Take Home Exam

Axel Sjöberg - ax387sj-s@student.lu.se

January 2021

1 Portfolio Optimization

In this section we develop dynamical trading strategies using a Markowitz mean-variance framework. The investment problem we are to solve is one where we combine the S&P 500 ETF (SPY) and cash. There are several ways in which we could model the cash asset, e.g use ICSH as proxy, but in this problem we only consider the risk in the SPY and assume that the cash account will be constant, i.e the expected return and standard deviation is zero. Portfolio optimization is a framework for finding a portfolio that maximizes the expected return given a certain level of risk. In this exercise we will find the weights for such a portfolio by solving (1):

$$\begin{aligned} &w = argmax\ \mu_t^T w - \frac{\gamma}{2} w^T \Sigma_t w \\ &w \in C \end{aligned} \tag{1}$$

for some set C. I will experiment on using some different C when developing my trading strategies. As the expected return and standard deviation of the cash asset is zero we can write (1) as (2):

$$w_{spy} = argmax\ \mu_{spy} w_{spy} - \frac{\gamma}{2} w_{spy}^2 \sigma_{spy}^2 \tag{2}$$

Which we get the solution to by derivation, see (3):

$$\begin{aligned} w_{spy} &= \frac{\mu_{spy}}{\gamma \sigma_{spy}^2} \\ w_{cash} &= 1 - w_{spy} \end{aligned} \tag{3}$$

Our first assignment in this section is to fit a simple model constant parameters such that $r_t \sim N(\mu, \Sigma)$ and optimize the risk aversion parameter γ such that we have 60 % in the risky asset. (3) gives us that:

$$\gamma = \frac{\mu_{spy}}{w_{spy} \sigma_{spy}^2} \tag{4}$$

Before calculating γ using some simple approach, we need to have a look at our data to see if there are any irregularities. In fact there are two times

mean	variance	γ
0.00039	0.00012	5.4

Table 1: Log returns model parameters and corresponding risk parameter

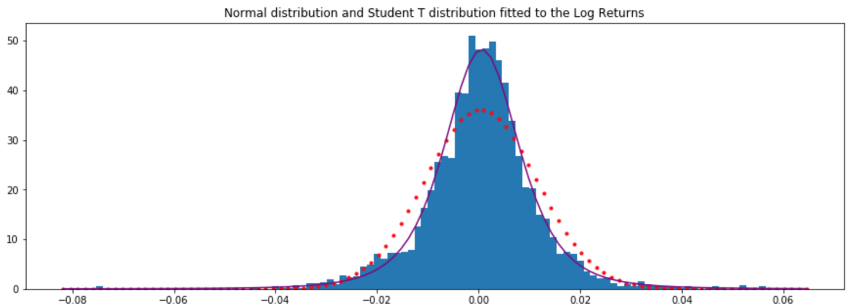

Figure 1: Histogram of Log returns with pdf of fitted Student-t (continuous line) and Gaussian distribution (dotted line) using maximum likelihood estimation.

in between 1993-2021 when the markets were forced to close, 9/11 and the days following the terrorist attack and the Sandy hurricane disaster in 2012. Of course the markets have been closed on other days as well like weekends, national holidays and days of mourning for ex-presidents that have passed away. We assume that these "missing" data-entries should not have too much of an impact and therefore disregard them, as is the standard approach in financial modelling.

To impute the data on the missing date we first have a look at the distribution of our log returns, r_t (on the training data naturally) , which is presented in fig 1 where we also have plotted two PDFs, one belonging to a Gaussian distribution and one belonging to a student t distribution. Both of them are fitted using an ML estimation on our training data. From the looks of it, the student-t distribution seems to be a better fit to our data than the Gaussian distribution. We aussume here that $r_t \in t_\nu$ with $\nu = 3.6, \mu = 0.00061, \sigma = 0.0077$. We use this distribution to impute data using a simple Brownian bridge. After this step there were no missing data and we can start the modelling.

The risk parameter was calculated by using (4) and together with $\mu_{spy} = E[r_t]$ and $\Sigma_t = var[r_t]$. This might seem strange as we imputed the data using parameters that were fitted to a student t distribution, but I found that using the parameters given by the student t fit to solve (4) to give an unreasonable large value to γ (16). Thus I chose to use the mean and variance instead, which would correspond to a normal distribution. We do not by any strict sense have to use the same to impute the data, we could have used some simpler interpolation method but I think this one works fine. The variance, mean and γ is presented in table 1

From looking at the log returns it seems like the changes of larger magnitude have a tendency to cluster. Thus a GARCH-type model seems appropriate for modelling the volatilises. There a few variations available that are more or less suitable in different settings. For this exercise I have tried out four different variations, GARCH, EGARCH, GJR-GARCH and TARCH. The GARCH model is given by (5).

$$
\begin{aligned}
&r_t = \mu + \epsilon_t \\
&\epsilon = \sigma e_t \qquad \sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_{t-k}\epsilon_{t-k}^2 + \sum_{k=1}^{q} \beta_{t-k}\sigma_{t-k}^2
\end{aligned} \tag{5}
$$

where e_t has some distribution, usually standard normal or Student-t.

The EGARCH, GJR-GARCH and TARCH model models the volatility in a slightly different way. The EGARCH volatility is given by (6) , The GJR-GARCH volatility is given by (7) and the TARCH volatility is given by (8). An advantage of the EGARCH model is that we have fewer restrictions on our parameters but we need to choose a suitable f(·) function for the error terms. THE GJR-GARCH and TARCH model have a nice intuitive interpretation which is that the estimated volatility is treated differently when $\epsilon < 0$. It is pretty well know that negative news tend to increase the volatility more than positive, losses loom larger than gains. In the GJR-GARCH(1,1,1) and TARCH(1,1,1) model this corresponds to a positive value of γ.

$$\log \sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_{t-k} f(\epsilon_{t-k}) + \sum_{k=1}^{q} \beta_{t-k} \log \sigma_{t-k}^2 \tag{6}$$

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_{t-k}\epsilon_{t-k}^2 + \sum_{k=1}^{u} \gamma_{t-k}\epsilon_{t-k}^2 I_{\epsilon_{t-k}<0} + \sum_{k=1}^{q} \beta_{t-k}\sigma_{t-k}^2 \tag{7}$$

$$\sigma_t = \omega + \sum_{k=1}^{p} \alpha_{t-k}|\epsilon_{t-k}| + \sum_{k=1}^{u} \gamma_{t-k}|\epsilon_{t-k}| I_{\epsilon_{t-k}<0} + \sum_{k=1}^{q} \beta_{t-k}\sigma_{t-k} \tag{8}$$

In the EGARCH model I selected $f(\cdot) = \sum_{k=1}^{p} \alpha_k(|\epsilon_{t-k}| - \sqrt{2/\pi}) + \gamma_k \epsilon_{t-k}$. I tried several variations e.g $f(\cdot) = \sum_{k=1}^{p} \alpha_k \epsilon_{t-k}$ and $f(\cdot) = \sum_{k=1}^{p} \alpha_k \epsilon_{t-k}^2$ but these were not as good. We also get an asymmetrical term γ which share some nice properties with the GJR-GARCH and TARCH model. If the γ is negative we have that bad news generate more volatility than good news. We also have the regular GARCH-effect from the α_k parameters. For this EGARCH variation the model can be specified in a similar fashion as we did with the EGARCH and TARCH, see (9)

$$\log \sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k(|\epsilon_{t-k}| - \sqrt{2/\pi}) + \sum_{k=1}^{u} \gamma_k \epsilon_{t-k} + \sum_{k=1}^{q} \beta_{t-k} \log \sigma_{t-k}^2 \tag{9}$$

The fashion in which I first selected which model orders were suitable for the different models was the following. I started with the lowest possible model

GARCH model parameter estimates							
Model	μ	ω	$\alpha1$	$\alpha2$	$\beta1$	$\beta2$	ν
(1,1)	0.0684	0.0052	0.0614	-	0.9360	-	7.6293
(1,2)	0.0684	0.0065[1]	0.0328[1]	0.0392[1]	0.9250	-	7.6023
(2,1)	0.0684	0.0052	0.0614	-	0.9360	3.4175e-10[1]	7.6290

Table 2: GARCH parameter estimates. Estimates with a footnote are not significant.

EGARCH model parameter estimates											
Model	μ	ω	$\alpha1$	$\alpha2$	$\alpha3$	$\gamma1$	$\gamma2$	$\gamma3$	$\beta1$	$\beta2$	ν
(1,1,1)	0.0444	0.0001[1]	0.1169	-	-	-0.1015	-	-	0.9855	-	8.7
(1,2,1)	0.0424	0.0005[1]	0.1078	-	-	-0.2175	0.1300	-	0.9883	-	8.9
(2,1,1)	0.0382	0.0006[1]	-0.0596[1]	0.1945	-	-0.1158	-	-	0.9813	-	8.8
(1,1,2)	0.0444	0.0001[1]	0.1224	-	-	-0.1077	-	-	0.9209	0.0640[1]	8.7
(2,2,1)	0.0408	0.0003[1]	-0.0937	0.2181	-	-0.2398	0.1422	-	0.9857	-	9.3
(2,2,2)	0.0408	0.0003[1]	-0.0936	0.2181	-	-0.2398	0.1422[1]	-	0.9857[1]	1.0626e-12[1]	9.3
(2,3,1)	0.0389	0.0005[1]	-0.0913	0.2053	-	-0.2293	0.0164	0.1321[1]	0.9885	-	9.3
(3,2,1)	0.0414	0.0003[1]	-0.0945	0.2363	-0.0187[1]	-0.2394	0.1428	-	0.9860	-	9.3

Table 3: EGARCH parameter estimates. Estimates with a footnote are not significant.

order. Then I added parameters 1 at the time. If the added parameter was significant I added parameters from this new baseline. I did this until I could not add any more parameters that were significant. The estimated parameters are presented in table 2, 3, 4 and 5. As we see it is only for the EGARCH model we have a more complex model order. In order to determine which one of these I should use, I used a likelihood ratio test approach. The deviance statistic which compare the likelihood ratios is given by (10). By the Wilks theorem this follows a χ^2 distribution where the degree of freedom is the positive difference between the two models we compare. If $\lambda_{LR} >$ c, where c is a p-quantile of the the χ^2 distribution (I use the standard 95 % quantile), we accept the more complex model.

$$\lambda_{LR} = -2(l(\theta_0) - l(\hat{\theta})) \tag{10}$$

The LR comparing an EGARCH(1,1,1) to an EGARCH(1,2,1), gave $\lambda_L R$ = 17.09 which is larger than the 95th quantile, 3.84. Thus EGARCH(1,2,1) is preferred over EGARCH(1,1,1). Similarly, comparing the EGARCH(1,2,1) with the EGARCH(2,2,1) gave λ_{LR} = 22.91 which is also larger than 3.84.

GJR-GARCH model parameter estimates								
Model	μ	ω	$\alpha1$	$\gamma1$	$\gamma2$	$\beta1$	$\beta2$	ν
(1,1,1)	0.0478	0.0095	7.9227e-13[1]	0.1185	0	0.9311	-	8.5591
(1,2,1)	0.0478	0.0095	2.0116e-11[1]	0.1186	8.7104e-12[1]	0.9311	-	8.5592
(1,1,2)	0.0478	0.0097	9.6863e-12[1]	0.1216	0	0.9004	0.0289[1]	8.5587

Table 4: GJR-GARCH parameter estimates. Estimates with a footnote are not significant.

TARCH model parameter estimates								
Model	μ	ω	$\alpha1$	$\gamma1$	$\gamma2$	$\beta1$	$\beta2$	ν
(1,1,1)	0.0427	0.0145	0.0068[1]	0.1137	0	0.1137	-	8.6550
(1,2,1)	0.0427	0.0145	0.0068[1]	0.1137	2.7792e-10[1]	0.9363	-	8.6547
(1,1,2)	0.0427	0.0149	0.0065[1]	0.1185	0	0.8914	0.0428[1]	8.6553

Table 5: TARCH parameter estimates. Estimates with a footnote are not significant.

EGARCH(2,2,1) is therefore chosen as the best order for the EGARCH model.

The TARCH(1,1,1) is preferred over the GJR-GARCH(1,1,1) as it has the same number of parameters but perform worse on AIC,BIC and likelihood. They are also somewhat similar models so I decided to drop the GJR-GARCH. The EGARCH(2,2,1) is preferred over the TARCH(1,1,1) and the GARCH(1,1) when using a likelihood ratio test and it has the lowest AIC and BIC. Everything therefore indicates that this is our best model. As a final check to see if this is the case, the QQ-plot of the log returns normalized by the predicted volatility are plotted for the three different models, see figure 2, 3 and 4. The QQ-plot looks good for all the models, however, just looking at them I don't notice any major difference between the EGARCH(2,2,1) and the TARCH(1,1,1) whereas the GARCH(1,1) looks slightly more off. As the EGARCH(2,2,1) is better in every other regard I dropped the other models and continued with the EGARCH(2,2,1) model.

The predicted volatility is used in the trading strategies. Each day, the volatility is predicted using the EGARCH(2,2,1) and then the optimal weights for the portfolio is computed using equation (3) with the risk parameter $\gamma = 5.4$ calculated above. Two strategies are used, both of them use (3), the difference is that the first strategies has $c \in [0{,}1]$ and while strategy two has $c \in [0.2{,}1.2]$. The reason for the somewhat arbitrary choice in c for strategy 2 is that after a large negative shock or a period of sharp decline, it is not unusual to see a quick bounce back, just look at this years stock market. We also allow in this strategy for a modest leveraged position in SPY. I think some restrictions on how much

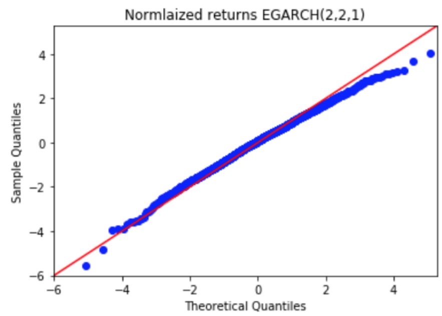

Figure 2: Log returns during the training period normalized by the predicted volatility from the EGARCH(2,2,1) model. The theoretical quantiles comes from a Student-T distribution with $\nu = 9.3$

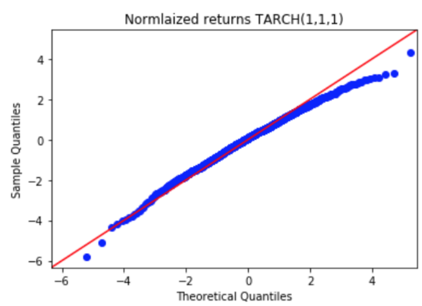

Figure 3: Log returns during the training period normalized by the predicted volatility from the TARCH(1,1,1) model. The theoretical quantiles comes from a Student-T distribution with $\nu = 8.7$

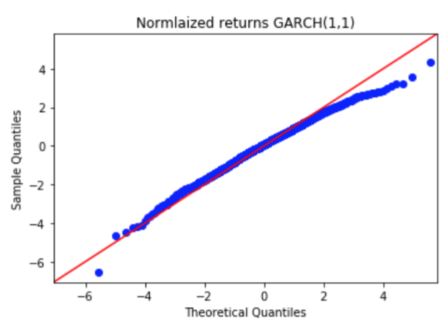

Figure 4: Log returns during the training period normalized by the predicted volatility from the GARCH(1,1) model. The theoretical quantiles comes from a Student-T distribution with ν =7.6

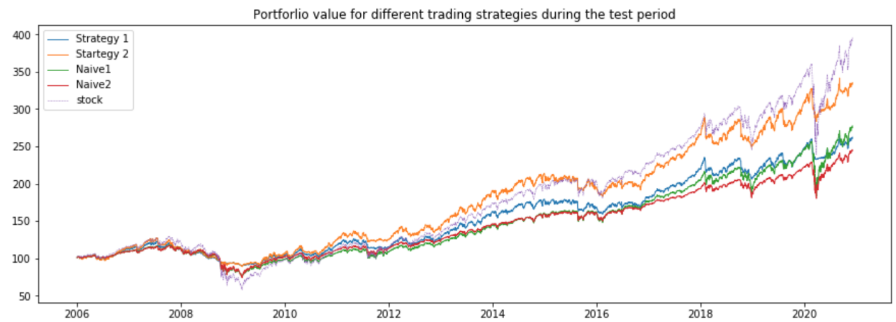

Figure 5: Development of portfolio value on evaluation data

we borrow needs to made, not to mention if and how much you are allowed to short a stock. Ideally this should be determined by fitting the endpoints in the interval c by treating them as hyper-parameters. Due to time-constraints, the strategies mentioned will have to do for now.

The two trading strategies are compared with two naive trading strategies: In the first of these, Naive 1, 60 % of the initial sum is invested in SPY and 40 % is kept as cash. No trades are made throughout the evaluation period. In Naive 2 we have a 60 % weight in SPY throughout the entire evaluation period by re-balancing the portfolio everyday. In figure (5), the value of the portfolio for the my two strategies, the two naive strategies and the value of SPY are plotted against each other. As we see both of my strategies perform rather well compared to the two naive approaches. At more uncertain times we can clearly see how my two strategies' development plateaus while the naive approaches plummet. This can be seen in the both during the longer period after the Lehman Brothers crash and this year when the markets crashed.

In figure 6 the development of the portfolio value, predicted volatility, Log Returns, and weights of the portfolio during the evaluation period is plotted in 4 separate sub-figures. We see that the size of the log returns are rather clustered and that the predicted volatility looks reasonable given the log-returns. Moreover, the figure highlights that when we have a a high predicted volatility the weights in SPY drop accordingly. The two strategies actually work as I had intended for them. During the financial crisis in 2008 we can see that..

In 6 the expected log-return, standard deviation of the log-returns and Sharpe Ratio for the different trading strategies are presented. Both strategy 1 and 2 perform a lot better than the two naive approaches. Strategy 2 has a higher expected return and a lower standard deviation than the naive approaches while Strategy 2 has a marginally lower expected return than Naive1. Strategy 2 however has a lot lower standard deviation than Naive 1 and therefore has a much higher Sharpe ratio. Naive 2 performed worse than Strategy 1 and 2 on all metrics. I argue that both of my strategies are better than the naive ones, and that modelling the volatility with an EGARCH(2,2,1) model is

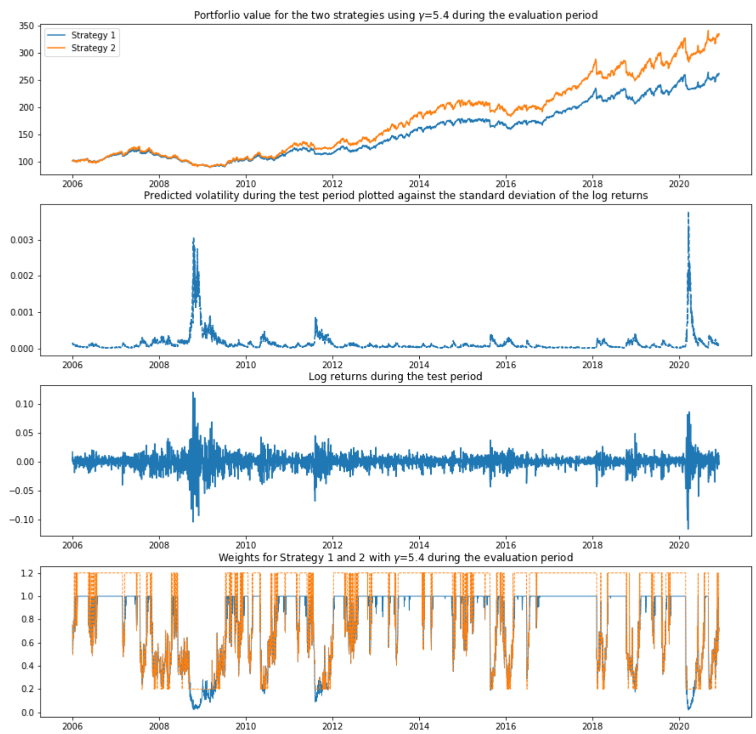

Figure 6: Development of the portfolio value, predicted volatility, log-returns (SPY), and weights (in SPY) of the portfolios during the evaluation period. If two plots in one sub-figure then blue corresponds to strategy 1 and orange to strategy 2

Evaluation results			
Strategy	μ	σ	Sharpe Ratio
100 % SPY	**0.00037**	0.01272	1.0172
Naive 1	0.00027	0.00847	1.13388
Naive 2	0.00024	0.00762	1.10776
Strategy 1	0.00026	**0.00636**	1.42615
Strategy 2	0.00032	0.00751	**1.51622**

Table 6: Expected log-return, standard deviation of the log-returns and Sharpe Ratio (recalculated as the period rate) for the different trading strategies.

better than assuming constant volatility as the Naive 2 strategy can be seen as a proxy for.

2 COVID-19

2.1 Introduction to the problem

In the assignment a SIR model is fitted to simulated data. This is done using the stochastic generalization presented in (11).

$$\begin{aligned} \begin{pmatrix} dS \\ dI \end{pmatrix} &= \begin{pmatrix} -\beta SI/N \\ \beta SI/N + \gamma I \end{pmatrix} dt + \\ &\begin{pmatrix} -\sqrt{SI/N} & 0 \\ \sqrt{SI/N} & -\sqrt{I} \end{pmatrix} d\boldsymbol{W} \end{aligned} \tag{11}$$

Where $d\boldsymbol{W}$ is a row vector containing two independent Brownian motions (BM).

This stochastic generalization varies from the basic form model given in the problem description. First I have removed the R parameter. I could do so as the population was roughly constant throughout the entire period. Thus, setting the total population as a constant number, N (10000009), the number of recovered (R) could be inferred via the number of susceptible (S), infected (I) and N by: $R = N - S - I$. Thus we reduce the dimension of the problem from 3 to 2, which simplifies our problem.

For this assignment we have low frequency data which makes the most simple approach, a discretized maximum likelihood approach infeasible as the estimator is not constant. There are a few approaches available that have consistent estimators, GMM-type estimators and approximate ML estimators. Of these methods I chose to use approximate ML estimator as GMM-type estimators require us to correctly specify the moments, which can be difficult. There are three approximate ML estimators we have covered in class. Of these, I chose to use simulated likelihood for this assignment together with a Euler Maruyama (EM) discretization scheme. The EM discretization of the stochastic generalization given in (11) is given by (12).

$$\begin{aligned} \begin{pmatrix} S_{t+1} \\ I_{t+1} \end{pmatrix} &= \begin{pmatrix} S_t \\ I_t \end{pmatrix} + \begin{pmatrix} -\beta S_t I_t/N \\ \beta S_t I_t/N + \gamma I_t \end{pmatrix} \Delta t + \\ &\begin{pmatrix} -\sqrt{S_t I_t/N} & 0 \\ \sqrt{S_t I_t/N} & -\sqrt{I_t} \end{pmatrix} d\boldsymbol{\Delta W} \end{aligned} \tag{12}$$

which we can write on the more compact form:

$$X_{\tau+1} = X_\tau + \alpha(X_\tau, \theta)\Delta\tau + \sqrt{\beta(X_\tau, \theta)}\Delta \boldsymbol{W_\tau} \tag{13}$$

Where $\boldsymbol{W_{\tau_k}} \sim N(0, \Delta\tau I)$ and I is the identity matrix. We assume no measurement noise, and with the notation in (13), the posterior density of one diffusion bridge is given by (14).

$$\pi(x_{(t,t+1)}|x_t, x_{t+1}, \theta) \propto \Pi_{\tau=0}^{m-1} \pi(x_{\tau+1}|x_\tau, \theta) \tag{14}$$

where,

$$\pi(x_{\tau+1}|x_\tau, \theta) = N(x_{\tau+1}; x_\tau + \alpha(x_\tau, \theta)\Delta\tau, \beta(x_\tau, \theta)\Delta\tau) \tag{15}$$

Between each real observations we introduce $m-1$ intermediary time points, $(x_\tau, x_{\tau+1}, ..., x_{\tau+m-1})$ where the value for each of these are generated by the density in (14). Between each consecutive observations, k bridges are simulated, and the likelihood of moving from $x_{\tau+m-1}$ to $x_{\tau+m} = x_{t+1}$ is averaged over all the bridges between those observations. The likelihood in moving from x_t to x_{t+1} is approximately proportional to:

$$L = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp(-\frac{1}{2}(X_{t+1} - X_{t_m})^T \Sigma^{-1}(X_{t+1} - X_{t_m})) \tag{16}$$

where $\Sigma = \sqrt{\beta(X_{t_{m-1}})}\sqrt{\beta(X_{t_{m-1}})}^T \Delta t$, which in our case is equal to eq 17

$$\Sigma = \begin{pmatrix} SI/N & -SI/N \\ -SI/N & SI/N + \gamma I \end{pmatrix} \Delta t \tag{17}$$

By Sylvesters criterion we have that in $\tilde{\Sigma} = \Sigma + \epsilon I_2$, I_2 being the two dimensional identity matrix and ϵ a small positive number, $\tilde{\Sigma}$ is a positive definite matrix and thus also invertible. This also ensures us that the determinant is strictly positive. We can therefore specify the negative log-likelihood (ll_-) according to eq 18:

$$ll_- = \frac{\log|\tilde{\Sigma}| + (X_{t+1} - X_{t_m})^T \tilde{\Sigma}^{-1}(X_{t+1} - X_{t_m})}{2} \tag{18}$$

where we have dropped some of the terms that are not effected by our choice of the parameters. Using $\tilde{\Sigma}$ instead of Σ should not impact the likelihood to much as the $\log(|\tilde{\Sigma}|)$ is generally small compared to the Mahalanobis distance term.

Between every consecutive observation a path is simulated k times and for each of them the negative log-likelihood is calculated using (18). These are then added together and divided by k. Before this summation, sample paths that have a very high log-likelihood are removed from the sample set. This can be seen as a very basic form of importance sampling where instead of taking all the samples multiplicated with a weight that have been calculated using a probability density function, we simple take all the samples that yield a very low likelihood times zero. Of course this could have been done a lot better but due to time constraints I did not have enough time to improve this step. I do recognize the drawbacks of my implementation, however it was also an implementation that proved to be good enough. The process in this paragraph is repeated between each consecutive observation and we sum all of the average likelihoods, which is the likelihood we will use as the loss function in the numerical optimization. Worth mentioning is that the paths are sampled using random common numbers to help the optimizer. This is essentially the Pedersen algorithm but with a simplified importance sampler.

β	2.3086 **2.3115** 2.3145
γ	0.3814 **0.3826** 0.3838
σ_1	26.8667 **27.0495** 27.2322
σ_2	30.8823 **30.8973** 30.9123

Table 7: Parameter estimates with confidence interval of SIR model

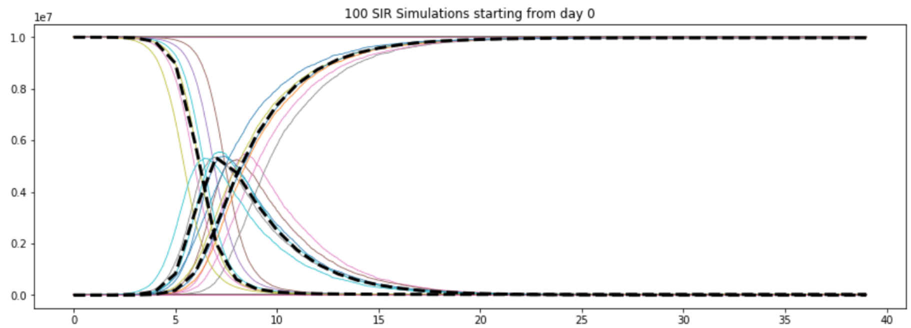

Figure 7: 100 simulations of epidemic starting from Day 0 using the fitted parameter estimates

The parameters are fitted using SciPy's optimizer, the BFGS method and $m = 52$. In table 7 the estimated parameters are presented and in figure 7, 100 simulations from the fitted model are plotted and in figures 8, 9, 10, 10 simulations from the fitted model are plotted. I present 100 simulations from day 0 because as we can see it is rather unlikely that the virus will propagate and become an epidemic with so few initially infected. As we see in the dynamics of the model in (11) the drift is heavily dependent on the SI term. With very few infected, the SI term will be smaller and the probability of a full-blown outbreak will be lower. If we instead start from day 1 we see that 6 out of the 10 simulations lead to a complete outbreak. Starting from day 2 all the simulations followed the true outbreak extremely well and when starting from day 3, the drift dominates so much of the dynamics that it almost seems deterministic. We also see for all of the plots that the shape of the curves are always identical they just "take off" at different times. This "take off" is for all plots rather symmetrical around the true development which of course is a nice property.

There is one thing that is especially interesting to observe in plots 7, 8, 9 and 10. As the number of infected increases, the probability of complete outbreak increase drastically. As we see in figure 7, very few of the simulations actually results in a complete breakout. By day 2 though, a complete outbreak is virtually bound to happen. In week 0, 10 people were infected, in week 2, there were 3645 people infected. If we think that the SIR model does a reasonable job at describing the real dynamics of an epidemic outbreak, we can see how important it is with early detection and isolation of the virus.

The Pedersen algorithm tend to give rather high variance, though looking

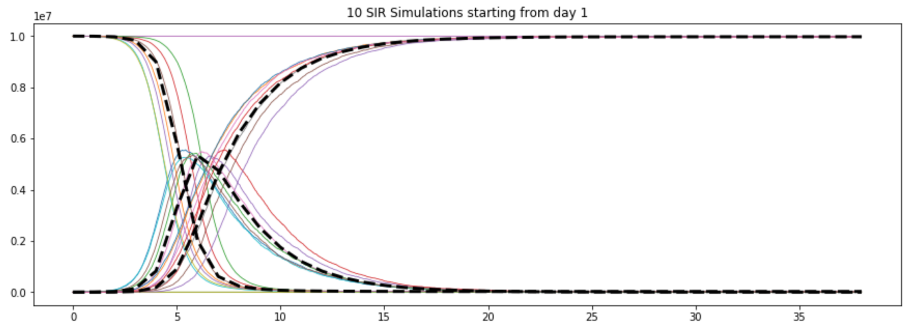

Figure 8: 10 simulations of epidemic starting from Day 1 using the fitted parameter estimates

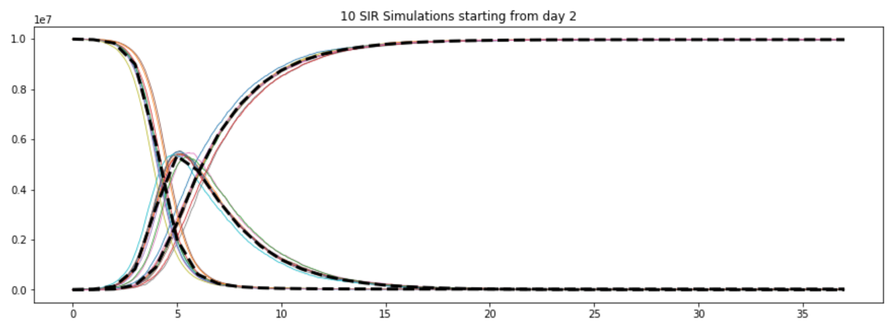

Figure 9: 10 simulations of epidemic starting from Day 2 using the fitted parameter estimates

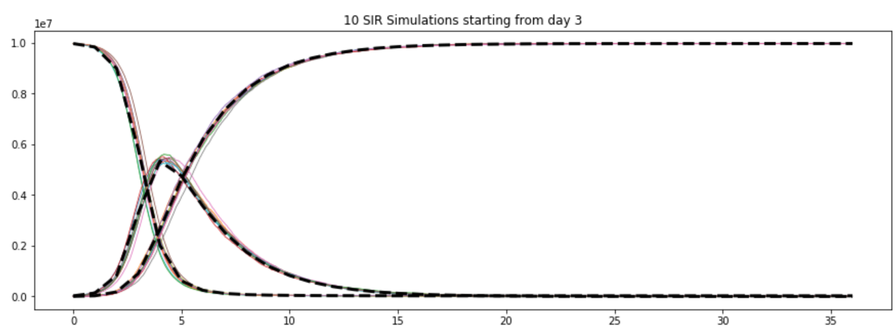

Figure 10: 10 simulations of epidemic starting from Day 3 using the fitted parameter estimates

at the simulations, I don't think that this has been to much of problem in this assignment. As the number of infected increases, we see already from week 2 that the simulations behave as if they were almost deterministic implying that the linear term clearly dominates the diffusion. We also used a EM discretzation which is a vanilla discretization scheme. We could have used a scheme with stronger order of convergence but since our model seems to fit the data really good, I see no need for a more advanced discretization scheme.

In conclusion we used both the most basic discretization scheme and the most basic bridge process but we were successful in doing so, most likely due to us having access to very nice and clean simulated data. When we move over to the main assignment, this will not be the case.

2.2 Main assignment

In this assignment we were to model Covid-19 in Sweden by making extensions to the basic SIR-model. We wanted to give β a periodic parametrization like the one present in 19:

$$\beta(t) = \beta_0 + \beta_1 \sin(2\pi wt) + \beta_2 \cos(2\pi wt) \tag{19}$$

where we have to find some suitable value for w.

We also wanted to proxy and probabilistic model for the state variables as the states are not observable. Unfortunately I only had time to do a simple model where I assumed the states were observable but where β were parameterized according to 19.

In figure 11 the Covid-19 cases reported in Sweden are presented in three ways, cases/day, cases 7 days rolling average, and cases per week collected every Sunday. As is clear from the first subplot, cases/day is extremely noisy and very unsuitable to use as data for continuous time modelling. Subplot 2 shows the 7 day rolling average of reported cases and the plot is a lot more smooth. Subplot 3 shows the reported cases per week and has an even smother plot. Which one of these makes sense to use? I argue that either the 7 day average or reported cases per week should be used as they make the plot more smooth and they counter noise and week-end effects. I chose to use the weekly observations here as the plot looked easier to fit using continuous time series modelling. If I would have more time, it would be interesting to analyse the different parameter estimates we get by using either the 7 day average or the weekly data.

Moving on, now that we are using weekly data, a suitable choice for w in (19) is 52. The reasoning behind this is that virus infections like the flu generally exhibits a yearly cycle with a lot of cases during the winter and very few cases during the summer. I choose to ignore all of the observations starting earlier than the 5th of March due to there being very few reported cases before then. I assume that the Covid-19 outbreak will be most severe in Jan-Feb 2021 and least severe in July-Aug 2021. Thus, I want the largest value for β during the late winter and a lowest value for β during the late summer. Making further the assumption that β_1 and β_2 are not of too different magnitude we can shift

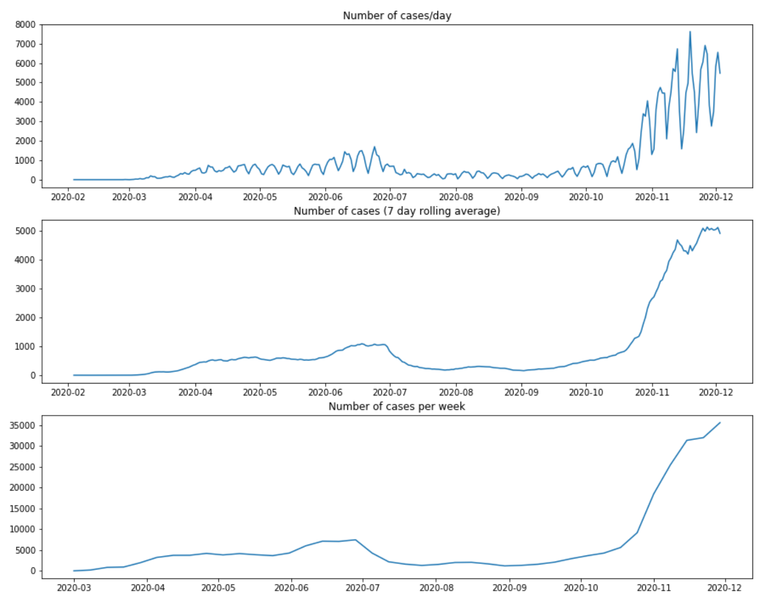

Figure 11: Covid cases reported in Sweden.

β_0	β_1	β_2	γ	σ_1	σ_2
0.3600	-0.1073	-0.2045	6.8692	5.1430	0.3034

Table 8: Parameter estimates with confidence interval

the sinusoidals in 19 by a factor so that β exhibits the seasonal effect we want. I found that shifting the sinusoidals $35\pi/52$ in the left direction and using a negative β_1 and β_2 to be successful in this endeavor, see 20

$$\beta(t) = \beta_0 + \beta_1 \sin(2\pi wt - \pi\frac{35}{52}) + \beta_2 \cos(2\pi wt - \pi\frac{35}{52}) \tag{20}$$

As I mentioned in the paper discussion, I wanted to do the parameter estimation in this assignment by using diffusion bridges based on residual process. Due to time constraints however I only had time to implement a solution which is based on simulated likelihood using modified diffusion bridge sampler. This sampler is usually not very good when the process we model exhibit strong non-linearity, which is the case for us as can clearly be seen in figure 11. I would make the argument however that even though the process is not globally linear, it is for most periods, locally linear. I thereby use a model which assumes that the states are observable (which is really not the case, but due to time-constraints I were not able implement a better solution) and where the simulated likelihood uses the MDB sampler. If the process we model is given in the form presented in (13) The transition densities under this approach is given by (21).

$$q(x_{\tau_{k+1}}|x_{\tau_k}, x, \Theta) = N(x_{\tau_{k+1}} : x_{\tau_k} + \frac{x_T - x_{\tau_k}}{T - \tau_k}\Delta\tau, \frac{T - \tau_{k+1}}{T - \tau_k}\beta(x_{\tau_k})\Delta\tau) \tag{21}$$

I used the discretization presented in (22), which is slighlty different than the one I used in assignment 2.1.

$$\begin{aligned}\begin{pmatrix} S_{t+1} \\ I_{t+1} \end{pmatrix} = \begin{pmatrix} S_t \\ I_t \end{pmatrix} + \begin{pmatrix} -\beta S_t I_t/N \\ \beta S_t I_t/N + \gamma I_t \end{pmatrix} \Delta t + \\ \begin{pmatrix} -\sqrt{\beta S_t I_t/N} & 0 \\ \sqrt{\beta S_t I_t/N} & -\sqrt{\gamma I_t} \end{pmatrix} d\boldsymbol{\Delta W}\end{aligned} \tag{22}$$

Where $\Sigma = \sqrt{\beta}\sqrt{\beta}'$ is given by (23)

$$\Sigma = \begin{pmatrix} SI/N & -SI/N \\ -SI/N & SI/N + \gamma I \end{pmatrix} \Delta t \tag{23}$$

This gives the Likelihood presented in (16), and the log-likelihood given by (18) (But for the log likelihood, instead of using $\tilde{\Sigma}$, I instead use Σ). The log-likelihood acts as the loss function for the numerical approximation. I used the BFGS method with SciPy's optimizer with $m = 70$ which gave the parameter estimates presented in table 8

10 simulations using these parameters starting from week 0, 1 and 2 are presented in figure 12, 13 and 14.

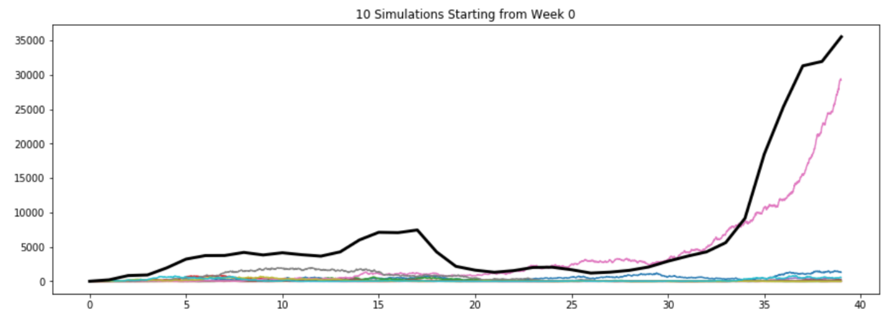

Figure 12: 10 COVID-19 (infections) simulations starting from week 0. Bold line is the reported number of weekly cases

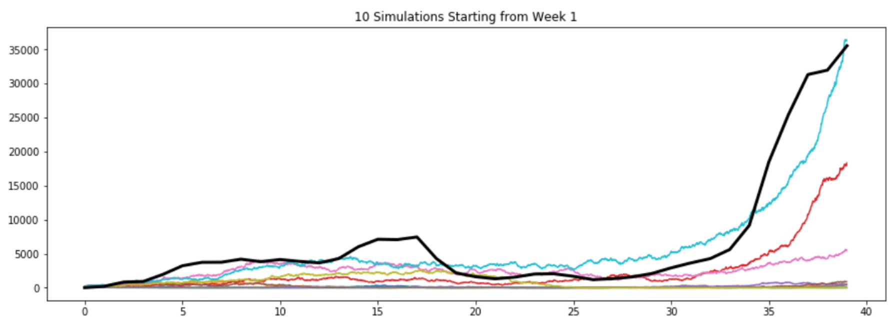

Figure 13: 10 COVID-19 (infections) simulations starting from week 1. Bold line is the reported number of weekly cases

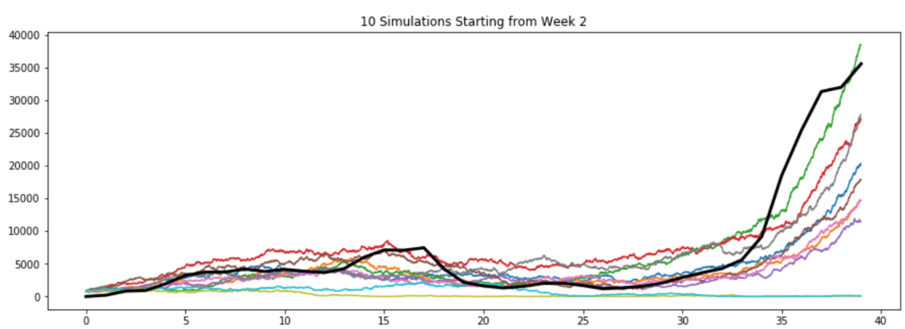

Figure 14: 10 COVID-19 (infections) simulations starting from week 2. Bold line is the reported number of weekly cases

As we see in the simulations the model does a poor job at simulating the epidemic. From week 2 the plots show that the model is able to take into account some level of seasonality however the changes are somewhat slow.

If I had more time on my hands I would try to use a bridge sampler that is based on residual process instead, preferably the one that subtracts both the drift and the linear noise approximation. I think it would to a lot better job to account for the non-linearities. Moreover, it is an extremely unreasonable assumption that we have observable states as there are many asymptotic carriers and limited testing. To make the problem even more complex the level of testing have been different during the period where the capacity to test is a lot higher now than it was during this spring. The recorded number of dead people is probably less contaminated by noise than the number of active cases. However it is not without problem to use this state as observable. We first have to determine the death rate of the disease and then we also have the problem that many people that die from the disease do so 10 days after the incubation period. In spite of this, using deceased as an observable state with a predetermined death rate and some measurement noise is probably the most sound strategy in my opinion. As many people die relatively late after the incubation period, a 10 day rolling average seems more appropriate than weekly observations. It would also be interesting the experiment with higher order discretization schemes.