Time Series Analysis: Assignment 2

Bowen Ma (12960780), Mianyun He (13605275), Shiyi Yang (13627295)

Part 1

a.

Given MSE for a one-step prediction $E\left[\left(y_{t+1}-\hat{y}_{t+1}\right)^2\right]$ and since $y_{t+1}=f\left(y_t\right)+\varepsilon_{t+1}$, and $\hat{y}_{t+1}=\hat{f}\left(y_t\right)$

$$\begin{aligned} E\left[\left(y_{t+1}-\hat{y}_{t+1}\right)^2\right] &= E\left[\left(f(y_t)+\varepsilon_{t+1}-\hat{f}\left(y_t\right)\right)^2\right] \\ &= E\left[\left(f(y_t)-\hat{f}\left(y_t\right)\right)^2+\varepsilon_{t+1}^2+2[(f(y_t)-\hat{f}(y_t)]\varepsilon_{t+1}\right] \end{aligned}$$

Given $E\left[\varepsilon_{t+1}\right]=0$, $\operatorname{Var}\left[\varepsilon_{t+1}\right]=\sigma^2$, and ε_{t+1} is independent of $f\left(y_t\right)$ and $f\left(\hat{y}_t\right)$,

$E[\varepsilon_{t+1}(f(y_t)-f(\hat{y_t}))]=E[\varepsilon_{t+1}]E[f(y_t)-f(\hat{y_t})]=0$

We can decompose the expression according to $E\left[X^2\right]=\operatorname{Var}[X]+(E[X])^2$ for a random variable X,

$$\begin{aligned} E\left[\left(y_{t+1}-\hat{y}_{t+1}\right)^2\right] &= \operatorname{Var}[f(y_t)-f(\hat{y_t})]+(E[f(yt)-\hat{f}(y_t)])^2+\operatorname{Var}\left[\varepsilon_{t+1}\right]+\left(E\left[\varepsilon_{t+1}\right]\right)^2+0 \\ &= \operatorname{Var}[(y_t)-\hat{f}(y_t)]+(E[f(y_t)-\hat{f}(y_t)])^2+\sigma^2 \end{aligned}$$

b.

The initial term, denoted as Variance, characterizes the volatility of the disparity between the actual model and the predictive model, contingent upon the training data set S and the test data (x, y). Models endowed with elevated capacity, such as neural networks featuring an extensive number of layers, exhibit heightened variance, whereas models with diminished capacity, exemplified by linear regression, manifest reduced variance.

The subsequent term encompasses the squared Bias, serving as an indicator of the fidelity of our predictor in approximating the authentic model. Models endowed with heightened capacity demonstrate diminished bias, while models characterized by diminished capacity manifest augmented bias. The third term, Noise, delineates the influence of observational noise, which is independent of any variables other than the inherent distribution of the noise. The reduction of this noise is inherently unattainable, rendering it irreducible.

A frequently encountered dilemma entails a trade-off between bias and variance. Elevating model complexity is associated with a decline in bias but a concurrent increase in variance. Conversely, simpler models characterized by reduced complexity tend to incur augmented bias but diminished variance.

c.

To show that the the conditional variance of the ARCH(m) model can be rewritten as an AR(m) model for the squared residuals, we can deduce it the other way around by AR(m) to ARCH(m). AR(m) model for the squared residuals can be written as $\varepsilon_t^2=\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2+\omega_t$, where $\omega_t \backsim WN$

Conditional variance of ε_t^2

$\operatorname{Var}\left(\varepsilon_t \mid Y_{t-1}\right)=E\left[\varepsilon_t^2 \mid Y_{t-1}\right]=E\left[\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2+w_t \mid Y_{t-1}\right]=\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2$

Assume $\varepsilon_t \backsim N\left(0,\sigma_{t^2}\right)$, thus $\sigma_t^2=\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2$

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

$$\Rightarrow \varepsilon_t^2=\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2=\alpha_0+\sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2+\underbrace{\varepsilon_t^2-\sigma_t^2}_{\omega_t}$$

$\omega_t = \varepsilon_t^2 - \sigma_t^2$ follows White Noise process,

- $E[\omega_t] = E[\sigma_t^2 + \eta_t^2] - E[\sigma_t^2] = E[E[\sigma_t^2 \cdot \eta_t^2]|\sigma_t^2] - E[\sigma_t^2] = E[\sigma_t^2 \underbrace{E[\eta_t^2|\sigma_t^2]}_{=1}] - E[\sigma_t^2] = 0$
- $Var(\omega_t) = E[\omega_t^2] = E[\varepsilon_t^4] - 2E[(\varepsilon_t\sigma_t)^2] + E[\sigma_t^4] = E[\sigma_t^4 \underbrace{E[v_t^4|\sigma_t^4]}_{=3}] - 2e[\sigma_t^4 \underbrace{E[v_t^2|\sigma_t^4]}_{=1}] + E[\sigma_t^4] = 2E[\sigma_t^4]$
- $Cov(\omega_t, \omega_{t-k}) = E[(\varepsilon_t^2 - \sigma_t^2)(\varepsilon_{t-k}^2 - \sigma_{t-k}^2)] - E[\varepsilon_t^2 - \sigma_t^2]E[\varepsilon_{t-k}^2 - \sigma_{t-k}^2]$

Hence, ω_t is a WN process, ε_t^2 is an AR(m) process.

d.

Since $f(y_t \mid Y_{t-1}) = f(y_t \mid y_{t-1}, \ldots, y_1) \cdot f(y_{t-1} \mid y_{t-2}, \ldots, y_1) \cdot f(y_1)$

Define $L = \prod_{i=1}^T f(y_t \mid Y_{t-1}) = \prod_{i=1}^T \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_t}{\sigma_t}\right)^2}$, where $y_t = \varepsilon_t = \sigma_t\nu_t$

hence $L = \prod_{i=1}^T \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\frac{1}{2}(\nu_t)^2}$

Log likelihood

$$\ell = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^T log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^T \nu_t^2$$

For $\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j\sigma_{t-j}^2$

The first-order conditions

$$\frac{\partial\ell}{\partial\alpha_0} = -\frac{1}{2}\sum_{t=1}^T \left(\frac{1}{\sigma_t^2} - \frac{\varepsilon_t^2}{(\sigma_t^2)^2}\right) = 0$$
$$\frac{\partial\ell}{\partial\alpha_1} = -\frac{1}{2}\sum_{t=1}^T \left(\frac{\varepsilon_{t-i}^2}{\sigma_t^2} - \frac{\varepsilon_t^2\varepsilon_{t-i}^2}{(\sigma_t^2)^2}\right) = 0$$
$$\frac{\partial\ell}{\partial\beta_j} = -\frac{1}{2}\sum_{t=1}^T \left(\frac{\sigma_{t-j}^2}{\sigma_t^2} - \frac{\varepsilon_t^2\sigma_{t-j}^2}{(\sigma_t^2)^2}\right) = 0$$

Please note that in the context of a GARCH(m, s) model, if we choose not to condition on $\mathcal{F}_0 = \{y_{-p+1}, \ldots, y_0\}$, it becomes crucial to condition on the initial $\max(m+1, s+1)$ observations when formulating the log-likelihood function. Therefore, the summation in the log-likelihood function should begin from $\max(m+1, s+1)$, without additional information.

e.

$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i\varepsilon_{t-j}^2 + \sum_{i=1}^s \beta_i\sigma_{t-j}^2$, positively we need : $\alpha_0 > 0, \begin{cases}\alpha_i \geq 0, i = 1, ..., m \\ \beta_j \geq 0, j = 1, ..., s\end{cases}$

We know from a) :

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j\sigma_{t-j}^2 + \left(\varepsilon_t^2 - \sigma_t^2\right)$$
$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j\left(\varepsilon_{t-j}^2 - \omega_{t-j}\right) + \omega_t$$
$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i\varepsilon_{t-j}^2 + \sum_{j=1}^s \beta_j\varepsilon_{t-j}^2 + \omega_t - \sum_{j=1}^s \beta_j\omega_{t-j}$$

ε_t is stationary if $\sum_{i=1}^m \alpha_i + \sum_{j=1}^s \beta_i < 1$, then $\sigma_t^2 = \varepsilon_t^2 - \omega_t$ is stationary since $\omega_t \sim WN$

Part 2

a.

The R script imports the data from the website of Yahoo Finance for the 10-year Treasury Yield (ticker: ˆTNX) from January 2, 2002 up to and including December 29, 2023, and extracts the adjusted closing prices to form a time series $\{y_t\}$. Any null values in the dataset are replaced with NA and then removed. The script further processes the data to extract prices and dates, ensuring they are converted to the appropriate data types. After this, the daily log-return at time t is computed as $\log(\frac{y_t}{y_{t-1}})$.

b.

The AR(p) models for $p = 1, 2, 3, 4, 5$ were estimated using the full sample. For each model, the log likelihood, AIC, and BIC were computed and reported in the following table:

lag(s)	Log-Likelihoods	AIC	BIC
1	12443.841	-24883.681	-24870.446
2	12455.222	-24904.444	-24884.591
3	12459.372	-24910.745	-24884.274
4	12465.561	-24921.122	-24888.033
5	12473.665	-24935.329	-24895.623

The in-sample log likelihood measures how well each model fits the data used for estimation, while the AIC considers both goodness-of-fit and model complexity. AIC penalizes complex models to prevent overfitting, making it a useful criterion for model selection. BIC imposes a stronger penalty on model complexity than AIC, particularly for larger sample sizes, favoring simpler models.

Observing the log-likelihood values, the model fit generally improves with higher lag orders. Furthermore, considering AIC, the preferred model is the AR(5) model, as it has the lowest AIC value, indicating a good balance between model fit and complexity.

c.

Based on the OLS estimation codes provided in assignment 1, we implement the rolling window method to compute one-step-ahead forecasts utilizing a fixed sub-sample size of 750 observations.

In these codes, lagged variables are initially generated for OLS and consolidated in the 'X' matrix. Following this, an empty vector is initialized to store the RMSE values. To enhance clarity, we execute five separate sets of codes for each AR(p) model, where $p = 1, 2, 3, 4, 5$. In the initial forecast, only past information is utilized. However, as forecasting progresses, both previous information and forecasted values might be necessary for subsequent forecasting, as discussed in the lecture.

Ultimately, the forecasts are compared with the actual data values using the RMSE, and the outcomes are presented in the first column of the table in part f.

d.

Now, the rolling window method has been adjusted. Instead of estimating coefficients and forecasting all observations using the first 750 observations, we now estimate the model based on observations 1-750 and forecast observation 751. Subsequently, we use observations 2-751 to forecast observation 752, and continue this process until all forecasted values are obtained.

During forecasting, for example, when predicting observation 752, we utilize past information up to observation 750, as well as the forecasted value for observation 751, adhering to the forecasting strategy outlined in the lecture. Following this logic, we forecast the remaining observations.

Finally, the forecasts are compared to the actual data values using the RMSE metric, and the results are displayed in the second column of the table in part f.

e.

Another approach to the rolling window method is the expanding window method, where the window length increases at each step. In contrast to the rolling window method, no observations are discarded, and the estimation subsample grows larger with each step.

For instance, in the first step, we estimate the model using the initial 750 observations and make the first forecast based on that model. Subsequently, we re-estimate the model using the first 751 observations and make another one-step forecast based on the updated model. We continue this process, incorporating both past information and previously forecasted values, until we reach the end of the sample.

Finally, we compare the forecasts with the actual data values using the RMSE metric, and the outcomes are presented in the third column of the table in part f.

f.

The RMSE's obtained in (c), (d) and (e) for the different AR(p) models are reported in the following table:

lag(s)	Rolling Window	Rolling Estimation Window	Rolling Window Expanded
1	0.02673197	0.02675663	0.02673533
2	0.02673212	0.02675628	0.02673542
3	0.02673196	0.02675679	0.02673542
4	0.02673200	0.02675692	0.02673546
5	0.02673206	0.02675729	0.02673554

We observed that the Root Mean Squared Error (RMSE) values across all scenarios exhibit negligible differences. For inquiries pertaining to questions c), d), and e), we advocate the selection of AR(3), AR(1), and AR(1) models, respectively. Notably, none of these chosen models align with the outcome obtained in question b), where the optimal model identified was AR(5). We will need other tests to identify the correct model.

Reflection.

We also want to provide some reflections based on our codes, results and the forecasting strategy implemented from part c to part e.

Firstly, the approach in parts c to e involves separate code segments for each AR(p) model. While this setup facilitates easier inspection, the code becomes lengthy and verbose. A more efficient solution would be to implement a loop that iterates over different lag orders. This loop should dynamically adjust the lag variables in our OLS method as the lag order changes, ensuring the model remains flexible.

Secondly, to maintain consistency with the lecture and tutorial exercises, we incorporated both past information and previously forecasted values for subsequent forecasts in parts d and e. However, an alternative approach could involve using only past information and excluding previous forecasted values for forecasting. In part d, where we shift our window, and in part e, where we expand our window, this alternative approach ensures that all necessary information for one-step ahead forecasts is available, and we provide the table as below. Although this alternative approach yielded more variation in the RMSE values in the second and third columns of the final table, the differences were negligible, leading us to report and analyze the results based on the methods introduced in parts d and e.

lag(s)	Rolling Window	Rolling Estimation Window	Rolling Window Expanded
1	0.02673197	0.02725159	0.02683636
2	0.02673212	0.02746576	0.02683397
3	0.02673196	0.02754266	0.02684586
4	0.02673200	0.02776557	0.02686620
5	0.02673206	0.02785689	0.02689480

Lastly, for model selection, in addition to information criteria, we can utilize the Breusch-Godfrey test to determine the optimal number of lags. For forecast evaluation, results from tests such as the Granger-Newbold test and Diebold-Mariano test should be considered if the goal is to identify the best fitting model. The Giacomini-White test can also be valuable when the aim is to determine which model provides the most accurate forecasts.