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:

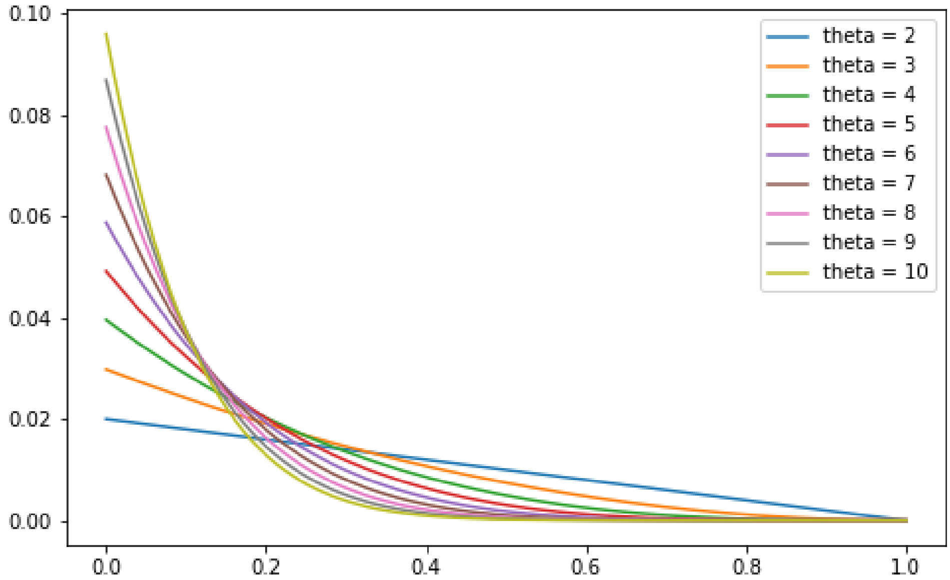

Figure 1: Plot of Beta Lag weighting function in equation 1 with K = 100, $\theta_1 = 1$ and $\theta_2 = 2, ..., 10$

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 Estimation

This simple model's parameter estimation happens through the sum of squared estimate of error:

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

As SSE is a quadratic function, we can find the optimal parameters by minimizing it:

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

where β_0, β_1 and $theta$ are the desired paremetrs to be estimated.

1.1.3 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.

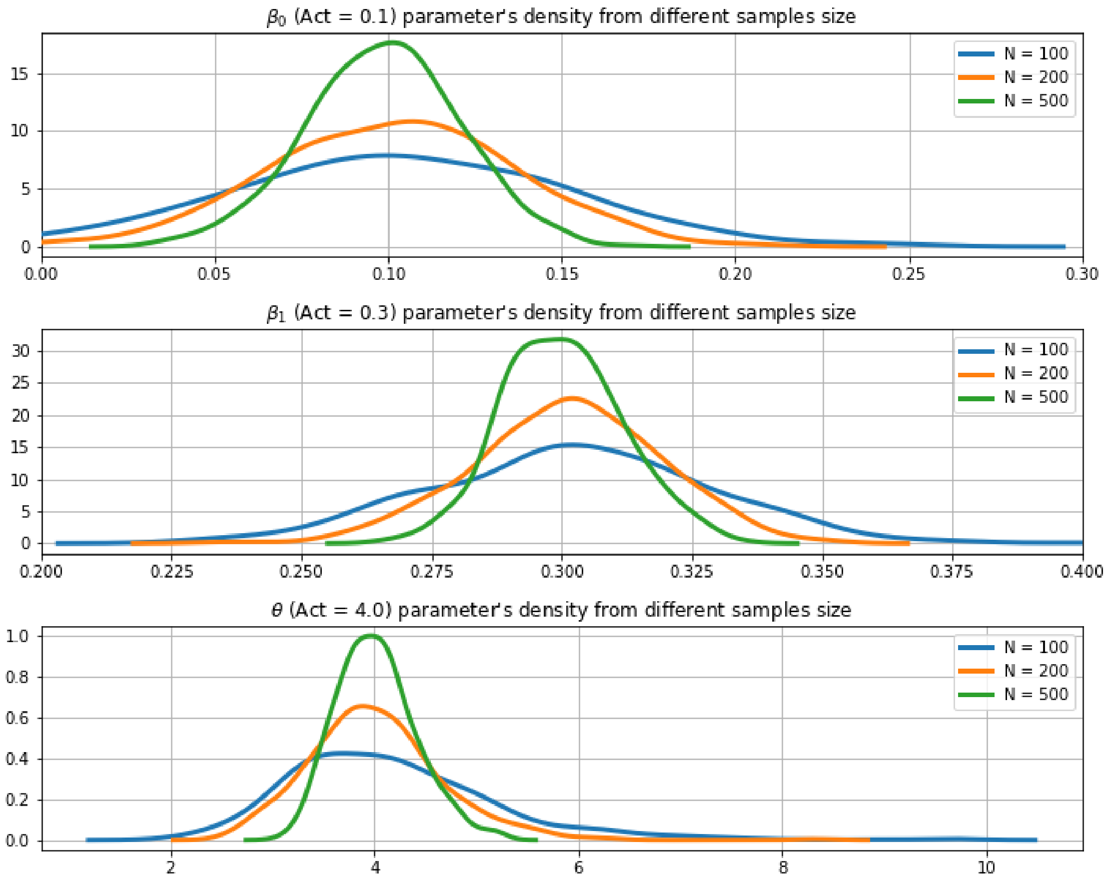

Figure 2: Plot of estimated parameter distributions with sample sizes of 100, 200, 500

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:

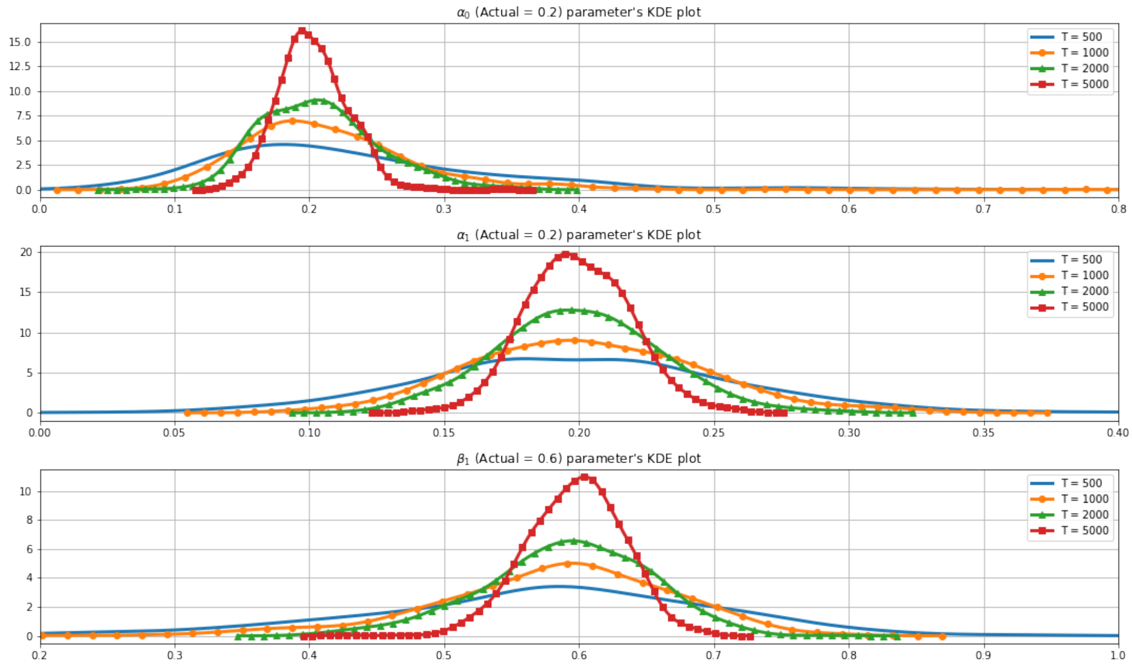

Figure 3: Plot of estimated parameter distributions with sample sizes of 500, 1000, 2000, 5000

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. [16] 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:

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

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

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

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

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

$$Var(r_{i,t}) = E(r^2_{i,t}) - E(r_{i,t})^2 = E(r_{i,t}) = E(\sigma^2_{i,t}Z^2_{i,t}) =$$

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,

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

since we assumed that these two factors contibute for the underlying volatility process, namely, for $\sigma^2_{i,t}$. 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.

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

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

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

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

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

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

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,

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

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.

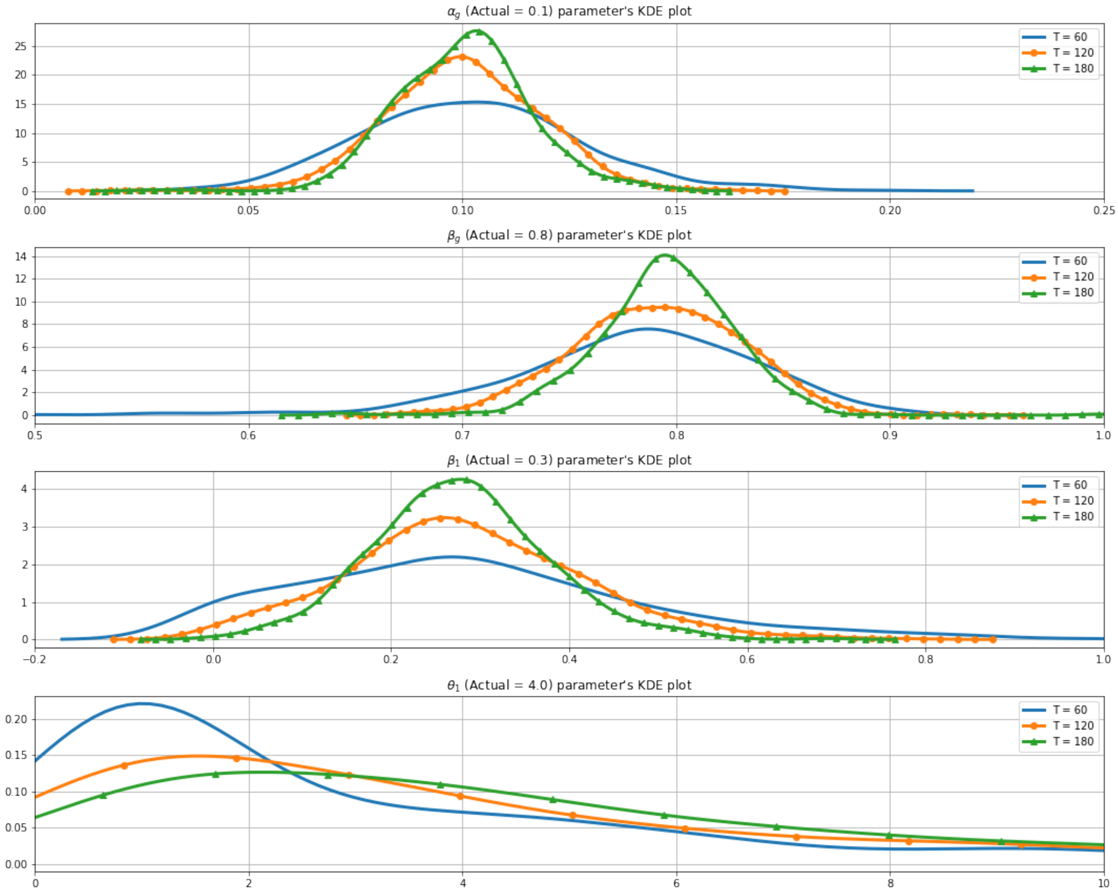

Figure 4: Plot of estimated parameter distributions with sample sizes of T = 60, 120, 180

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 I would like to introduce a new application of the MIDAS framework, namely, the panel version of the model. This approach not only provide better understanding about the underlying long term volatility, but it is parsimonious. The main idea was that there are a panel of stock returns and we would like to say something about the common long-term volatility component. In our application, we used the macroeconomic explanatory variables to model the volatility.

Let $r_{i,t}^{(j)}$ denote the j-th stock's daily return for j = 1, ... N, at i-th day for i = 1, ..., I_t of t-th month for t = 1, ..., T. The indexes came from the GARCH-MIDAS section and we try to use the same symbols to be consistent. We assumed that these stocks share the same underlying volatility component that will be marked as τ_t. The long-term volatility component contains all the explanatory variables, which can describe the volatility. If we choose $I_t = 1$ means we would like to calculate the monthly returns volatility, so both r and τ will be at the same frequency. Furthermore, we can apply weekly, daily or even intra daily explanatory variables in modeling the volatility. In the lack of intra day data and to keep the modeling parsimonious as possible we used only monthly sampled data.

We describe τ_t as we did in GARCH-MIDAS section:

$$\tau_t = \sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \psi_k(1.0, \theta_m) X_{t-k}^{(m)} \tag{15}$$

where m refers to the m-th explanatory variable, K is the lag parameter and ϕ_k refers to the weighting scheme, which we can specify. We used the Beta weighting scheme for modeling as most of the papers suggested. 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. We used the exponentail specification of τ_t as we used explanatory variables which can take negative values:

$$\tau_t = exp\left(\sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \phi_k(1.0, \theta_m) X_{t-k}^{(m)}\right) \tag{16}$$

1.4.1 Parameter Estimation

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

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

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:

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

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

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

where $\beta_1 = 0.3$ and $\theta = 4.0.\tau_t$ 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 $I_t = 22$.

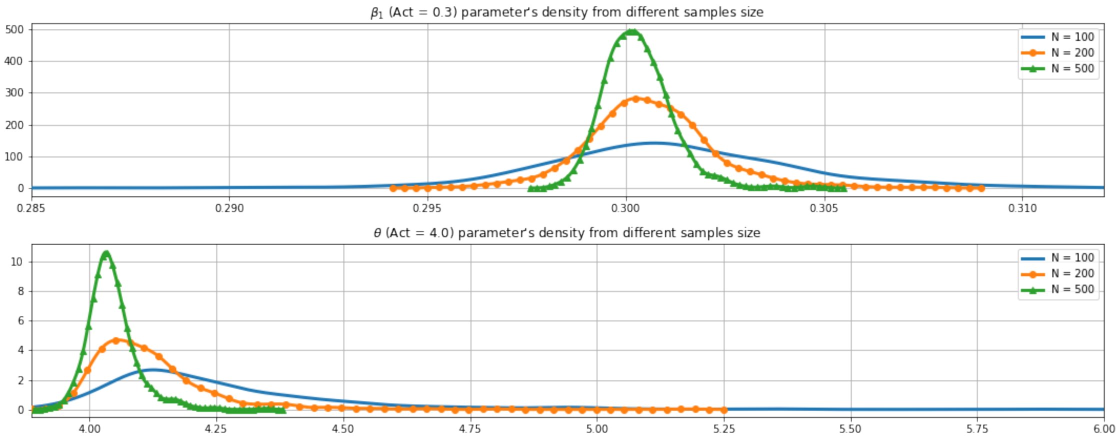

Figure 5: Plot of estimated parameter distributions with sample sizes of N = 100, 200, 500

The distributions show that as we increase the panel size we get better parameter estimation. We simulated 100 month of data, which means around 2200 day's of returns. In the case of N = 500, we can say the model can estimate the parameters which we used for simulations.

1.5 Panel GARCH

In this section we would like to present the panel version of the GARCH model we described above. Our main goal was here to make the model as parsimonious as we could to become the benchmark model for our further analysis.

Let $r_{i,t}^{(j)}$ denote to the j-th stock's daily return for j = 1, ..., N at i-th day for i = 1, ..., I_t of t-th month for t = 1, ..., T. In order to be consistent with the indexation, we decided to choose the ones that we used in Panel MIDAS section. Not only it will make easier to describe Panel GARCH-MIDAS model, but it also create consistency in indexes. We used the assumption of that the parameters for the dynamics of the volatilites are common to every stocks. In addition, the unconditional means of the volatilies are asset specific. The daily returns follow:

$$r_{i,t}^{(j)} = \varepsilon_{i,t}^{(j)} = \sigma_{i,t}^{(j)} Z_{i,t}^{(j)} \tag{22}$$

where $Z_{i,t}^{(j)}$ the innovations which identiacally independent distributed random variables with mean 0 and variance 1. As we described in GARCH section, we specified the distributions for this innovation term, the first one is the Normal distribution and the other is the Student-t distribution to capture more extreme returns. The volatility equation can be written as:

$$\sigma_{i,t}^{(j)^2} = \mu^{(j)}(1-\alpha_1-\beta_1) + \alpha_1 \varepsilon_{i-1,t}^{(j)^2} + \beta_1 \sigma_{i-1,t}^{(j)^2} \tag{23}$$

where $\mu^{(j)}$ refers to the unconditional variance and the parameters of α_1 and β_1 satisfies $\alpha_1 \geq 0, \beta_1 \geq 0$ and $1 > \alpha_1 + \beta_1$ for wide-sense stacionarity. If we would like to estimate α and β for each individual stock, we have to estimate N + 2 number of parameters. This can be challenging to estimate as the number of the assets increase, hence we assumed that the parameters in the volatility equations are common as we said earlier.

1.5.1 Parameter estimation

First of all we take advantage of the moment matching to calculate $\mu^{(j)}$. Not only make it easier the estimation, but it will be more parsimonious. As $\mu^{(j)}$ is the unconditional variance of the returns, we can estimate by averaging the squared returns:

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

In the second step given the unconditional variance estimates, the parameter space will reduce into just two parameter $\Theta = \alpha_1, \beta_1$. We used the QMLE to minimize the negative log likelihood function given by:

$$\log \mathscr{L}^{(j)}(\Theta) = -\frac{1}{TI_t} \sum_{t=1}^{T} \sum_{i=1}^{I_t} \left(\frac{1}{2} \log 2\pi + \frac{1}{2} \log \sigma_{i,t}^{(j)^2} + \frac{1}{2} \frac{\varepsilon_{i,t}^{(j)^2}}{\sigma_{i,t}^{(j)^2}} \right) \tag{24}$$

where $\sigma_{i,t}^{(j)^2}$ is the function of the α_1, β_1

$$\arg\min_{\Theta} \sum_{j=1}^{N} \log \mathscr{L}^{(j)}(\Theta)$$

1.5.2 Simulations

We applied the same simulation scheme as we previously did in GARCH section. The only difference is we simulated matrix of returns. Here we sampled returns in arrays these are the rows in the matrix and in the end the columns mean the individual stock returns. As we simulated the returns we matched the individual volatility component which related to the desired "stock" return, so in each row every return has its unique variance.

$$r_{i,t}^{(j)} \sim \mathscr{N}\left(0, \sigma_{i,t}^{(j)}\right)$$

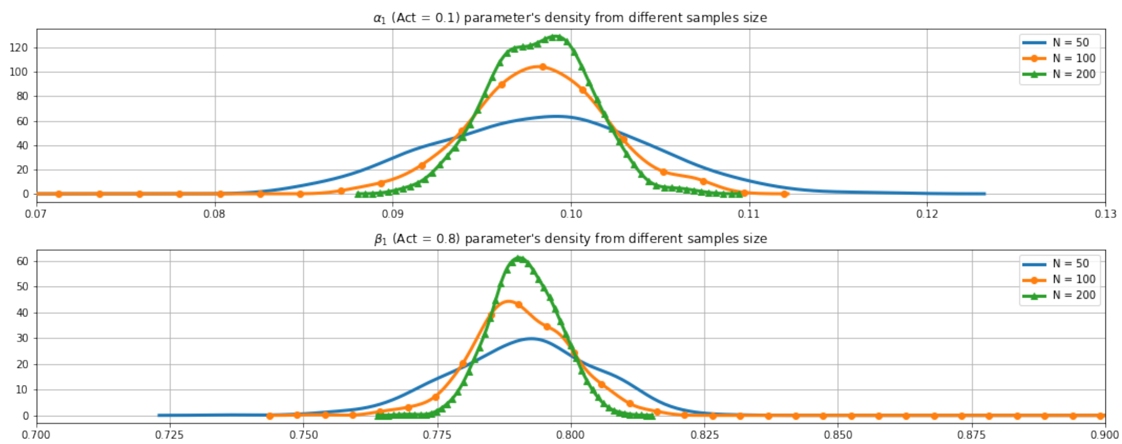

Figure 6: Plot of estimated parameter distributions with sample sizes of 500 and N = 50, 100, 200

In the simulations we simulated panels with 50, 100, 200 number of individual stocks and 500 number of returns for each. In the above figure, we can see, as we increase the number of stocks in the panel the accurate will be the parameter estimations.

1.6 Panel GARCH with cross sectional adjustment

In this section, I would like to give you an overview of this unique model. The key concept of the model lays on the cross sectional adjustment part, where we adjust the individual volatilites with the $c_{i,t}$ component. We used the same notations as we did in the Panel GARCH section, to be consistent and the reader can easily get familiar with the changes we made. Let $r_{i,t}^{(j)}$ denote to the daily return of the j-th stock j = 1, ..., N at i-th day for i = 1, ..., I_t of t-th month for t = 1, ..., T. The daily returns follow:

$$r_{i,t}^{(j)} = \sigma_{i,t}^{(j)} c_{i,t} \varepsilon_{i,t}^{(j)} \tag{25}$$

where $\varepsilon_{i,t}^{(j)}$ is the innovation, which identically independent distributed random variables with mean 0 and variance 1. The $c_{i,t}$ component is the cross sectional adjusment term, which can interpret as the common short-term volatility throughout the panel. If we examine the equation of this component, we can see that as the market or panel, which we investigate, produce more and more abnormal returns, so the $c_{i,t}$ component's value is increasing. This increment can be used to better forecast volatilities for the panel. On the other hand, if the market conditions generates low volatility period, then $c_{i,t}$ will decrease.

$$c_{i,t} = (1-\phi) + \phi \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{r_{i-1,t}^{(j)}}{\sigma_{i-1,t}^{(j)} c_{i-1,t}} - \frac{1}{N}\sum_{j=1}^{N}\frac{r_{i-1,t}^{(j)}}{\sigma_{i-1,t}^{(j)} c_{i-1,t}}\right)^2} \tag{26}$$

where $1 \geq \phi \geq 0$ is the parameter of the adjustment. We can see if we observe ϕ equal to zero, then it is identically the same model with the Panel GARCH. In our invastigations, we found out that as the higher the ϕ the better the model performance in contrast with the vanilla Panel GARCH model. This whole models lie on the $c_{i,t}$ last term, which is tecnically is the standard deviation of the innovations throughout the panel. For illustration purposes, we would like to present the an estimated $c_{i,t}$ about a panel, which we will use in the modelling section:

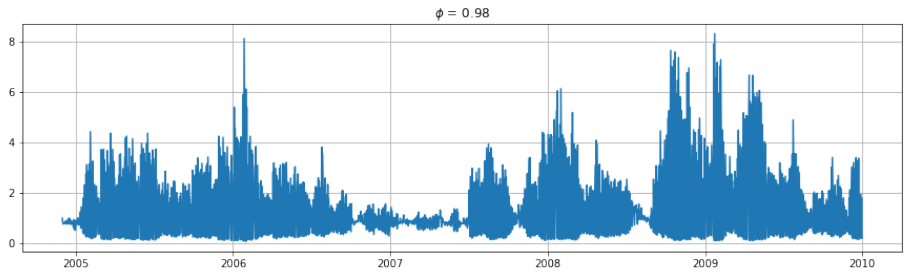

Figure 7: Plot of cross sectional adjustment componet

In the above figure, we used a panel that consist of 492 stocks and the investigated period is between 2014-12-1 and 2009-12-31. You can clearly spot the higher volatility regimes in the plot. The equation for

$\sigma_{i,t}^{(j)^2}$ given by

$$\sigma_{i,t}^{(j)^2} = \mu^{(j)}(1-\alpha_1-\beta_1)+\alpha_1\varepsilon_{i-1,t}^{(j)^2}+\beta_1\sigma_{i-1,t}^{(j)^2} \tag{27}$$

where we applied the moment matching that we did in Panel GARCH section. The parameters of α_1 and β_1 satisfies $\alpha_1 \geq 0, \beta_1 \geq 0$ and $1 > \alpha_1 + \beta_1$

1.6.1 Parameter estimation

As we mentioned in previously, we approximated the unconditional variance of the returns by averaging the squared returns. Moreover, we applied the same parameter estimation sheme for this model, namely the QMLE where the parameter space is then $\Theta = \phi, \alpha, \beta$. We asssumed that the innovations are normally distributed, but as we would assume Student-t distribution, the parameter space would be expand with the parameter of the Student-t distribution's degree of freedom. We minimize the negative log likelihood function, which is given by:

$$\log \mathscr{L}^{(j)}(\Theta) = -\frac{1}{TI_t}\sum_{t=1}^{T}\sum_{i=1}^{I_t}\left(\frac{1}{2}\log 2\pi + \frac{1}{2}\log\sigma_{i,t}^{(j)^2} + \frac{1}{2}\frac{\varepsilon_{i,t}^{(j)^2}}{\sigma_{i,t}^{(j)^2}}\right) \tag{28}$$

$$\arg\min_{\Theta}\sum_{j=1}^{N}\log\mathscr{L}^{(j)}(\Theta)$$

1.6.2 Simulations

The simulations are made by a fairly simulare way, which we did in Panel GARCH. The returns are generated randomly from normal distribution with variance of $\sigma_{i,t}^{(j)}$:

$$r_{i,t}^{(j)} \sim \mathscr{N}\left(0, \sigma_{i,t}^{(j)}\right) \tag{29}$$

The $c_{i,t}$ component effect is take place in $\varepsilon_{i,t}^{(j)}$. We simulated different size of simulated panels with the same sample size, T = 500. The parameters which we used for simulation purposes are ϕ = 0.9, α_1 = 0.2 and β_1 = 0.6. The results are the following:

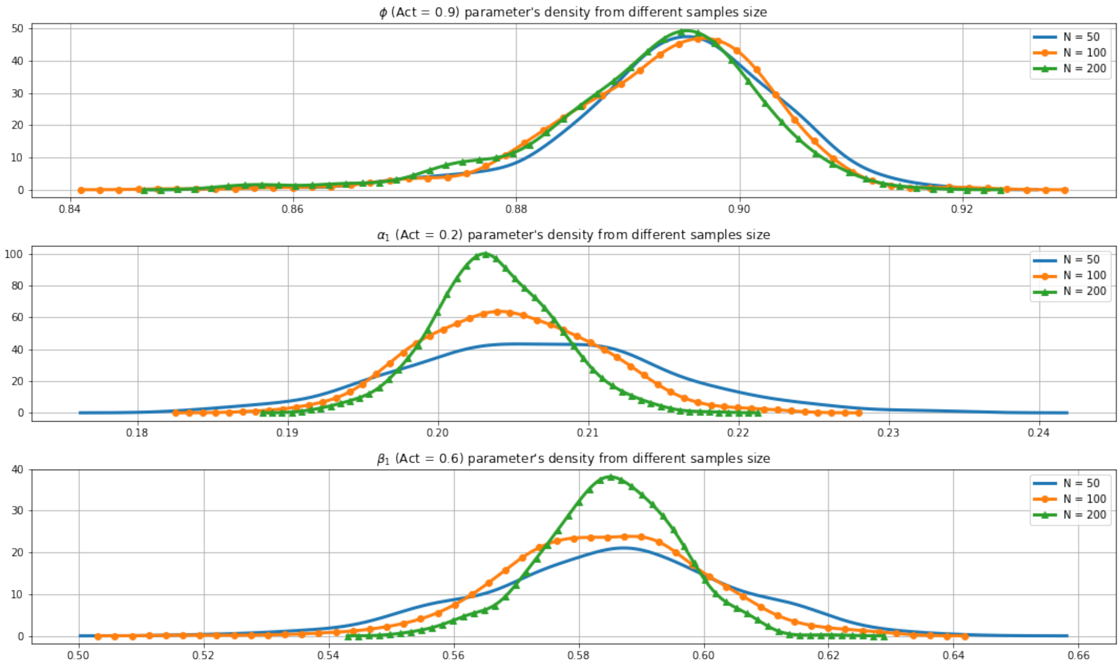

Figure 8: Plot of estimated parameter distributions with N = 50, 100, 200

We can clearly see, as we increase the panel size, the less the variance of the estimated parameters. Unfortunatelly, in the case of ϕ and β_1 the median estimated parameters are less, then what we used for simulations. In the case of α_1 the median estimation is higher. These miss estimation can be easily come from the relatively small sample size, the high α_1 and low β_1. In the modeling section we will point out, this model required quite huge size of samples and the $\alpha_1 + \beta_1$ tend to be close to 1.0.

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. Let $\tau_{i,t}$ to be:

$$\tau_{i,t} = \beta_0 + \sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \psi_k(1.0, \theta_m) X_{t-k}^{(m)} \tag{30}$$

where m refers to the m-th explanatory variable, K is the lag parameter and ϕ_k refers to the weighting scheme. We used the exponentail specification of $\tau_{i,t}$ as we used explanatory variables which can take negative values:

$$\tau_{i,t} = exp\left(\beta_0 + \sum_{m=1}^{M} \beta_m \sum_{k=0}^{K} \phi_k(1.0, \theta_m) X_{t-k}^{(m)}\right) \tag{31}$$

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 $\tau_{i,t}$:

$$\hat{r}_{i,t}^{(j)} = \frac{r_{i,t}^{(j)}}{\sqrt{\tau_{i,t}}} \tag{32}$$

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

$$\hat{r}_{i,t}^{(j)} = \varepsilon_{i,t}^{(j)} = \sigma_{i,t}^{(j)} Z_{i,t}^{(j)} \tag{33}$$

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:

$$r_{i,t}^{(j)} = \sqrt{\tau_{i,t}\sigma_{i,t}^{(j)^2}} Z_{i,t}^{(j)} \tag{34}$$

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:

$$\sigma_{i,t}^{(j)^2} = \mu_j(1-\alpha-\beta) + \alpha\varepsilon_{i,t-1}^{(j)^2} + \beta\sigma_{i,t-1}^{(j)^2} = \mu_j(1-\alpha-\beta) + \alpha\frac{r_{i,t-1}^{(j)^2}}{\tau_{i,t-1}} + \beta\sigma_{i,t-1}^{(j)^2} \tag{35}$$

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:

$$\log \mathscr{L}(\Theta_1) = -\frac{T*I_t}{2}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{2} \log 2\pi + \log \hat{\tau}_{i,t}(\Theta_1) + \frac{r_{i,t}^{(j)^2}}{\hat{\tau}_{i,t}(\Theta_1)} \tag{36}$$

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

$$\log \mathscr{L}(\Theta_2) = -\frac{T*I_t}{2}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{2} \log 2\pi + \log \hat{g}_{i,t}(\Theta_2) + \frac{\hat{r}_{i,t}^{(j)^2}}{\hat{g}_{i,t}(\Theta_2)} \tag{37}$$

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

This model basically used the same simulation framework, which we previously described in GARCH-MIDAS section. In the simulations, we assumed that the common long-term volatility component is generated from X_t an AR(1) process:

$$X_t = \psi X_{t-1} + u_t \tag{38}$$

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,

$$log(\tau_t) = 0.3 \sum_{k=0}^{K} \phi_k(1.0, 4.0) X_{t-k} \tag{39}$$

where we chose $\beta_1 = 0.3, \theta = 4.0$ and K = 12. In the short-term volatility component $\alpha = 0.06$ and $\beta = 0.8$. As we described above, the returns generated in the following way:

$$r_{i,t} \sim \mathcal{N}(0, g_{i,t}\tau_t) \tag{40}$$

The simulations ran in T = 100, 200, in other words they can be interpeted as 8-year, or 16-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.

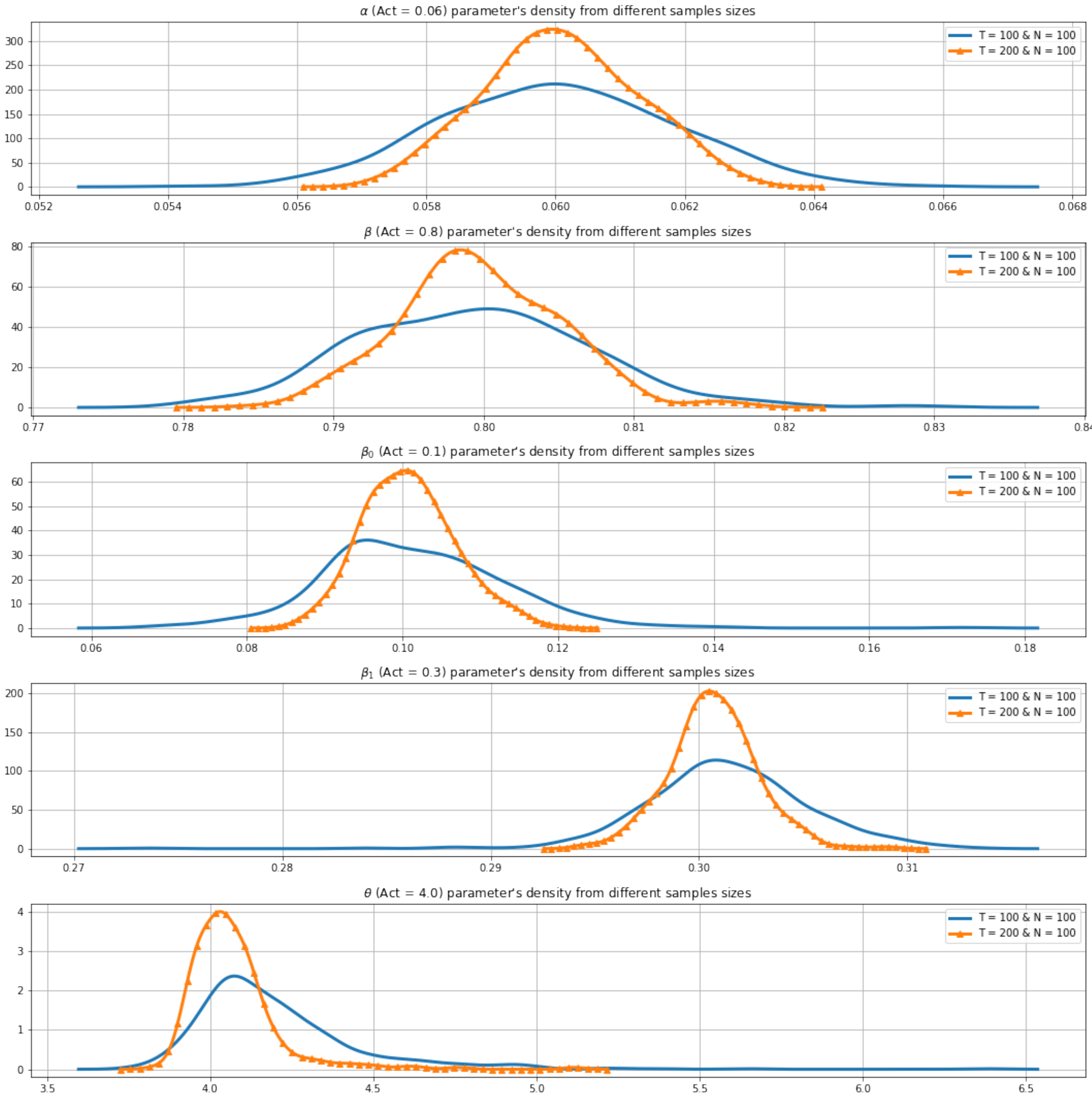

Figure 9: Plot of estimated parameter distributions with sample sizes of 500 and N = 50, 100, 200

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.

1.8 Panel EWMA

In this section we present the industry standard volatility model, namely the Exponentially Weighted Moving Average (henceforth EWMA) model. In [13] used EWMA with $\lambda = 0.94$. Let r_t denote to 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. We assumed conditional normality for the distribution of returns r_t, with the volatility eqution is the following

$$\sigma_t^2 = (1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2 \tag{41}$$

where λ is our only parameter to be estimated, which can only take $1 \geq \lambda \geq 0$. The above equation can be familiar with the one I described in GARCH section, this is not a coincidence. The EWMA is a special case of the GARCH model, namely the Integrated-GARCH (IGARCH), where we choose $\alpha_0 = 0, \alpha_1 = 1-\lambda$ and $\beta_1 = \lambda$, so $1 = \alpha_1 + \beta_1$ which import a unit root to the GARCH process. For the unconditional variance of the returns we use the mean of the squared returns.

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} r_t^2 \tag{42}$$

Let us describe the panel version of the EWMA model, where Let $r_t^{(j)}$ denote to the j-th stock's daily return for j = 1, ..., N at time t for t = 1, ..., T. We also assumed conditional normality for the distribution of returns $r_t^{(j)}$, with the volatility eqution is the following

$$\sigma_t^{(j)^2} = (1-\lambda) r_{t-1}^{(j)^2} + \lambda \sigma_{t-1}^{(j)^2} \tag{43}$$

1.8.1 Parameter estimation

As I previously mentioned the only parameter we would like to estimate is λ, so the parameter space is $\Theta = \lambda$. We used the QMLE to minimize the negative log likelihood function given by:

$$\log \mathscr{L}^{(j)}(\lambda) = -\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{2} \log 2\pi + \frac{1}{2} \log \sigma_t^{(j)^2} + \frac{1}{2} \frac{\varepsilon_t^{(j)^2}}{\sigma_t^{(j)^2}} \right) \tag{44}$$

where $\sigma_t^{(j)^2}$ is the function of the λ

$$\arg\min_{\Theta} \sum_{j=1}^{N} \log \mathscr{L}^{(j)}(\Theta)$$

1.8.2 Simulations

We applied the same simulation scheme as we previously did in Panel GARCH section. The returns are generated randomly from a normal distribution with variance $\sigma_t^{(j)}$:

$$r_{i,t}^{(j)} \sim \mathscr{N}\left(0, \sigma_t^{(j)}\right) \tag{45}$$

We simulated different size of simulated panels with the same sample size, namely T = 500. The parameter which we used for simulation purposes is $\lambda = 0.94$. The results are in the following figure:

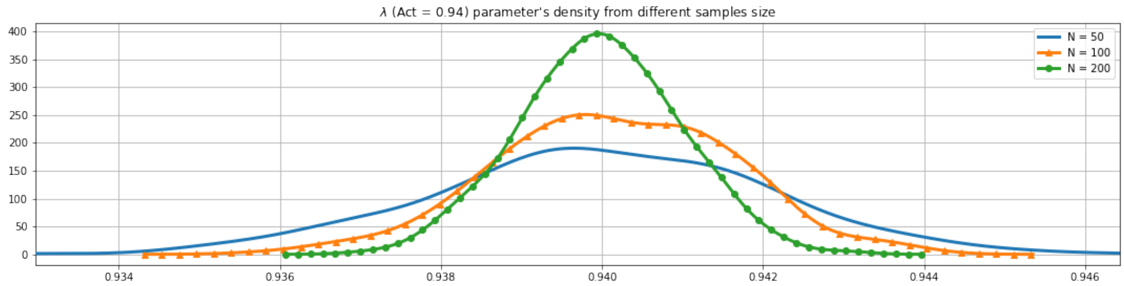

Figure 10: Plot of estimated parameter distributions with N = 100 and the sample size is T = 100, 200

We can observe from the simulation results, that the increment of panel size reduce the variance of the parameter estimation.

2 Empirical Results

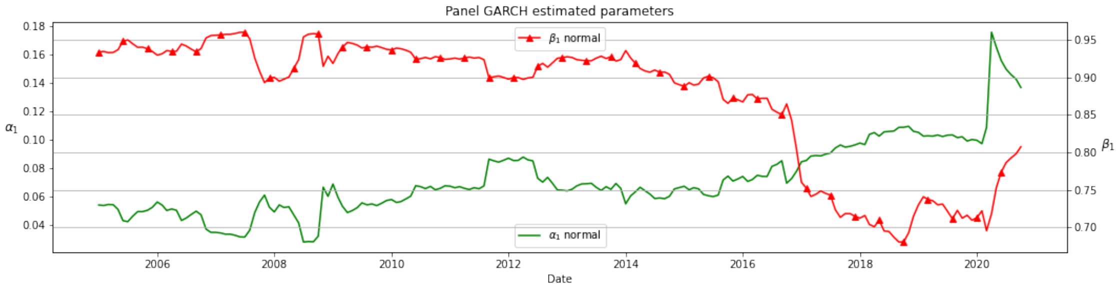

Figure 11: Plot to-do

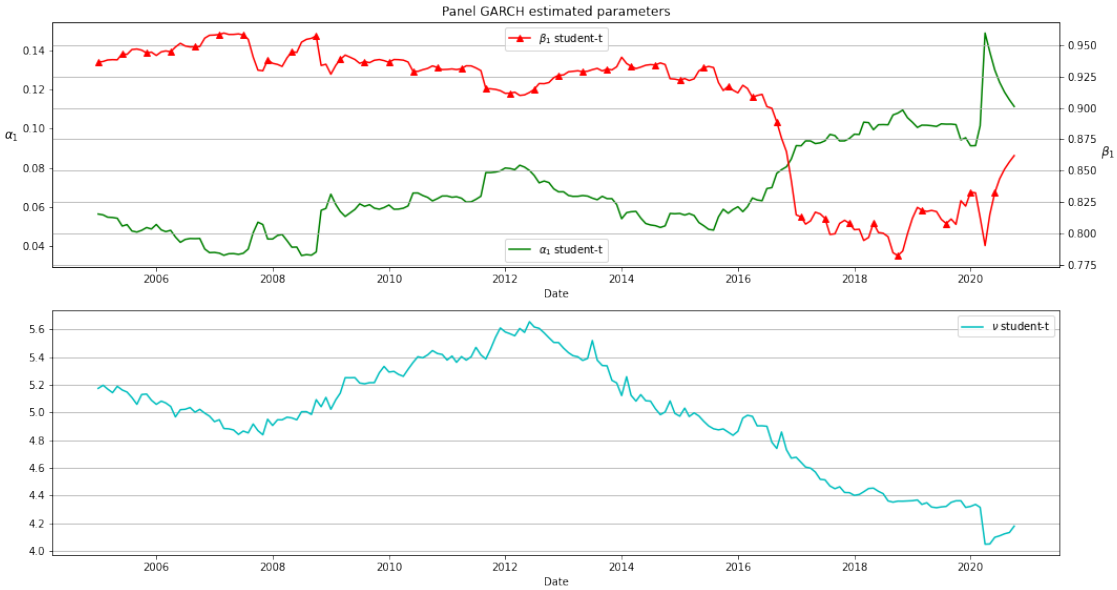

Figure 12: Plot to-do

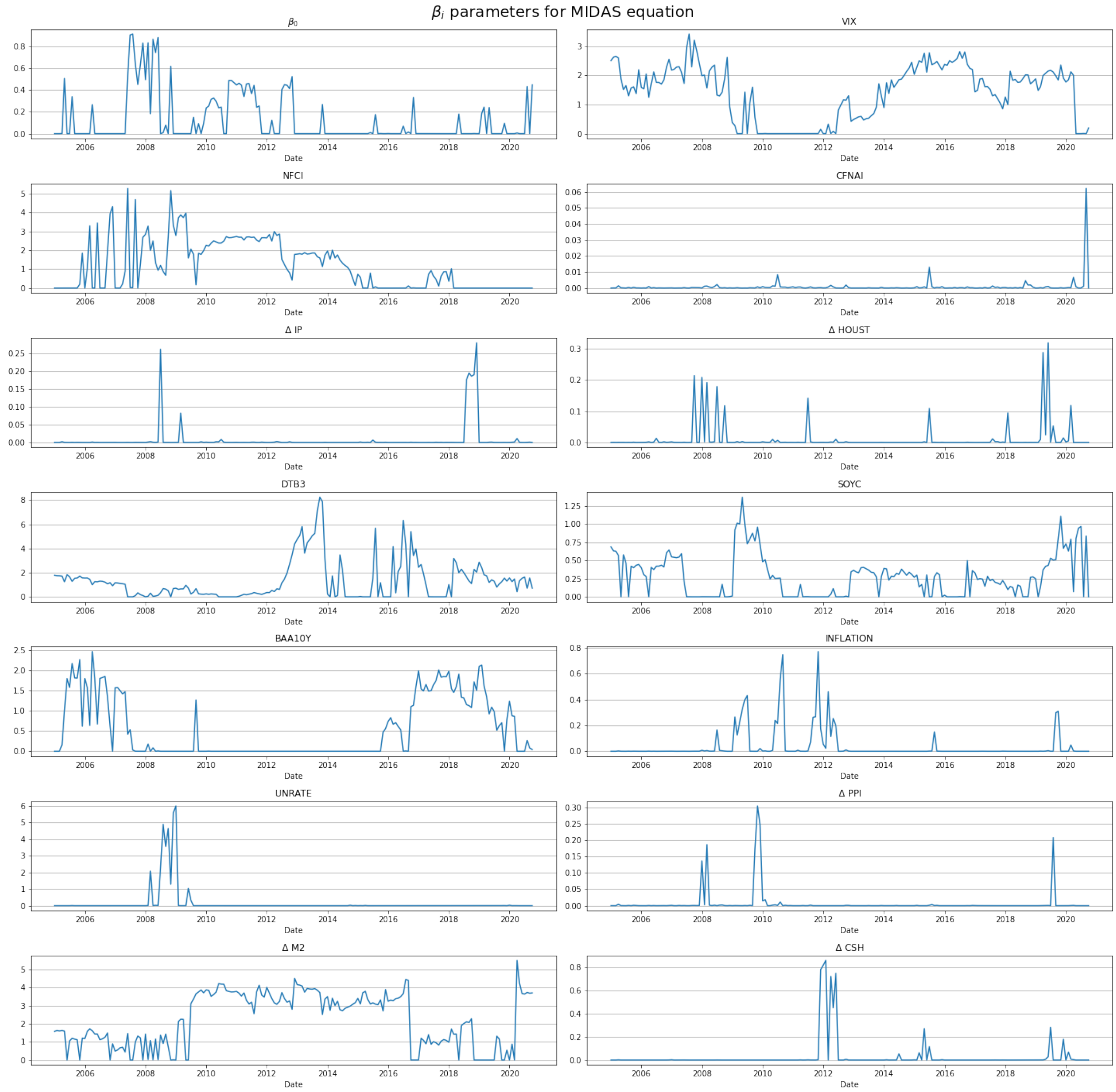

Figure 13: Plot to-do

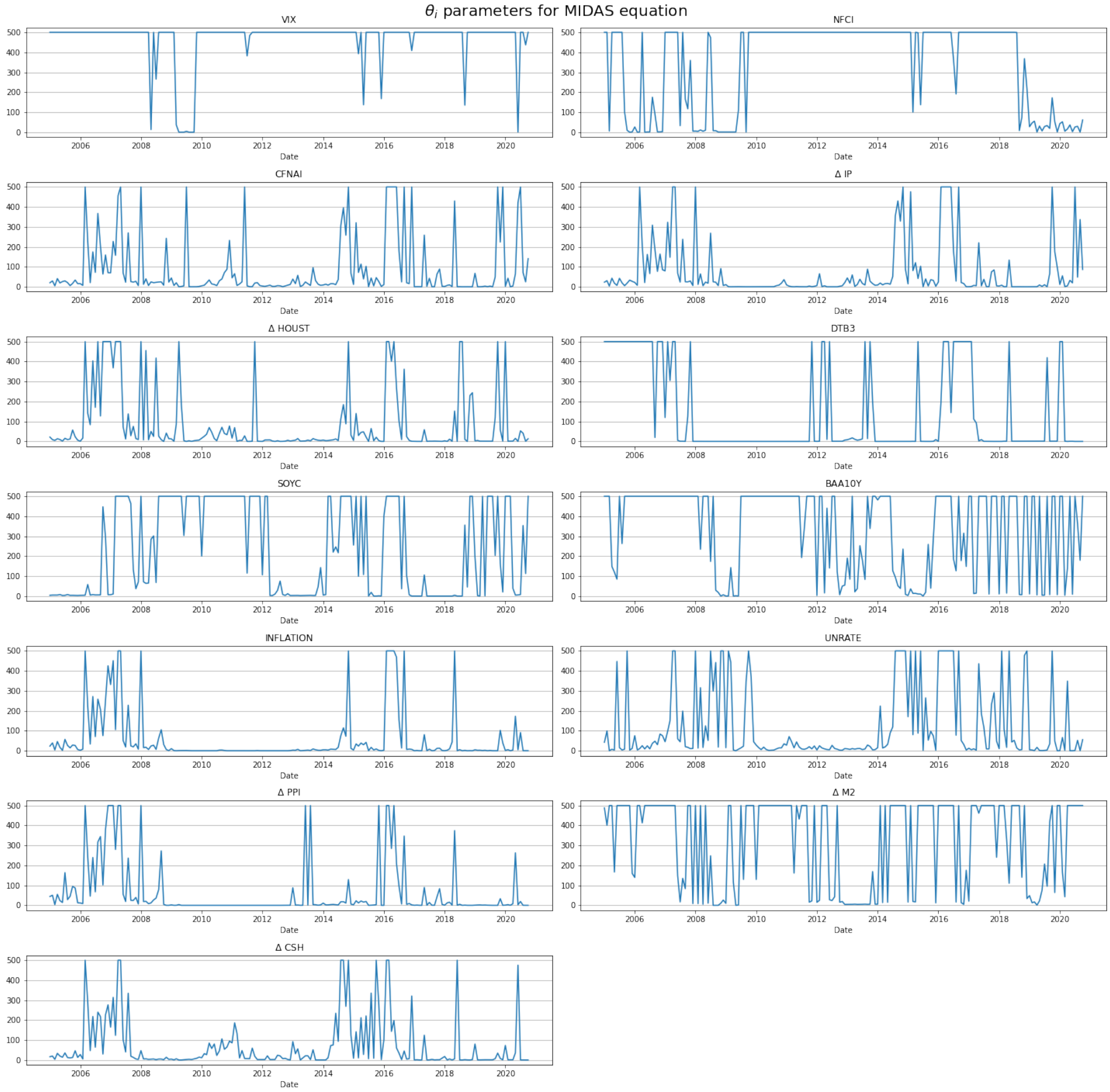

Figure 14: Plot to-do

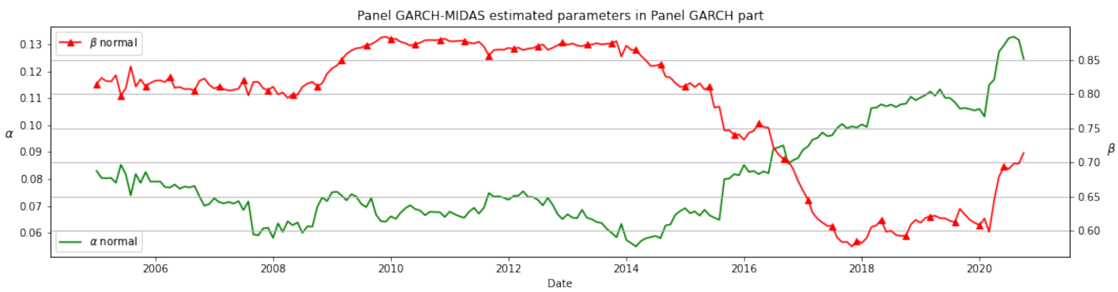

Figure 15: Plot to-do

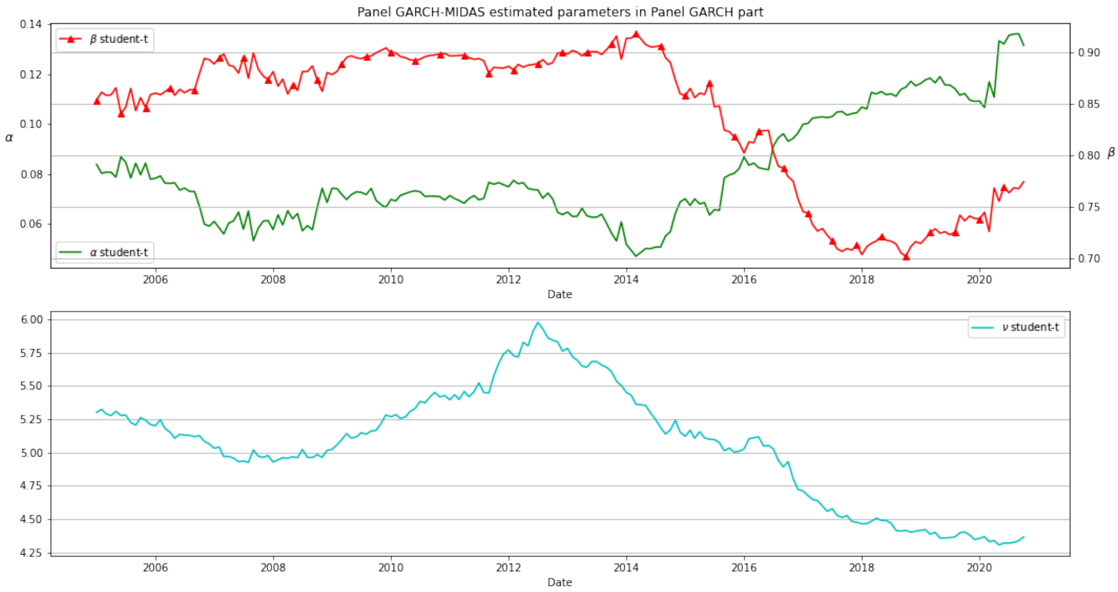

Figure 16: Plot to-do

	mean	std	min	25%	50%	75%	max
β_0	0.123	0.221	0.000	0.000	0.000	0.185	0.910
β_{VIX}	1.408	0.931	0.000	0.516	1.631	2.143	3.412
θ_{VIX}	465.062	119.194	0.005	500.000	500.000	500.000	500.000
β_{NFCI}	1.163	1.285	0.000	0.000	0.800	2.181	5.267
θ_{NFCI}	339.100	221.789	0.000	43.462	500.000	500.000	500.000
β_{CFNAI}	0.001	0.005	0.000	0.000	0.000	0.000	0.062
θ_{CFNAI}	88.756	152.055	0.000	4.645	20.172	69.518	500.000
$\beta_{\Delta IP}$	0.008	0.039	0.000	0.000	0.000	0.000	0.280
$\theta_{\Delta IP}$	74.894	140.399	0.001	0.592	15.074	64.989	500.000
$\beta_{\Delta HOUST}$	0.012	0.045	0.000	0.000	0.000	0.001	0.319
$\theta_{\Delta HOUST}$	89.362	162.229	0.000	1.672	9.590	69.569	500.000
β_{DTB3}	1.299	1.605	0.000	0.113	0.863	1.702	8.244
θ_{DTB3}	136.342	216.862	0.000	0.059	0.859	479.935	500.000
β_{SOYC}	0.278	0.287	0.000	0.000	0.250	0.416	1.366
θ_{SOYC}	233.717	230.906	0.000	3.891	113.813	500.000	500.000
β_{BAA10Y}	0.544	0.735	0.000	0.000	0.001	1.153	2.470
θ_{BAA10Y}	340.812	210.839	0.000	120.335	500.000	500.000	500.000
$\beta_{INFLATION}$	0.038	0.118	0.000	0.000	0.000	0.001	0.770
$\theta_{INFLATION}$	58.145	132.918	0.000	0.139	3.351	28.239	500.000
β_{UNRATE}	0.168	0.836	0.000	0.000	0.000	0.000	5.992
θ_{UNRATE}	124.791	184.858	0.000	8.003	23.292	123.233	500.000
$\beta_{\Delta PPI}$	0.007	0.038	0.000	0.000	0.000	0.000	0.304
$\theta_{\Delta PPI}$	65.414	140.406	0.000	0.055	2.281	28.113	500.000
$\beta_{\Delta M2}$	2.192	1.476	0.000	0.999	2.270	3.576	5.505
$\theta_{\Delta M2}$	324.795	220.850	0.000	29.210	500.000	500.000	500.000
$\beta_{\Delta CSH}$	0.029	0.133	0.000	0.000	0.000	0.000	0.857
$\theta_{\Delta CSH}$	72.195	129.914	0.000	2.221	11.025	70.402	500.000

Table 1: Estimated parameters for the Panel MIDAS model

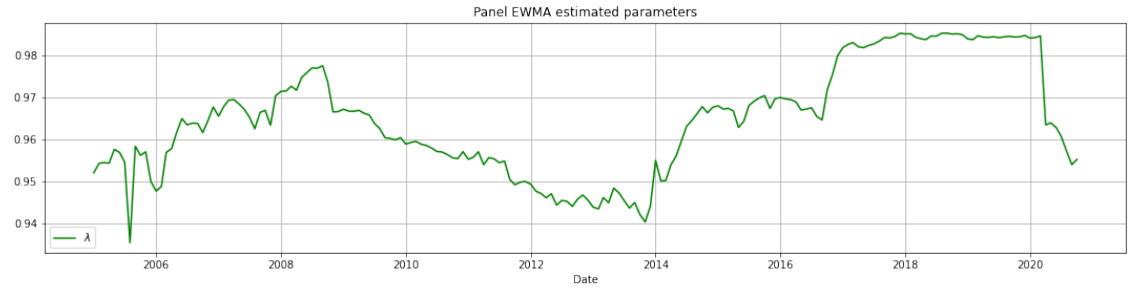

Figure 17: Plot to-do

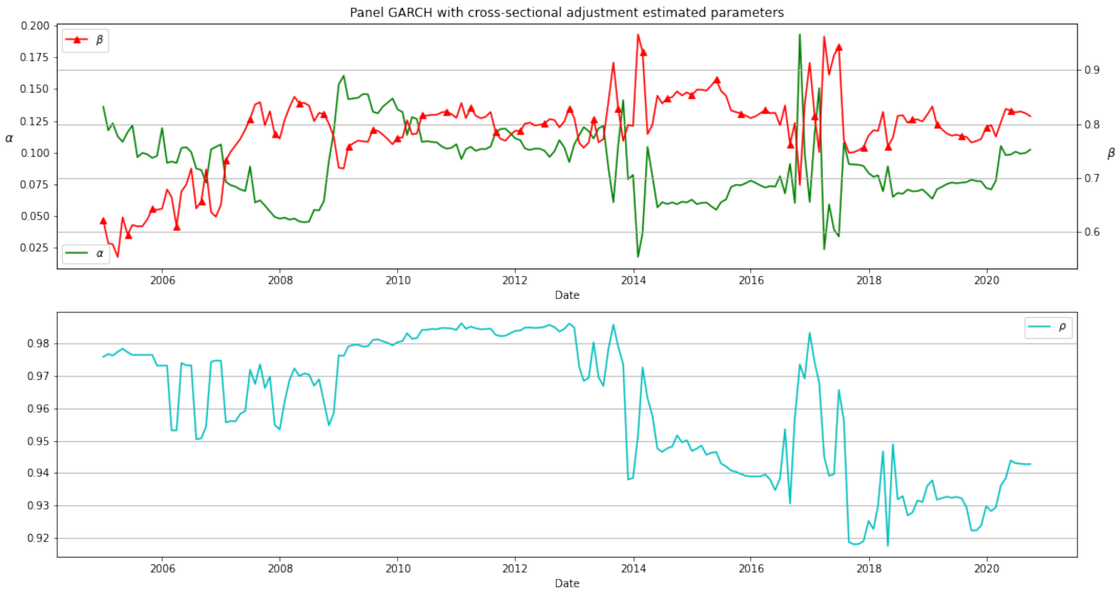

Figure 18: Plot to-do

3 Implementation strategy

In this section I would like to describe our framework, how we implemented and tested out certain models. I have to highlight the fact that every single model, what you can find in this thesis are implemented by myself, so we didn't used other specific python packages for modeling purpuses. One of our main objective was to create a so-called metaclass, where we can declaire all the non-model specific function. The metaclass is a class instances are classes in object-oriented programming. This approach provided us the advantage to test and debug our models more easily. We relied on the Python's abc package to create this metaclass. We decided to use the L-BFGS-B method for minimizing the objective function, in our case the negative loglikelihood function. The L-BFGS-B method relies on the approximation to the Hessian matrix of the loss function, so as we take advantage of information matrix equality we can calculate the standard errors easily.

The first issue we faced in the optimization process was the mishandling of bounds and constraint. In order to avoid such failure of optimization, we implemented a technic to transform model parameters.

3.1 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 θ denote to the parameter, we want to transform:

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

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

In the log likelihood function instead of calculting with the actual θ, the estimation will take place with $\tilde{\theta}$. Then the optimization finished, we can easily transform back. The only issue, which we had to handle was that the estimation of standard errors is not correct, so we impelemented a function called gradient. In this function, you can see we calculated the first derivatives of the possible transformation. θ^* marked as the estimated parameters that were previously transformed.

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

In order to calculate the standard errors of the estimated parameters, now we can calculate their gradients, the inverse Hessian matrix was estimated throughout the optimization process, so we have to take the square root of the diagonal of the inverse Hessian and multiple by the gradients.

3.2 Model definition framework

In the following figure you can see how we defined certain parts of the model definition. As I mentioned previously, the main objective was to create a metaclass, where we define all the common functions. It will be denote with "Base" in the following figure.

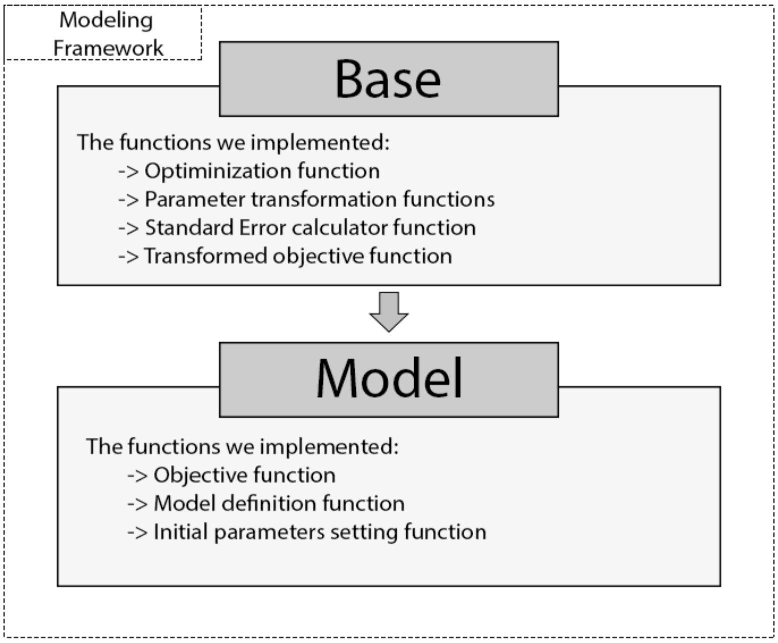

Figure 19: Modelling framework

The sections above described all the necessary informations about the ”Base” class, so I would like to write about the ”Model” class. This class is unique for each and every model, so this mean all the model what I described in Models section are one-one model class. The initial parameters setting function is the one, which I haven't desribed yet. When you use certain optimization method, you have to declare initial values for parameters, which serve as a starting point for the optimization algorithm. As you can imagine it is crutual to find good initial values, because the optimization routine can fail. Our approach, for find them, was kind of a trial-error, what I meant to mean is for example in the GARCH model you know that most of the cases approximately $\alpha_1 \approx 0.1$ and $\beta_1 \approx 0.9$, so we decided to initialize them as [0.05, 0.85].

4 Predictive ability tests

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 [14] 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. He described the following family of loss functions, indexed by the scalar parameter b:

$$L\left(\hat{\sigma}^2, h; b\right) = \begin{cases} h - \hat{\sigma}^2 + \hat{\sigma}^2 log\frac{\hat{\sigma}^2}{h} & \text{, b = -1} \\ \frac{\hat{\sigma}^2}{h} - log\frac{\hat{\sigma}^2}{h} - 1 & \text{, b = -2} \\ \frac{1}{(b+1)(b+2)}\left(\hat{\sigma}^{2b+4} - h^{b+2}\right) - \frac{1}{b+1}h^{b+1}\left(\hat{\sigma}^2 - h\right) & \text{, otherwise} \end{cases}$$

In order to define different loss functions with only one parameter, namely with b, we implemented the above family of loss functions in our DM test. The author pointed out, the MSE loss function is obtained when b = 0 and the QLIKE when b = -2. These two function are the only robust ones, moreover the QLIKE is less sensitive with respect to extreme observation than the MSE loss. DM test relies on assumptions made directly on the forecast error loss differential [5]. In DM test we compare two rival models, by taking there loss differencies in the following way:

$$d_{12t} = L\left(\hat{\sigma}^2, h^{(1)}; b\right) - L\left(\hat{\sigma}^2, h^{(2)}; b\right) \tag{46}$$

where $\hat{\sigma}^2$ is the volatility proxy variable. [5] presented that DM assumes:

$$DM = \begin{cases} E(d_{12t}) = \mu & \text{, } \forall \text{ t} \\ cov(d_{12t}, d_{12(t-\tau)}) = \gamma(\tau) & \text{, } \forall \text{ t} \\ 0 < var(d_{12t}) = \sigma^2 < \infty & \text{, otherwise} \end{cases}$$

$$DM = \frac{E\left(d_{12t}\right)}{\sqrt{var\left(d_{12t}\right)}} \sim \mathcal{N}(0,1) \tag{47}$$

where $E(d_{12t})$ is the sample mean loss differential and $\sqrt{var\left(d_{12t}\right)}$ is a consistent estimate of the standard deviaton of d_{12t}.

The DM-test is a handy when we compare two forecasts for a single asset, what if we would like to compare two forecasts for panels. [15] gave an approach us to do that. They developed new methods for testing forecasting powers in panels Let's start with the calculation of the mean loss throughout the whole panel, where we denote $\bar{L}_m$ as

$$\bar{L}_h = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} L\left(\hat{\sigma}_t{}^{(j)^2}, h_t^{(j)}; b\right) \tag{48}$$

Their first hypothesis was that the panel's mean loss for a pair of forecasts, $h^{(1)}$ and $h^{(2)}$ is equal in expectation:

$$H_0^{panel} : E\left(\bar{L}_{h^{(1)}}\right) = E\left(\bar{L}_{h^{(2)}}\right) \tag{49}$$

In order to test the null hypothesis let's use the above defined d_{12t}, which is now a matrix and not an array, so the loss differential between foecasts $h^{(1)}$ and $h^{(1)}$ as

$$d_{12t} = L\left(\hat{\sigma}^2, h^{(1)}; b\right) - L\left(\hat{\sigma}^2, h^{(2)}; b\right) \tag{50}$$

We can test the null hypothesis in (37) using the test statistic

$$DM = (NT)^{-\frac{1}{2}} \frac{\sum_{j=1}^{N} \sum_{t+h=1}^{T} d_{12t}}{\hat{\sigma}\left(d_{12t}\right)} \tag{51}$$

where $\hat{\sigma}\left(d_{12t}\right)$ is a consistent estimator of standard deviation of d_{12t}.

Furthermore, we used the Mincer-Zarnowitz Regression [12], which is basically a simple regression to be estimated:

$$\sigma_{t+1}^2 = \beta_0 + \beta_1 \hat{\sigma}_{t+1}^2 + \varepsilon_t \tag{52}$$

where $\hat{\sigma}_{t+1}^2$ is the predicted value of the volatility for t+1. If the prediction is unbiased, then the coefficients are $\beta_0 = 0$ and $\beta_1 = 1$. We also investigate the R^2 aswell.

Another measure, what we investigated is the RMSE (root mean squared error)

$$RMSE = \sqrt{\frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} \left(\sigma_t^{(j)^2} - \hat{\sigma}_t^{(j)^2}\right)^2} \tag{53}$$

where we can compare the RMSE values to eachother.

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. A. Mincer and V. Zarnowitz. The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pages 3–46. National Bureau of Economic Research, Inc, October 1969.

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

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

[15] A. Timmermann and Y. Zhu. Comparing forecasting performance with panel data. May 2019.

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

A 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:

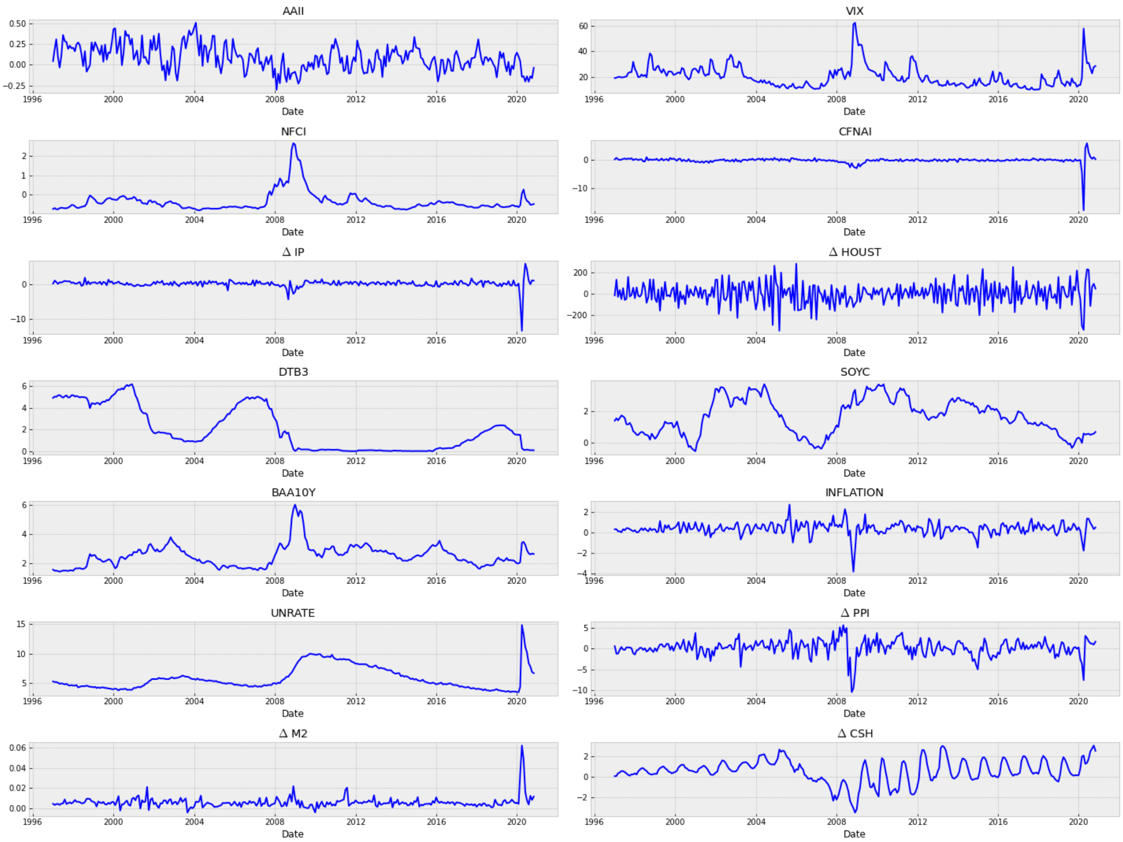

Figure 20: Time-series of macroeconomic variables

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