(14 pts) Assume that you want to build a simple stock market trading engine using R’s Reinforcement Learning library that we demo’ed in class. How would you create a training data set in this case? Please make sure to specify what you would use as the current state, the action, the reward and the next state to form the training tuples (si, ai, ri+1, si+1).

Current state – Current Stock Price

Action – Trading decision (Buy, Hold, Sell)

Reward – Change in Portfolio Value. If action is sell then calculate the profit it is positive then reward should be “1” for other two actions it should be “0” since we cannot calculate the realise profit when holding or buying stocks. This is just simple reward system and the reward system can get more complexed based on the trading strategy they use. If the trading strategy is high risk, high return then reward would be different than mentioned one.

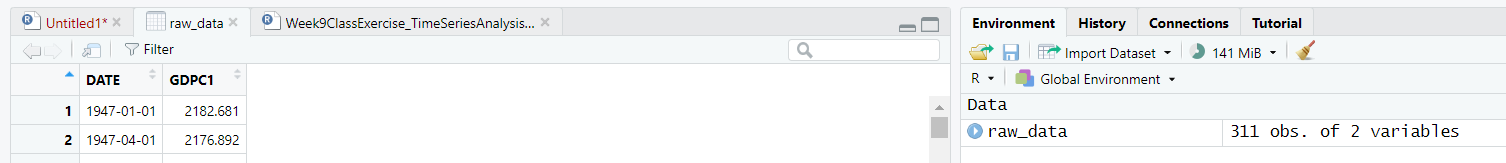
Next state – Updated market and portfolio information

To generate a dataset :

1. Use historical stock price data to create a sequential dataset.
2. For each time step in historical data :
   * **Extract si**: Calculate the position status for the current time step.
   * **Determine ai**: Select an action (randomly or based on a policy during training).
   * **Calculate ri+1**: Compute the reward based on the action taken.
   * **Form si+1**: Record the next state by moving to the next time step.
3. Store each tuple (si,ai,ri+1,si+1)(si, ai, ri+1, si+1)(si,ai,ri+1,si+1) as a row in the training dataset.

Question 2 (100 pts)

1. (4 pts) Import the Real Gross Domestic Product (GDP) data from the FRED website by visiting https://fred.stlouisfed.org/series/GDPC1. Upon a successful import you should see 2 variables and 311 observations.

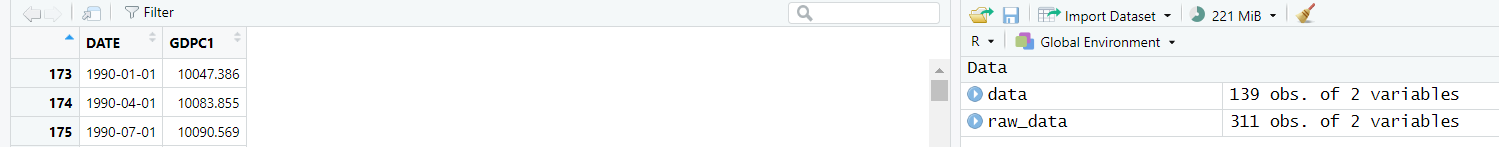


2. (6 pts) Use the lubridate library to create a date column that has the data type Date. This is a required step which will allow us to use time series functions necessary for the next set of questions.

A number and numbers on a white background

Description automatically generated

3. (6 pts) We will be using real GDP data from post 1990. Create a new data frame that only contains observations greater than or equal to 1990-01-01.



4. (4 pts) Create a time series plot of real GDP, with date on the horizontal axis and real GDP on the vertical axis.



5. (6 pts) Does real GDP look stationary? Why or why not? Briefly explain

Stationarity in time series data means that the statistical properties (such as mean and variance) do not change over time. In this plot, the GDP series shows a clear upward trend over time, indicating that both the mean and potentially the variance are increasing rather than remaining constant.

6. (8 pts) Create a new variable in the data frame called GDPGrowth using the following annualized real GDP growth formula: (GDPt − GDPt−1)/GDPt−1 ∗ 400. The reason we multiply by 400 is to get to the annualized growth rate from quarterly data.

A table with numbers and letters

Description automatically generated

7. (4 pts) Create a time series plot of real GDP growth, with date on the horizontal axis and GDPGrowth on the vertical axis.



8. (6 pts) Does real GDP growth look stationary? Why or why not? Briefly explain.

Yes, this series does appear more stationary. The GDP growth rate shown here fluctuates around a relatively constant mean, and there is no clear trend or systematic change in the variance over time.

9. (4 pts) Perform an Adjusted Dickey-Fuller test on real GDP growth. Report the p-value from the test. What does the test tell us about the stationarity of real GDP growth?

A screenshot of a computer

Description automatically generated

As the P-value is less than 0.01, there is strong evidence indicating that the GDP growth series is stationary.

10. (8 pts) Plot the Auto Correlation Function (ACF) of real GDP growth. How many lags exhibit statistically significant spikes?



**only one lag (lag 1)** exhibits a statistically significant autocorrelation for the GDP growth series. All other lags fall within the confidence intervals, suggesting that their correlations are not statistically significant.

11. (8 pts) Plot the Partial Auto Correlation Function (PACF) of real GDP growth. How many lags exhibit statistically significant spikes?



No lags exhibit statistically significant spikes. Except the lag 1 but it does not go below the blue line so there is no evidence of significant partial autocorrelation of any lag.

12. (6 pts) What do the ACF and PACF plots suggest about the order of the ARMA model? Estimate the suitable ARMA model using R’s Arima() function. Report the AIC of the estimated model?

In PACF plot, there are no significant lags, meaning there’s no strong indication of an autoregressive (AR) component. This absence of a cut-off in the PACF aligns with the characteristics of an MA process rather than an AR process.

the ACF plot has only one significant lag, it suggests that an MA model might be appropriate, where *q = 1*.

A close up of numbers

Description automatically generated

13. (8 pts) Calculate 4 period ahead forecasts of real GDP growth using the estimated ARMA model above. Report the forecasted values.



14. (6 pts) Plot the forecasted values with their confidence intervals.



15. (8 pts) Estimate another ARMA model using R’s auto.arima() function. What is the AIC of the estimated model? How does it compare to the AIC of the model you estimated above?



Both the models have same AIC value.

16. (8 pts) Calculate 4 period ahead forecasts of real GDP growth using the estimated ARMA model above. Report the forecasted values.



17. (6 pts) Plot the forecasted values with their confidence intervals



Question 3 (16 pts) The following questions are based on the economist article titled How Scientists are Using Artificial Intelligence which has been posted on D2L.

1. (4 pts) Who is Regina Barzilay? What are some of her significant scientific contributions?

Regina Barzilay is a computer scientist at MIT who has made significant contributions to the field of AI-assisted drug discovery. She was instrumental in developing AI models that identified two new antibiotics, halicin and abaucin, which are promising candidates against antibiotic-resistant bacteria. Barzilay’s work demonstrates how AI can dramatically accelerate drug discovery that efficiently scans vast chemical databases to pinpoint viable drug candidates.

2. (4 pts) How has AlphaFold helped biomedical research?

AlphaFold has produced a database of over 200 million predicted protein structures, used by over 1.2 million researchers. It enabled researchers to identify a malaria-related protein target in mosquitoes, aiding potential drug development, and assisted in mapping protein structures related to liver cancer, facilitating targeted treatment research

3. (4 pts) Does the AI revolution threaten to put all our scientists out of business?

I don’t think it will replace scientists it enhances scientists capability, speed and accuracy, particularly in tasks that require creative extrapolation beyond current knowledge. It does not yet explain the underlying mechanisms. The article states that "It's not a revolution that puts all of our scientists out of business".

4. (4 pts) Can you cite some examples where AI is ”acting as a multiplier for human ingenuity”?

AI is transforming scientific research by significantly enhancing efficiency and speed. In materials science, researchers have leveraged AI to reduce the pool of potential battery materials from thousands to just five, saving both time and resources. Weather forecasting has also improved, with AI models like Pangu-Weather and FourCastNet generating predictions much faster and at a lower cost than traditional methods, crucial for disaster preparedness. In nuclear fusion research, AI control systems allow scientists to experiment with plasma shapes more effectively, optimizing their experiments. Additionally, "self-driving laboratories" automate the planning and execution of experiments, making the discovery of new compounds up to a thousand times quicker than conventional methods.