Assignment 2: Stochastic Volatility Model
Time Series Models

Valentina Guldberg
Benjamin Weggelaar
Daniel Cedrés Garrido
Sofía Belén Romero

March 2023

A Present graphs and descriptives of log daily returns

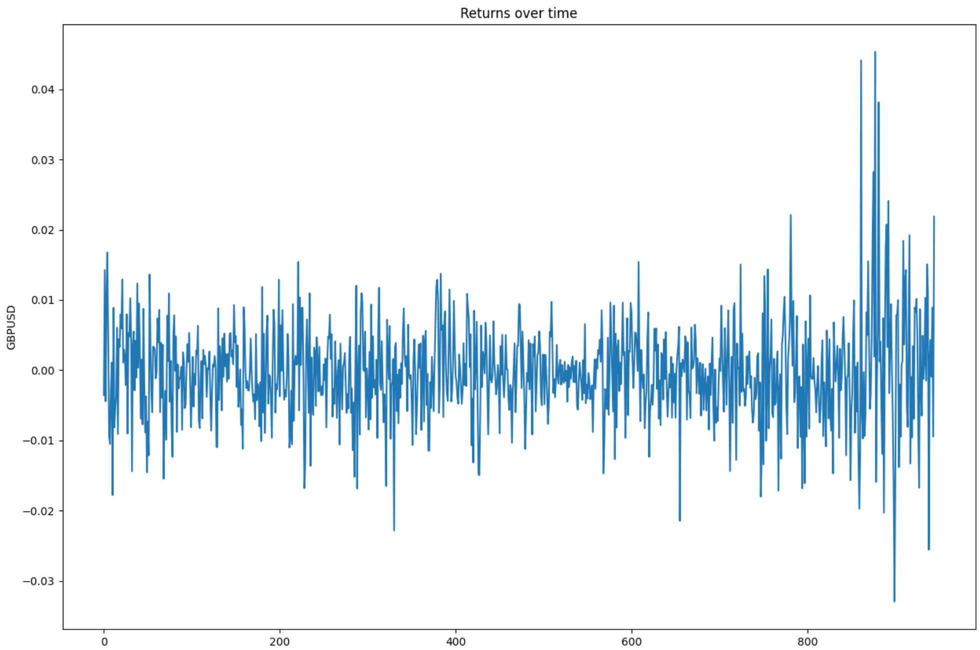

Figure 1: Daily returns over time: Log of returns (pound-dollar exchange rate) in the Y-axis. Time indexes in the X-axis.

Figure 1 showcases the log of returns, with the returns being the expected percentage change in the value of the investor's portfolio. Such return is explained by the exchange rate between the pound and dollar over a period of time. We observe the series being non-stationary: although there seems to be a constant mean, there are periods of high and low volatility. This phenomenon is called volatility clustering.

	GBPUSD
count	945.000000
mean	-0.035310
std	0.711089
min	-3.296118
25%	-0.439391
50%	-0.045708
75%	0.364447
max	4.534522

Table 1: Summary statistics for log daily returns of the pound-dollar exchange rate.

The returns series has a mean close to 0 and standard deviation equal to 0.71. This means that, on average, the returns lay 0.7 units away from the mean of the dataset. Both the minimum and the maximum values are among the final observations, meaning the data becomes more volatile at the end of the period.

B Perform the data transformation and compute xt and present a graph

By using the mean μ of the returns data calculated in the previous step, we are able to compute x_t for the time series returns data y_t:

$$x_t = log(y_t - \mu)^2$$

With this formula we can then compute x_t for each value of y_t. A plot of x_t is shown below:

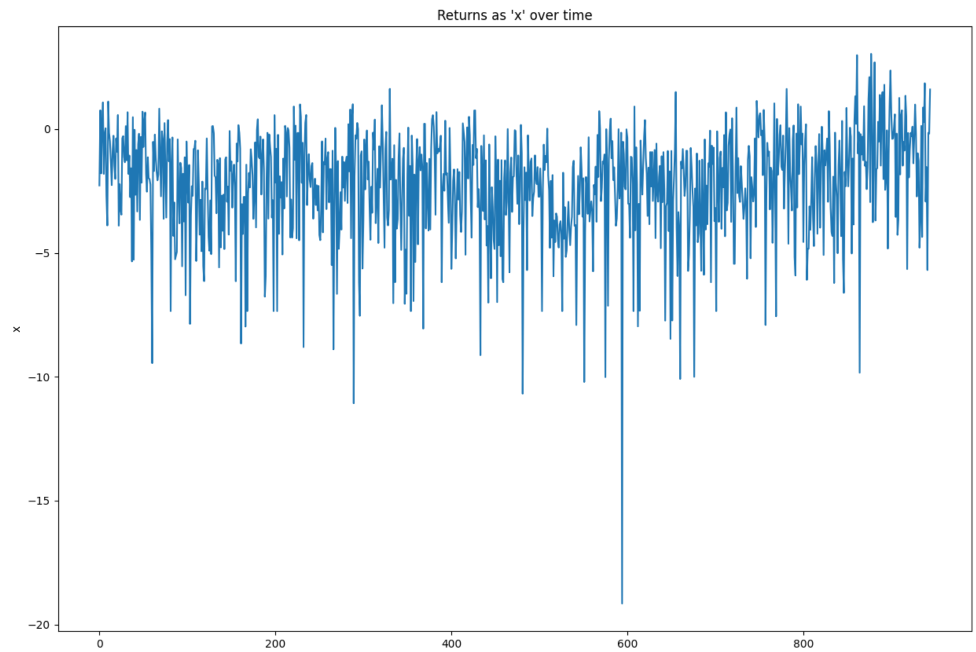

Figure 2: Transformed returns over time: Squared log of demeaned returns (pound-dollar exchange rate) in the Y-axis. Time indexes in the X-axis.

C Use the QML approach based on the linearized model with observations xt.The Kalman filter can be used for (approximate) maximum likelihood estimation. Present the estimates in a table.

The SV model we will model is:

$$y_t = \mu_t + exp\left(\frac{\xi + \tilde{h}_t}{2}\right)\varepsilon_t \qquad \varepsilon_t \sim N(0,1)$$
$$\tilde{h}_{t+1} = \phi\tilde{h}_t + \eta_t \qquad \eta_t \sim N(0, \sigma_\eta^2)$$

This is a non-linear Gaussian state space model. Using the transformed data x_t, we can rewrite the model, resulting in the following linearized specification:

$$x_t = h_t + u_t \qquad u_t \sim p(u_t)$$
$$h_{t+1} = \omega + \phi h_t + \eta_t \qquad \eta_t \sim N(0, \sigma_\eta^2)$$

where $\omega = 1/(1-\phi)\xi$ and $h_t = \tilde{h}_t + \xi$ and $u_t = log\varepsilon^2$. This model is linear non-Gaussian. The disturbance u_t follows a $log\chi^2$ distribution. For simplicity, we assume for now that u_t follows a normal distribution, with mean and variance from the $log\chi^2$ distribution.

To estimate the parameters ω, ϕ, and σ_η using the QML approach based on the linearized model, we first specified the likelihood function, assuming that the error term u_t is Gaussian with mean -1.27 and variance $\pi^2/2$.

We then applied the Kalman Filter to compute the predicted values of h_t, and then, by maximizing the log-likelihood function we obtained the estimates of the parameters.

To maximize the log-likelihood function we used the L-BFGS- algorithm as our optimization technique. The estimated parameters are shown in the table below:

Estimated σ_η^2	Estimated ω	Estimated ϕ
0.007017	-0.006987	0.991211

Table 2: QML-Kalman filter estimation results.

D Compute the smoothed values of ht based on the approximate model for xt by using the Kalman filter and smoother, and present them in a graph along with the transformed data xt. In addition, present both the filtered and smoothed estimates of $\tilde{h}_t$ in a graph.

From the optimization above we got the filtered and smoothed states ht. To obtain both the filtered and smoothed estimates of Ht, we used the QML estimated parameters and calculated ξ using the following equation:

$$\xi = \omega/(1-\phi)$$

We then computed $\tilde{h}_t$ by subtracting ξ from h_t. We do this for the filtered and the smoothed estimate.Below we present the graphs of both filtered and smoothed $\tilde{h}_t$:

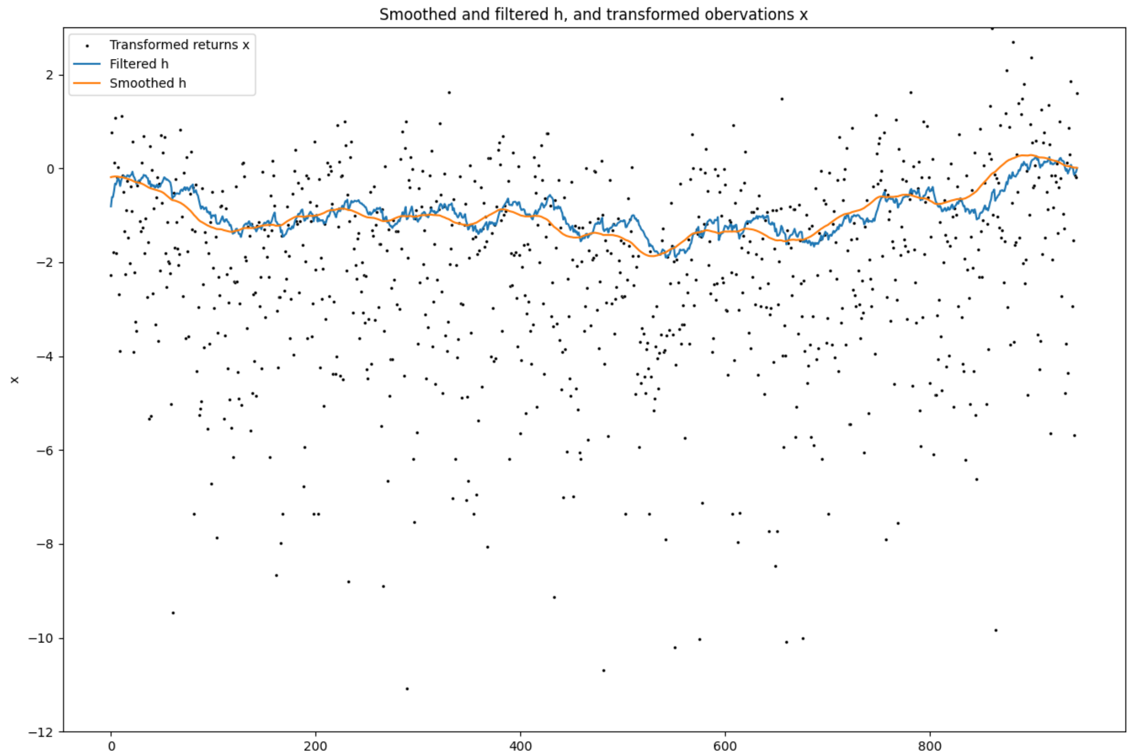

Figure 3: Filtered and smoothed estimates of ht, with observations of the transformed data. Some negative outliers were omitted in the graph for better representation

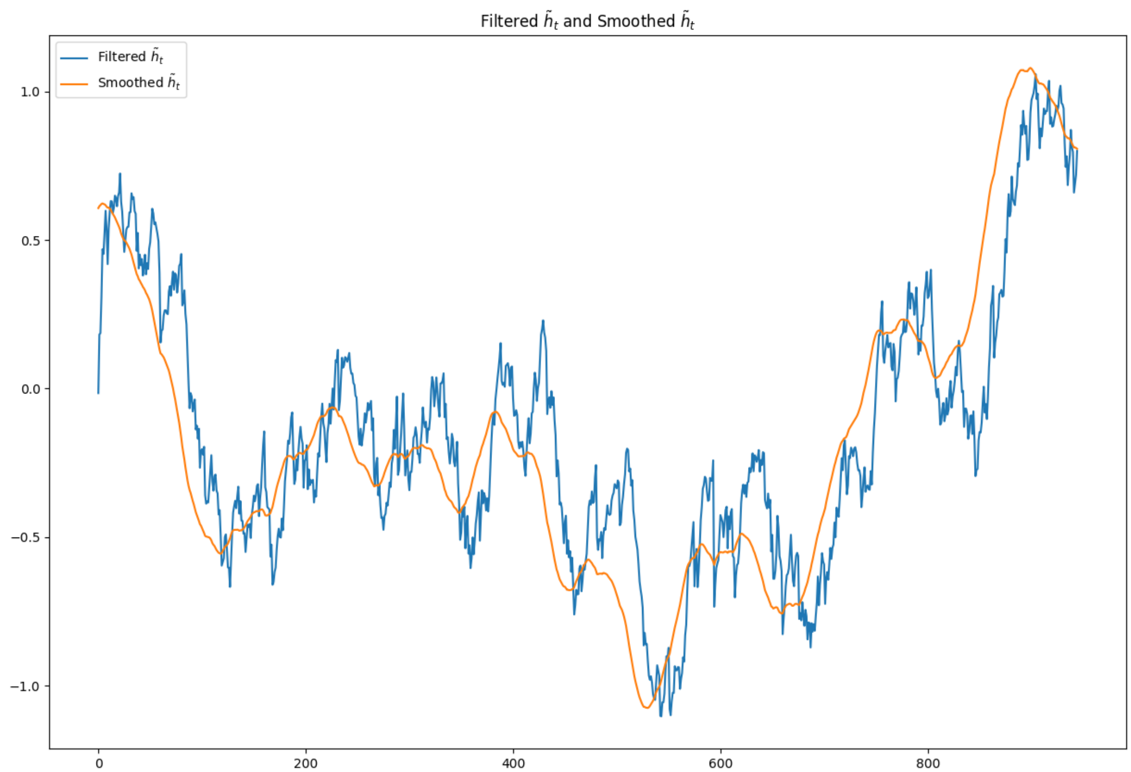

Figure 4: Comparison of filtered and smoothed h tilde.

E For a period of at least five years, consider the daily returns for SP500 index (or another stock index) that you can obtain from Yahoo Finance, and re-visit the analysis from questions (a) - (d). Next, to improve the performance of the SV model, you can extend your analysis with a Realized Volatility measure of your choice that can be be found on Canvas. For this purpose, consider the extended model

We decided to work with a subset of five years of the SP500 data, going from 01/01/2016 to 31/12/2020. To re-visit the analysis of the previous questions we first computed the log return in the following way:

$$y_t = log(P_t/P_{t-1})$$

where P_t denotes the closing price of the trading day.
Below we present the results of questions a) to d) applied to the SP500 data.

E.1 SP500: question a)

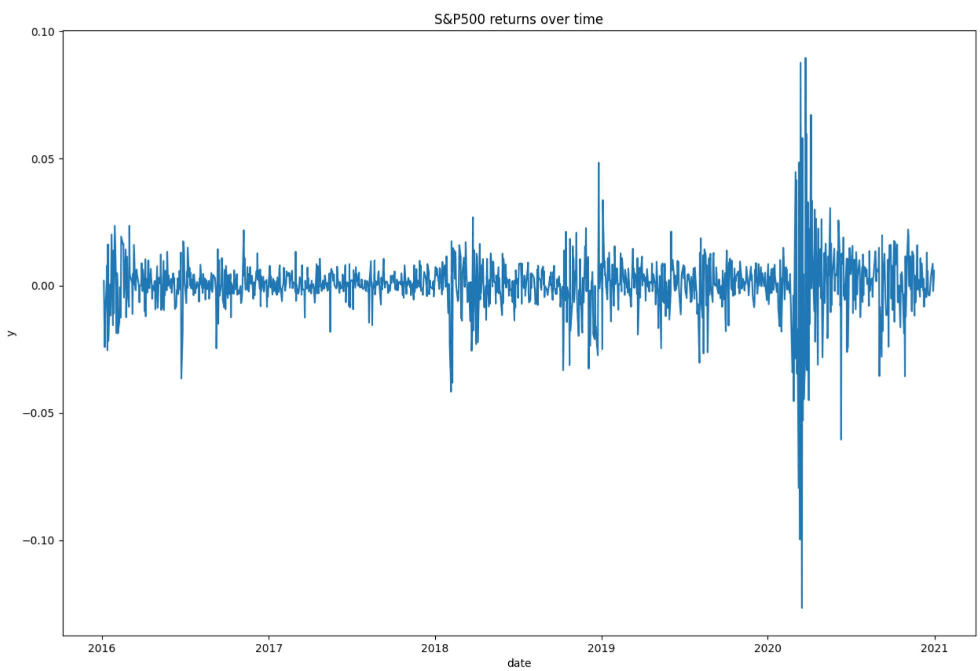

Figure 5: SP500 daily returns over time,in the Y-axis. Time indexes in the X-axis.)

The SP500 returns series is non-stationary, as it showcases a constant mean but a volatile variance over time. Moreover, there is an abrupt peak of returns between 2020 and 2021. This might suggests the presence of a structural break that is changing the underlying distribution of the series in that period.

	y
count	1249.000000
mean	0.000499
std	0.012177
min	-0.126703
25%	-0.002895
50%	0.000707
75%	0.005373
max	0.089612

Table 3: Summary statistics for SP500 returns.

	Estimated σ_η^2	Estimated ω	Estimated ϕ
Optimization results	0.085785	-0.251426	0.974868

Table 4: Estimated parameters for QML-Kalman Filter, using the SP500 data.

E.2 SP500: question b)

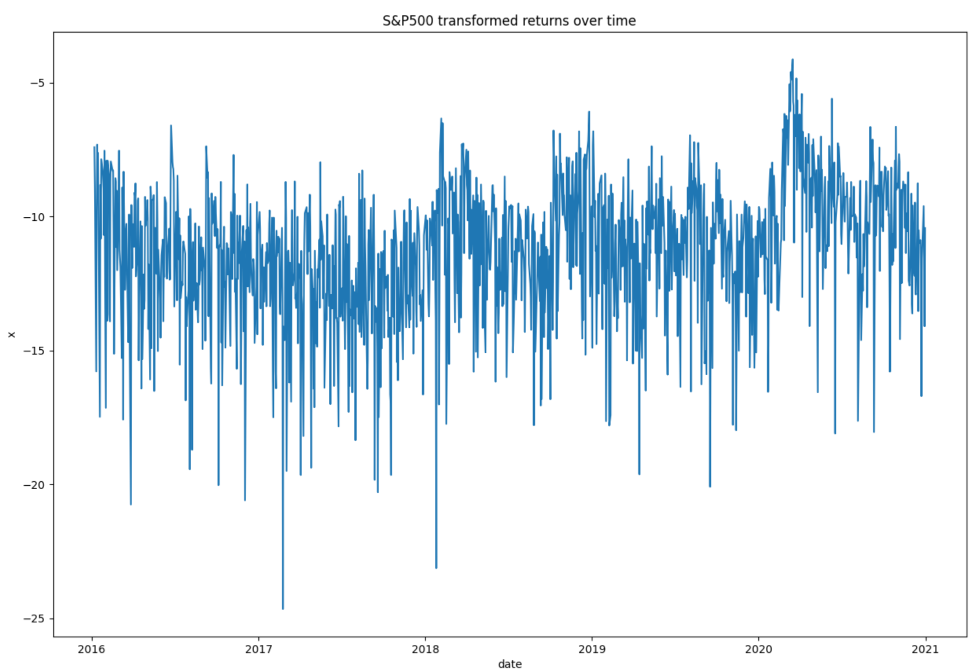

Figure 6: SP500 transformed returns over time: Squared log of demeaned returns (pound-dollar exchange rate) in the Y-axis. Time indexes in the X-axis.

E.3 SP500: question c)

The estimated parameters are shown in the table 4.

We see that the parameter estimates change, compared to the original dataset. The effect of ϕ becomes lower, while the effect of both σ_η^2 and ω become larger (in absolute value). These changes were expected, because:

1. Since ϕ is smaller, the value of the state in the previous time period is less important, and thus, the state is less persistent.
2. A higher σ_η^2 is more accurate for stock indices, because volatility is higher in stock markets than exchange rates between two of the most important currencies, which is expected to be relatively stable.
3. A higher ω contributes to a higher mean in the state, which is consistent because we expect a higher return in the stock market compared to the currency exchange market.

E.4 SP500: question d)

Same as for the original model, from the optimization above we got the filtered and smoothed states ht. To obtain both the filtered and smoothed estimates of Ht, we used the QML estimated parameters and calculated ξ.

Below we present the graphs of both filtered and smoothed $\tilde{h}_t$:

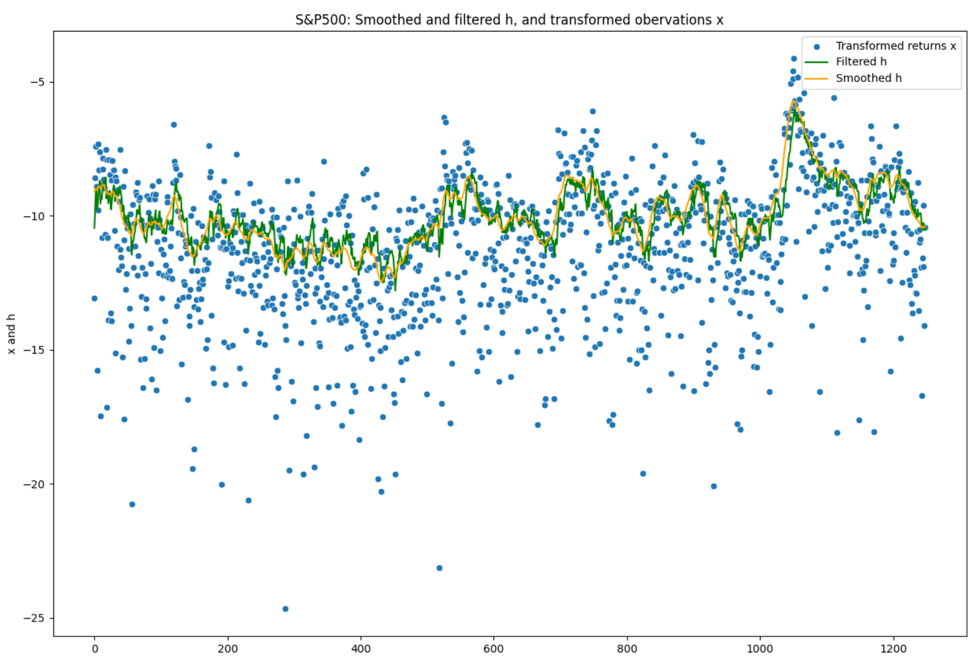

Figure 7: Filtered and smoothed estimates of ht, with observations of the SP500 transformed returns.)

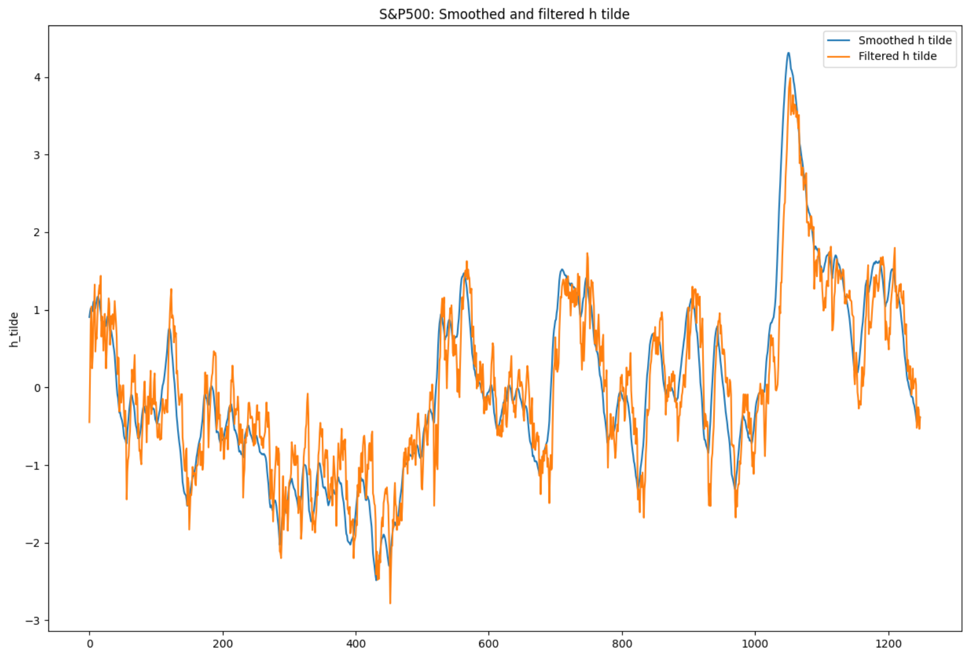

Figure 8: Comparison of filtered and smoothed h tilde.

E.5 Extended model analysis

The extended model is:

$$x_t = \beta logRV_t + h_t + u_t \qquad u_t \sim p(u_t)$$
$$h_{t+1} = \omega + \phi h_t + \eta_t \qquad \eta_t \sim N(0, \sigma_\eta^2)$$

For the purpose of this analysis we took the column named 'rv5' as our Realized Volatility, as in forecasting applications, it is found that a low frequency "truncated" RV outperforms most other realized measures (Liu, Patton Sheppard, 2012). Overall, we conclude that it is difficult to significantly beat 5-minute RV. In the extended scenario we aimed at estimating the three parameters from the original model, ω, ϕ, and σ_η, as well as β. To estimate β we decided to follow the Augmented Kalman Filter approach, which consists of three steps:

1. Running the Kalman Filter for the linearized model as before and storing v* and F.
2. Running the Kalman Filter using log(RV)-1.27 as observations and storing the prediction errors X*.
3. Computing β using the following equation:

$$\hat{\beta} = \left(\sum_{t=1}^{n} X_t^{*\prime} F_t^{-1} X_t^* \right)^{-1} \sum_{t=1}^{n} X_t^{*\prime} F_t^{-1} v_t^*$$

The obtained estimation is: $\hat{\beta} = 0.7615$.

Lastly, we re-ran Kalman Filter using y-$\hat{\beta}$*log(RV) as observations, as for this model with adjusted observations, the sum of squared prediction errors can be minimized analytically with respect to β.

E.6 Contrasting original and extended models

Below we compare the original model to the extended model by plotting the prediction errors and calculating MSE and RMSE error metrics:

Using our estimate for β, the logarithm of the Realized Volatility (rv5) and the estimated signal, we obtained the estimated transformed returns 'x' for each model.

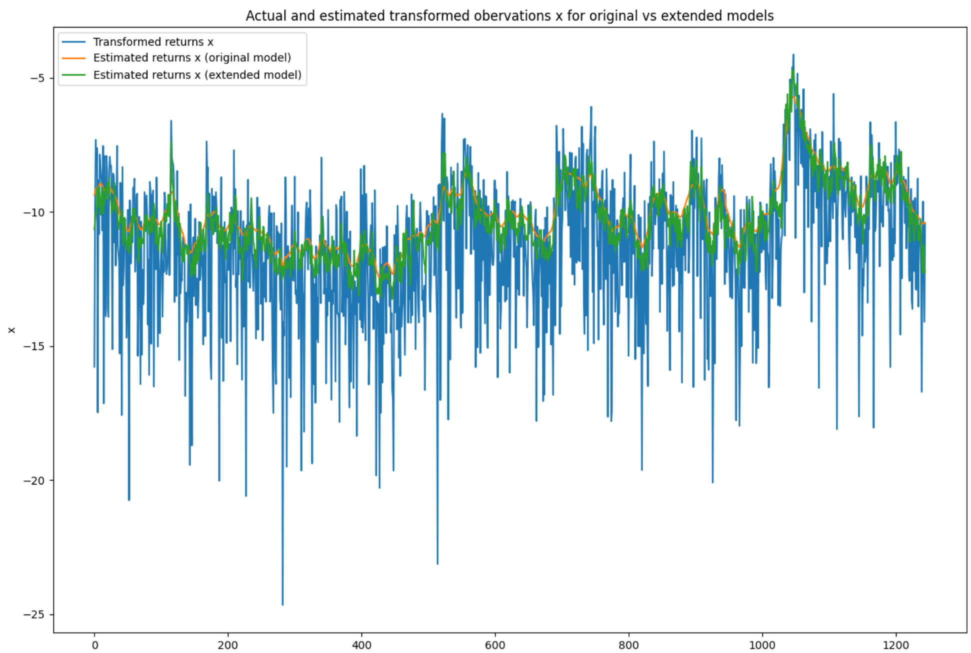

Figure 9: Actual transformed returns (blue). Predicted transformed returns for the original model (orange). Predicted transformed returns for the extended model (green). Transformed returns in the Y-Axis and time index in the X-Axis.

From the plot above we can conclude that the estimation of returns in the extended model is more accurate than in the original one, as the model has an extra term β*log(RV) that is able to model the persistence of the Realized Volatility effect over time. Adding this term to the original model results in a better adjustment to the original series and therefore, a lower predictive error when comparing the actual and predicted series, which is shown in the figure below.

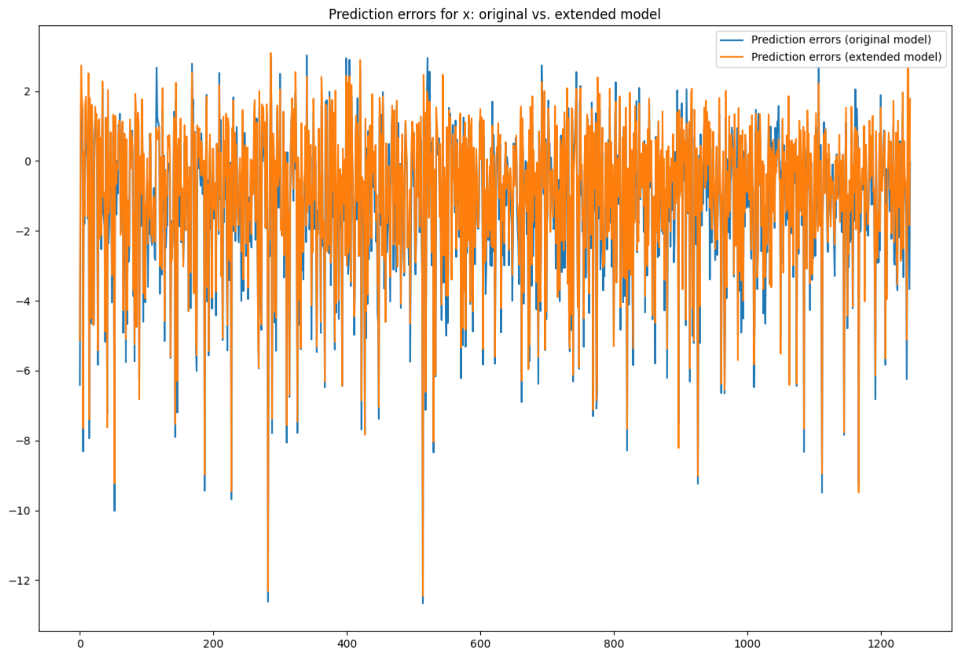

Figure 10: Prediction errors calculated as the difference between the actual and predicted transformed returns 'x', for each model. Errors in the Y-Axis and time index in the X-Axis.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

Figure 11: MSE and RMSE formulas used to compute table below.

	MSE original model	MSE extended model	RMSE original model	RMSE extended model
Errors	6.509721	5.590129	2.551416	2.364345

Table 5: MSE and RMSE for both original and extended models

From the error plot and the MSE and RMSE metrics shown above, we conclude that the extended model performs better as the prediction errors are lower in such case.

F Returning to the first model specification (non-linear Gaussian), compute the filtered estimates of $\tilde{h}_t$, and compare to the earlier filtered QML estimates in a graph. Repeat as well for the stock index

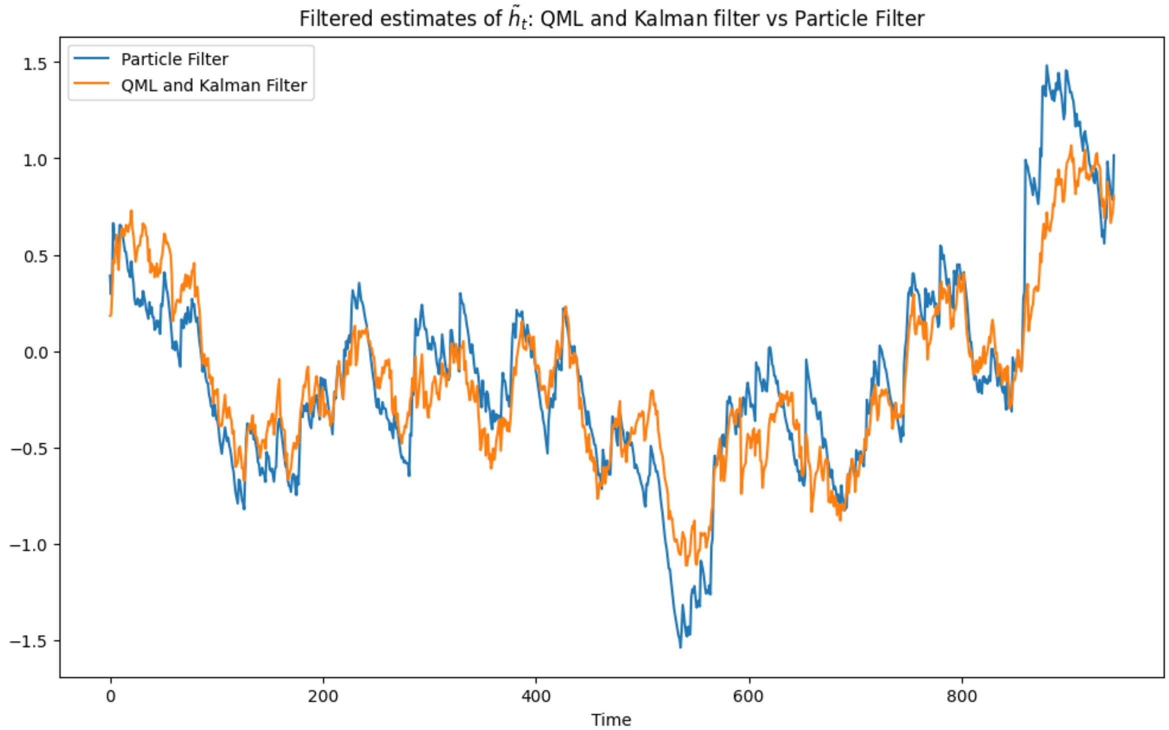

Figure 12: Filtered estimates: QML and Kalman filter, using the linearized model, and the Particle filter, using the non-linear Gaussian model.

Figure 12 shows the results estimates of the filtered state $\tilde{h}_t$, using the two main model specifications. The results of the particle filter (in blue) oscillated around the results of QML- Kalman filter. This was expected, since the QML-Kalman filter methods is an unbiased estimator of the state.

Figure 13 shows the difference between estimates of the state using the QML-Kalman filter and Particle filter, for SP500 data. We observe the very high volatility during the beginning of the Covid-19 pandemic, in 2020. Similar o the previous figure, we see that the estimates of the two methods are quite similar.

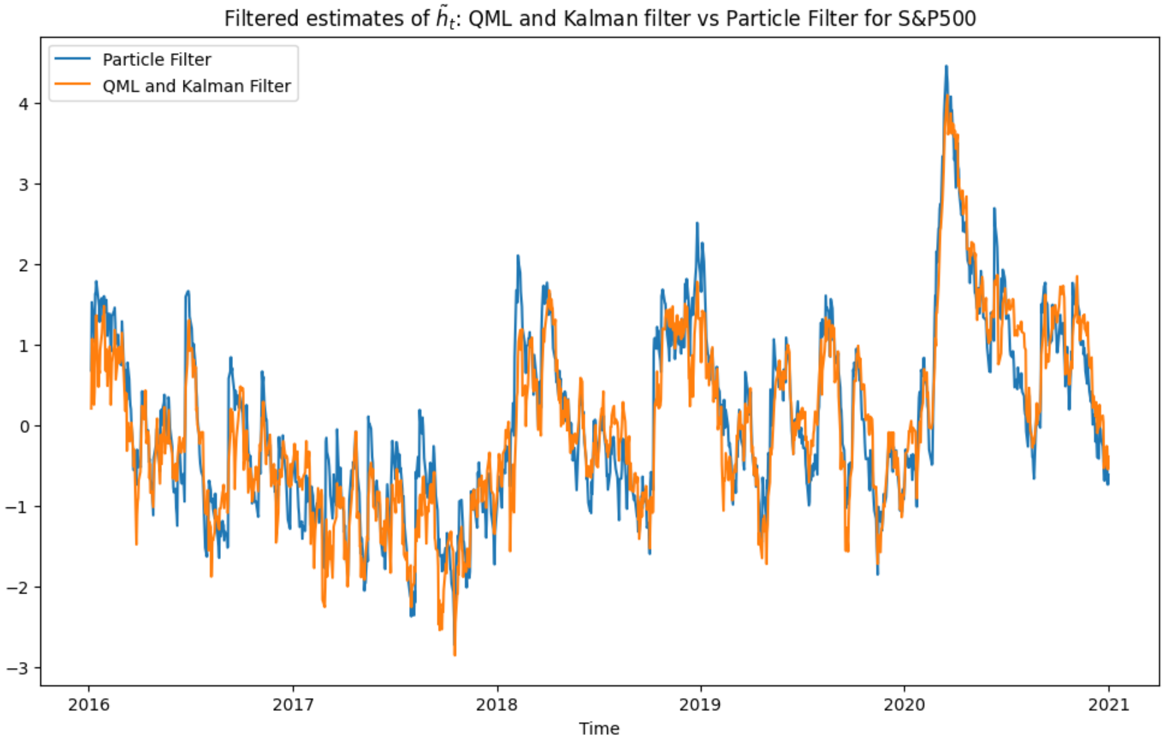

Figure 13: Filtered estimates: QML and Kalman filter, using the linearized model, and the Particle filter, using the non-linear Gaussian model.

G References

Liu, Sheppard, amp; Patton. (2012). Does anything beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes - Duke University and University of Oxford. Retrieved March 20, 2023, from https://public.econ.duke.edu/ ap172/Liu$_Patton_Sheppard_dec12.pdf$