

Introduction

This case study delves into the application of the AutoRegressive Integrated Moving Average (ARIMA) model in forecasting stock prices. By analyzing time series data for both stable and volatile stocks, specifically Procter & Gamble (P&G) and Minerva Neurosciences, this study evaluates the accuracy and reliability of ARIMA models across different market conditions.

Hypothesis

We hypothesize that the ARIMA model will accurately predict future stock prices within a confidence interval, provided the model parameters are optimally selected based on historical data.

Objectives

The primary objective of this case study is to determine the valid timeframe over which the ARIMA model provides reliable predictions and to assess whether ARIMA is a suitable model for stock price forecasting.

Methodology

Our approach involves:

Utilizing time series data from the New York Stock Exchange (NYSE) for P&G and Minerva.

Employing ARIMA modeling techniques to forecast future stock prices.

Analyzing the model's performance by comparing predicted values against actual stock prices to evaluate prediction accuracy.

Data Description

The data consists of daily stock prices including open, high, low, and close values, along with trading volume. This data was sourced from official NYSE datasets, focusing on the most recent four months for both training and testing purposes.

Analysis

Model Training: The ARIMA model parameters (p , d , q) were identified using historical data.

Validation: The model's predictions were tested against actual stock prices to measure accuracy.

Adjustments: Parameters were fine-tuned in response to the model's performance, especially in handling the inherent volatility observed in the Minerva stock data.

Results

The ARIMA model demonstrated high accuracy in predicting stock prices for P&G, a relatively stable stock, over the tested period. However, predictions for Minerva, which is more volatile, showed variability in accuracy, indicating the model's sensitivity to market fluctuations.

Conclusions

Our findings suggest that while the ARIMA model can be an effective tool for forecasting stock prices in stable conditions, its performance is less reliable in volatile markets without additional adjustments and considerations.

Next Steps

Future work will focus on refining the ARIMA model's parameters to enhance its robustness, particularly for volatile stocks. Additionally, expanding the dataset and incorporating more granular temporal resolutions may improve predictive accuracy.

Case Study Presentation Rubric

DS 4002 – Spring 2024 - Instructors: Javier Rasero and Harsh Anand

Due: Tuesday of Presentation week

Submission format: PDF

Group Assignment

General Description: Submit to canvas a pdf version of the PowerPoint presentation presented in class by the group leader. Upload this presentation to the highest level of the Github repository as well.

Why am I doing this?

If your project is more on the science side of the spectrum:

When you conduct research, the objective is to come into possession of knowledge no one else has. For a moment you and your colleagues are in sole possession of new knowledge... and the responsibility to make sure that the energy you expended is not in vain. You have the responsibility to share your knowledge... and do so responsibly. You will undertake this exercise as the first part of practicing the sharing of knowledge. Without this step all the work you have done is for nothing.

What am I going to do? Your group will produce a set of slides for the presentation, and this presentation will be given by your group leader (see PR1). The presentation need not be long, we are aiming for **7 minutes**, which means about seven slides. It makes it harder to be brief, there is a famous quip that is along the lines of “If I had more time, I would have written you a shorter letter”. It does need to cover all the essential pieces, from motivation and context, to describing your data, through analysis, and on to the next steps. The key is to think carefully about the important pieces for the audience and then present those in a coherent way. Making the supporting visual materials is a critical step.

Tips for success:

- Read the github repo and supplemental materials carefully.
- Submit slides early – if you submit slides early, we, the instructors, can give comments.
- Make your figures as large as possible, and with the highest resolution possible.
- Relatedly, pay attention to the readability of your slides. For example, try to increase the label sizes.
- PRACTICE. The more you practice, the better. This is also helpful for timing the presentation.

How will I know I have Succeeded? You will meet expectations on MI4 Prepare Presentation when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none">• About 7 slides• PDF format for submission to collab• Generate the slides through the program of your choice• Slide numbers (except for title slide)• Order

	<ul style="list-style-type: none"> ○ Title & Outline ○ Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/Etc. ○ Data Explanation/Acquisition ○ Analysis Plan and Justification ○ Tricky Analysis Decision ○ Bias and Uncertainty Validation ○ Results/Conclusions ○ Next Steps ○ References/Resources/Acknowledgements ○ Closing Slide ● General note: <ul style="list-style-type: none"> ○ Each section can take as many slides as it needs to unless specifically indicated as 1 slide. ○ Do not take the exact slide count as critical. When you practice you can time out the talk to hit the 7 minute mark. Some slides need more, and some slides need less time.
Title & Outline	<ul style="list-style-type: none"> ● 1 slide ● <u>Goal</u>: Hook your audience and make the group that did the work clear ● Include: Group number and members and leader ● Include: Course and Date ● Title <ul style="list-style-type: none"> ○ Package your most important finding in headline format ● Outline <ul style="list-style-type: none"> ○ Indicate the major parts of your presentation
Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/etc.	<ul style="list-style-type: none"> ● 1-2 slides ● <u>Goal</u>: Explain why your hypothesis is an important question ● Motivation/Context <ul style="list-style-type: none"> ○ Share relevant information with the audience so that they can understand what you are investigating. ○ Share relevant information so that they can understand why this matters. ● Restate Hypothesis ● Restate Research Question ● Restate Modeling Approach
Data Acquisition/Explanation	<ul style="list-style-type: none"> ● <u>Goal</u>: Your audience understands the key features of your data set and challenges if they are relevant. ● High-level summary of data set: <ul style="list-style-type: none"> ○ Share the information from your data dictionary in a condensed way. ○ Leave out most details but be prepared to answer follow-up questions. ○ State the format (e.g.: text, image, etc.) ○ State the size and necessary techniques if applicable ○ Discuss how your data set was acquired. Include relevant licensing or ethical concerns.

Analysis Plan and Justification	<ul style="list-style-type: none"> ● <u>Goal</u>: The audience will understand the pipeline you created. ● Show the diagram for your analysis pipeline. ● Highlight any particularly challenging components. ● Show any diagnostic or validation work done.
Tricky Analysis Decision	<ul style="list-style-type: none"> ● <u>Goal</u>: Identify and describe a decision you made in the analysis that was non-trivial. ● Every project has moments in analysis where you just need to make a decision, but the 'right' answer is not obvious or may not even exist. This slide is about one of those moments. ● Think about a judgment call you had to make in your project and describe why it matters and why it was tricky. ● Describe the impact of your decision. ● E.G.: What evaluation metric to choose, how many epochs to use, what data to exclude from analysis. ● N.B. If this decision made you deviate from the analysis plan registered in MI2, this is the moment to mention it.
Bias and Uncertainty Validation	<ul style="list-style-type: none"> ● <u>Goal</u>: This slide is designed to explain the difficulties in your data set ● List biases you consider and ones you correct for. ● Describe how you came to determine the uncertainty in the values you report.
Results and Conclusions	<ul style="list-style-type: none"> ● 1 slide ● <u>Goal</u>: Summarize the results ● Address the hypothesis ● Share relevant figures and findings
Next Steps	<ul style="list-style-type: none"> ● 1 slide ● Describe new lines of exploration discovered during the project ● Describes improvements that could be made ● Describe New questions that have arisen
References/Resources/Acknowledgements	<ul style="list-style-type: none"> ● Provide citations as necessary ● Provide links to resources used (e.g.: your GitHub page) ● Acknowledge contributions and assistance
Closing slide	<ul style="list-style-type: none"> ● 1 slide ● <u>Goal</u>: Finish on a slide with an image you want people to remember ● This slide will sit up while people are asking questions, pick something memorable ● It can be a repeat from earlier ● Don't be afraid to return to an earlier slide if a question takes you there.

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from [Streifer & Palmer \(2020\)](#).