Medical Inflation Case Study Rubric

DS4002 – Spring 2025 – Camila Gutierrez
Due: TBD
Submission Format: Upload link of GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?
This case study allows you to show your understanding of time series analysis to answer how medical care CPI differs from headline CPI. You will ultimately visualize what effect it has had on historical and future health expenditures through SARIMA (Seasonal Autoregressive Integrated Moving Average) time series forecasting in Python.

What am I going to do?
The GitHub repository for this case study can be found at https://github.com/camigutie/DS4002-CS3/tree/main. The modeling approach will consist of three main steps: data collection and cleaning, historical trend analysis, and Seasonal ARIMA trend forecasting. You will obtain the datasets consisting of yearly trends in headline (general) CPI and medical care CPI over the past 25 years. These datasets have been pre-processed to classify all quantitative data as integers and remove any rows or columns that do not need to be used. You will perform moving-average time analysis through SARIMA trend forecasting to analyze how time series data of medical care and headline CPI are anticipated to change in comparison to the health expenditures price index. The model output will be analyzed through the SARIMA regression summary to evaluate the dependency between observations and residual errors.

Tips for success:

- Get ahead. Read through the hook and rubric as soon as possible to start thinking about the time needed to complete this case study. Although all the cleaned data and scripts are provided, it is better to build in extra time to work through questions you may have.
- Be bold. This is your chance to dive into a new topic with the guidance of past students. Do not shy away from questioning the reasoning behind analysis and data decisions.
- Talk to the instructors. The components of this project align with what past students have done and asked for feedback on. The challenges faced may be the same speedbumps the instructors guided past students through one year ago.
- Talk to your fellow students. Your peers may be doing similar time series data case studies using SARIMA trend forecasting. If this is the case, it is beneficial to

Your final deliverables should include:
● Visualizations of CPI and MCPI Forecasts with Error Bars
● A data dictionary
● A commented source code
● A GitHub repository containing all materials used

How will I know I have succeeded? You will meet expectations when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	- One GitHub repository (submitted via link on Canvas) - Create a new GitHub repository for this assignment titled 'CS2_MedicalInflation' that contains - README.md - LICENSE.md - Source Code File - Your data (i.e., the images you chose) - REFERENCES.md
README.md	Concise summary of what you produced for the case study. This provides enough general information to orient someone to your project repository.
Source Code File	Well-documented Python Script that contains the code used to execute your CPI and health expenditures analyses. In the source code, you must include: - The Health Expenditures Price Index, Headline CPI, and Medical Care CPI datasets used - Data cleaning steps for each dataset - Pearson correlation coefficients between the indices - Linear regression results predicting Medical Care CPI from the Health Expenditures Price Index - SARIMA model definition steps, including p, d, q parameter selection - Model training and fitting procedures - Future forecasts at 1-year, 5-year, and 10-year horizons - Comments throughout, and especially in the Python Script when interpreting regression and SARIMA model outputs
References	A markdown file titled "REFERENCES.md" with citations of any resources (journal articles, websites, etc.) referenced in helping you create your model in IEEE Documentation style.

Acknowledgements: Thanks to Professor Alonzi for providing the rubric outline. This structure is pulled from Streifer & Palmer (2020).