1 Models

In this section we will describe the models that have been implemented throughout the process and highlight the changed and simplifications. To make a framework for model development we created a metaclass, so called BaseModel, where we created all the common functions that all these models share. We manage to implement a strategy to create bounds with existing parameter's transformation. Not only give us a unique opportunity to modeling with various parameter transformations, but in optimizations brought as a more stable results. The optimizer that we use the Python's scipy.optimize.minimize function with the setting of the L-BFGS-B method.

1.1 MIDAS

The first model was the MIDAS that was first introduced [10], where they compare distributed lag models with MIDAS regression. The MIDAS model can be described with the expressions used in [11]. Suppose y_t is the low-frequency dependent variable that can be observed once within a time-step t (say, monthly), then $x_t^{(m)}$ the high-frequency explanatory variable can be observed m times during one time-step (say, daily or m = 22). We want to describe the relationship between y_t and $x_t^{(m)}$, in the sense of using lagged observations of $x_t^{(m)}$. The model is the following:

$$y_t = \beta_0 + \beta_1 B(L^{\frac{1}{m}}, \theta) x_t^{(m)} + \varepsilon_t^{(m)} \tag{1}$$

where $B(L^{\frac{1}{m}}, \theta) = \sum_{k=0}^{K} B(k, \theta) L^{\frac{k}{m}}$, where $L^{\frac{k}{m}}$ is a lag operator such that $L^{\frac{1}{m}} x_t^{(m)} = x_{t-\frac{1}{m}}^{(m)}$. The lag coefficients in $B(k, \theta)$ of the corresponding lag operator $L^{\frac{k}{m}}$ are parameterized as a function of a small-dimensional vector of parameters Θ. β_1 is a scale parameter for the lag coefficients

1.1.1 Specification of Weighting Function

In the MIDAS literature there is one weighting function that used the most, namely "Beta" Lag. [11, 8, 9]. For completeness, I mention the others, these are the Exponential Weighting and the Exponential Almon Lag. Beta Lag involves two parameters, $\Theta = (\theta_1, \theta_2)$, and the parametrization:

$$B(k, \theta_1, \theta_2) = \frac{f(\frac{k}{K}), \theta_1, \theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K}), \theta_1, \theta_2)} \tag{2}$$

where

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \tag{3}$$

$$\Gamma(a) = \int_0^{\infty} e^{-x} x^{a-1} dx \tag{4}$$

The following figure will deonstrate how flexiable it is correspond to different parameters:

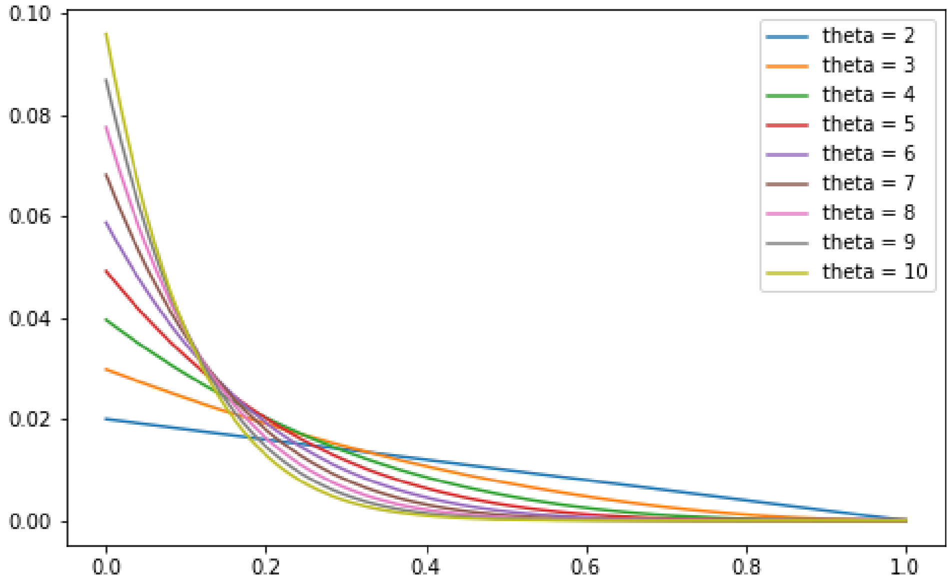

Figure 1: Plot of Beta Lag weighting function in equation 1 with K = 100, $\theta_1 = 1$ and $\theta_2 = 2, ..., 10$

We can see that if we choose to fix $\theta_1 = 1$ and in the case of $\theta_2 > 1$ cause a monoton decliyin weighting structure. This weight function specification provide us positive coefficients, which is crutual when we want to modeling volatility.

1.1.2 Parameter Transformation

In this section we will describe an approach to make parameter estmation more consistant and stabil, it is so called parameter transformation. The main idea behind this strategy is that estimators can treat bounds, but in practice it is much more convenient to transform our parameters. With this approach we can create bounds without explicitly programming to the estimator function. First we describe the transform and the back-transform function, then show how they incoperate to the function that will be estimated. Let denote θ with the parameter that we want to work with:

$$\tilde{\theta} = \begin{cases} \log(\theta) & \text{,if 'pos'} \\ \log(\theta) - \log(1-\theta) & \text{,if '01'} \\ \theta & \text{otherwise.} \end{cases}$$

$$\theta = \begin{cases} \exp(\tilde{\theta}) & \text{,if 'pos'} \\ \frac{1}{1+\exp(-\tilde{\theta})} & \text{,if '01'} \\ \tilde{\theta} & \text{otherwise.} \end{cases}$$

In the log likelihood function instead of calculting with the actual θ, we will make the estimation with θ. Than we transform back as the estimation finished. One issue raise from this estimation strategy, is that the standard error won't be correct. So there is another function called gradient. θ^* marked as the estimated parameters that were previously transformed.

$$gradient = \begin{cases} \exp(\theta^*) & \text{,if 'pos'} \\ \frac{\exp(\theta^*)}{(1+\exp(\theta^*))^2)} & \text{,if '01'} \\ 1 & \text{otherwise.} \end{cases}$$

As the L-BFGS-B method relies ont he approximation to the Hessian matrix of the loss function, se as we take advantage of information matrix equality we can calculate the standard errors easily.

1.1.3 Parameter Estimation

In the parameter estimation we will use the Python's function from scipy.optimize library, called minimize. I applied L-BFGS-B method, this method allow us to define bounds for parameters, and the biggest advantage is that approximate the inverse Hessian matrix. The estimation is happening throughout the sum of squared estimat of error:

$$SSE = \varepsilon^T \varepsilon = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 B(L^{\frac{1}{m}}, \theta) x_t^{(m)})^2 \tag{5}$$

$$\arg \min_{\beta_0, \beta_1, \theta_2} SSE$$

1.1.4 Simulations

The Monte-Carlo simulation was from [[3]]. Suppose we have X_t is an AR(1) process, that:

$$X_{i,t} = \phi X_{i-1,t} + \varepsilon_t$$

where t = 1, ..., T show the low-frequency time-steps, i = 1, ...,I_t is the high-frequency. Set the I_t equals to 22, $\phi = 0.9$ and $\varepsilon_t \sim \mathcal{N}(0,1)$ standard normal variable. The MIDAS equation will be:

$$y_t = \beta_0 + \beta_1 \sum_{k=0}^{K} \xi_k(1.0, \theta) X_{i-k,t} + z_t$$

with the parameters $\beta_0 = 0.1, \beta_1 = 0.3, \theta = 4.0$ and $z_t \sim \mathcal{N}(0, 0.5)$. We made simulations with T = 100, 200, 500. A A simulare approach was described in [[9]] with the difference of simulating quarterly/monthly data and they found out, as they increase the sample ssize the more accurate their parameter estimations will be, furthermore the more parsimonious will be the model's computational cost.

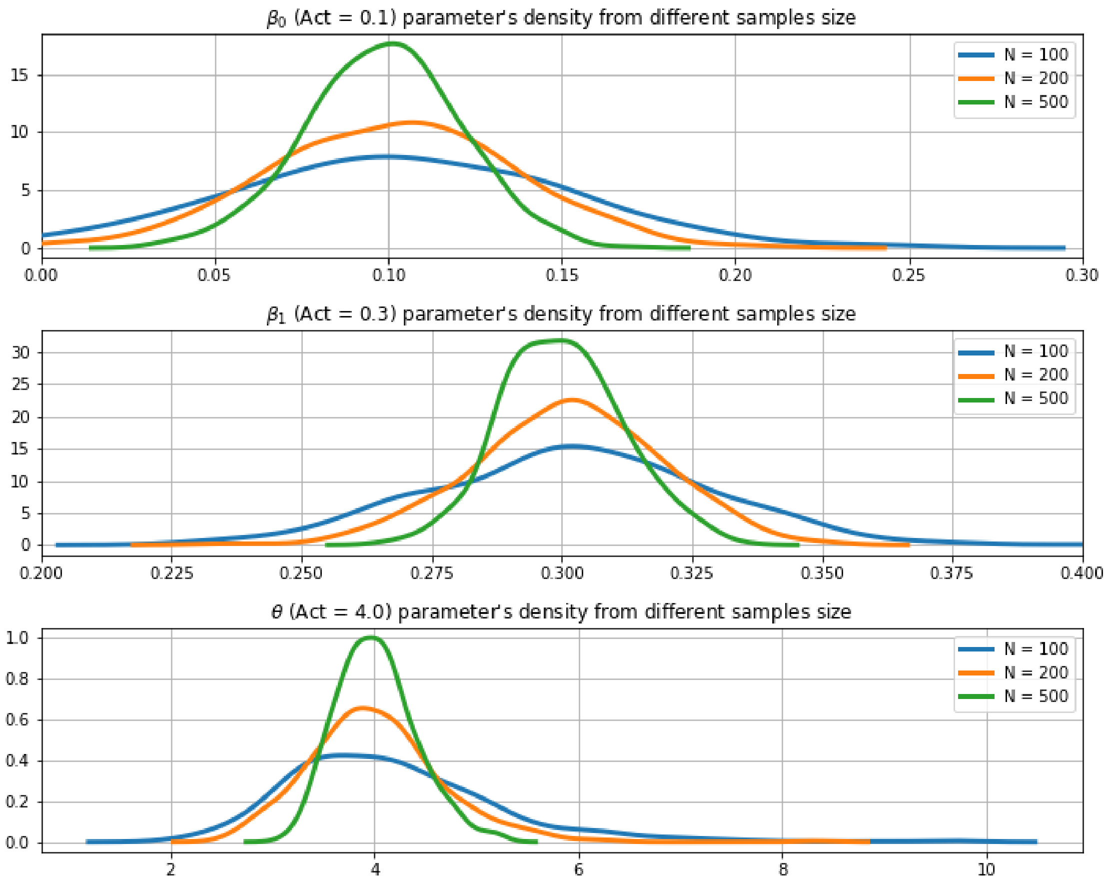

Figure 2: Plot of estimated parameter distributions with sample sizes of 100, 200, 500

1.2 GARCH

In this section, I would like to give a brief overview about GARCH model. The underlying concept was first developed in [7], the ARCH model, where we associated r_t with the daily log return ($r_t = \log P_t - \log P_{t-1}, P_t$is the stock price at time t) for t = 1, ..., T, and assume that it can be written as $r_t = \mu_t \varepsilon_t, \varepsilon_t$ are modelled with ARCH model:

$$\varepsilon_t = \sigma_t Z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$$

where the innovation Z_t are iid random variables with mean 0 and variance 1. Suppose $Z_t \sim \mathcal{N}(0,1)$. The innovation's distribtion can be modelled with various ways, such as Student-t distributed or, the most common, Normally distributed. The parameter constraints are: $\alpha_0 > 0, \alpha_i \geq 0$ This model was extended in [2] to the Generalized ARCH model, where previous values of σ_t^2 are added to the volatility process. This extension create phenomenons that can be observed in markets, such as volatility clustering, where high volatility peroids tends to persist. The GARCH(1, 1) process is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where ε_t are real-valued discrete-time stochastic process and $\mathscr{F}_t$ is the information set of all information up to time t.

$$\varepsilon_t \mid \mathscr{F}_{t-1} \sim \mathscr{N}(0, \sigma_t^2)$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ and $1 > \alpha_1 + \beta_1$ enough wide-sense stacionarity.

$$E(r_t) = 0$$

$$Var(r_t) = E(r_t^2) - E(r_t)^2 = E(\sigma_t^2 Z_t^2) = E(\sigma_t^2)E(Z_t^2)$$

Since $Z_t \sim N(0,1)$, so $E(Z_t^2) = 1$. Then,

$$E(\sigma_t^2) = E(\alpha_0 + \alpha_1 r_t^2 + \beta_1 \sigma_{t-1}^2) = \alpha_0 + \alpha_1 E(r_{t-1}^2 + \beta_1 E(\sigma_{t-1}^2)$$

From $r_t = \sigma_t Z_t$, it is known that $E(r_t^2) = E(\sigma_t^2)$ and for the process to be stationary $E(\sigma_t^2)$ must be a constant for all t:

$$E(\sigma_t^2) = E(\sigma_{t-1}^2) = E(r_{t-1}^2) = \sigma^2$$

the unconditional mean of the volatility process and the unconditional variance of the returns hence,

$$E(\sigma_t^2) = Var(r_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

To examine the tail behavior we have to examine the excess kurtosis, which implies that the forth moment to exist and be finite. The excess kurtosis of r_t with normally distributed innovations is then

$$\begin{aligned}
\frac{E(r_t^4)}{Var(r_t)^2} - 3 &= \frac{3(1 + \alpha_1 + \beta_1)\alpha_0}{(1 - \alpha_1 - \beta_1)(1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2)(\frac{\alpha_0}{1 - \alpha_1 - \beta_1})^2} - 3 \\
&= 3\frac{1 - (\alpha_1 + \beta_1)^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} - 3 \\
&= \frac{2\alpha_1^2}{1 - \alpha_1^2 - (\alpha_1 + \beta_1)^2} > 0
\end{aligned}$$

which means that r_t is fat-tailed, in other words extreme returns can be observed more frequently than they would with normally distributed innovations.

1.2.1 Parameter Estimation

We applied the Quasi-Maximum Likelihood Estimation (henceforth QMLE) for parameter estimation. In general, we assume an underlying distribution, which has some kind of probability density function, and we have θ the set of parameters to be estimated. In the assumption of normal distribution:

$$f(\varepsilon_t|\theta) = \frac{1}{\sqrt{2\pi\hat{\sigma}_t{}^2(\theta)}} exp\left(-\frac{\varepsilon_t^2}{2\hat{\sigma}_t{}^2(\theta)}\right)$$

where ε_t are the innovations in GARCH type models, and $\hat{\sigma_t^2}(\theta)$ are the volatility estimates with given θ parameters. The loglikelihood function will be

$$\mathscr{L}(\theta) = f(\varepsilon_1, ..., \varepsilon_T|\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\hat{\sigma}_t{}^2(\theta)}} exp\left(-\frac{\varepsilon_t^2}{2\hat{\sigma}_t{}^2(\theta)}\right)$$

since the logarithm is a monotonically increasing function, the value which maximize the likelihood function will also maximize its logarithm as well. We can write a sum instead of a product in the log-likelihood function:

$$\log \mathscr{L}(\theta) = -\frac{T}{2}\sum_{t=1}^{T}\left(\log 2\pi + \log \hat{\sigma_t}^2(\theta) + \frac{\varepsilon_t^2}{\hat{\sigma_t}^2(\theta)}\right) \tag{6}$$

The QMLE is then

$$\hat{\theta} = \arg\max_{\theta} \log \mathscr{L}(\theta) = \arg\min_{\theta} -\log \mathscr{L}(\theta) \tag{7}$$

If a function that maximum is a negative number, then we can mupltiple by minus 1 to minimize it. In practical application there are several algorithms to minimize functions, so we use the negative log-likelihood to estimate the models parameters. Another usefully specification of the probility density function is the Student-t distribution. The log-likelihood of a Student-t distributed specification is the following:

$$\log \mathscr{L}(\theta) = -\sum_{t=1}^{T} -\log\Gamma\left(\frac{\nu+1}{2}\right) + \log\Gamma\left(\frac{\nu}{2}\right) + \log\sqrt{2\pi(\nu-2)} + \frac{1}{2}\log \hat{\sigma_t}^2(\theta)$$
$$+\frac{\nu+1}{2}\log\left(1+\frac{\varepsilon_t^2}{\hat{\sigma_t}^2(\theta)(\nu-2)}\right)$$

1.2.2 Simulations

To ensure our implementation will truely estiamte the desired parameters, we perform a Monte-Carlo simulation for the GARCH(1, 1) model with different sample sizes. In the simulation process we assume that, the log returns follow a normal distribution with the variance from the current state's σ_t^2. The results from the simulations estimated by QMLE as it was described in the previously. Formally we can write as

$$\varepsilon_t \sim \mathscr{N}(0, \sigma_t^2) \tag{8}$$

for t = 1, ..., T, where T is the length of the sample size. The results of the parameter estimations is shown:

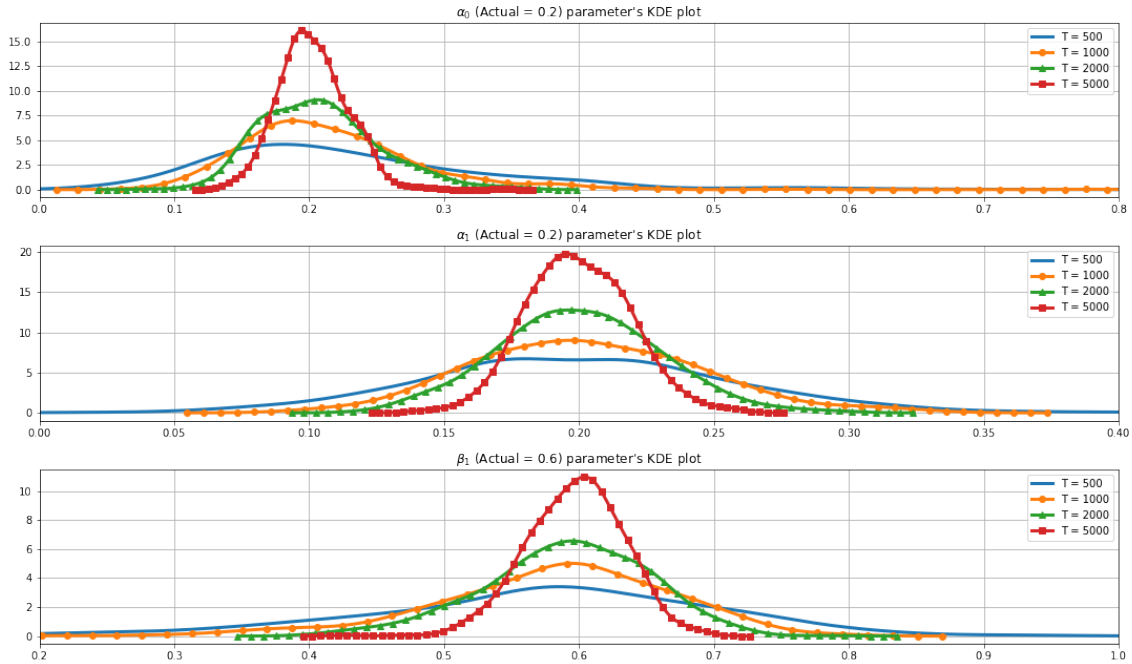

Figure 3: Plot of estimated parameter distributions with sample sizes of 500, 1000, 2000, 5000

The parameter distributions are presented with Kernel Density Estimation (KDE) for the better visualization. The peaks of the distributions are in the theoretical values, moreover, we can see that as we increase the sample size, the more accurate will be the parameter estimations and decrease the variances of the parameters.

1.3 GARCH-MIDAS

In this section we present a a new class of component GARCH model based on the MIDAS regression. This GARCH-MIDAS framework gives us the possibility to incorporate macroeconomic variables sampled at a different frequency. All macroeconomic variables will be in the specification of the long-term component.

Several papres have been publicated in the recent years about he topic of GARCH-MIDAS model. [8] was one of the first to discucc about such a model. They rely on long historical time series and examined what was the impact of adding economic variables to the GARCH model. [1] used this framework to predict future volatility, they incorporate with a principal component approach as well to reduce the dimensions of the explanatory variables. To conduct their results they found that GARCH-MIDAS forecast volatility bettter than a GARCH model. In [4] found that long-term financial volatility behaves counter-cyclically and certain macroeconomic variables help GARCH-MIDAS model to predict better the long-term volatility. [14] showed thet GARCH-MIDAS model has fat-tailed marginal distribution.

The GARCH-MIDAS framework give us the opportunity to incorporate macroeconomic variables sampled at a different frequency. In the recent years many studies showed the effectiveness of this approach, the only pitfall, we found, is the amount of underlying data that we provide for the algorithm. Let r_t be the daily log-returns, for t = 1, ..., T refers to certain period (say a month) and the index i = 1, ..., I_t (say days) within that period. Assume that the daily log-returns follows $r_{i,t} = \varepsilon_{i_t}$ and

$$\varepsilon_{i,t} = \sqrt{g_{i,t}\tau_t}Z_{i,t} \tag{9}$$

where $\varepsilon_{i,t} \mid \mathscr{F}_{i-1,t} \sim mathcalN(0, g_{i_t}, \tau_t)$ with $\mathscr{F}_{i-1,t}$ is the information set up to day i-1 of period t. $g_{i,t}$ denote to the short-term component of conditional variance and follows a unit-variance GARCH(1, 1) process:

$$g_{i,t} = \alpha_0 + \alpha\frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t} \tag{10}$$

where $\alpha_0 > 0, \alpha \geq 0, \beta \geq 0$ and $1 > \alpha + \beta$ enough wide-sense stacionarity. τ_t is defined as a function of the explanatory variables $X_t^{(m)}$, where m refers to the m-th explanatory variable. This will serve as the long-term component that varies at lower frequency. We specified with the MIDAS regression as

$$\tau_t = \sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \phi_k(1.0, \theta_m)X_{t-k}^{(m)} \tag{11}$$

where K is the lag parameter and ϕ_k refers to the weighting scheme, which we can specify. We chose the Beta weighting scheme such as most of the papers suggested. As we fixed the first paramter of the Beta weighting scheme to 1.0, then it will allow us to get monotoneously declining or increasing weights, as it was showed in the MIDAS section. If we want to use explanatory variables that can take positive or negative values, we used the exponential specification of the τ_t

$$\tau_t = exp(\sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \phi_k(1.0, \theta_m)X_{t-k}^{(m)}) \tag{12}$$

We tried to minimize the number of estimated parameters, so we changed the short-term volatity component's equation with a few assumptions.

$$E(r_{i,t}) = 0$$

$$Var(r_{i,t}) = E(r_{i,t}^2) - E(r_{i,t})^2 = E(r_{i,t}) = E(\sigma_{i,t}^2 Z_{i,t}^2) =$$

where they are indepent, since $Z_{i,t} \sim \mathscr{N}(0,1)$, so the expected value of squared $Z_{i,t}$ is equal to 1. Then,

$$E(\sigma_{i,t}^2) = E(g_{i,t}\tau_t) = E(g_{i,t})E(\tau_t) =$$

since we assumed that these two factors contibute for the underlying volatility process, namely, for $\sigma_{i,t}^2$. The other assumption, that we used, is the expected values of the macroeconimic variables is equal to 0, namely, $E(X_t^{(m)}) = 0$. Hence, the logarithm of τ has an expected value 0 and τ_t is equal to 1.

$$E(g_{i,t}) = E(\alpha_0 + \alpha\frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t}) = \alpha_0 + \alpha E(g_{i-1,t}) + \beta E(g_{i-1,t})$$

From $r_{i,t} = \sigma_{i,t}Z_{i,t}$, it is known that $E(r_{i,t}^2) = E(\sigma_{i,t}^2)$ and as we see in the GARCH section the unconditional variance of the returns henceforth

$$E(\sigma_{i,t}^2) = Var(r_{i,t}) = \frac{\alpha_0}{1-\alpha-\beta}$$

We used the moment matching to eliminate the α_0 parameters in the following way:

$$\hat{\mu} = \frac{1}{TI_t} \sum_{t=1}^{T} \sum_{i=1}^{I_t} r_{i,t}^2$$

then we can rewrite the short-term volatility equation:

$$g_{i,t} = \hat{\mu}(1-\alpha-\beta) + \alpha \frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t} \tag{13}$$

1.3.1 Parameter Estimation

The GARCH-MIDAS estimation made by maximum likelihood estimation, where the underlying two volatility component's parameters are estimated by a single-step. It was neceserry to highlight the single-step estimation, because in the further sections we will introduce the two-step method, which worked better for us in panel models. The loglikelihood function, then

$$\log \mathscr{L}(\Theta) = \frac{1}{TI_t} \sum_{t=1}^{T} \sum_{i=1}^{I_t} \frac{1}{2} \log 2\pi + \frac{1}{2} \log\left(g_{i,t}\tau_t\right) + \left(\frac{\varepsilon_{i,t}^2}{2g_{i,t}\tau_t}\right) \tag{14}$$

$$\arg\min_{\Theta} \log \mathscr{L}(\Theta)$$

1.3.2 Simulations

In the simulations, we assumed that the long-term volatility component is generated from X_t an AR(1) process:

$$X_t = \psi X_{t-1} + u_t$$

where u_t is a random process with mean zero and variance h^2 and $X_0 = 0$. We set them to $\psi = 0.8$ and $h^2 = 0.3$. Hence,

$$\log(\tau_t) = 0.3 \sum_{k=0}^{K} \psi_k(1.0, 4.0) X_{t-k}$$

where we chose $\beta_1 = 0.3, \theta = 4.0$ and K = 12. In the short-term volatility component $\alpha = 0.1$ and $\beta = 0.8$. As we described above, the returns generated in the following way:

$$r_{i,t} \sim \mathscr{N}(0, g_{i,t}\tau_t)$$

The simulations ran in T = 60, 120, 180, in other words they can be interpeted as 5-year, 10-year or 15-year length of data. The main objective was to about to see the increase the accuracy and the decrease of the variance of the parameter estimation.

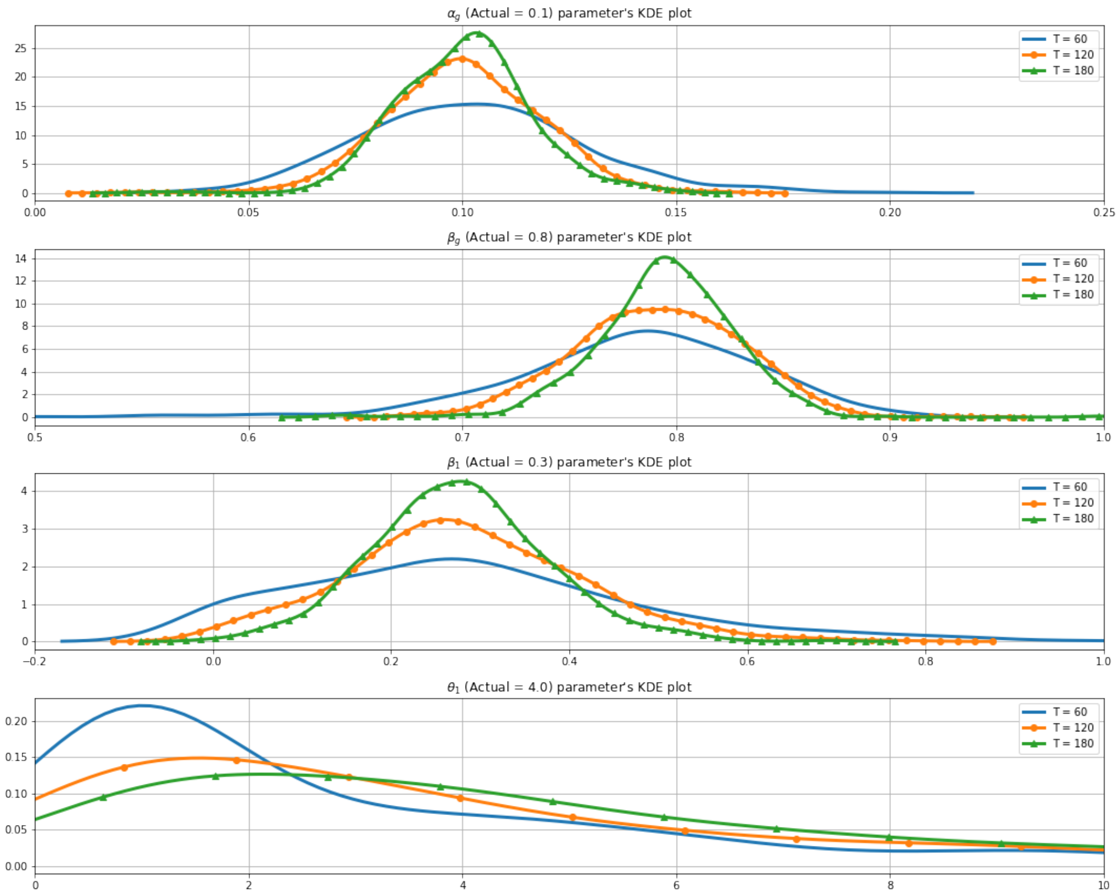

Figure 4: Plot of estimated parameter distributions with sample sizes of T = 60, 120, 180

In the the case of the first three parameters, we can see the expected tendency, what we waited for. Not only we can experinance huge deviations in parameter's accuary, but the θ estimated parameter's are not accurate at all. We conducted from these simulations, that we need to provide more data for this algorithm to excel in the parameter estimations, so in the following sections we will describe panel models.

1.4 Panel MIDAS

In this section we would like to introduce a new application of the MIDAS framework. This is a panel model, where all of the stocks in the panel share the same underlying explanatory variables for their volatility component. Let denote $r_{i,t}^{(j)}$ is the j-th stock's daily log return for j = 1, ... N. We mark the index t = 1, ..., T refers to a given period (say month) and the index i = 1, ..., I_t refers to (say days) in that period, as it was signed previously in GARCH-MIDAS section. Applying the assumption of these stocks share the same underlying long-term volatility component that will be marked as τ_t. This long-term volatility component contains all the explanatory variables, which can describe the volatility. Not only we can calculate long-term volatility component, but we can short-term or mixed component. As we choose $I_t = 1$, then the log returns and the explanatory variables will share the same sampling frequency. Even if we use both low and high frequency data, that updates their values in a certain period, say one of the explanatory variable is daily and the others are weekly. To keep our model as simple as possible we used only monthly sampling explanatory variables.The calculation of τ_t is handy for us when we are about to describe the Panel GARCH-MIDAS model. The equation of τ_t is given by:

$$\tau_t = \beta_0 + \sum_{m=1}^{M} \beta_m \sum_{k=1}^{K_m} \xi_k(1.0, \theta_m) X_{t-k}^{(m)} \tag{15}$$

where M refers to the number of explanatory variables and K is the number of lags. For simplicity we choose $K \equiv K_m$. As previously did, we choose the Beta weighting scheme and as the explanatory variables can take positive or negative values, we choose the exponentail specification of τ_t to be strictly positive:

$$\tau_t = \exp(\beta_0 + \sum_{m=1}^{M} \beta_m \sum_{k=1}^{K_m} \xi_k(1.0, \theta_m) X_{t-k}^{(m)}) \tag{16}$$

1.4.1 Parameter Estimation

The Panel MIDAS model estimated by QMLE that was described previously. We assumed that the stock's log returns mean are equal to zero, then the negative loglikelihood function will be:

$$\log \mathscr{L}(\Theta) = -\frac{T}{2} \sum_{j=1}^{M} \sum_{t=1}^{T} (\log 2\pi + \log \tau_t + \frac{(r_{i,t}^{(j)})^2}{\tau_t}) \tag{17}$$

The log likelihood of each individual stock is summed up to be minimized:

$$\arg\min_{\Theta} \log \mathscr{L}(\Theta) \tag{18}$$

1.4.2 Simulations

The simulation was conducted in the spirit of MIDAS simulation with same changes. Let suppose we have one explanatory variable that define the volatility say X_t is an AR(1) process:

$$X_t = \phi X_{t-1} + \varepsilon_t \tag{19}$$

where t = 1, ..., T, $\phi = 0.9$ and $\varepsilon_t \sim \mathscr{N}(0,1)$ standard normal variable, than the MIDAS model will be:

$$\log \tau_t = \beta_0 + \beta_1 \sum_{k=0}^{K} \xi_k(1.0, \theta) X_{t-k} \tag{20}$$

where $\beta_0 = 0.1, \beta_1 = 0.3$ and $\theta = 4.0$.τ_t remains the same throughout the whole period. The τ_t will determine the return's volatility, the returns are generated from normal distribution with zero mean and τ_t variance:

$$r_{i,t} \sim \mathscr{N}(0, \tau_t) \tag{21}$$

as τ_t is set to be a monthly variable, we generate daily return, so i mark mean that the i-th day of t-th month, i = 1, ..., I_t, where $I_t = 22$.

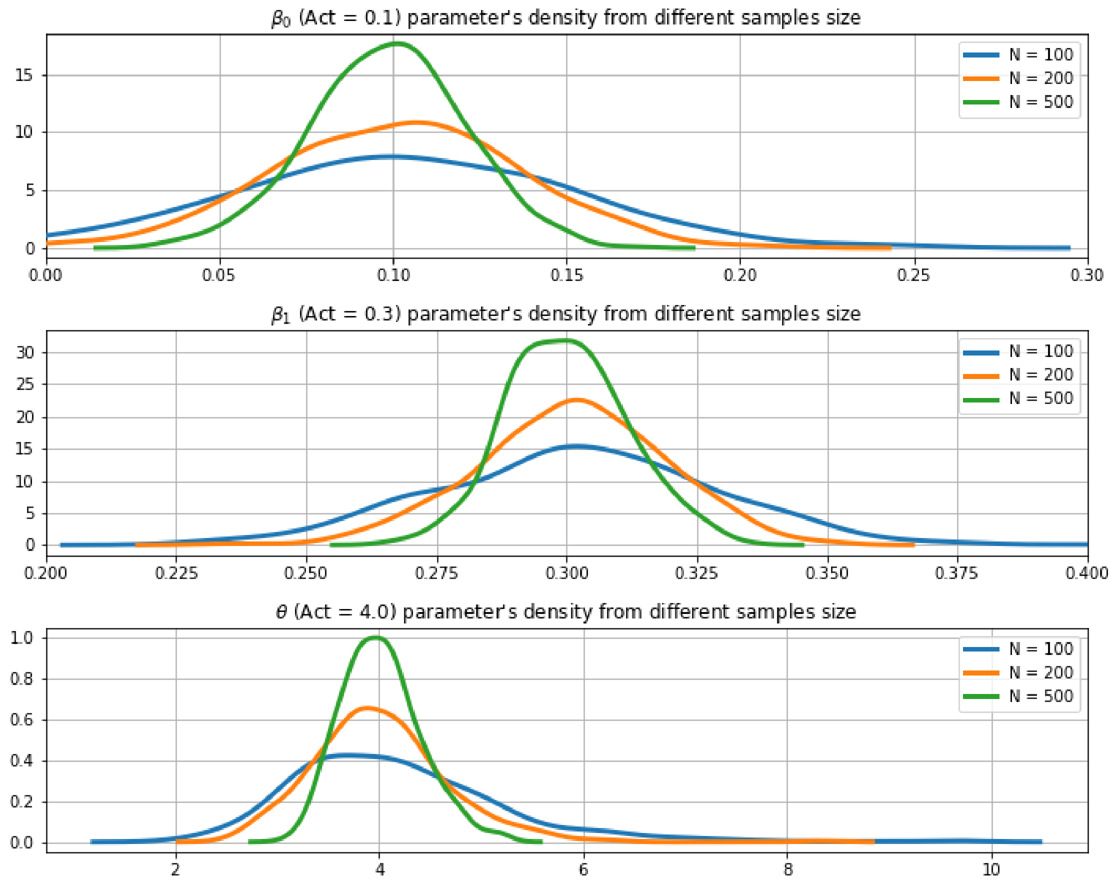

Figure 5: Plot of estimated parameter distributions with sample sizes of T = 100, 200, 500

1.5 Panel GARCH

In this section we present th panel version of the GARCH(1, 1) model. This model will serve for us as a benchmark model to compare the forecast accuracies, moreover this model is the rest of the Panel GARCH-MIDAS model.

Let denote $r_{i,t}$ the i-th stock's daily log return the index t = 1, ..., T refers to the days. We assume that the parameters which drive the dynamics of the volatilities are common to every stocks. However the unconditional means of the volatilites are asset specific. The daily log returns follow:

$$r_{i,t} = \sigma_{i,t}\varepsilon_{i,t}$$

where $\varepsilon_{i,t} = \sigma_{i,t}Z_{i,t}$, the innovations $Z_{i,t}$ are identiacally independent distributed random variables with mean 0 and variance 1. We considered two distributions for the innovation term, the first one is the normal distribution and the Student-t distribution to capture more extreme returns. The $\sigma^2_{i,t}$ can be written as

$$\sigma^2_{i,t} = \mu_i(1-\alpha-\beta) + \alpha\varepsilon^2_{i,t-1} + \beta\sigma^2_{i,t-1} \tag{22}$$

where μ_i refers to the uncoditional variances and the parameters of α and $beta$ satisfies $\alpha \geq 0, \beta \geq 0$ and $1 > \alpha + \beta$ for wide-sense stacionarity. This means we have N + 2 number of parameter to estimate. This can be challenging to estimate as the number of assets increases. To tackle this issue we use the following procedure

1.5.1 Parameter estimation

First of all we take advantage of the moment matching to calculate μ_i. As μ_i is the unconditional variance of the returns we can estimate the μ_i parameters by averaging the squared returns.

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}^2$$

In the second step given the unconditional variance estimates, the parameter space will be just $\Theta = \alpha, \beta$. We used the QMLE to minimize the negative log likelihood function given by:

$$\log \mathscr{L}(\Theta) = -\frac{T}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\log 2\pi + \log \sigma_{i,t}^2 + \frac{r_{i,t}^2}{\sigma_{i,t}^2})$$

where $\sigma_{i,t}^2$ is the function of the α, β.

$$\arg\min_{\Theta} \log \mathscr{L}(\Theta)$$

1.5.2 Simulations

We applied the same simulation technic, that we described in the GARCH section. The aim of the simulation was still the same, but now we examined the impact of the increment in the size of N. The results can be seen in the following figures:

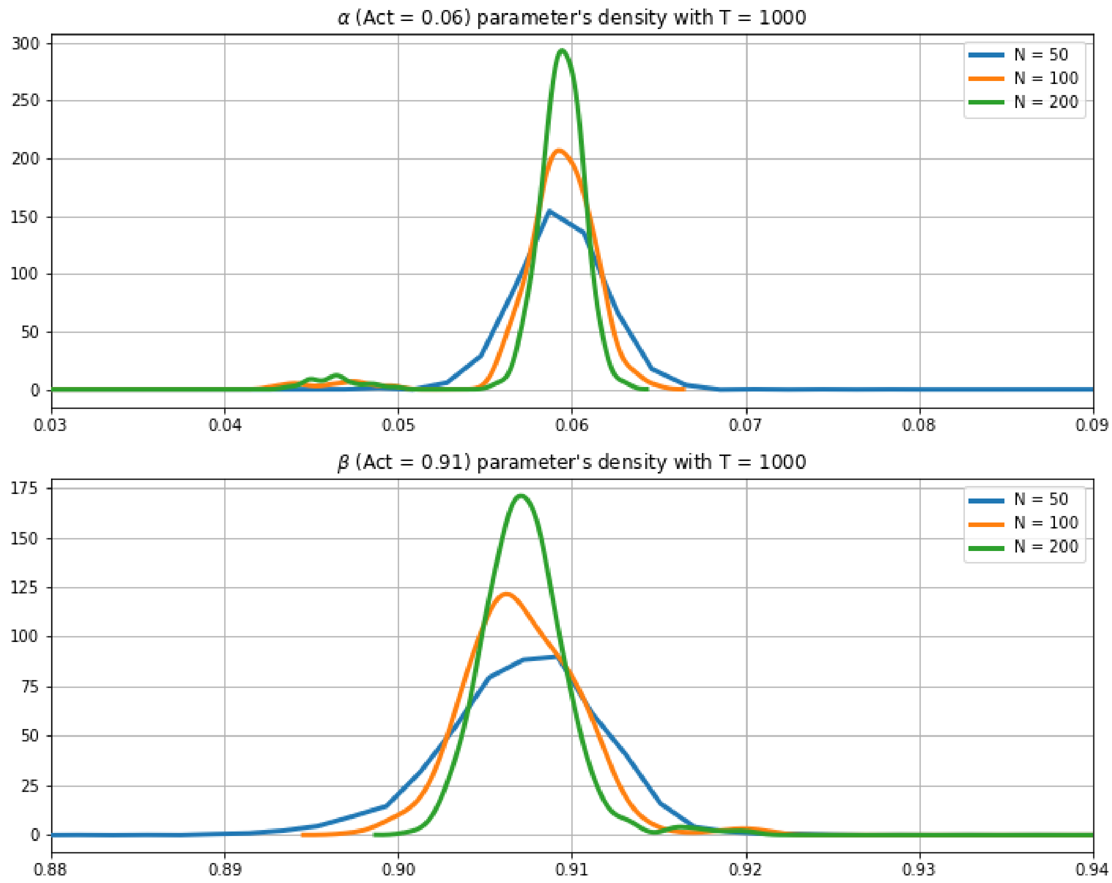

Figure 6: Plot of estimated parameter distributions with sample sizes of 1000 and N = 500, 1000, 2000

1.6 Panel GARCH with cross sectional adjustment

In this section we present a modified version of the above specified Panel GARCH model, called the Panel GARCH with cross sectional adjustment. We used the same notations and suppose the daily log returns follow:

$$r_{i,t} = \sigma_{i,t} c_t \varepsilon_{i,t}$$

where the innovation $Z_{i,t}$ is the same as in preivous section. The key difference is the c_t component which describe the cross sectional adjustment, as

$$c_t = (1-\phi) + \phi \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{r_{i,t-1}}{\sigma_{i,t-1}c_{t-1}} - \frac{1}{N}\sum_{i=1}^{N}\frac{r_{i,t-1}}{\sigma_{i,t-1}c_{t-1}}\right)^2} \tag{23}$$

where $1 \geq \phi \geq 0$ is the parameter of adjustment. We can see if we set ϕ equal to zero izd identically same with the Panel GARCH model, se we can clearly see how much impact have been made for forecasting accuracy by this component.Another element is the last term with the square root, this is technically the standard deviation of the innovations for the whole panel. The equation for $\sigma^2_{i,t}$ given by

$$\sigma^2_{i,t} = \mu_i(1-\alpha-\beta) + \alpha\varepsilon^2_{i,t-1} + \beta\sigma^2_{i,t-1} \tag{24}$$

where we applies the moment matching that we did in section above and the constraint for α and β stand for as well.

1.6.1 Parameter estimation

The parameter estimation is made by QMLE where the parameter space is then $\Theta = \phi, \alpha, \beta$ for the case of the innovations normally distributed. As we want to estimate with Student-t distributed we have tot estimate 4 parameters, the existing ones and the degreee of freedom for Student-t distribution. We also take the negative of the log likelihood to minimize its values

$$\arg\min_{\Theta} \log \mathscr{L}(\Theta)$$

1.6.2 Simulations

The simulations remain the same, so the random generation happens through ε_t. We examined the cases of increase the sample sizes and the number of stocks. The results are the following:

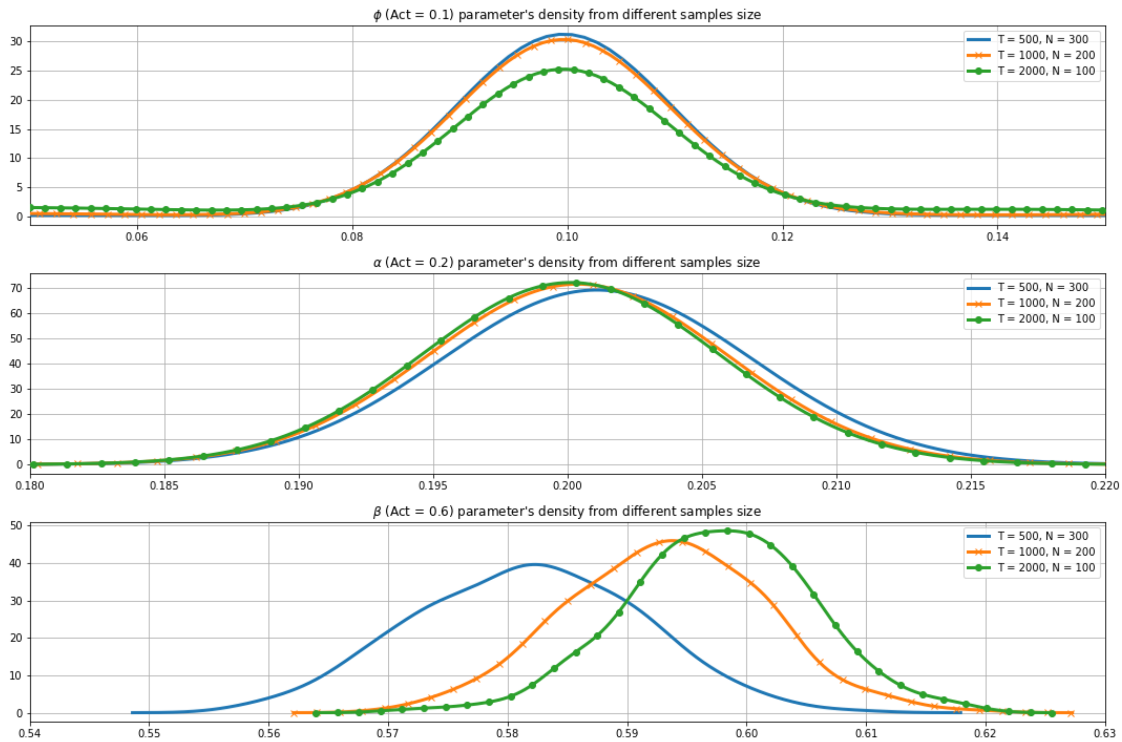

Figure 7: Plot of estimated parameter distributions with sample sizes of 1000 and N = 500, 1000, 2000

We can see that for the first to parameters (ϕ and α) the distributions of the parameter estimation's median is equal to the given parameters. In the case of β as we increase the sample size and number of stocks, we get closer and closer to the theoretical parameter.

1.7 Panel GARCH-MIDAS

In this section we specify the Panel version of the GARCH-MIDAS model. Durint the implementation process of this model some issues arised such as identification issues in th estimation of parameters. Finally we decided that we will make a two-step estimation and combine the best of the two world namely Panel MIDAS and Panel GARCH. The implementation is design to handle both multiple asset and single assets, in order to compare the accuracy of parameter estimation with the original single-step GARCH-MIDAS model. The first step is to calculate the long-term volatility component by the Panel MIDAS model. The estimation will provide us the $\tau_{i,t}$, the i index refers to have the same length of the returns, but τ is constant between intra periods. With the long-term component we can rescale the returns, by dividing them with the square root of $\tau_{i,t}$:

$$\hat{r}_{i,t}^{(j)} = \frac{r_{i,t}^{(j)}}{\sqrt{\tau_{i,t}}}$$

This rescaled return will be modeled by Panel GARCH model to get the short-term volatility component. The daily rescaled log returns follow:

$$\hat{r}_{i,t}^{(j)} = \varepsilon_{i,t}^{(j)} = \sigma_{i,t}^{(j)} Z_{i,t}^{(j)}$$

where $Z_{i,t}^{(j)}$ is the innovations, and we can rewrite this equation as the original literatures suggested if we replace the rescaled returns:

$$r_{i,t}^{(j)} = \sqrt{\tau_{i,t}\sigma_{i,t}^{(j)^2}} Z_{i,t}^{(j)}$$

where the returns are driven by the short- and long-term volatility components. Let's take a look at the short term volatility component's equation:

$$\sigma_{i,t}^{(j)^2} = \mu_j(1-\alpha-\beta) + \alpha\varepsilon_{i,t-1}^{(j)^2} + \beta\sigma_{i,t-1}^{(j)^2} = \mu_j(1-\alpha-\beta) + \alpha\frac{r_{i,t-1}^{(j)^2}}{\tau_{i,t-1}} + \beta\sigma_{i,t-1}^{(j)^2} \tag{25}$$

where $\alpha \geq 0, \beta \geq 0$ and $1 > \alpha + \beta$.

1.7.1 Parameter Estimation

As we discussed, the estimation is a two-step QMLE estimation. In which we first optimize the parameter's of the long-term volatility component where assumed the normal distribution, then the negative log - likelihood function looks like:

$$\log \mathscr{L}(\Theta_1) = -\frac{T * I_t}{2} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{i=1}^{2} \log 2\pi + \log \hat{\tau}_{i,t}(\Theta_1) + \frac{r_{i,t}^{(j)^2}}{\hat{\tau}_{i,t}(\Theta_1)} \tag{26}$$

In order to get the optimal parameters we minimize the argument's of the negative log likelihood

$$\arg\min_{\Theta_1} \log \mathscr{L}(\Theta_1)$$

Then in the case of short-term volatility component, we calculate with the rescaled return, so

$$\log \mathscr{L}(\Theta_2) = -\frac{T * I_t}{2} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{i=1}^{2} \log 2\pi + \log \hat{g}_{i,t}(\Theta_2) + \frac{\hat{r}_{i,t}^{(j)^2}}{\hat{g}_{i,t}(\Theta_2)} \tag{27}$$

This approach is slower in respect of take two optimization instead of one, but we found out that it can estimate the best parameters better in that case.

1.7.2 Simulations

To-Write

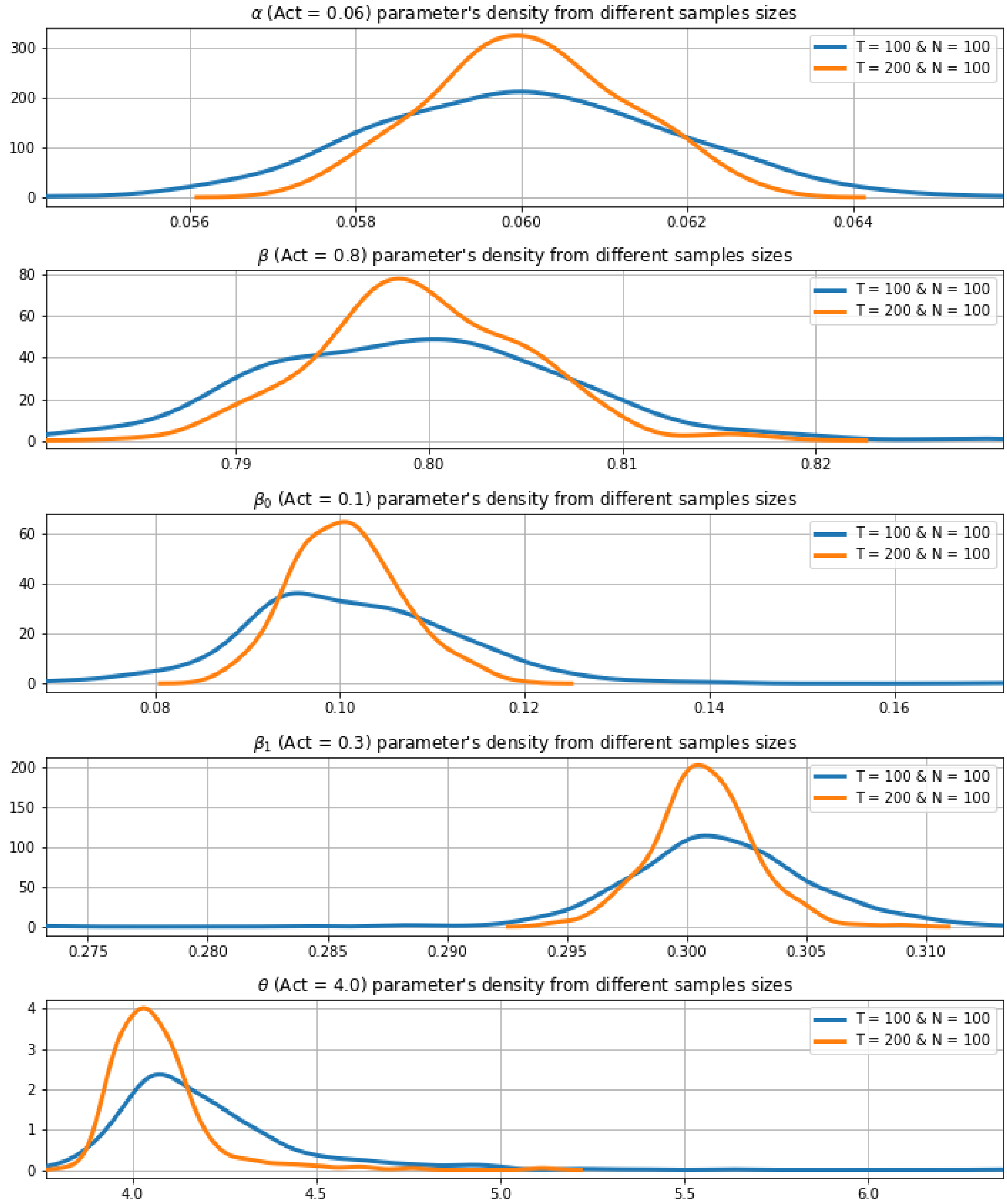

Figure 8: Plot of estimated parameter distributions with N = 100 and the sample size is T = 100, 200

1.8 EWMA

In this section we present one of the most commonly used model which is the Exponentially Weighted Moving Average (henceforth EWMA). [12] refers to EWMA ($\lambda = 0.94$) as $RiskMetrics^{TM}$. The model for volatility is given by

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2 \tag{28}$$

where λ is the only parameter to be estimated, which can take $1 \geq \lambda \geq 0$. This implies $\lambda + (1-\lambda) = 1$, which can be familiar as we think aobut the Integrated GARCH (1, 1) model with $\omega = 0$. We also implemented the panel version of EWMA, where we can describe the whole market's volatility with only one parameter or we can set to certain number as [12] did. Both cases, namely the EWMA and Panel EWMA are estimated via QMLE as it was previously described.

2 Data

In this section, we will go through the macroeconomic variables we used, and what transformation or changes we made. We want to mention, all of our data came from resources that free for everyone. In the selection of macroeconomic variable we mainly rely on those that was previously used in research papers, such as [3] where they used several variables, that will be presented. Both of these time series data start at 1997-01-01 and end at 2020-11-01. We make use of the following time series:

- The AAII Investor Sentiment Survey (AAII) measures the percentage of individual investors who are bullish, bearish, and neutral on the stock market for the next months. The series reported on a weekly basis.
  https://www.aaii.com/files/surveys/sentiment.xls

- Moody's Seasoned BAA Corporate Bond Yield Relative to Yield on 10 Year Treasury Constant Maturity (BAA10Y) is a daily series.
  https://fred.stlouisfed.org/series/BAA10Y

- The Chicago Fed National Activity Index (CFNAI) is a weighted average of 85 monthly filtered and standardized economic indicators. Whereas positive CFNAI values indicate an expanding US-economy above its historical trend rate, negative values indicate the opposite. [3]
  https://alfred.stlouisfed.org/series?seid=CFNAI

- Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL) is a measure of the average monthly change in the price for goods and services paid by urban consumers between any two time periods.
  https://alfred.stlouisfed.org/series?seid=CPIAUCSL

- Case-Shiller U.S. National Home Price Index (CSUSHPINSA) is a monthly index the leading measures of U.S. residential real estate prices, tracking changes in the value of residential real estate nationally.
  https://fred.stlouisfed.org/series/CSUSHPINSA

- 10-Year Treasury Constant Maturity Rate (DGS10) is a daily percent.
  https://fred.stlouisfed.org/series/DGS10

- 3-Month Treasury Bill: Secondary Market Rate (DTB3) is a daily percent. [1]
  https://alfred.stlouisfed.org/series?seid=DTB3

- Housing Starts Total: New Privately Owned Housing Units Started (HOUST) is a monthly unit. [3]
  https://fred.stlouisfed.org/series/HOUST

- Industrial Production: Total Index (INDPRO) is a monthly economic indicator that measures real output for all facilities located in the U.S. [3]

https://alfred.stlouisfed.org/series?seid=INDPRO

- M2 Money Stock (M2SL) is a monthly value in units of dollar billions.
  https://fred.stlouisfed.org/series/M2SL

- Chicago Fed National Financial Conditions Index (NFCI) provides a weekly update on U.S. financial conditions in money markets. Positive values of the NFCI indicate financial conditions that ar tighter than average, negative values indicate financial conditions that are looser than average. [3]
  https://fred.stlouisfed.org/series/NFCI

- Producer Price Index by Commodity: All Commodities (PPIACO) is a monthly index.
  https://alfred.stlouisfed.org/series?seid=PPIACO

- Unemployment Rate (UNRATE) represents the number of unemployed as a monthly percentage of the labor force. [1]
  https://fred.stlouisfed.org/series/UNRATE

- CBOE Volatility Index: VIX (VIXCLS) is a daily close index that measures market expectation of near term volatility conveyed by stock index option prices. [3]
  https://fred.stlouisfed.org/series/VIXCLS

The two most commonly used variables for calculating inflation are the differences of Consumer Price Index and Producer Price Index, as we will show in the correlation matrix they have a solid correlation between them. We are going to mark Inflation with the differences of CPIAUCSL and Δ PPI is the differences of PPIACO. We make the same transformation for the M2 Money Stock, Case-Shiller U.S. National Home Price Index, Housing Starts Total and Industrial Production. We wanted to measure the slope of the yield curve we subtracted the 10-Year Treasury Constant Maturity Rate with the 3-Month Treasury Bill as [1] they specified. Those variables the are observed weekly or daily we finally take there monthly mean. In order to keep our models as simply as possible, we will only use monthly macroeconomic variables for modeling, but we didn't want to miss them out, so we took those variables monthly average. The time-series of final dataset we will use for modeling:

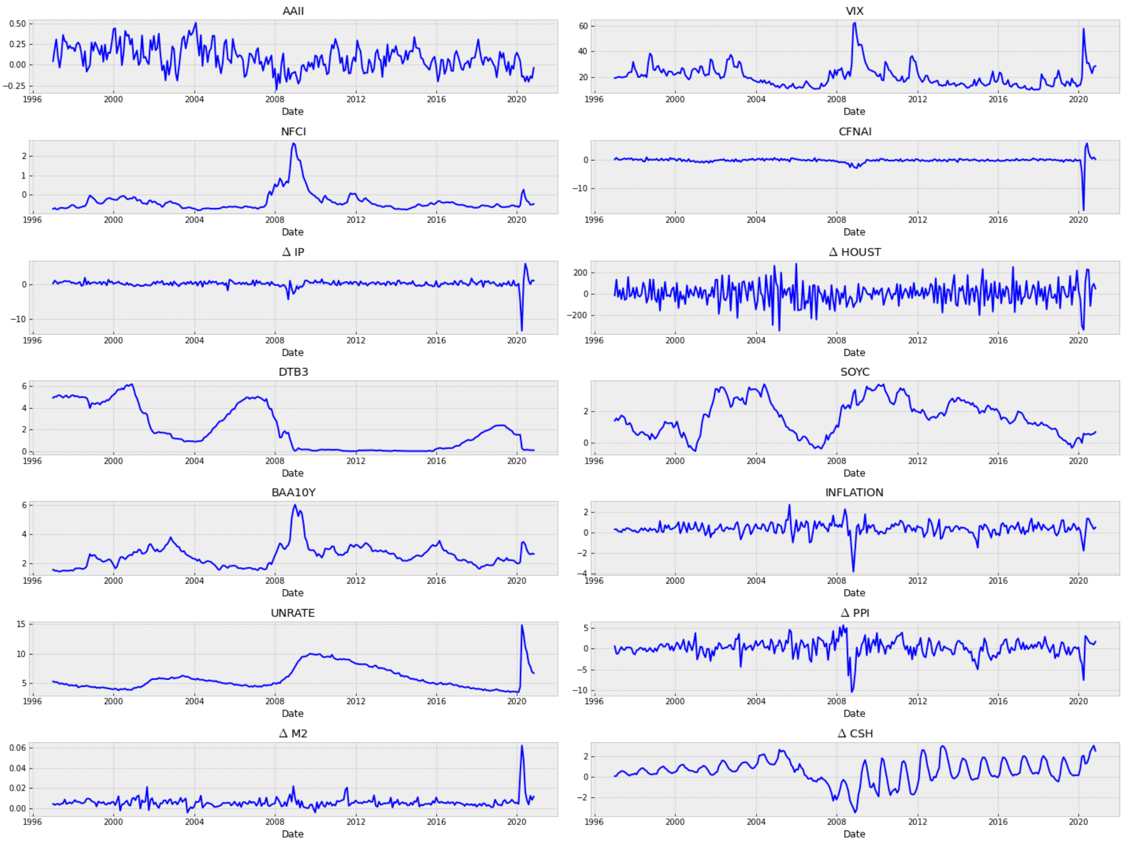

Figure 9: Time-series of macroeconomic variables

The stock prices we used for modeling are the prices of the S&P 500 Index components between 1999-12-01 and 2020-10-31, so due to the recent COVID-19's selloffs in the first quarter of 2020, we can examine two stressed period with our models. These data were downloaded by Python's package called *yfinance*.

	Min.	Max.	Mean	Median	Sd.	Skew.	Kurt
AAII	-0.30	0.51	0.08	0.08	0.15	0.21	-0.30
VIX	10.12	62.25	20.43	19.17	8.27	1.87	5.56
NFCI	-0.80	2.68	-0.36	-0.51	0.51	3.35	13.73
CFNAI	-17.73	5.96	-0.10	-0.01	1.28	-8.89	128.41
Δ IP	-13.26	5.74	0.09	0.14	1.13	-5.80	72.17
Δ HOUST	-343.00	279.00	0.73	-3.00	99.74	-0.26	0.79
DTB3	0.01	6.17	2.00	1.41	1.97	0.61	-1.14
Soyc	-0.53	3.69	1.64	1.62	1.11	0.03	-1.04
BAA10Y	1.45	6.01	2.53	2.50	0.75	1.57	4.73
Inflation	-3.84	2.70	0.36	0.39	0.59	-1.48	10.54
UNRATE	3.50	14.80	5.81	5.10	1.92	1.34	1.72
Δ PPI	-10.50	5.70	0.24	0.30	1.98	-1.28	5.75
Δ M2	-0.01	0.06	0.01	0.01	0.01	5.64	46.91
Δ Csh	-3.53	3.04	0.52	0.51	1.14	-0.60	0.92

Table 1: *Notes:* The table presents summary statistics for the variables.

3 Empirical Results

3.1 Predicting Abiliyt Test

In this section we describe the most commonly used to evaluate the volatility predictions. We mainly rely on previous research papers that used this methodology for testing predicting capability. This is the so called Diebold-Mariano Test (henceforth DM Test), which was first developed by [6].

In the research of [3] refers to [13] paper about to compare the most commonly used loss functions for volatility forecast comparison. He found that there are only two robust loss functions, namely the mean squared error (MSE) and the QLIKE. They are defined as the following:

$$QLIKE(\sigma^2_{k,t+1}, h_{k,t+1|t}) = \frac{(\sigma^2_{k,t+1}}{h_{k,t+1|t}} - \log(\frac{(\sigma^2_{k,t+1}}{h_{k,t+1|t}})\text{–}1 \tag{29}$$

$$MSE((\sigma^2_{k,t+1}, h_{k,t+1|t}) = (\sigma^2_{k,t+1} - h_{k,t+1|t})^2 \tag{30}$$

The QLIKE and MSE are the only robust loss function. In [13] it was written that Qlike is less sensitive with respect to extreme observations than the mean squared error loss. Furthermore, the comparison took place in the test of [6]. DM relies on assumptions made directly on the forecast error loss differential [[5]]. We used the two evaluation criterion that defined above (QLIKE, MSE). Let denote:

$$d_t^{(QLIKE)} = QLIKE(\sigma^2_{k,t+1}, \hat{h^1_{k,t+1|t}}) - QLIKE(\sigma^2_{k,t+1}, \hat{h^2_{k,t+1|t}}) \tag{31}$$

$$d_t^{(MSE)} = MSE(\sigma^2_{k,t+1}, \hat{h^1_{k,t+1|t}}) - MSE(\sigma^2_{k,t+1}, \hat{h^2_{k,t+1|t}}) \tag{32}$$

where $\sigma^2_{k,t+1}$ is the volatility proxy variable, as we don't have intraday data we choose the daily squared returns. [5] presented that DM assumes:

$$DM = \begin{cases} E(d_t) = \mu & , \forall \text{ t} \\ cov(d_t, d_{t-\tau}) = \gamma(\tau) & , \forall \text{ t} \\ 0 < var(d_t) = \sigma^2 < \infty & \text{, otherwise} \end{cases}$$

$$DM = \frac{\bar{d_t}}{\hat{\sigma d_t}} \sim \mathcal{N}(0,1) \tag{33}$$

where $\bar{d_t} = E(d_t)$ is the sample mean loss differential and $\hat{\sigma d_t}$ is a consistent estimate of the standard deviaton of $\bar{d_t}$.

References

[1] H. Asgharian, A. J. Hou, and F. Javed. The importance of the macroeconomic variables in forecasting stock return variance: A garch-midas approach. *Journal of Forecasting*, 32(7):600–612, 2013.

[2] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986.

[3] C. Conrad and O. Kleen. Two are better than one: Volatility forecasting using multiplicative component garch-midas models. *Journal of Applied Econometrics*, 35(1), 2019.

[4] C. Conrad and K. Loch. Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, 2014.

[5] F. Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. *Journal of Business and Economic Statics*, 33(1), 2015.

[6] F. Diebold and R. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statics*, 1995.

[7] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[8] R. F. Engle, E. Ghysels, and B. Sohn. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3):776–797, 2013.

[9] E. Ghysels and H. Qian. Estimating MIDAS regressions via OLS with polynomial parameter profiling. *Econometrics and Statistics*, 9(C):1–16, 2019.

[10] E. Ghysels, P. Santa-Clara, and R. Valkanov. The MIDAS Touch: Mixed Data Sampling Regression Models. CIRANO Working Papers 2004s-20, CIRANO, May 2004.

[11] E. Ghysels, A. Sinko, and R. Valkanov. Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007.

[12] J. Morgan. Riskmetrics: technical document. *Morgan Guaranty Trust Company of New York*, 1996.

[13] A. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011.

[14] F. Wang and E. Ghysels. Econometric analysis of volatility component models. *Econometric Theory*, 31(2):362–393, 2015.