FINANCIAL ECONOMETRICS

EXAM

Andreas Kracht Frandsen*

201506176

* Department of Mathematics, Aarhus University, andreas.kracht.frandsen@post.au.dk.

Typeset using Palatino Linotype

Contents

List of Figures

List of Tables

Preface

This document answers the Exam of Financial Econometrics. The following conventions are used throughout the document:

Well known rules from probability theory are marked as *italic*.

R packages are marked as **bold**.

R functions are marked as ***bold and italic***.

References to articles or books are marked as (Engle and Gallo, 2006).

Matrices including some chosen vectors are written in bold like $\boldsymbol{Y}$.

The code used in the document is available here Code_Flow_38.R.

1 Exercise 1

In this exercise we examine the Volaility Index (also known as: CBOE VIX) which is an index of volatility computed by the Chicago Board Options Exchange that is freely available to the public quoted in percentage points. We obtain it through Yahoo Finance, which is available through the **quantmod** R package. It is a commonly known measure of volatility in the US equity market and is used by both institutional and private investors. It is restrained to only take on positive values since it is a volatility measure. Market models are highly dependent on the modelling and prediction of the VIX.

1.1 Theoretical Part

We have that Y_t is the VIX at time $t > 0$ and consider the following model

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{G}(\mu_t, a), \qquad t = 1, \dots, T,$$

where $\mathcal{G}(\mu_t, a)$ is the Gamma distribution with mean $\mu_t > 0$ and scale $a > 0$, with probability density function

$$p(y_t \mid \mathcal{F}_{t-1}) = \frac{1}{\Gamma(a)} a^a y_t^{a-1} \mu_t^{-a} \exp -a \frac{y_t}{\mu_t}.$$

Where we use the parameterization of the Gamma distribution proposed by (Engle and Gallo, 2006). $\Gamma(\cdot)$ is the Gamma function. In this parameterization we have the following

$$\mathbb{E}(Y_t \mid \mathcal{F}_{t-1}) = \mu_t, \quad \mathbb{V}(Y_t \mid \mathcal{F}_{t-1}) = \frac{\mu_t^2}{a}.$$

Utilizing the Generalized Autoreggresive Score Model, the updating equation or so called filter for the above mean, μ_t, becomes

$$\mu_t = \omega + \alpha u_{t-1} + \beta \mu_{t-1}, \tag{1.1}$$

where $u_t = S_t \nabla_t$.

It shall be noted that no link function is used, as no link function is given in the assignment. This could potentially create problems, as μ_t could become negative for certain values of ω, α and β. To obtain u_t we calculate the Score function, $\nabla_t = \nabla(y_t; \mu_t)$, but first we notice that

$$\begin{aligned}
\ln\left(p\left(y_t \mid \mathcal{F}_{t-1}; \mu\right)\right) &= \ln\left(\frac{1}{\Gamma\left(a\right)} a^a y_t^{a-1} \mu_t^{-a} \exp\left(-a\frac{y_t}{\mu_t}\right)\right) \\
&= \ln\left(a^a y_t^{a-1} \mu_t^{-a} \exp\left(-a\frac{y_t}{\mu_t}\right)\right) - \ln\left(\Gamma\left(a\right)\right) \\
&= a \ln\left(a\right) + \left(a-1\right) \ln\left(y_t\right) - a\left(\frac{y_t}{\mu_t} + \ln\left(\mu\right)\right) - \ln\left(\Gamma\left(a\right)\right) \\
&= \ell_t. \qquad (1.2)
\end{aligned}$$

Now we take the first derivative of Equation (1.2) wrt. μ_t, to obtain the Score function

$$\begin{aligned}\nabla_t &= \frac{\partial \ln\left(p\left(y_t \mid \mathcal{F}_{t-1};\mu\right)\right)}{\partial \mu_t} \\ &= \frac{\partial \ell_t}{\partial \mu_t} \\ &= \frac{a\left(y_t-\mu_t\right)}{\mu_t^2}. \end{aligned} \tag{1.3}$$

This is the Score of the conditional distribution $p(y_t \mid \mathcal{F}_{t-1};\mu_t)$, it gives the direction of the update to μ_t, as in the well known Newton-Raphson algorithm. Next we calculate the Fisher Information Matrix, $\mathcal{I}_t(\mu_t)$, since the scaling is defined as

$$S_t = \mathcal{I}_t^{-d} = \mathcal{I}_t^{-\frac{1}{2}}.$$

The parameter d can be selected on the basis of several reasonings: likelihood criteria, theory arguments, computational arguments, etc. We set $d=\frac{1}{2}$ on the basis of the assignment guidelines. This quantity scales the Score in order to account for the curvature of the likelihood at time t. Using Equation (1.3) we obtain

$$\begin{aligned}\mathcal{I}_t &= \mathbb{E}\left(\nabla_t^2 \mid \mathcal{F}_{t-1}\right) \\ &= \mathbb{E}\left(\frac{a^2(y_t-\mu_t)^2}{\mu_t^4} \mid \mathcal{F}_{t-1}\right) \\ &= \mathbb{E}\left(\frac{a^2y_t^2}{\mu_t^4}+\frac{a^2\mu_t^2}{\mu_t^4}-\frac{2a^2y_t\mu_t}{\mu_t^4} \mid \mathcal{F}_{t-1}\right) \\ &= \mathbb{E}\left(\frac{a^2y_t^2}{\mu_t^4}+\frac{a^2}{\mu_t^2}-\frac{2a^2y_t}{\mu_t^3} \mid \mathcal{F}_{t-1}\right) \\ &= \frac{a^2\mathbb{E}\left(y_t^2 \mid \mathcal{F}_{t-1}\right)}{\mu_t^4}+\frac{a^2}{\mu_t^2}-\frac{2a^2\mathbb{E}\left(y_t \mid \mathcal{F}_{t-1}\right)}{\mu_t^3} \\ &= \frac{a^2\frac{\mu_t^2(1+a)}{a}}{\mu_t^4}+\frac{a^2}{\mu_t^2}-\frac{2a^2\mu_t}{\mu_t^3} \\ &= \frac{a^2\frac{(1+a)}{a}}{\mu_t^4}+\frac{a^2}{\mu_t^2}-\frac{2a^2}{\mu_t^2} \\ &= \frac{a^2}{\mu_t^2}+\frac{a}{\mu_t^2}+\frac{a^2}{\mu_t^2}-\frac{2a^2}{\mu_t^2} \\ &= \frac{a}{\mu_t^2}.\end{aligned}$$

Where we use the fact that a is constant for all t, μ_t is measurable wrt. $\mathcal{F}_{t-1}$, and in the sixth equality uses that $\mathbb{E}(Y_t^2 \mid \mathcal{F}_{t-1}) = \frac{\mu_t^2(1+a)}{a}$. Now it follows that the scaling becomes

$$S_t = \mathcal{I}_t^{-\frac{1}{2}} = \left(\frac{a}{\mu_t^2}\right)^{-\frac{1}{2}} = \frac{\mu_t}{\sqrt{a}}.$$

And we can obtain the term u_t as

$$u_t = S_t\nabla_t = \frac{\mu_t}{\sqrt{a}}\frac{a(y_t-\mu_t)}{\mu_t^2} = \frac{\sqrt{a}(y_t-\mu_t)}{\mu_t}. \tag{1.4}$$

We can insert Equation (1.4) into Equation (1.1) to obtain the final filter

$$\mu_t = \omega + \alpha\left(\frac{\sqrt{a}(y_{t-1}-\mu_{t-1})}{\mu_{t-1}}\right)+\beta\mu_{t-1}.$$

Thus the updating equation consist of: a intercept, a score coefficient times the scaled direction, and a term consisting of an autoregressive coefficient times the previous value. We have know established the GAS model in our above setup.

To find the log likelihood we use the common density $p(y_t \mid \mathcal{F}_{t-1})$ and it's logarithm as found above

$$\begin{aligned}\mathcal{L}\left(Y_{1:T} \mid \phi\right) &= \ln\left(p\left(y_1;\phi\right)\right) + \sum_{t=2}^{T} \ln\left(p\left(y_t \mid \mathcal{F}_{t-1};\phi\right)\right) \\ &= \sum_{t=1}^{T} \ell_t \\ &= T \cdot \ln\left(\frac{a^a}{\Gamma\left(a\right)}\right) + \left(a-1\right)\sum_{t=1}^{T} \ln\left(y_t\right) - a\sum_{t=1}^{T} \ln\left(\mu_t\right) + \frac{y_t}{\mu_t}.\end{aligned}$$

Where $\mu_t = \mu_t(\phi)$ and with parameter vector $\phi' = (\alpha, \beta, a, \omega)$.

1.2 Computational Part

The relevant code can be found in Appendix file 'Code_Flow_38.R', attached with the document through WiseFlow. The code is fully documented with comments throughout all the functions.

To estimate the GAS-GAMMA model as discussed in the above Theoretical Part (and similiarly for the following models in the Empirical Part), we create three functions ***GASGAMMA_Filter***, ***NegLogLikelihood*** and ***Estimate_GASGAMMA***.

GASGAMMA_Filter is the filter for the GAS-GAMMA model. It takes the vector of values and a vector of parameters as input. Here we take use of the updating equation for μ_t established in the Theoretical Part. Through a loop we compute the next value of μ_t, from $t = 1, \ldots, T$, where we set

$$\mu_1 = \mathbb{E}(\mu_t) = \frac{\omega}{1-\beta},$$

thus the first value is set to the unconditional expectation of μ_t. Because of the possibility of negative and/or zero values we make an *if* statement checking for negative and/or zero values. Further we compute the log likelihood associated with the parameters and the values of μ_t, using the established function $L\left(Y_{1:T} \mid \phi\right)$ as deduced in the Theoretical Part.

NegLogLikelihood is the helper function for finding the maximum likelihood estimates of our parameters. It takes the vector of values and a vector of parameters as input. It computes the negative log likelihood

$$N\mathcal{L} = -\mathcal{L}\left(Y_{1:T} \mid \phi\right).$$

Estimate_GASGAMMA estimates the GAS-GAMMA model by first finding maximum likelihood estimates of our parameters, which are obtained by optimizing the negative log likelihood. The used optimizer, ***gosolnp***, is available through the **Rsolnp** R package, it randomly chooses starting values for the parameter vector constrained to the lower and upper bounds. After convergence of a solution the final parameters are then feeded to the ***GASGAMMA_Filter*** to obtain the final filtered values of μ_t and the log likelihood value. We also compute the Bayesian Information Criterion which penalizes models with many parameters it is defined as

$$\text{BIC} = \ln(n)k - 2\hat{\mathcal{L}}\left(Y_{1:T} \mid \hat{\phi}\right).$$

Thus it serves as a measure of choosing the 'best' model. A lower BIC is favourable to a high BIC.

Similar functions are created for the rest of the models described in the Empirical Part.

1.3 Empirical Part

Using Yahoo Finance data of the VIX spanning from 2010-01-01 to 2019-01-01 we estimate the GAS model using the functions written in code 'Code_Flow_38.R' and described in Section 1.2. It shall be noted that the

Table 1.1: The first 5 observations of VIX.

VIX.Open	VIX.High	VIX.Low	VIX.Close	VIX.Volume	VIX.Adjusted
21.68	21.68	20.03	20.04	0	20.04
20.05	20.13	19.34	19.35	0	19.35
19.59	19.68	18.77	19.16	0	19.16
19.68	19.71	18.70	19.06	0	19.06
19.27	19.27	18.11	18.13	0	18.13

maximum likelihood estimates obtained are highly dependent upon the initial values set for each parameter, though using the ***gosolnp*** function which randomly sets starting values, as described above, seem to converge always.

Table 1.1 show the five first observations. We will only use the adjusted column.

Figure 1.1 show the evolution of the time series.

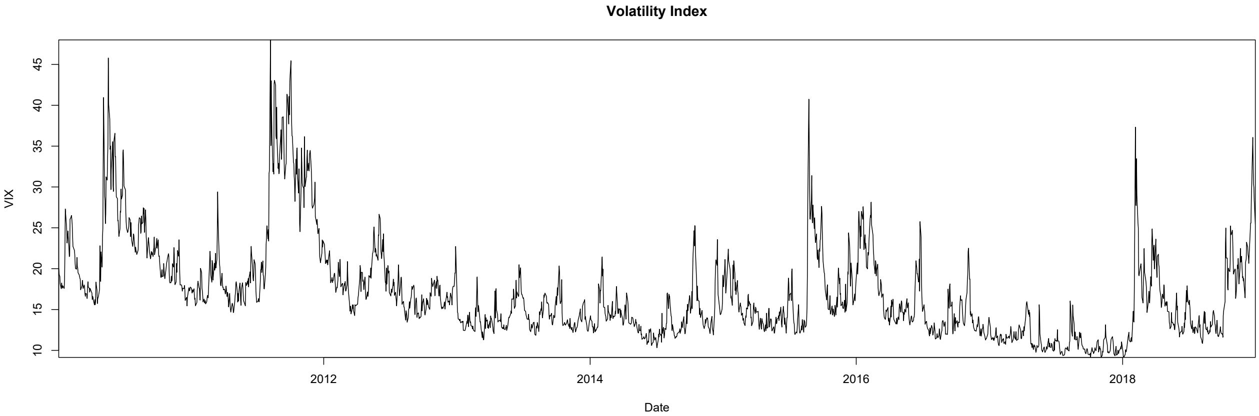

Figure 1.1: VIX from 2010-01-01 to 2019-01-01.

The spikes occur at events such as the Eurozone Crisis, the U.S. Debt-Ceiling Crisis, Brexit, the 2016 U.S. Election, Trump impeachment, Trade-war uncertainty, etc.

1.3.1 GAS

To obtain the GAS-GAMMA model for our VIX data, we feed the relevant quantities to the relevant functions, using the constraints

$$\omega \in [-0.5, 0.5], \quad \alpha \in [0.001, 1.5], \quad \beta \in [0.01, 0.999] \quad \text{and} \quad a \in [0.1, 300],$$

as given in the assignment text. The maximum likelihood estimates of our parameters are

$$\hat{\omega} = 0.499, \quad \hat{\alpha} = 1.198, \quad \hat{\beta} = 0.972 \quad \text{and} \quad \hat{a} = 154.287.$$

One may see that the intercept $\hat{\omega}$ is close to the upper bound, though experimenting with a higher bound, didn't yield better results. We obtain a log likelihood of -3817.897 and a BIC of 7666.694.

1.3.2 Constrained GAS

Considering the constrained version of the GAS-GAMMA model, we have to set the extra constraint

$$\mu_t = \mu \quad \forall t.$$

We set the value of μ to the unconditional expectation

$$\mu := \frac{\omega}{1-\beta} = \mathbb{E}(\mu_t),$$

Thus in this constrained model there are three parameters to estimate ω, β and a, and we never update μ. Functions as the ones desribed above in Section 1.2 have been made to accomplish the task of estimating in this model.

Using the same constraints as in the GAS-GAMMA model for the relevant parameters

$$\omega \in [-0.5, 0.5], \quad \beta \in [0.01, 0.999] \quad \text{and} \quad a \in [0.1, 300].$$

We obtain the following maximum likelihood estimates of our parameters are

$$\hat{\omega} = 0.393, \quad \hat{\beta} = 0.977 \quad \text{and} \quad \hat{a} = 10.396.$$

It shall be noted that the intercept estimate is a bit volatile. We obtain a log likelihood of -6905.022 and a BIC of 13833.22. Thus the model performance is worse than the GAS-GAMMA model. This was also expected since the mean is constant throughout time, and at the same time we only estimate one parameter less. Thus the BIC penalizes the model slightly less, but still becomes higher than in the GAS-GAMMA model.

1.3.3 MEM

The Multiplicative Error Models suggested by (Engle and Gallo, 2006) are used to model positive valued series such as the VIX, and it serves well to use as a comparison to the GAS models above. (Engle and Gallo, 2006) show that one-month-ahead forecasts made from a MEM model match well the market-based volatility measure provided by the VIX.

We again assume that

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathcal{G}(\mu_t, a),$$

and the updating equation for the mean, μ, is given as

$$\mu_t = \kappa + \eta y_{t-1} + \phi \mu_{t-1}.$$

Thus the updating equation consist of: a intercept, a coefficient times the value of VIX, and a term consisting of an autoregressive coefficient times the previous value.

We set the value of μ_1 to the unconditional expectation

$$\mu_1 := \frac{\kappa}{1 - \eta - \phi} = \mathbb{E}(\mu_t),$$

Thus in this model there are four parameters to estimate κ, η, ϕ and a. Functions as the ones desribed above in Section 1.2 have been made to accomplish the task of estimating in this model.

Using the constraints

$$\kappa \in [0.1, 10], \quad \eta \in [0.01, 0.99], \quad \phi \in [0.01, 0.99] \quad \text{and} \quad a \in [0.1, 300].$$

as given in the assignment text. The maximum likelihood estimates of our parameters are

$$\hat{\kappa} = 0.410, \quad \hat{\eta} = 0.945, \quad \hat{\phi} = 0.033 \quad \text{and} \quad \hat{a} = 162.832.$$

We obtain a log likelihood of -3756.546 and a BIC of 7543.991. Compared to the GAS-GAMMA model we observe a slightly better BIC that yields the logic that more information is obtained directly through the values of Y_t than through the scaled direction u_t.

1.3.4 Comparison

Table 1.2 gives a summary of the log likelihood, estimates and BIC of every model examined. Comparing the GAS-GAMMA models solely the highest log likelihood is reported by the GAS-GAMMA with time varying μ. Comparing all of the models the highest log likelihood is associated with the MEM-GAMMA model.

Comparing the parameters in the GAS-GAMMA models, the contrained version has similar parameter estimates except for the scale a, which is compressed to overcome the non-flexibility of the model.

Table 1.2: Summary of all models.

	$\hat{\mathcal{L}}$	$\hat{\omega}\ (\hat{\kappa})$	$\hat{\alpha}\ (\hat{\eta})$	$\hat{\beta}\ (\hat{\phi})$	$\hat{a}$	BIC
GAS-GAMMA	-3817.898	0.4999508	1.1979337	0.9718808	154.2635	7666.695
GAS-GAMMA-C	-6905.022	0.3713394	0.0000000	0.9781855	10.3961	13833.219
MEM-GAMMA	-3756.546	0.4099280	0.9454647	0.0325607	162.8638	7543.991

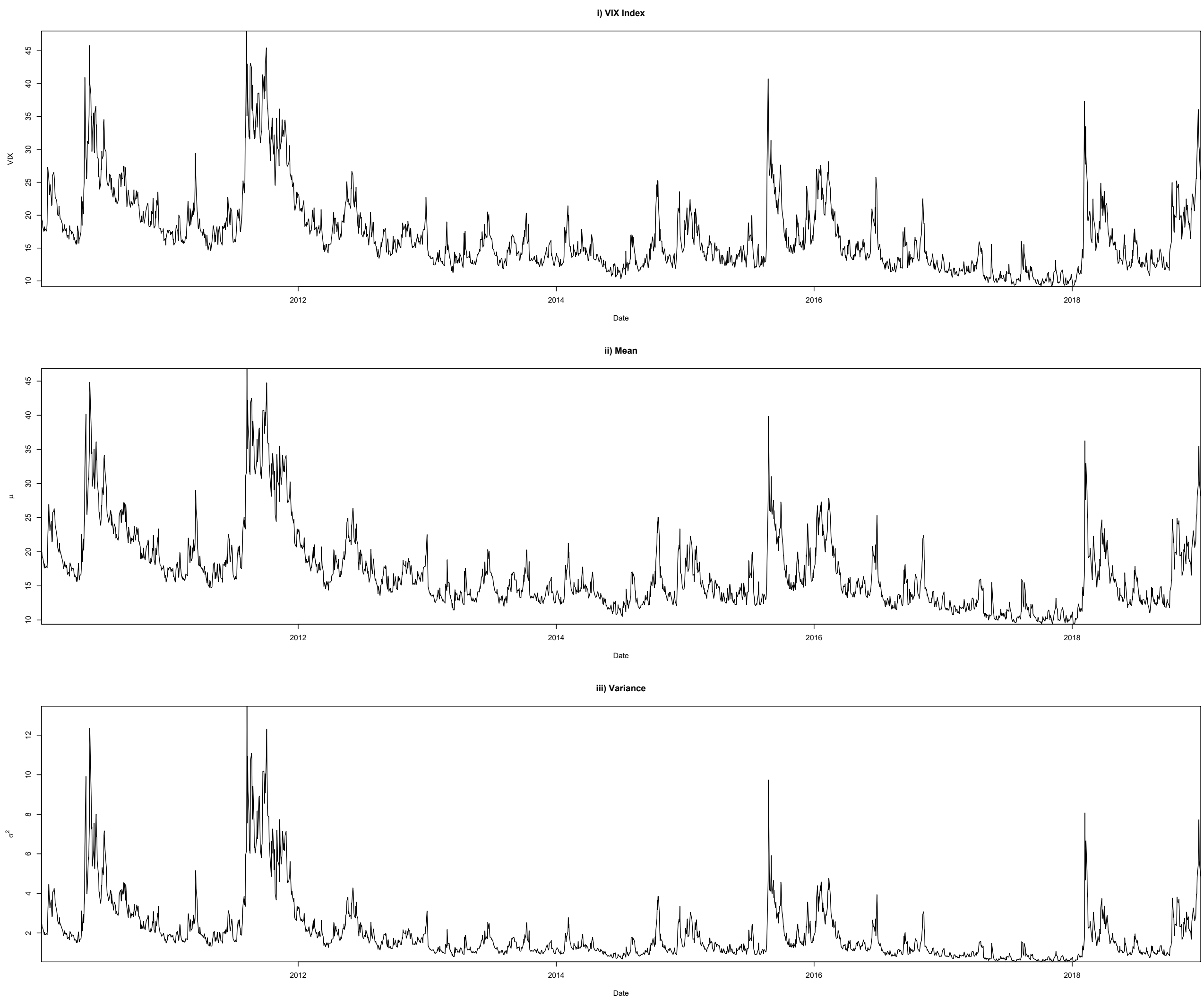

Figure 1.2: i) VIX, ii) Filtered mean, iii) Filtered variance.

Comparing the autoregressive coefficients we see that ϕ from the MEM-GAMMA model is noticeably lower than the β's in the GAS-GAMMA models.

Given the fact that the MEM-GAMMA model yields best performance from a BIC perspective we have chosen to continue with this model. Thus Figure 1.2 gives the summary of the MEM-GAMMA model.

Here the conditional variance at time t is computed as

$$\mathbb{V}(Y_t \mid \mathcal{F}_{t-1}) = \frac{\mu_t^2}{a}.$$

All the three time series plots are similar as expected.

2 Exercise 2

In this exercise we examine two indices: S&P 500 (commonly known as S&P500, with ticker ˆGSPC) which is a stock market price index that that measures the stock performance of 500 large companies listed on exchanges in the United States. Many institutional and private investors consider it to be one of the best representations of the U.S. stock market. Dow Jones Industrial Average (commonly known as DOW, with ticker ˆDJI) which is a stock market price index that that measures the stock performance of 30 large companies listed on exchanges in the United States. Many consider the DOW to not be a good representation of the U.S. stock market and consider the S&P500, which also includes the 30 components of the DOW, to be a better representation of the U.S. stock market. We obtain both through the dataset given in the assignment, spanning from 2007-01-03 to 2019-01-01, thus the Financial Crisis is included.

2.1 Theoretical Part

We consider the bivariate random vector at time $t > 0$

$$\boldsymbol{Y}_t = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix}, \qquad t = 1, \ldots, T,$$

where $Y_{1,t}$ and $Y_{2,t}$ are the given SP&500 and DOW returns. We assume that Y_t is a bivariate zero mean Gaussian

$$\boldsymbol{Y}_t \mid \mathcal{F}_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t).$$

We want to derive the Dynamic Conditional Correlations model for the above setup, which parameterize the conditional correlations directly. Using DCC models have a computational advantage over multivariate GARCH models, in the sense that the number of parameters to be estimated in the correlation process is independent of the number of series to be correlated.

Empirical findings by (Engle, 2002) show that the DCC is often the most accurate when compared to a simple multivariate GARCH and several other estimators including the exponential moving average and 100-day moving average. His findings show that this is true whether the performance criterion is mean absolute error, diagnostic tests, or tests based on value at risk calculations.

To obtain the DCC model in our setup we rewrite the covariance matrix, $\boldsymbol{\Sigma}_t$, such that we obtain

$$\boldsymbol{Y}_t \mid \mathcal{F}_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{D}_t^{\frac{1}{2}} \boldsymbol{R}_t \boldsymbol{D}_t^{\frac{1}{2}}),$$

here $\boldsymbol{D}_t$ is the diagonal matrix with elements

$$\sigma_{it}^2 = \mathbb{V}(Y_{it} \mid \mathcal{F}_{t-1}),$$

which we assume follows a Generalized Autoregressive Conditional Heteroskedasticity process of order $p = q = 1$ or more commonly known as the GARCH$(1,1)$ process for $i = 1, 2$. The updating equation for σ_{it}^2 thus becomes

$$\sigma_{it}^2 = \omega + \alpha Y_{it-1} + \beta \sigma_{it-1}^2,$$

where ω, α and β are unknown coefficients. $\boldsymbol{R}_t$ is the correlation matrix with elements

$$\rho_{ijt} = \text{cor}(Y_{it}, Y_{jt} \mid \mathcal{F}_{t-1}).$$

Lastly the DCC model assumes that the correlation matrix, $\boldsymbol{R}_t$, can be decomposed into

$$\boldsymbol{R}_t = \tilde{\boldsymbol{Q}}_t^{-\frac{1}{2}} \boldsymbol{Q}_t \tilde{\boldsymbol{Q}}_t^{-\frac{1}{2}},$$

here $\tilde{\boldsymbol{Q}}_t$ is a diagonal matrix which contains the diagonal elements of the conditional covariance matrix, $\boldsymbol{Q}_t$, which is given as

$$\boldsymbol{Q}_t = \bar{\boldsymbol{Q}}(1-a-b) + a\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}' + b\boldsymbol{Q}_{t-1},$$

where a and b are unknown coefficients, $\boldsymbol{\eta}_t = \boldsymbol{D}_t^{-\frac{1}{2}}\boldsymbol{Y}_t$ thus the standardized residuals and $\bar{\boldsymbol{Q}}$ is fixed to to the empirical correlation of $\boldsymbol{\eta}_t$. To ensure that $\boldsymbol{Q}_t$ is positive definite we muse impose the constraint $a + b < 1$, which will also guarantee that our model is mean reverting.

To obtain the log likelihood we use the Gaussianity of our returns to obtain

$$\begin{aligned}
\ln(L_T) &= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{\Sigma}_t|) + \boldsymbol{Y}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{Y}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{D}_t^{\frac{1}{2}}\boldsymbol{R}_t\boldsymbol{D}_t^{\frac{1}{2}}|) + \boldsymbol{Y}_t'\boldsymbol{D}_t^{-\frac{1}{2}}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-\frac{1}{2}}\boldsymbol{Y}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{R}_t|) + \boldsymbol{\eta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{D}_t|) + \ln(|\boldsymbol{R}_t|) + \boldsymbol{\eta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t + \boldsymbol{Y}_t'\boldsymbol{D}_t^{-\frac{1}{2}}\boldsymbol{D}_t^{-\frac{1}{2}}\boldsymbol{Y}_t - \boldsymbol{\eta}_t'\boldsymbol{\eta}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{D}_t|) + \boldsymbol{Y}_t'\boldsymbol{D}_t^{-1}Y_t) - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\eta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t - \boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \ln(|\boldsymbol{R}_t|)).
\end{aligned}$$

Where we use that $|\boldsymbol{ABC}| = |\boldsymbol{A}||\boldsymbol{B}||\boldsymbol{C}|$ and that $\ln(|\boldsymbol{AB}|) = \ln(|\boldsymbol{A}||\boldsymbol{B}|) = \ln(|\boldsymbol{A}|) + \ln(|\boldsymbol{B}|)$.

Thus we have factorized the log likelihood into two parts: a volatility component and a correlation component. These are

$$\begin{aligned}
L_V(\theta) \equiv \ln(L_{V,T}(\theta)) &= -\frac{1}{2}\sum_{t=1}^{T}(2\ln(2\pi) + \ln(|\boldsymbol{D}_t|) + \boldsymbol{Y}_t'\boldsymbol{D}_t^{-1}\boldsymbol{Y}_t) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{2}\left(\ln(2\pi) + \ln(\sigma_{it}^2) + \frac{r_{it}^2}{\sigma_{it}^2}\right),
\end{aligned}$$

and

$$L_C(\theta, \Phi) \equiv \ln(L_{C,T}(\theta, \Phi)) = -\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\eta}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t - \boldsymbol{\eta}_t'\boldsymbol{\eta}_t + \ln(|\boldsymbol{R}_t|)).$$

Here θ denotes the parameter vector in $\boldsymbol{D}_t$ and Φ denotes the parameter vector in $\boldsymbol{R}_t$. Thus

$$L(\theta, \Phi) = L_V(\theta) + L_C(\theta, \Phi).$$

Maximizing the log likelihood becomes a two-step approach. In the first step we find

$$\hat{\theta} = \arg\max\{L_V(\theta)\},$$

and then use this value as given in the next step

$$\max_{\phi} L_C(\hat{\theta}, \Phi).$$

Under reasonable regularity conditions (Engle, 2002), consistency of the first step will ensure consistency of the second step.

We can derive the Constant Conditional Correlations model as a special case of the DCC model. DCC only differs in allowing $\boldsymbol{R}$ to be time dependent. Thus the CCC model assumes

$$\boldsymbol{Y}_t \mid \mathcal{F}_{t-1} \sim N(\boldsymbol{0}, \boldsymbol{D}_t^{\frac{1}{2}} \boldsymbol{R} \boldsymbol{D}_t^{\frac{1}{2}}),$$

where $\boldsymbol{R}$ is the correlation matrix containing the conditional correlations

$$\mathbb{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' \mid \mathcal{F}_{t-1}) = \boldsymbol{D}_t^{-\frac{1}{2}} \boldsymbol{\Sigma}_t \boldsymbol{D}_t^{-\frac{1}{2}}.$$

To obtain the same $\boldsymbol{Q}_t$ throughout time we then set $a = b = 0$, such that $\boldsymbol{Q}_t = \boldsymbol{Q}$.

2.2 Computational Part

The relevant code can be found in Appendix file 'Code_Flow_38.R', attached with the document through WiseFlow. The code is fully documented with comments throughout all the functions.

To estimate the GARCH(1,1) model as discussed in the above Theoretical Part, we create four functions ***GARCHFilter***, ***ObjectiveFunction***, ***ineqfun_GARCH_WS*** and ***EstimateGARCH***.

GARCHFilter is the filter for the GARCH(1,1) process. It takes a vector of values and the parameters as input. Here we take use of the updating equation for σ_{it}^2 established in the Theoretical Part. Through a loop we compute the next value of σ_{it}^2, from $t = 1, \dots, T$, where we set the first variance to the empirical variance of the first 10% of the observations. Further we compute the log likelihood associated with the parameters and the values of σ_{it}^2, using the established function $L_V(\theta)$ as deduced in the Theoretical Part.

ObjectiveFunction is the helper function for finding the maximum likelihood estimates of our GARCH(1,1) parameters. It takes the vector of values and a vector of parameters as input. It computes the negative log likelihood

$$N\mathcal{L} = -L_V(\theta).$$

ineqfun_GARCH_WS serves as a basis to evaluate the inner part of the inequality constraints that need to be satisfied to impose weak stationarity, which is

$$0 < \alpha + \beta < 1.$$

EstimateGARCH estimates the GARCH(1,1) model by first finding maximum likelihood estimates of our parameters, which are obtained by optimizing the negative log likelihood. The used optimizer, ***solnp***, is available through the **Rsolnp** R package, we set initial starting values, essentially we set starting value for α and β and set ω to target the unconditional variance of the GARCH(1,1) model. After convergence of a solution the final parameters are then feeded to the ***GARCH_Filter*** to obtain the final filtered values of σ_{it}^2 and the log likelihood value. We also compute the Bayesian Information Criterion and the standardized residuals $\boldsymbol{\eta}_t$.

To estimate the DCC (CCC) model as discussed in the above Theoretical Part (and similiarly for the following models in the Empirical Part), we create two functions ***DCCFilter*** and ***Estimate_DCC***.

DCCFilter is the filter for the DCC (CCC) model. It takes the vector of standardized residuals, the parameters and the unconditional correlation as input. Here we take use of the equations derived for $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ in the Theoretical Part. Further we compute the log likelihood associated with the parameters and the values of $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$, using the established function $L_C(\theta, \Phi)$ as deduced in the Theoretical Part.

Estimate_DCC estimates the DCC (CCC) model by first finding maximum likelihood estimates of our parameters, if we are in the DCC model, which are obtained by optimizing the negative log likelihood. The used optimizer, ***solnp***, is available through the **Rsolnp** R package, we set initial starting values, essentially we set starting value for a and b. After convergence of a solution the final parameters are then feeded to the ***DCCFilter*** to obtain the final filtered values of $\boldsymbol{Q}_t$, $\boldsymbol{R}_t$ and the log likelihood value. We also compute the Bayesian Information Criterion.

2.3 Empirical Part

Using the updated dataset of the S&P500 and DOW returns spanning from 2007-01-03 to 2019-01-01 we estimate the DCC and CCC model using the functions written in code 'Code_Flow_38.R' and described in Section 2.2.

Table 2.1 show the five first observations.

Table 2.1: The first 5 observations of the returns of SP500 and DOW.

GSPC	DJI
0.1038474	0.0287223
-0.6292226	-0.6854062
0.2028799	0.1845763
-0.0705868	-0.0762133
0.1749414	0.1849128

Figure 2.1 show the evolution of the S&P500 time series.

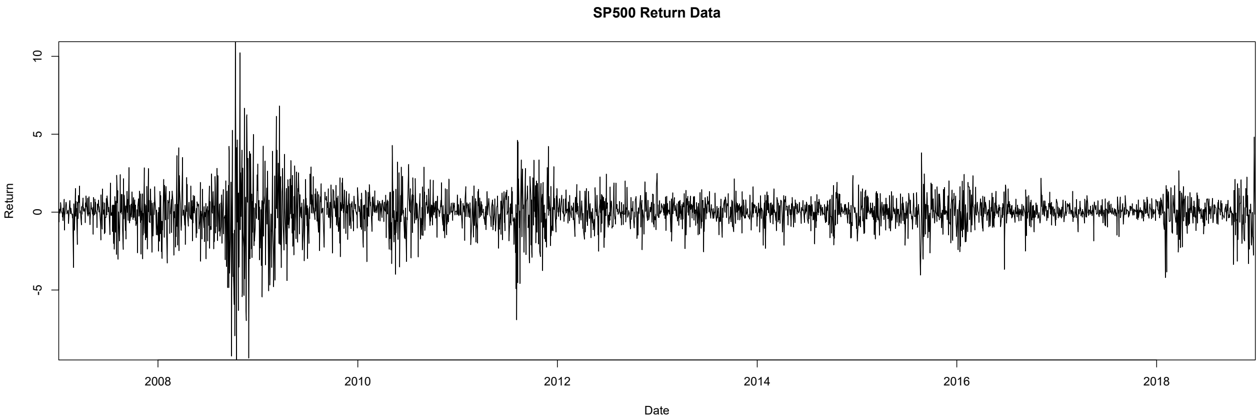

Figure 2.1: SP500 returns from 2007-01-03 to 2019-01-01.

Figure 2.2 show the evolution of the DOW time series.

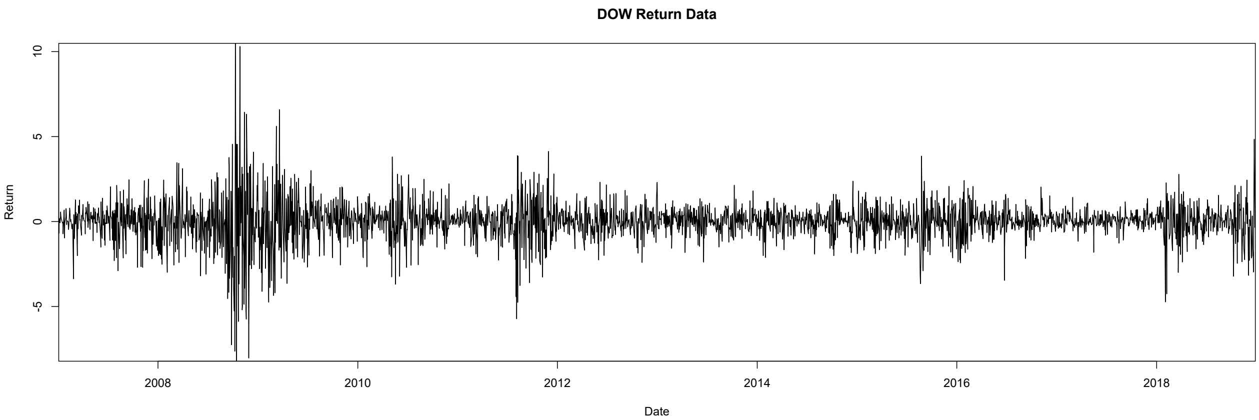

Figure 2.2: DOW returns from 2007-01-03 to 2019-01-01.

2.3.1 DCC

To obtain the DCC model for our data, we feed the relevant quantities to the relevant functions. The maximum likelihood estimates of our parameters are

$$\hat{\omega} = \begin{pmatrix} 0.025 \\ 0.023 \end{pmatrix}, \quad \hat{\alpha} = \begin{pmatrix} 0.125 \\ 0.13085261 \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} 0.857 \\ 0.851 \end{pmatrix}, \quad \hat{a} = 0.070 \quad \text{and} \quad \hat{b} = 0.902.$$

We obtain a log likelihood of -3688.162 and a BIC of 7440.425. Since the correlation matrix $\boldsymbol{R}$ is time-varying we will not post the results here. One can check the values throughout time using 'Code_Flow_38.R'.

2.3.2 CCC

To obtain the CCC model for our data, we feed the relevant quantities to the relevant functions setting the boolean parameter 'CCC' to TRUE. Which evaluates our filter using $a = b = 0$ as argued earlier. The maximum likelihood estimates of our parameters are exactly the same as for the DCC, since they are the outcome of the same two GARCH(1,1) processes.

The constant correlation matrix $\boldsymbol{R}$ is given as

$$\boldsymbol{R} = \begin{pmatrix} 1.0000000 & 0.9671498 \\ 0.9671498 & 1.0000000 \end{pmatrix}.$$

Which show a very high correlation. We obtain a log likelihood of -3894.175 and a BIC of 7852.45. Compared to the DCC model we observe a slightly worse BIC that yields the logic that allowing time-dependency in the correlation matrix yields better performance.

Continues on next page.

2.3.3 MVP

To obtain the weights associated with the Minimum Variance Portfolio we consider the risk-averse investors problem at day t

$$\min_{\omega_{t|t+h}} \omega'_{t|t+h}\boldsymbol{\Sigma}_{t|t+h}\omega_{t|t+h}, \qquad \text{s.t.} \quad \omega'_{t|t+h}\ell = 1.$$

The optimal solution can be deduced to

$$\omega'^{*}_{t|t+h} = \frac{\boldsymbol{\Sigma}^{-1}_{t|t+h}\ell}{\ell'\boldsymbol{\Sigma}^{-1}_{t|t+h}\ell}. \tag{2.1}$$

Thus we use the above Equation (2.1) to obtain the weights for the S&P500 Index and DOW throughout time. Figure 2.3 shows the Minimum Variance Portfolio weights using the estimates obtain from the DCC model. Functions and documentation can be found in 'Code_Flow_38.R'.

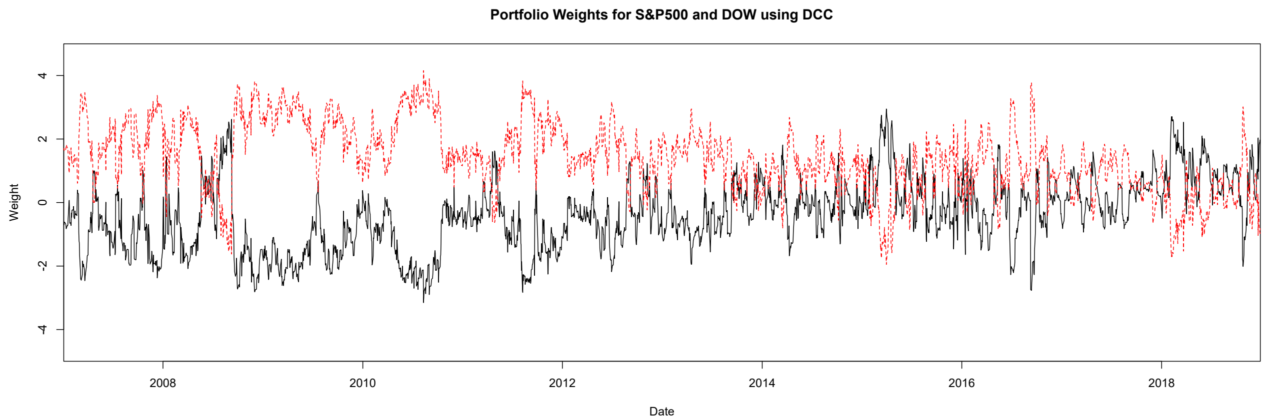

Figure 2.3: Minimum Variance Portfolio Weights using DCC Model. Black line represents SP500 weights, red line represents DOW weights.

One may notice the relatively high shorting positions. Figure 2.4 shows the Minimum Variance Portfolio weights using the estimates obtain from the CCC model.

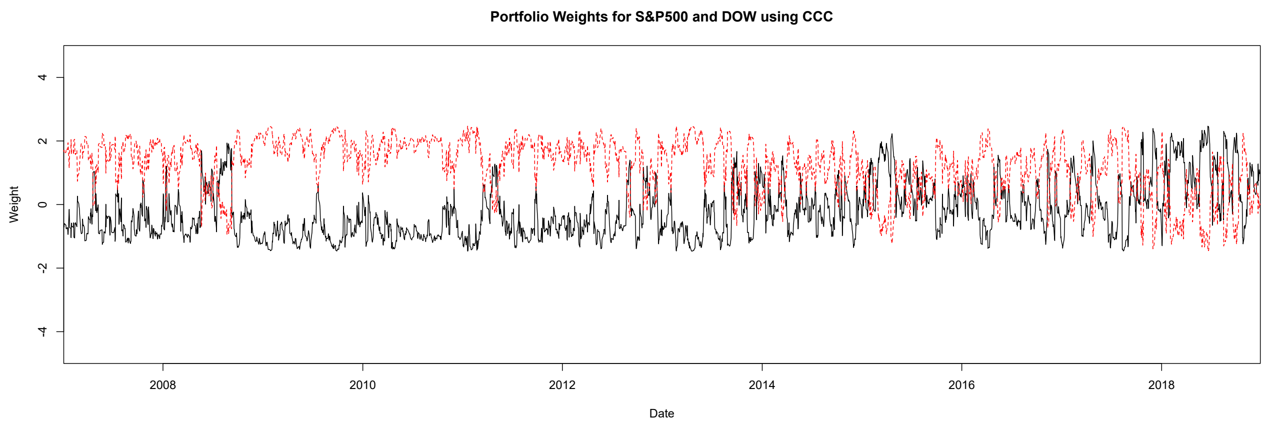

Figure 2.4: Minimum Variance Portfolio Weights using CCC Model. Black line represents SP500 weights, red line represents DOW weights.

Using the CCC model we observe somewhat more stable positions with diminished shorting positions. This makes sense due to the constant correlation matrix. That doesn't allow the returns to be more or less correlated through time.

2.3.4 CoVaR

We want to compute the Conditional Value at Risk, which is the value at risk (VaR) of a financial system conditional on institutions being under distress. It is defined as the α-quantile of the conditional distribution

$$Y_{1t} \mid Y_{2t} \leq \mathrm{VaR}_{Y_{2t}}(\alpha).$$

And can also be defined as

$$\mathrm{CoVaR}_{Y_{1t}|Y_{2t}}(\alpha, \alpha) = F^{-1}_{Y_{1t}|Y_{2t} \leq \mathrm{VaR}_{Y_{2t}}(\alpha)}(\alpha).$$

We can find it as the solution to the equality below

$$F_{Y_{1t},Y_{2t}}(\mathrm{CoVaR}_{Y_{1t}|Y_{2t}}(\alpha, \alpha), \mathrm{VaR}_{Y_{2t}}(\alpha)) = \alpha^2.$$

This is obtained using the *uniroot* function from the **stats** package in R. Figure 2.5 shows the CoVaR using the estimates obtain from the DCC model. Functions and documentation can be found in 'Code_Flow_38.R'.

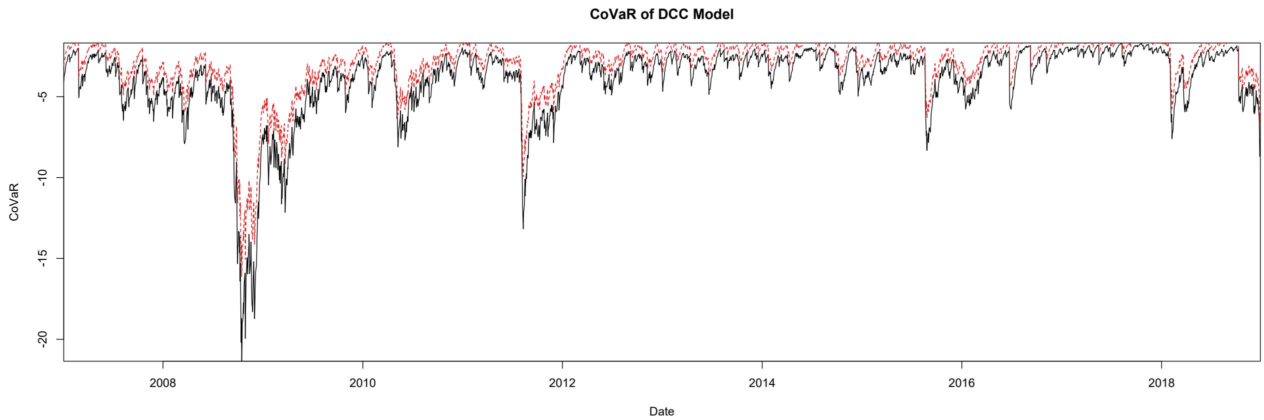

Figure 2.5: CoVaR using DCC Model. Black line represents CoVaR with significance level 0.01, red line represents CoVaR with signifiance level 0.05.

One may notice the Financial Crisis, given the very big drop in CoVaR in 2008. Figure 2.6 shows the CoVaR using the estimates obtain from the CCC model.

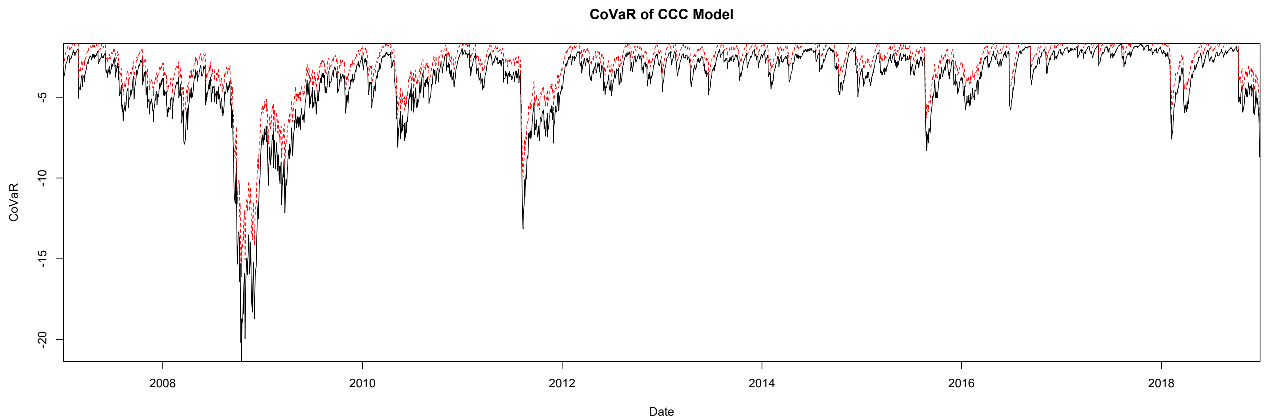

Figure 2.6: CoVaR using CCC Model. Black line represents CoVaR with significance level 0.01, red line represents CoVaR with signifiance level 0.05.

The CoVaR of the CCC model is indeed very similar to DCC model and it's hard to tell where they differ. Thus we have now fully compared the DCC and CCC models.

Bibliography

Catania, L. (2020). Lecture Slides and Theoretical Notes from Financial Econometrics site on Blackboard, Aarhus University.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 3:339–350.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131:3,27.

Genz, A. and Bretz, F. (2009). *mvtnorm: Multivariate Normal and t Distributions*. R Foundation for Statistical Computing.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer-Verlag London Limited.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ryan, J. A. and Ulrich, J. M. (2019). *quantmod: Quantitative Financial Modelling Framework*. R package version 0.4-15.

Ye, Y. (1987). *Interior Algorithms for Linear, Quadratic, and Linearly Constrained Non-Linear Programming.*