\varphi}_t) \\ s_t(y_t, \nu, \tilde{\varphi}_t) &= \nu \frac{|y_t|^\nu}{2\exp\left(\nu \tilde{\phi}_t\right)} - 1\end{aligned}$$

Now we have everything needed in order to formalize the updating step for the GAS-GED model:

$$\begin{aligned}\varphi_t &= \exp \tilde{\varphi}_t \\ \tilde{\varphi}_t &= \omega + \alpha s_{t-1} + \beta \tilde{\varphi}_{t-1} \\ s_t &= \nu \frac{|y_t|^\nu}{2 \exp\left(\nu \tilde{\phi}_t\right)} - 1\end{aligned}$$

In order to find the unconditional expected value of s_t, $E(s_t)$, we consider s_t in its general expression $s_t = \frac{\partial \log p(y_t | \mathbf{y}_{1:t-1}; \varphi, \nu, \mu)}{\partial \tilde{\varphi}_t}$.

$$E(s_t) = E\left(\frac{\partial \log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)}{\partial \tilde{\varphi}_t}\right)$$

We compute the expectation by using the PDF of y_t in the conditional distribution, since s_t is the score of the conditional distribution.

$$\begin{aligned} E(s_t) &= \int_y \frac{\partial \log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)}{\partial \tilde{\varphi}_t} p(y_t \mid \mathbf{y}_{1:t-1}; \varphi_t, \nu)\, dy \\ \text{(Using the chain rule)} &= \int_y \frac{\partial p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)}{\partial \tilde{\varphi}_t} \frac{1}{p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)} p(y_t \mid \mathbf{y}_{1:t-1}; \varphi_t, \nu)\, dy \\ &= \int_y \frac{\partial p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)}{\partial \tilde{\varphi}_t} dy \\ \text{(Leibniz integral rule)} &= \frac{\partial \int_y p(y_t \mid \mathbf{y}_{1:t-1}; \varphi, \nu)\, dy}{\partial \tilde{\varphi}_t} \\ &= \frac{\partial 1}{\partial \tilde{\varphi}_t} \\ &= 0 \end{aligned}$$

Similarly, we can find $E(s_t \mid \mathcal{F}_{t-1})$, by using the argument that s_t is the score of the conditional distribution, meaning that conditioning again on the same information set will give us $E(s_t)$ by the law of iterated expectation. That is

$$E(s_t \mid \mathcal{F}_{t-1}) = E(s_t) = 0.$$

1.2 Log-likelihood and parameter restrictions

In question 1, we have already established the individual log-likelihood contributions as

$$\log p(y_t \mid \mathbf{y}_{1:t-1}; \varphi_t, \nu) = -\left[(1 + 1/\nu)\log(2) + \log(\varphi_t) + \log\left(\Gamma(1 + 1/\nu)\right)\right] - \frac{|y_t/\varphi_t|^\nu}{2}.$$

Adding up, we get the log-likelihood function, where we note that φ_t is a function of (ω, α, β) such that the log-likelihood is also a function of these parameters:

$$\log L(\omega, \alpha, \beta, \nu) = -\sum_{t=1}^{T} \left[(1 + 1/\nu)\log(2) + \log(\varphi_t(\omega, \alpha, \beta)) + \log\left(\Gamma(1 + 1/\nu)\right)\right] + \frac{|y_t/\varphi_t(\omega, \alpha, \beta)|^\nu}{2}.$$

This function must by numerically maximized with respect to the parameters ω, α, β and ν. Let us now look at the parameter constraints needed.
In order to establish covariance stationarity of the sequence $\{\tilde{\varphi}_t\}$, we must have that the mean and variance of $\tilde{\varphi}_t$ is constant over time

$$\begin{aligned} E\left(\tilde{\varphi}_t\right) &= E\left(\tilde{\varphi}_{t+h}\right) = \mu_{\tilde{\varphi}} < \infty \\ \text{Var}\left(\tilde{\varphi}_t\right) &= \text{Var}\left(\tilde{\varphi}_{t+h}\right) = \sigma^2_{\tilde{\varphi}} \in (0;\infty) \end{aligned}$$

The expected value of $\tilde{\varphi}_t$ is

$$\begin{aligned} E\left(\tilde{\varphi}_t\right) &= E\left(\omega + \alpha s_{t-1} + \beta \tilde{\varphi}_{t-1}\right) \\ &= \omega + \beta E\left(\tilde{\varphi}_{t-1}\right) \\ E\left(\tilde{\varphi}\right) &= \frac{\omega}{1-\beta} \end{aligned}$$

In order for this to be finite, we need to impose $\beta \neq 1$. The variance of $\tilde{\varphi}_t$ is

$$\begin{aligned} \text{Var}\left(\tilde{\varphi}_t\right) &= \text{Var}\left(\omega + \alpha s_{t-1} + \beta \tilde{\varphi}_{t-1}\right) \\ &= \alpha^2 \text{Var}\left(s_t\right) + \beta^2 \text{Var}\left(\tilde{\varphi}_{t-1}\right) \\ \text{Var}\left(\tilde{\varphi}\right) &= \frac{\alpha^2 E\left(s_t^2\right)}{1-\beta^2} \end{aligned}$$

If we assume that $E\left(s_t^2\right) < \infty$, we need to restrict $|\beta| < 1$ in order to have a finite positive variance. This means that the two restrictions we need to make are

$$\nu > 0 \text{ and } |\beta| < 1.$$

1.3 Scale parameter response to special cases

Let us write the two functions

$$\begin{aligned} f_1\left(y_t\right) &= s\left(y_t, 1, 0\right) = \frac{|y_t|}{2} - 1 \\ f_2\left(y_t\right) &= s\left(y_t, 2, 0\right) = y_t^2 - 1 \end{aligned}$$

Let us compare the value of the two and see when one is bigger than the other. We have $f_1\left(y_t\right) > f_2\left(y_t\right)$ if

$$\begin{aligned} \frac{|y_t|}{2} - 1 &> y_t^2 - 1 \\ \frac{|y_t|}{2} &> y_t^2 \\ |y_t| &< \frac{1}{2} \end{aligned}$$

Intuitively, this makes sense as ν indicates the fatness of the tails. For $\nu = 1$ the tails are fatter than for $\nu = 2$. In a distribution with fatter tails, we will have that the scale parameter will respond the most (for given α at least) if we observe $|y_t| < \frac{1}{2}$ since it has less probability mass close to the mean (of zero), which means that it will correct its scale/variance relatively more than the distribution with thinner tails and, hence, more probability mass around its mean. In contrast, we see that for $|y_t| > \frac{1}{2}$, the thinner-tailed distribution will have to correct its scale (and hence variance) the most, in order to make the observation more "likely" to occur.

2 Computational part

2.1 Estimation of GAS-GED

In order to estimate a GAS-GED model, I have written two central codes. One that filter the scale parameter and, hence, the standard deviation for given parameter values and data. Another that uses maximum likelihood to estimate the parameter values in the filter. I have commented a bit on the input and outputs of each function in the R code.
However, I want to comment on a particular decision of mine here. I have initialized the recursions for the scales by setting the standard deviation equal to the sample standard deviation estimated from the first 150 observations of the process. This is done for two reasons: Firstly, in the empirical part, we want to compare the filtered volatilities from a GAS-GED model with those from a GARCH(1,1) estimated by the *ugarchfit* function in the *rugarch* package. This function initializes the process, by default, to the sample variance based on the whole sample. However, as mentioned later, I have changed this as well. Secondly, the variance seem to increase significantly towards the financial crises such that initializing the process by the sample variance based on the whole sample, will be a very misleading guess on the initial variance.

2.2 Estimation of the Value-at-Risk (VaR)

In order to estimate the Value-at-Risk, I have chosen to code up functions to compute it by either alternative. However, I have only used the function from option i) in the description, in my empirical analysis. It really does not matter, which function I use, as the results are the same.

3 Empirical part

For this part of the exercise, I have used the the *rugarch* package in R, as well as coded some functions of my own. See the "appendix" in wiseflow for the codes.

(a) GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1)

For this exercise, we are considering three types of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, the regular GARCH(1,1) model, the EGARCH(1,1) and the GJR-GARCH(1,1).
The regular GARCH(1,1) model is

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

as we have assumed zero mean of y_t.
The EGARCH(1,1) (exponential GARCH) model is

$$\log\left(\sigma_t^2\right) = \omega + \alpha \varepsilon_{t-1} + \gamma\left(|\varepsilon_{t-1}| - E\left(|\varepsilon_{t-1}|\right)\right) + \beta \log\left(\sigma_{t-1}^2\right),$$

where $\varepsilon_{t-1} = \frac{y_{t-1}}{\sigma_{t-1}}$ such that it is consistent with the exercise description. We see that this model incorporates the information about both the sign and size of the lagged residuals, whereas the GARCH only include information of the size as we square the "forcing" variables.
The GJR-GARCH reads

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \gamma S_{t-1}^- y_{t-1}^2 + \beta \sigma_{t-1}^2$$

with

$$S_{t-1}^{-} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}.$$

This model differs from the regular GARCH by assuming that there is asymmetric effects from positive and negative shocks, just as the EGARCH model does. It incorporates this assumption via the indicator S_{t-1}^{-}.

(i) Estimation of the models

Now let us estimate the three models introduced above. I have done so by using the *rugarch*-package available in R. One noticeable thing here is that I have decided how the package initializes the recursions of the volatility. If I go with the default option, it sets the starting value of the recursion to the sample variance. However, if we look at the plots in Figure 1, we see that setting the initial variance equal to the sample variance over the whole sample, we will make a inappropriate guess as we incorporate the observations over the financial crisis, which had unusual extreme observations. Instead, I have chosen to base the initial variance on the sample variance over the first 150 observations only. I do so by using the $fit.control$ option in *ugarchspec* and setting $rec.init = 150$. This way to initializing the recursions actually gives me a higher likelihood value and, hence, smaller BIC for each of the three specifications. Using this procedure I get the estimates for the three models as in Table 1 below.

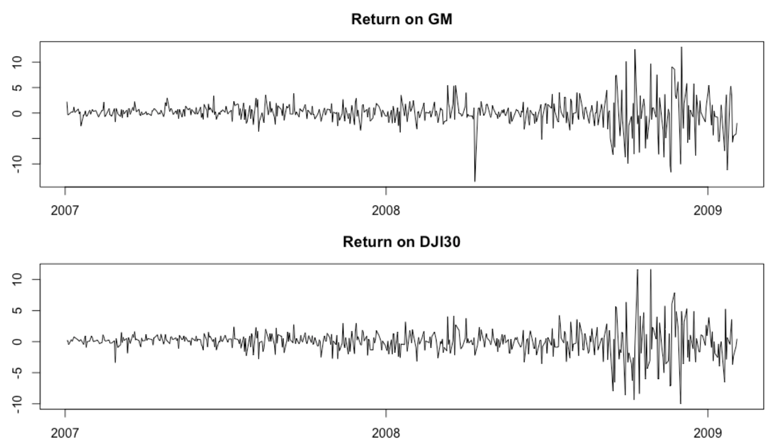

Figure 1: Return on GM and DJI30

	GARCH(1,1)		EGARCH(1,1)		GJR-GARCH(1,1)	
	GM	*DJI30*	*GM*	*DJI30*	*GM*	*DJI*
ω	0.0479	0.0200	0.0230	0.0314	0.0619	0.0291
α	0.1172	0.0822	-0.0537	-0.1662	0.0863	0.0000
β	0.8818	0.9168	0.9891	0.9735	0.8730	0.9150
γ			0.1634	0.0886	0.0794	0.1649
ν	1.2356	1.3364	1.2407	1.4405	1.2211	1.3974

Table 1: Coefficient estimates for each model

(ii) Comparison of filtered volatilities

Now we have estimated each of the models. Let us then look at the filtered volatilities of the three models. Figure 2 and 3 below shows the filtered volatilities (in terms of standard deviations) from the three models for GM and DJI30, respectively.

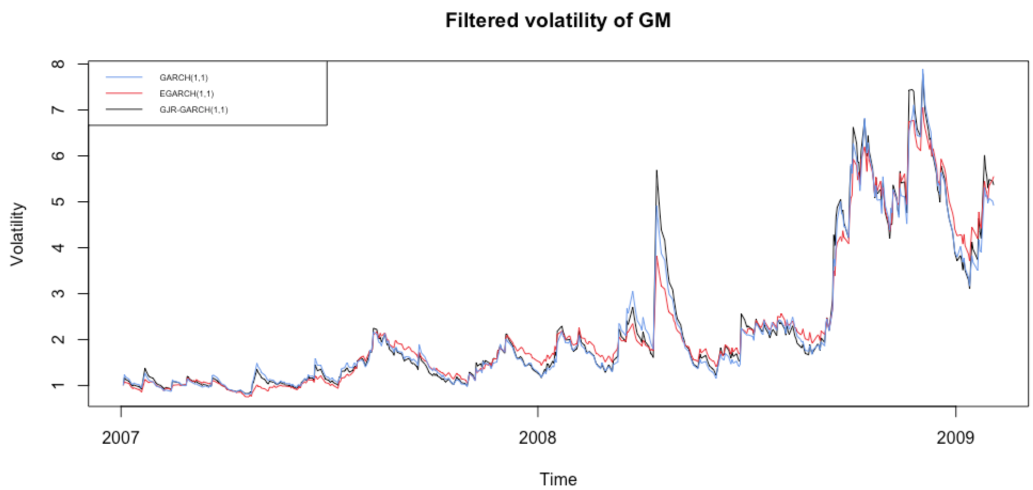

Figure 2: Filtered volatilities of GM

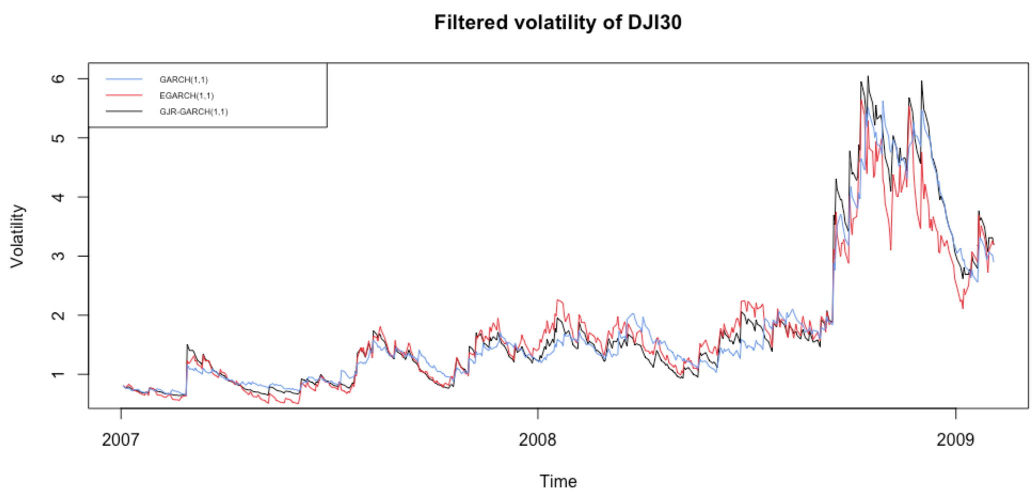

Figure 3: Filtered volatilities of DJI30

We see that the three specifications produce very similar volatilities. They do off course filter slightly different values due to the properties mentioned in the previous section . One thing to note is that all model suggest higher volatility during the start of the financial crisis, which is also to be expected when looking at the returns in Figure 1.

(iii) Selecting the best model

In order to select the best model we compute the average BIC for each model. Table 2 shows exactly this information criteria. Comparing the values, we see that for the GM, we have that the EGARCH(1,1) model is the best, as it has the lowest BIC. Actually, we also have that the EGARCH(1,1) has the best fit for DJI30.

		GARCH(1,1)	EGARCH(1,1)	GJR-GARCH(1,1)
BIC	*GM*	4.108	4.099	4.117
	DJI30	3.656	3.625	3.636

Table 2: Average BIC values

(b) Estimation of GAS-GED

(i) Comparison of filtered volatilities with GAS-GED and GARCH(1,1)

For this question, I have estimated the GAS-GED model by using the code written in the computational part. The estimated coefficients of the GAS-GED model is showed in table 3.

	GAS-GED	
	GM	*DJI30*
ω	0.0030	0.0022
α	0.0507	0.0473
β	0.9952	0.9916
ν	1.1702	1.3219

Table 3: Coefficient estimates for the GAS-GED model

On thing to note here is that the the model estimates fatter tails for GM than for DJI30 for their conditional distributions. The reason is that GM is a single asset, while the DJI30 is an index, i.e. a weighted average of multiple assets. Hence, ceteris paribus, it will be more likely to find extreme values of the single asset than for the index. This result is also seen if we look at the estimates from each of the three GARCH specifications in Table 1. To extract the filtered volatilities of the GAS-GED using these estimates, I use the filtering-codes written in the computational part. Then I compare these volatilities with those reported by the use of *rugarch* package. The comparison is plotted in Figure 4 . I will interpret on the plots in the last subquestion for this exercise.

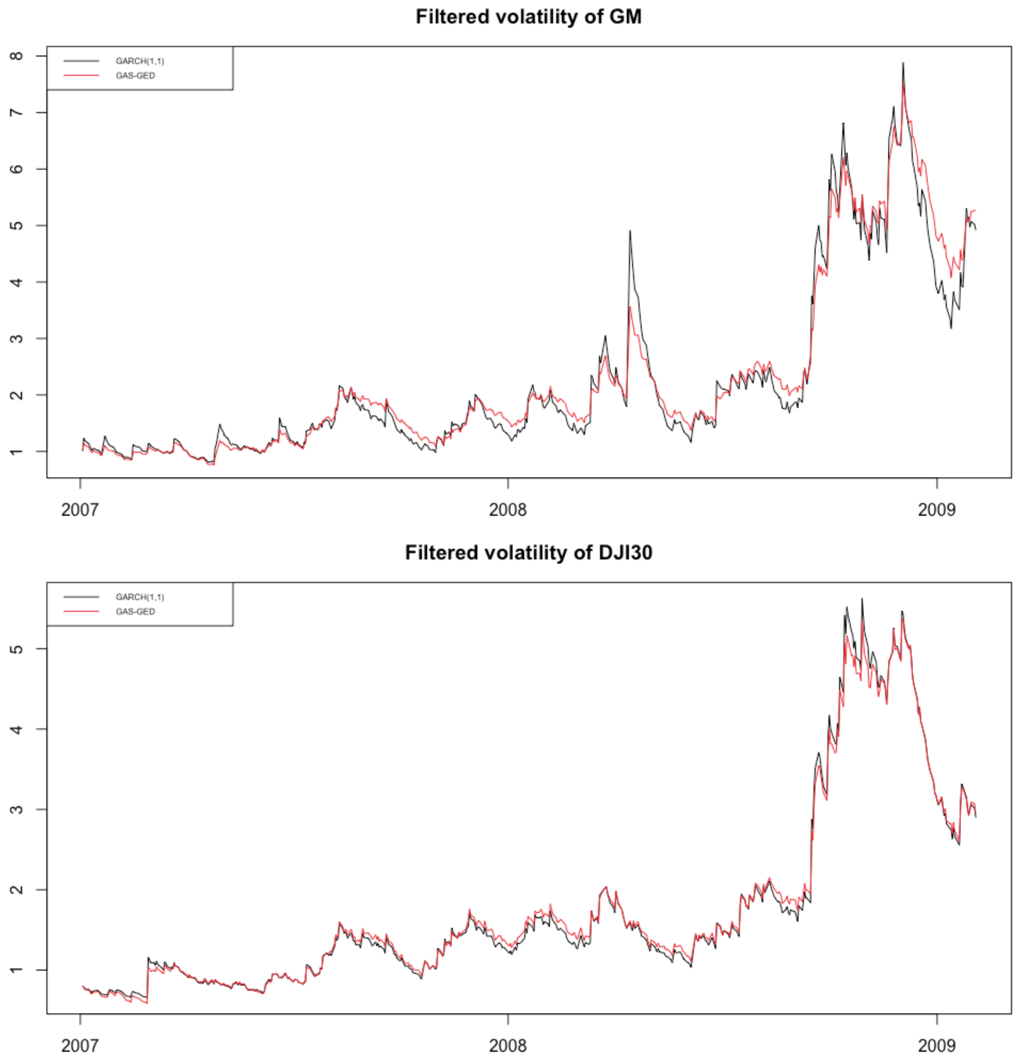

Figure 4: Filtered Volatilities of GM and DJI30 using GAS-GED and GARCH(1,1)

(ii) Comparison of VaR with GAS-GED and GARCH(1,1)

Now let us use the filtered volatilities to compute a VaR at each point in time for both models and assets at a confidence level of $\alpha = 0.01$ and $\alpha = 0.05$. Note that I have used the definition as given in the question. That is, we define the VaR as the maximum loss that the asset will give in 95% of the times, or equivalently, the minimum loss it will give in 5% of the times. I have used the codes written in the second question in the computational part. Figure 5 contains plots of the estimated VaRs.

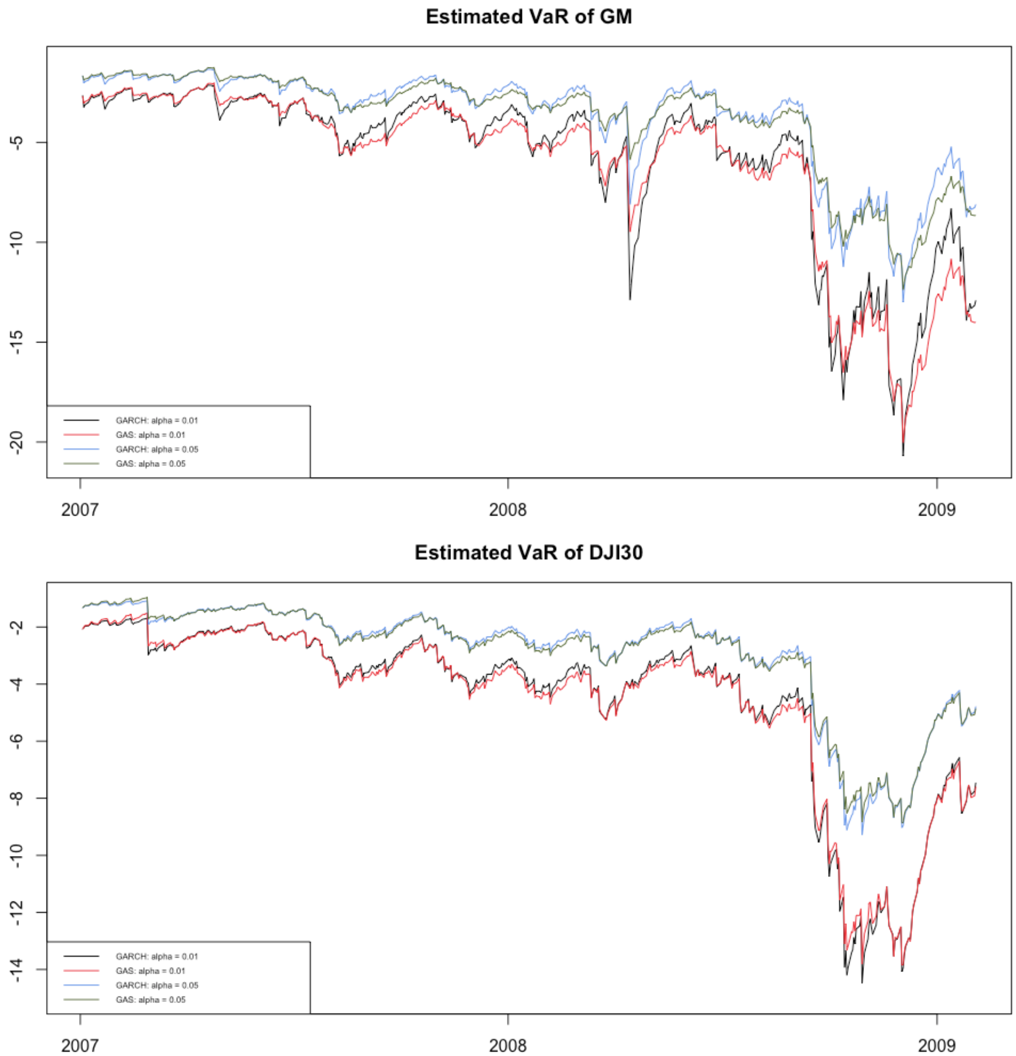

Figure 5: Estimated VaR of GM and DJI30 using GAS-GED and GARCH(1,1)

We see that the VaR resembles the patterns (just in the opposite direction) of the filtered variances from Figure 4 pretty well. This is also rather intuitive as we have assumed zero-mean of the returns, meaning that the parameters driving the VaR is the volatility and the tail-thickness, since the VaR is basically a percentile of the distribution and the tail-thickness parameter is very similar across the two model specifications. Furthermore, we also see that the lower α level gives a lower VaR, which is also obvious due to the definition of percentiles.

(iii) Selection of the best model and general conclusions on the difference between the models

Now that we have seen both the filtered volatilities and associated VaR, let us determine whether the GAS-GED or the GARCH(1,1) perform the best for each of the assets. In order to do so, let us look at the average BIC for each model.

	GM		*DJI30*	
	GAS-GED	GARCH(1,1)	GAS-GED	GARCH(1,1)
Average BIC	4.0992	4.1083	3.6574	3.6557

Table 4: Average BIC

From Table 4 we see that for GM, the GAS-GED is the preferred model as it has the lowest average BIC. For the DJI30 the preferred model is the GARCH(1,1). However, the differences between the values are very small, in particular for DJI30.

The general difference between the two model is the forcing variable of the update equation. In the GAS-GED model the forcing variable is the score, which will give the direction of the update, i.e. the slope of the likelihood at time $t-1$, s_{t-1}. This mean that the GAS-GED incorporates information of the distribution. In contrast, the GARCH(1,1) has the forcing variable y_{t-1}^2. We saw in the last question of the theoretical part that the score in the GAS-GED model reduces to y_{t-1}^2 (at least $y_{t-1}^2 - 1$ is proportional to y_{t-1}^2) when $\nu = 2$. Then we saw that this score has a higher impact on the scale parameter for observations, y_{t-1}, over a given threshold, at least for a given α. That is, for more extreme observations of y_{t-1} compared to its variance, the higher the effect will be on the variance in the following period in the GARCH(1,1) model compared to the GAS-GED model. This is also what we see in Figure 4. For GM we see that every time there is a jump in y_t compared to its variance, the GARCH(1,1) makes a bigger jump in the volatility than the GAS-GED does. For DJI30, the result is not as clear since the ν is estimated to a larger value and an index like DJI30 does not make as high jumps as an individual asset does.

c) Computation of the covariance structure and MVP between GM and DJI30

i) + ii) Computation of the covariance matrix for the assets using each model

For this exercise, I have written a function that compute the covariance matrix over time assuming a constant correlation. I then use the function on the filtered volatilities of each of the models. In Figure 6 I have plotted the covariance between the assets, i.e. the off-diagonal element in the covariance matrix, which is off course a symmetric matrix.

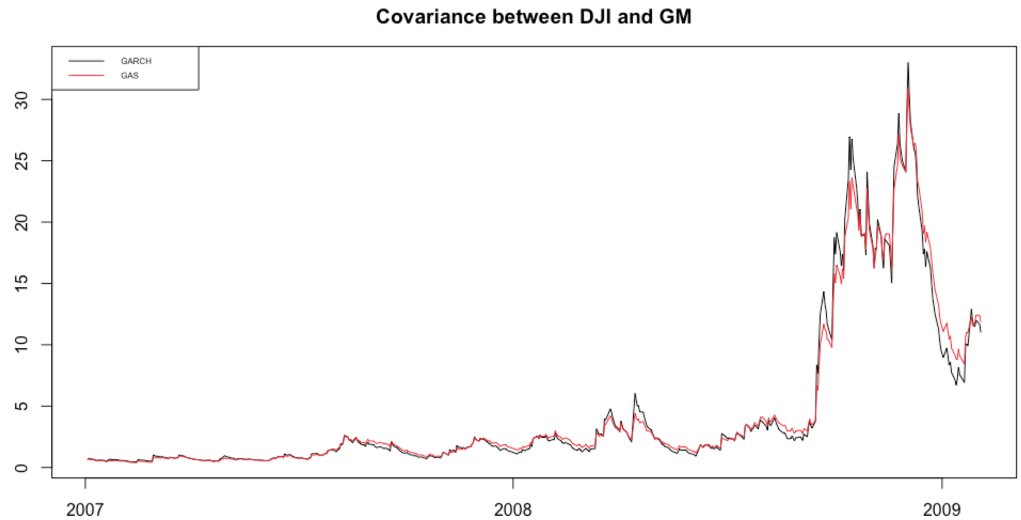

Figure 6: Covariance between GM and DJI30

We see that the covariance between the asset increases in the last part of 2008. This is only due to the increases in volatility that we saw earlier as we have assumed constant correlation. One could have expanded the analysis to see how the correlation changes over the financial crisis by using a Dynamic Conditional Correlation model instead.

iii) + iiii) The MVP of the two assets under each model specification

For this part of the exercise, I have written a function that computes the MVP of the assets based on the covariance matrices over time. In Figure 7, I have plotted the portfolio weight on GM, ω_t, in

$$y_t = \omega_t y_t^{GM} + (1 - \omega_t)\, y_t^{DJI30}$$

such that the weight on DJI30 is just $(1 - \omega_t)$. The figure contains the portfolio weight from both model specifications as well.

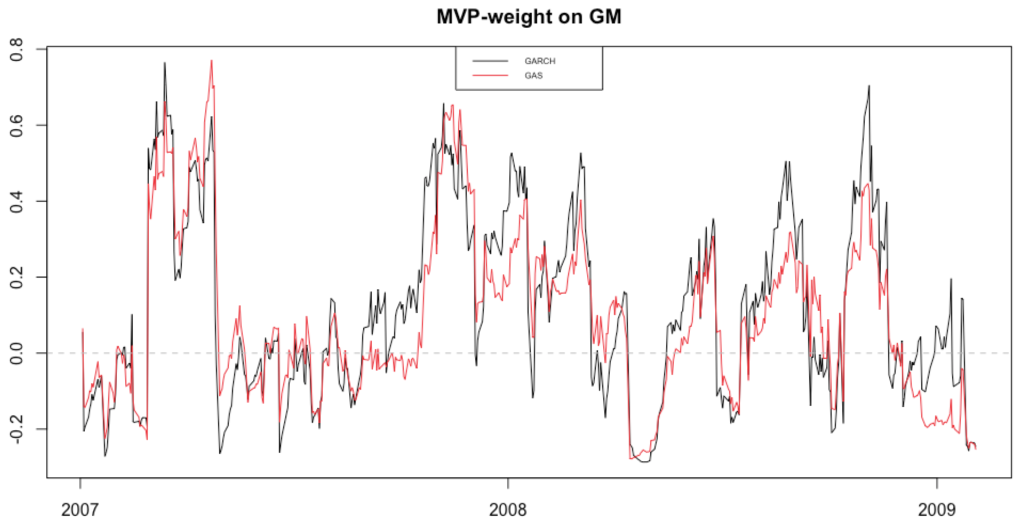

Figure 7: Minimum variance portfolio weight on GM