

Advanced Financial Data Analytics

Coursework Assessment 2024/25

Table of contents

1	Module Information	2
1.1	Overview	2
1.2	Learning Outcomes	2
2	Project Requirements	2
2.1	Glossary of Key Terms	2
2.2	Data Sources	3
2.3	Research Questions	3
2.3.1	Section 1: Exploratory Analysis (30 marks)	3
2.3.2	Section 2: Time Series Modeling (40 marks)	4
2.3.3	Section 3: Advanced Analysis (30 marks)	5
2.4	Required Analysis Components	5
2.4.1	Section 1: Exploratory Analysis	5
2.4.2	Section 2: Time Series Modeling	5
2.4.3	Section 3: Advanced Analysis	6
2.5	Deliverables	6
2.5.1	Written Report (70% of total mark)	6
2.5.2	Code Submission (20% of total mark)	7
2.5.3	Video Presentation (10% of total mark)	7
2.6	Addressing Fairness Concerns	7
3	Technical Requirements	8
3.1	Software Requirements	8
3.2	Example Code Structure	8
4	Assessment Criteria	9
4.1	Marking Scheme Summary	9
4.2	Report Assessment (80 marks)	9
4.2.1	Research Design and Literature (15 marks)	9
4.2.2	Section 1: Exploratory Analysis (15 marks)	10
4.2.3	Section 2: Time Series Modeling (25 marks)	10
4.2.4	Section 3: Advanced Analysis (30 marks)	11
4.2.5	Presentation and Structure (10 marks)	12

4.3	Code Assessment (10 marks)	12
4.3.1	Code Functionality (5 marks)	12
4.3.2	Documentation (3 marks)	13
4.3.3	Code Organization (2 marks)	13
4.4	Video Presentation (10 marks)	13
4.4.1	Content and Structure (4 marks)	13
4.4.2	Technical Demonstration (4 marks)	14
4.4.3	Time Management (2 marks)	14
4.5	Additional Assessment Notes	14
5	Submission Guidelines	15
5.1	File Requirements	15
5.2	Deadlines	15
6	Academic Integrity	15
6.1	Plagiarism Policy	15
6.2	Code Attribution	15
6.3	AI Policy	16

1 Module Information

1.1 Overview

This module assessment helps you understand and apply different ways to analyze financial data that changes over time (like stock prices or exchange rates). You'll start with simple techniques to explore and visualize the data, then move on to more advanced methods to build models and make forecasts. Even if you're not a math whiz, this module focuses on helping you understand what the results mean and how they can be used in finance.

1.2 Learning Outcomes

By completing this project, you will be able to:

- Understand the basic characteristics of financial time series data (like trends, patterns, and how much it fluctuates).
- Explore and visualize data using charts and graphs.
- Build statistical models to forecast future values.
- Explain what the analysis tells us about the data and what it means in practice.
- Communicate your research findings in a written report and a short video.

2 Project Requirements

2.1 Glossary of Key Terms

To help you understand the technical language, here's a glossary of key terms:

- **Time Series:** A sequence of data points collected over time (e.g., daily stock prices, monthly inflation rates).
- **Stationarity:** Imagine a river flowing steadily. A stationary time series is like that - its statistical properties (like average and variability) don't change over time.
- **Autocorrelation:** Think of it like ripples on water. Autocorrelation measures how much a data point is related to previous data points in the same time series.
- **Volatility:** This is how much the data jumps around. High volatility means big price swings!
- **ARIMA Model:** A common model used to forecast future values in time series data. It's like using a special recipe that considers past values and errors to predict what might happen next.
- **GARCH Model:** A model that helps us understand and predict how volatility changes over time. Think of it like a weather forecast for how "stormy" the market might get.
- **Cointegration:** This means that two time series have a long-term relationship, even if they move apart in the short term. Imagine two friends who always end up at the same place, even if they take different routes.
- **Machine Learning:** A way to get computers to learn from data without being explicitly programmed. It's like teaching a dog a trick with treats - the computer learns by finding patterns in the data.

2.2 Data Sources

The `tsfe` R package (a collection of data and functions for analyzing financial time series) provides several datasets you can choose from:

1. Market Indices:

- Russell 2000 Price Index: Measures the performance of small-cap US companies.
- FTSE 100 Index: Measures the performance of the 100 largest companies on the London Stock Exchange.
- China A-Market Index: Measures the performance of stocks traded on the Shanghai and Shenzhen stock exchanges in China.

2. Exchange Rates:

- USD/GBP exchange rates: The value of the US dollar relative to the British pound.

3. Company Data:

- Carnival Plc earnings per share: A measure of the company's profitability.
- Historical stock prices: The price of a company's stock over time.

2.3 Research Questions

2.3.1 Section 1: Exploratory Analysis (30 marks)

Choose **one** of the following questions and use graphs and simple statistics to explore your chosen data:

1. **“Do stock prices for smaller US companies (Russell 2000) behave in the way theory suggests?”**
 - Look at how the returns are spread out.
 - Check if the returns follow a typical bell-shaped curve.
 - See if periods of high volatility tend to happen together.
 - Compare what you see in the data to what economic theory predicts.
2. **“What patterns can we find in Carnival Plc’s quarterly earnings?”**
 - Identify any long-term trends or repeating seasonal patterns.
 - Check if the statistical properties of the earnings stay the same over time.
 - See how the earnings are related to past earnings.
 - Look for any sudden changes in the earnings patterns.
3. **“How has the volatility of the USD/GBP exchange rate changed over time?”**
 - Measure how much the exchange rate fluctuates over different periods.
 - See if periods of high fluctuation tend to happen together.
 - Look at how the changes in the exchange rate are spread out.
 - Identify any distinct periods of high or low volatility.

2.3.2 Section 2: Time Series Modeling (40 marks)

Choose **one** of the following questions and build a statistical model to help you answer it:

1. **“Can we use a statistical model (ARIMA) to predict how the stock market will perform?”**
 - Build an ARIMA model to describe stock market returns.
 - Check if the model is a good fit for the data.
 - See how well the model predicts future returns.
 - Compare the model’s performance to simpler prediction methods.
2. **“Which techniques are best for smoothing out noisy financial data and making predictions?”**
 - Use different smoothing techniques to reduce noise in the data and find underlying patterns.
 - Compare how well the different techniques predict future values.
 - Look at the prediction errors made by each technique.
 - Discuss the pros and cons of each technique for real-world use.
3. **“What’s the best GARCH model for figuring out how volatile the market is?”**
 - Compare different versions of GARCH models used to estimate volatility.
 - Find the best values for the model’s parameters.
 - Assess how well the model predicts future volatility.
 - Discuss how the model can be used to assess financial risk.

2.3.3 Section 3: Advanced Analysis (30 marks)

Choose **one** of the following questions and use more advanced techniques to answer it:

1. **“Is there a long-term connection between major stock markets?”**
 - Test if there’s a long-term relationship between different stock market indices.
 - Build models that describe how the indices adjust to keep this relationship.
 - Analyze how the indices move in relation to each other.
 - Discuss how this information can be used to build investment portfolios.
2. **“Can machine learning predict financial markets better than traditional methods?”**
 - Use machine learning algorithms to predict market movements.
 - Compare the performance of machine learning to traditional statistical models.
 - Assess how well the models predict future values.
 - Discuss the challenges and limitations of using machine learning in finance.
3. **“Can we use VAR models to see how events in one market affect other markets?”**
 - Build VAR models to see how different markets influence each other.
 - Test if one market can be used to predict another.
 - Analyze how a shock in one market affects other markets.
 - Discuss what these spillover effects mean for the economy.

2.4 Required Analysis Components

2.4.1 Section 1: Exploratory Analysis

1. **Data Properties and Visualization:**
 - Create graphs showing how the data changes over time, with labels and explanations.
 - Use tools like ACF and PACF to analyze how data points are related at different times.
 - Use tests (like ADF and KPSS) to see if the data’s properties stay the same over time.
 - Use QQ plots to compare the distribution of the data to a normal distribution.
2. **Statistical Analysis:**
 - Create tables showing key features of the data (like average, middle value, and spread).
 - Use tests to see if the data follows a normal distribution.
 - Calculate statistics over different time windows to see how they change.
 - Measure the relationship between different variables.

2.4.2 Section 2: Time Series Modeling

1. **Basic Time Series Models:**
 - Choose the right type of ARIMA model.
 - Find the best values for the model’s parameters.
 - Do diagnostic checks to see if the model fits the data well.

- Assess how well the model predicts future values.

2. Volatility Modeling:

- Build a GARCH model to estimate volatility.
- Look at the errors made by the model.
- Predict future volatility.
- Calculate measures of financial risk.

2.4.3 Section 3: Advanced Analysis

The requirements depend on the topic you choose:

1. Multivariate Analysis:

- Define the relationships between multiple variables.
- Find the best values for the model's parameters.
- Use hypothesis tests to answer your research questions.
- Discuss what the results mean for the economy.

2. Machine Learning Applications:

- Build and train a machine learning model.
- Use cross-validation to check how well the model works on new data.
- Measure how well the model predicts future values.
- Compare different machine learning models and traditional statistical models.

2.5 Deliverables

2.5.1 Written Report (70% of total mark)

You need to submit a professional research report (maximum 3000 words) with these sections:

1. Introduction (10 marks):

- Give background information on the topic.
- Clearly state what you want to achieve with your research.
- Summarize previous research on the topic.
- Briefly describe the methods you'll use.

2. Data and Methods (20 marks):

- Describe the data you'll use in detail.
- Explain any steps you take to clean or prepare the data.
- Justify why you chose the methods you used.
- Describe how you implemented the methods.

3. Results (25 marks):

- Present your results clearly using tables and figures.
- Explain what the statistical results mean in simple terms.
- Check if the models meet the necessary assumptions.

- Compare the performance of different models or methods.

4. Discussion (15 marks):

- Summarize the key findings of your analysis.
- Discuss any limitations of your data or methods.
- Explain how your findings can be used in real-world situations.
- Suggest areas for further research.

2.5.2 Code Submission (20% of total mark)

Submit your R code files, making sure to:

- Include all the code you used for your analysis.
- Add clear comments to explain what the code does.
- Write your code in a clear and organized way.
- Use the required R packages.

2.5.3 Video Presentation (10% of total mark)

Create a 5-minute video that:

- Gives a quick overview of your project.
- Explains the methods you used.
- Highlights your key findings.
- Shows how you used R to do your analysis.

2.6 Addressing Fairness Concerns

This assessment is designed to be fair and inclusive for all students, regardless of their technical background. Here's how:

1. **Focus on Understanding, Not Just Coding:** The primary goal of this assessment is to evaluate your understanding of financial time series analysis and your ability to interpret and communicate results. While coding is a useful tool, it's not the only way to demonstrate your understanding.
2. **All R Code Provided:** To help you with the coding component, all the necessary R code for the core analyses has been made freely available in the GitHub repository for the module slides. You can adapt and use this code in your own analysis, even if you have limited coding experience. This ensures that everyone has access to the same resources and that no one is disadvantaged due to a lack of coding skills.
3. **Support for All Learners:** Extensive support is available to help you succeed in this assessment, regardless of your technical background:
 - **Coding Workshops:** Introductory workshops on R programming will be offered for those who need them.

- **Code Examples:** Well-commented code examples are provided in the GitHub repository, demonstrating how to perform the core analyses.
 - **Office Hours:** You can attend office hours to get help with coding or any other aspect of the project.
 - **Online Resources:** Links to helpful online tutorials and resources for learning R and financial time series analysis are available on the module Canvas page.
4. **Emphasis on Interpretation and Critical Thinking:** Your ability to interpret results, think critically about the data, and communicate your findings clearly is highly valued in this assessment. These skills are essential for success in finance, regardless of your technical background.
 5. **Transparent Grading:** A clear and detailed grading rubric is provided to help you understand how your work will be evaluated. Feedback will be provided on both the technical and non-technical aspects of your work.

3 Technical Requirements

3.1 Software Requirements

- R (version 4.0.0 or higher)
- RStudio (version 2023.06.0 or higher)
- Required R packages:
 - tsfe
 - tidyverse
 - tidyquant
 - fpp2
 - forecast
 - rugarch
 - vars

3.2 Example Code Structure

```
# Load the necessary packages
library(tsfe)
library(tidyverse)
library(tidyquant)
library(fpp2)

# Access and prepare the data
data(indice)

# Your analysis code goes here
```


4 Assessment Criteria

4.1 Marking Scheme Summary

Component	Weighting
Written Report	80%
Code Submission (mandatory)	10%
Video Presentation	10%

4.2 Report Assessment (80 marks)

4.2.1 Research Design and Literature (15 marks)

Grade	Marks	Criteria
Outstanding	13-15	You chose a great research question, explored relevant research thoroughly, and showed how your project fits into the bigger picture.
Good	10-12	You chose a good research question, explored relevant research, and demonstrated a good understanding of the background.
Satisfactory	7-9	You chose an okay research question, did some background research, and showed a basic understanding of the topic. You made some connections between your question and the analysis you plan to do.
Poor	4-6	Your research question wasn't ideal, you didn't do enough background research, or you didn't show a good understanding of the topic. The connection between your question and the analysis was unclear.

Grade	Marks	Criteria
Very Poor	0-3	Your research question wasn't appropriate, you didn't do enough background research, or your understanding of the topic was lacking. There was no clear connection between your question and the analysis.

4.2.2 Section 1: Exploratory Analysis (15 marks)

Grade	Marks	Criteria
Outstanding	13-15	You used advanced statistical methods and created excellent visualizations to explore the data, tested your ideas thoroughly, and interpreted your findings in a clear and insightful way.
Good	10-12	You followed a clear process, used effective visualizations, did the right statistical tests, and interpreted your findings well.
Satisfactory	7-9	You did the basic analysis, used standard visualizations, did some appropriate tests, and interpreted your findings reasonably well.
Poor	4-6	Your analysis was incomplete, your visualizations weren't great, or you used the wrong statistical tests.
Very Poor	0-3	You missed key parts of the analysis or used statistical methods incorrectly.

4.2.3 Section 2: Time Series Modeling (25 marks)

Grade	Marks	Criteria
Outstanding	22-25	You built an advanced model, tested it thoroughly, evaluated its predictions in a sophisticated way, and compared it effectively to other models.
Good	18-21	You chose the right model, did good diagnostic tests, evaluated its predictions clearly, and compared it to other models.
Satisfactory	14-17	You built a basic model, did some diagnostic tests, made some forecasts, and compared it to other models in a simple way.
Poor	9-13	You chose the wrong model, didn't do enough diagnostic tests, or didn't evaluate the predictions properly.
Very Poor	0-8	You built the wrong model, didn't do the necessary tests, or didn't make any forecasts.

4.2.4 Section 3: Advanced Analysis (30 marks)

Grade	Marks	Criteria
Outstanding	26-30	You used advanced techniques in a creative way, connected your analysis to the previous sections, and interpreted your findings exceptionally well.
Good	21-25	You used appropriate advanced methods, connected your analysis to the other parts of the project, and interpreted your findings well.
Satisfactory	15-20	You did some basic advanced analysis, used standard methods, connected your analysis to the other parts of the project reasonably well, and interpreted your findings adequately.

Grade	Marks	Criteria
Poor	10-14	Your advanced analysis wasn't strong enough, you didn't use the right methods, or you didn't connect it well to the other parts of the project.
Very Poor	0-9	Your advanced analysis was incorrect or you used inappropriate methods.

4.2.5 Presentation and Structure (10 marks)

Grade	Marks	Criteria
Outstanding	9-10	Your report is professionally formatted, well-written, perfectly cited, and clearly structured.
Good	7-8	Your report is well-formatted, clearly written, properly cited, and logically structured.
Satisfactory	5-6	Your report is adequately formatted, reasonably well-written, mostly correctly cited, and has a generally clear structure.
Poor	3-4	Your report has formatting issues, unclear writing, citation errors, or an unclear structure.
Very Poor	0-2	Your report has major formatting problems, poor writing, missing citations, or lacks a clear structure.

4.3 Code Assessment (10 marks)

4.3.1 Code Functionality (5 marks)

Grade	Marks	Criteria
Excellent	5	Your code works perfectly and implements the core analyses correctly.

Grade	Marks	Criteria
Good	4	Your code works and implements most of the core analyses correctly.
Poor	0-3	Your code has errors, doesn't work properly, or doesn't implement the core analyses correctly.

4.3.2 Documentation (3 marks)

Grade	Marks	Criteria
Excellent	3	Your code has clear comments explaining the key parts.
Good	2	Your code has some comments, but they could be more informative.
Poor	0-1	Your code has very few or no comments.

4.3.3 Code Organization (2 marks)

Grade	Marks	Criteria
Excellent	2	Your code is well-organized and easy to read.
Good	1	Your code is reasonably organized.
Poor	0	Your code is poorly organized and difficult to read.

4.4 Video Presentation (10 marks)

4.4.1 Content and Structure (4 marks)

Grade	Marks	Criteria
Excellent	4	You presented a comprehensive overview of your project with a clear and engaging structure.
Good	2-3	You covered the main points of your project with a reasonably clear structure.
Poor	0-1	Your presentation was missing key information or had a confusing structure.

4.4.2 Technical Demonstration (4 marks)

Grade	Marks	Criteria
Excellent	4	You clearly explained your code, effectively presented your results, and thoroughly discussed your methodology.
Good	2-3	You gave a basic explanation of your code, presented your results adequately, and discussed your methodology to some extent.
Poor	0-1	Your explanation of the code was unclear, your presentation of the results was poor, or you didn't discuss your methodology enough.

4.4.3 Time Management (2 marks)

Grade	Marks	Criteria
Excellent	2	Your presentation was perfectly timed within the 5-minute limit, with well-balanced sections.
Good	1	Your presentation was close to the 5-minute limit, with mostly balanced sections.
Poor	0	Your presentation was significantly over or under the time limit, with unbalanced sections.

4.5 Additional Assessment Notes

1. Section Integration (applies across all sections):

- Make sure there are clear connections between the research questions you chose.
- Your analysis should flow logically from exploratory analysis to more advanced techniques.
- Use a consistent methodology throughout your report.
- Tell a coherent story with your analysis.

2. Research Question Alignment:

- Each section of your report should clearly address the research question you chose for that section.
- Make sure the methods you use are appropriate for the research questions.
- Your results should directly relate to what you set out to investigate.

3. Standard Penalties:

- Late submission: 5% penalty per day (up to a maximum of 5 days late)
- Word count:
 - Up to 10% over: no penalty
 - 10-20% over: 5% deduction
 - More than 20% over: 10% deduction
- Missing sections: You may lose up to 100% of the marks for any section that is missing.

4. Academic Integrity:

- Any suspected plagiarism will be investigated.
- There are serious consequences for academic misconduct.
- Make sure to cite all your sources correctly using the APA 7th edition format.

5 Submission Guidelines

5.1 File Requirements

Submit the following files through Canvas:

1. Report files:

- `surname_firstname_report.pdf` (a PDF of your report)
- `surname_firstname_report.qmd` (the Quarto Markdown file used to create your report)

2. Code files:

- `surname_firstname_analysis.R` (your main R script)
- Any other R scripts you used

3. Video file:

- `surname_firstname_presentation.mp4`
- Maximum 5 minutes long
- Standard video format (MP4)

5.2 Deadlines

- Submission deadline: [Date] at [Time]
- Late submissions are penalized at 5% per day.
- You can submit up to 5 working days late.

6 Academic Integrity

6.1 Plagiarism Policy

All work must be your own. Make sure to reference all your sources properly using the APA 7th edition format. The university's academic integrity policy applies to everything you submit.

6.2 Code Attribution

If you adapt any code from external sources (like websites or textbooks), clearly comment on it in your code and provide the appropriate citations.

6.3 AI Policy

I expect you to use generative AI in this class. Some assignments will require it. Learning to use AI is an emerging skill, and I provide some tutorials in Canvas about how to use it. I am happy to meet and help with these tools during my coding clinics or after class.

Responsible AI Use:

- **Effort Matters:** The quality of your AI-generated output depends on the effort you put into crafting your prompts. Generic or low-effort prompts will likely result in low-quality outputs. Take the time to refine your prompts to achieve meaningful results.
- **Verification is Essential:** Don't blindly trust the information provided by AI tools. Always double-check facts and figures, especially those you are unfamiliar with. You are responsible for any errors or omissions in the AI-generated content you use.
- **Acknowledge AI Assistance:** Transparency is key. When you use AI tools in your work, include a paragraph at the end of your assignment explaining:
 - Which AI tools you used (e.g., ChatGPT, Bard).
 - How you used them (e.g., generating code, summarizing research, brainstorming ideas).
 - The prompts you used to obtain the results.

Failure to acknowledge AI assistance is considered a violation of academic integrity policies.

- **Use AI Judiciously:** AI is a powerful tool, but it's not always the right tool for every task. Think critically about when AI is appropriate and helpful for your analysis. Avoid over-reliance on AI, and ensure that your own understanding and insights are central to your work. See this [article](#) for how to be judicious.

Module Coordinator's Expertise:

Please be aware that the module coordinator has extensive research expertise in the use of Large Language Models (LLMs) and generative AI. This means they are well-equipped to identify instances of inappropriate or unacknowledged AI use in student work.

Consequences of Academic Misconduct:

Any violation of the university's academic integrity policies, including plagiarism, improper code attribution, or failure to acknowledge AI assistance, will be subject to investigation and may result in serious penalties.