For this assignment, you may work in groups of two or three students. Part 1 is theoretical and part 2 is an application of your results from part 1. For each sub-question, you can earn three points. It is crucial to motivate your answers: No points will be awarded for answers without motivation. Furthermore, reports that are not typeset as a single PDF file in LaTeX will not be considered. Please also make sure to submit your R (and/or Python) codes for question 2, separately, together with a CSV file of the data. For question 2, your code should reproduce all reported results without returning errors. Results that are not returned by your code cannot be graded.

The deadline for this assignment is Monday, February 19, 15:00.

Part 1. (a) Suppose the Mean Squared Error (MSE) for a one-step prediction is given by:

$$\text{MSE} = \mathbb{E}[(y_{t+1} - \hat{y}_{t+1})^2]$$

where y_{t+1} and $\hat{y}_{t+1}$ are the actual and predicted value of y one step ahead, defined as:

$$\begin{aligned} y_{t+1} &= f(y_t) + \varepsilon_{t+1} \\ \hat{y}_{t+1} &= \hat{f}(y_t) \end{aligned}$$

Here, $f(\cdot)$ denotes the true model specification for the process y_t (apart from the disturbance) and $\hat{f}(\cdot)$ denotes the estimated model specification.
Show that the one-step MSE can be written as:

$$\text{MSE} = \mathbb{V}[f(y_t) - \hat{f}(y_t)] + (\mathbb{E}[f(y_t) - \hat{f}(y_t)])^2 + \sigma^2$$

You may assume that $\mathbb{E}[\varepsilon_{t+1}] = 0$, $\mathbb{V}[\varepsilon_{t+1}] = \sigma^2$ and that ε_{t+1} is independent of $f(y_t)$ and $\hat{f}(y_t)$.

(b) The result from (a) shows that the MSE can be decomposed into three terms: the variance of the estimate, the squared bias of the estimate and the variance of the disturbance term. Explain how the three different terms are influenced by the model we choose and how this creates a trade-off.

The remainder of this part considers (cases) of the GARCH(m, s) model

$$
\begin{aligned}
y_t &= \mu + \varepsilon_t \\
\varepsilon_t &= \sigma_t \nu_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

where $\nu_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.

(c) Show that the ARCH(m) model for the conditional variance can be rewritten as an AR(m) model for the squared residuals $\{\varepsilon_t^2\}$.

(d) Assume that $\mu = 0$. Using the error decomposition of Q1c in Assignment 1, derive the likelihood of the GARCH(m, s) model. Additionally, derive the first-order conditions.

(e) Which restrictions on $\alpha_0, \alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_s$ are needed to ensure positivity for the conditional variance? And under which restrictions does the conditional variance satisfy stationarity?

Part 2. (a) Download daily data from the website of Yahoo Finance for the 10-year Treasury Yield (ticker: ˆ TNX) from January 2, 2002, up to and including December 29, 2023. Save as a CSV file and enable your script to read the data. Use the adjusted closing prices as our time series. Note that the data may contain days without a value (non-trading days). Make sure you remove these empty entries. Calculate the daily log-returns.

(b) Use the full sample to estimate five AR(p) models, namely for $p = 1, 2, 3, 4, 5$ and report the log likelihood, AIC and BIC based on the full sample for each of the five AR specifications. Explain the difference between the in-sample log-likelihoods and AICs. Also report the optimal lag order based on the AIC.

To compare different models, we often would like to test the out-of-sample performance, where the forecasts are compared to the actual observations. In time series it is customary to use rolling window forecasts. With this method, a subsample is used to calculate the one-step-ahead forecast. Then the subsample moves one observation forward, but remains of the same size, so the first observation in the previous subsample is dropped. Again the one-step-ahead forecast is calculated. This is repeated until the end of the full sample is reached. Note that your first forecast corresponds to the first observation after the subsample size m. Rolling the window through the data produces a sequence of $N - m$ forecasts $\hat{y}_i$ which can be compared with their actual values f_i by the RMSE, defined as

$$
\text{RMSE} = \left(\frac{1}{N - m} \sum_{i=m+1}^{N} (y_i - \hat{y}_i)^2 \right)^{1/2}
$$

(c) Extend your code for OLS estimation from assignment 1 and use the rolling window method to obtain forecasts for the log-returns, for AR(p) models for $p = 1, 2, 3, 4, 5$. Use a rolling window for forecasting of size $m = 750$. Use the first 750 observations to estimate the models and use the resulting model to calculate the rolling window forecasts. Calculate the RMSE's and save the results for the different AR(p) specifications.

(d) Repeat (c), but now re-estimate the models at each step. Hence, also use a rolling estimation window. At each step, use the previous 750 observations to estimate a new model and calculate the one-step-ahead forecasts, instead of estimating the model only once (based on the first 750 observations) and using this model to obtain all rolling window forecasts.

Another variation on the rolling window method is the expanding window method, in which the window length expands at each step. So as opposed to in the rolling window method, no observations are dropped and the estimation subsample becomes larger each step. For example, the first step is to estimate the model based on the first 750 observations and then use that model to make the first forecast. The next step would be to re-estimate the model based on the first 751 observations and then make the one-step-forecast based on the new estimated model. This continues until the end of the sample is reached.

(e) Use the expanding window method to obtain forecasts for the log-returns, for AR(p) models for $p = 1, 2, 3, 4, 5$. Calculate the RMSE's and save the results for the different AR(p) specifications.

(f) Show the RMSE's obtained in (c), (d) and (e) for the different AR(p) models in one table. Discuss results, what is the optimal AR model? Are there differences between the methods? Does the conclusion correspond to the one in (b)?