Advanced Financial Data Analytics

Coursework Assessment 2024/25

Table of contents

0.1 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
- Effectively present complex analytical findings through concise verbal and visual communication
- Develop communication skills that differentiate human expertise from AI-generated content
- Practice professional presentation methods essential for careers in financial analytics

0.2 Communication Skills in the Age of AI

In today's professional environment, the ability to effectively communicate complex financial findings is as valuable as the technical analysis itself. With AI increasingly able to generate basic reports and analyses, human professionals must demonstrate skills that go beyond what artificial intelligence can produce:

1. **Contextualized interpretation**: While AI can generate analyses, human experts provide critical context, insight, and judgment that connects findings to real-world implications.

2. **Compelling storytelling**: The 5-minute presentation component challenges you to distill complex information into a compelling narrative—a distinctly human skill that AI struggles to replicate authentically.

3. **Balanced judgment**: AI can present data but lacks the nuanced judgment to determine which findings merit emphasis for specific audiences or business decisions.

4. **Trust and credibility**: Financial stakeholders need to trust both the analysis and the analyst. Your presentation skills build the foundation for establishing professional credibility in ways AI cannot.

The 20% allocation to the video presentation reflects the growing importance of these communication abilities in financial analytics careers. As AI tools become more prevalent for generating basic analyses, employers increasingly value professionals who can effectively interpret, contextualize, and communicate findings in accessible ways.

1 Project Requirements

1.1 About Quarto Documents

This assessment requires you to submit your work as a Quarto document (`.qmd` file). Quarto is a modern publishing system that allows you to:

- Write text and code in the same document
- Generate high-quality PDF and HTML outputs
- Include interactive visualizations and tables
- Show or hide code as needed

Your Quarto document should include:

- All your analysis code in clearly labeled code chunks
- Your written report with integrated figures and tables
- References and citations

You do not need to submit separate R scripts if all your code is included in the Quarto document. The code chunks in your document should be well-organized and commented.

1.2 Posit Cloud Requirement

This project must be completed on Posit Cloud (formerly RStudio Cloud), similar to all the tutorials and workshops in this course. You will have access to:

- A Quarto template for the project
- A pre-configured environment with all the relevant packages already installed
- The datasets from the `tsfe` package

Using Posit Cloud ensures consistency in your computing environment and makes it easier for you to focus on the analysis rather than setting up the technical infrastructure. It also ensures that your project will be reproducible when evaluated.

1.2.1 Accessing the Project Template

To access and use the project template:

1. **Log in to Posit Cloud**: Navigate to posit.cloud and sign in with your QUB account credentials.

2. **Access the Course Workspace**: Select our course workspace from your workspace list.

3. **Create a New Project from Template**:

   - Click the "New Project" button
   - Select "New Project from Template"
   - Look for the "Advanced Financial Data Analytics Project Template" under "Your Content" or "Templates"
   - Click on the template to create your new project

4. **Rename Your Project**: Right after creation, rename your project to "Surname_Firstname_Financial_Ana for easy identification.

5. **Explore the Template**: The template includes:

   - A pre-configured Quarto document with all required sections
   - Example code for data loading and basic analyses
   - Pre-installed packages required for the analyses
   - Access to datasets in the `tsfe` package

6. **Save Your Work Regularly**: Posit Cloud automatically saves your work, but it's good practice to manually save important changes.

If you encounter any issues accessing the template, please contact the module coordinator immediately.

1.3 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.

1.4 Data Sources

The `tsfe` R package provides several financial datasets that 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.

1.5 Research Questions and Report Structure

Your report should investigate three related research questions (one from each section below) that form a coherent analysis. For example, you might examine market volatility through increasingly sophisticated lenses: first exploring its basic properties, then modeling it, and finally conducting advanced analysis.

1.5.1 Example of Connected Research Questions:

- **Section 1:** "How has the volatility of the USD/GBP exchange rate changed over time?" (Exploratory Analysis)
- **Section 2:** "What's the best GARCH model for figuring out how volatile the market is?" (Time Series Modeling applied to the same exchange rate)
- **Section 3:** "Can machine learning predict financial markets better than traditional methods?" (Comparing ML predictions of exchange rate volatility against your GARCH model)

This progression builds a comprehensive analysis: first understanding what volatility looks like, then modeling it with traditional methods, and finally assessing if advanced techniques offer improvements.

1.5.2 Section 1: Exploratory Analysis (24 marks - approximately 750 words)

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.

1.5.3 Section 2: Time Series Modeling (32 marks - approximately 1000 words)

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.

1.5.4 Section 3: Advanced Analysis (24 marks - approximately 750 words)

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.

1.6 Report Organization (80% of total mark)

Your report should be structured as follows, integrating your three chosen research questions into a coherent narrative:

1. **Introduction (8 marks, ~300 words):**

- Background covering all three research questions
- How your chosen questions connect/complement each other
- Overall research objectives
- Brief literature review

2. **Data and Methods (16 marks, ~750 words):**

   - Description of your chosen dataset(s)
   - Data cleaning and preparation steps
   - Methods for:
     - Section 1 (Exploratory Analysis)
     - Section 2 (Time Series Modeling)
     - Section 3 (Advanced Analysis)

3. **Results (40 marks, ~1500 words):**

   - Section 1 findings (12 marks)
   - Section 2 findings (16 marks)
   - Section 3 findings (12 marks)

   Each section should include:

   - Clear presentation of results with appropriate visualizations
   - Statistical interpretation
   - Model diagnostics where applicable

4. **Discussion (12 marks, ~450 words):**

   - Integration of findings from all three sections
   - Practical implications
   - Limitations and future research
   - Concluding remarks

5. **Technical Implementation (4 marks):**

   - Code quality and organization
   - Appropriate use of Quarto features
   - Documentation and reproducibility

Total word limit: 3000 words (excluding code, figures, tables, and references)

1.7 Project Workflow

To help organize your work on this project, consider following this workflow:

1. **Data Selection & Exploration**

   - Choose a dataset aligned with your research interests
   - Perform initial exploration to understand its characteristics
   - Formulate your three connected research questions

2. **Analysis Implementation**

   - Conduct exploratory analysis (Section 1)
   - Develop time series models (Section 2)
   - Perform advanced analysis (Section 3)

3. **Integration & Interpretation**

   - Connect findings across all three analyses
   - Draw overarching conclusions
   - Identify practical implications

4. **Report Preparation**

   - Draft your written report with integrated visuals
   - Prepare your video presentation
   - Review and refine both deliverables

Each step builds on the previous one, creating a logical progression from data selection to final submission.

1.8 Video Presentation (20% of total mark)

Create a 5-minute video that:

1. **Introduction (1 minute):**

   - Overview of your research questions
   - Why they're connected/important

2. **Methods & Results (3 minutes):**

   - Key findings from each section
   - How the analyses build on each other

3. **Conclusions (1 minute):**

   - Main takeaways
   - Practical implications

- **Technical Requirements:**

- MP4 format

- Clear audio

- Screen sharing of key visualizations

- Webcam footage of presenter

2 Assessment Criteria

2.1 Written Report (80 marks)

2.1.1 1. Introduction (8 marks)

Grade	Marks	Criteria
Outstanding	7-8	Excellent background coverage of all three research questions with clear connections between them. Strong justification for research objectives. Concise but comprehensive literature review demonstrating deep understanding.
Good	5-6	Good background coverage with clear links between questions. Clear research objectives and adequate literature review.
Satisfactory	4	Basic background provided for all questions with some connections made. Research objectives stated with basic literature review.
Poor	2-3	Incomplete background, weak connections between questions, or inadequate literature review.
Very Poor	0-1	Missing key background elements, no clear connections, or missing literature review.

2.1.2 2. Data and Methods (16 marks)

Grade	Marks	Criteria
Outstanding	14-16	Comprehensive description of data and preparation steps. Methods for all three sections are clearly explained and justified. Excellent integration of approaches.

Grade	Marks	Criteria
Good	11-13	Clear data description and preparation steps. Methods for all sections well explained with good justification.
Satisfactory	8-10	Adequate data description and basic preparation steps outlined. Methods explained but may lack some detail or justification.
Poor	4-7	Incomplete data description, unclear preparation steps, or inadequate methods explanation.
Very Poor	0-3	Major omissions in data description, preparation steps, or methods explanation.

2.1.3 3. Results (40 marks)

2.1.3.1 Section 1: Exploratory Analysis (12 marks)

Grade	Marks	Criteria
Outstanding	10-12	Sophisticated exploratory analysis with excellent visualizations. Clear, insightful interpretation of findings.
Good	7-9	Good exploratory analysis with effective visualizations. Clear interpretation of findings.
Satisfactory	5-6	Basic exploratory analysis with adequate visualizations. Reasonable interpretation.
Poor	3-4	Weak analysis, poor visualizations, or unclear interpretation.
Very Poor	0-2	Inadequate analysis, missing visualizations, or incorrect interpretation.

2.1.3.2 Section 2: Time Series Modeling (16 marks)

Grade	Marks	Criteria
Outstanding	14-16	Advanced modeling techniques properly implemented. Excellent model diagnostics and interpretation.
Good	11-13	Appropriate modeling with good diagnostics and clear interpretation.
Satisfactory	8-10	Basic modeling with adequate diagnostics and interpretation.
Poor	4-7	Poor model choice, weak diagnostics, or unclear interpretation.
Very Poor	0-3	Incorrect modeling, missing diagnostics, or incorrect interpretation.

2.1.3.3 Section 3: Advanced Analysis (12 marks)

Grade	Marks	Criteria
Outstanding	10-12	Sophisticated advanced analysis with excellent implementation and interpretation.
Good	7-9	Good advanced analysis with clear implementation and interpretation.
Satisfactory	5-6	Basic advanced analysis with adequate implementation and interpretation.
Poor	3-4	Weak analysis, poor implementation, or unclear interpretation.
Very Poor	0-2	Inadequate analysis, incorrect implementation, or missing interpretation.

2.1.4 4. Discussion (12 marks)

Grade	Marks	Criteria
Outstanding	10-12	Excellent integration of all findings. Insightful discussion of implications. Thoughtful consideration of limitations.
Good	7-9	Good integration of findings. Clear discussion of implications and limitations.
Satisfactory	5-6	Basic integration of findings. Adequate discussion of implications and limitations.
Poor	3-4	Poor integration, weak discussion of implications, or missing limitations.
Very Poor	0-2	No integration, missing implications, or no discussion of limitations.

2.1.5 5. Technical Implementation (4 marks)

Grade	Marks	Criteria
Outstanding	4	Excellent code organization, comprehensive documentation, perfect reproducibility. Creative use of Quarto features.
Good	3	Good code organization, clear documentation, reproducible analysis. Effective use of Quarto features.
Satisfactory	2	Adequate code organization and basic documentation. Basic use of Quarto features.
Poor	1	Poor code organization, weak documentation, or limited use of Quarto features.
Very Poor	0	Disorganized code, missing documentation, or no use of Quarto features.

2.2 Video Presentation (20 marks)

2.2.1 Content and Structure (10 marks)

Grade	Marks	Criteria
Outstanding	9-10	Excellent coverage of all three analyses with clear connections. Perfect time management and engaging delivery.
Good	7-8	Good coverage of analyses with clear structure. Good time management and clear delivery.
Satisfactory	5-6	Adequate coverage of analyses with basic structure. Reasonable time management.
Poor	3-4	Incomplete coverage, poor structure, or significant time management issues.
Very Poor	0-2	Missing key analyses, no clear structure, or major time management problems.

2.2.2 Technical Demonstration (6 marks)

Grade	Marks	Criteria
Outstanding	5-6	Excellent explanation of technical aspects. Clear demonstration of code and results.
Good	3-4	Good explanation of technical aspects. Clear presentation of main results.
Satisfactory	2	Basic explanation of technical aspects and results.
Poor	1	Weak technical explanation or unclear results presentation.
Very Poor	0	Missing technical explanation or no results shown.

2.2.3 Presentation Quality (4 marks)

Grade	Marks	Criteria
Outstanding	4	Professional presentation style. High-quality audio/visual elements.
Good	3	Clear presentation style. Good quality audio/visual elements.
Satisfactory	2	Basic presentation style. Adequate audio/visual quality.
Poor	1	Poor presentation style or low-quality audio/visual elements.
Very Poor	0	Unprofessional presentation or technical issues preventing comprehension.

3 Submission Requirements

Submit through Canvas:

1. `surname_firstname_report.qmd` (Quarto source file)
2. `surname_firstname_report.html` (HTML output)
3. `surname_firstname_presentation.mp4` (Video presentation)

You do not need to submit separate R scripts if all your code is in your Quarto document.

4 Academic Integrity

4.1 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.

4.2 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.

4.3 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.

4.4 Word Count Guidelines

The 3000-word limit excludes:

- Code chunks and outputs
- Figures and tables
- Figure/table captions
- References
- Appendices
- YAML header

The word count includes:

- All main body text
- Section headings
- In-text citations
- Footnotes

5 FAQ

Q: Do my three research questions need to use the same dataset? A: No, but they should form a coherent narrative. If using different datasets, explain the connection.

Q: How should I acknowledge AI usage in my Quarto document? A: Include an appendix section titled "AI Usage" detailing which tools you used and how.

Q: Can I use data sources not in the tsfe package? A: Yes, but discuss with the module coordinator first.

6 Project Timeline

To help you manage this project effectively within the 4-5 week timeframe, consider the following suggested schedule:

Timeframe	Milestone
Week 1	• Select your datasets • Choose your three research questions • Complete initial data exploration • Begin Section 1 analysis (Exploratory Analysis)
Week 2	• Complete Section 1 analysis • Begin Section 2 analysis (Time Series Modeling) • Plan your approach for Section 3
Week 3	• Complete Section 2 analysis • Conduct Section 3 analysis (Advanced Analysis) • Start drafting the written report
Week 4	• Finalize all analyses • Complete written report draft • Create visualizations • Prepare video presentation
Week 5 (if available)	• Review and refine all deliverables • Get feedback if possible • Submit completed project

This compressed timeline requires careful planning and consistent work. Focus on completing one section per week, with some overlap as needed. If you only have 4 weeks, compress the last two weeks' activities into your final week.

7 Deadlines

- Submission deadline: April 11th at 9am.
- Late submissions are penalized at 5% per day.
- You can submit up to 5 working days late.

8 Academic Integrity

8.1 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.

8.2 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.

8.3 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.