Tenure Tracker - Executive Summary

Overview:

As graduate students near the completion of their PhD journeys, many face the pivotal decision of whether to pursue careers in academia. While this choice is straightforward for some, it can be complex for others, with job security often playing a significant role. This project aims to address a key question: *How many faculty positions will be available next year?* By leveraging multiple data sources and applying advanced regression models, we seek to develop a robust framework to estimate future trends in the academic job market. This work provides a data-driven perspective to inform decision-making for prospective academics.

Objectives:

- Does the number of faculty hired across the US in postsecondary institutions increase/decrease with increasing economic indicators?
- What is the nest model to predict the number of faculty positions offered in a given year?
- What is the trend of the number of faculty hires over time?

Data Collection:

The data collection process involves gathering the annual number of faculty in the US from 1970 to 2022 from the National Center for Education Statistics (NCES). The difference between two consecutive years is used as an indicator of the number of faculty hires this year. Although this is not exact, it should be a representative number of the new faculty hires. Relevant economic indicators, such as GDP, were collected from the Federal Reserve Economic Data (FRED) and other sources. The economic features that we expected to be important are:

- GDP
- Inflation rate (CPI)
- Labor force participation for BS degree holders
- Unemployment rate for BS degree holders
- Amount of NSF awards per year
- Federal budget for postsecondary education
- The difference between the proposed and the actual federal budget for postsecondary education
- Labor market conditions.

This combined dataset enables analysis of the relationship between economic factors and faculty hiring trends, providing a basis for predictive modeling.

Data Analysis:

Time series analysis was performed on the number of faculty hires through different models. We used LOWESS (Locally Weighted Scatterplot Smoothing), a technique that smooths data while minimizing the influence of social fluctuations, providing clearer underlying trends. We also used ARIMA (Auto-Regressive Integrated Moving Average) models to predict faculty numbers by incorporating past values and external shocks. We employed Gaussian Processes under the assumption that the combined effects of factors follow a normal distribution. Additionally, we used Fourier analysis to identify and analyze cyclic patterns in the data.

Economic Variable Regression was performed to find the most correlated features affecting the number of faculty hires by computing the maximal lag correlation for each economic factor to determine their delayed effects, followed by shifting variables accordingly. Principal Component Analysis (PCA) is applied for feature selection, and validated with methods such as Mutual Information and LASSO to ensure robustness. Regression models, including linear and polynomial forms, combine time-series data with PCA-derived components for improved prediction accuracy.

*KPI*s included k-fold cross-validation to assess model performance across multiple data splits and AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) for hyperparameter tuning and model selection.

Conclusions:

- GDP, CPI, and Labor force participation for BS degree holders were the most correlated to the number of faculty hires. The first two were positively correlated whereas the third was negatively correlated with the number of faculty hires.

- Gaussian Process Regression outperformed the other analysis techniques with mean absolute percentage error (MAPE) of 1.15%. ARIMA(3,3,5) came in second place. (MAPE: 1.94%)

- Both the long-term trend and the cyclical patterns in faculty hiring numbers indicate a steady upward trajectory and therefore it is a good time to apply for faculty positions.

Accessibility:

To make our findings accessible and allow users to explore the data intuitively, we developed an interactive API using the Python library Streamlit. This API enables users to investigate correlations between input data and faculty hiring trends through a user-friendly interface.

The platform includes dropdown menus for selecting specific datasets and analysis models, providing flexibility for users to tailor their exploration. Once selections are made, the API applies a chosen model to the data, generating dynamic visualizations and displaying key

summary statistics along with a prediction for the number of faculty positions in the upcoming year. With interactive plots and a streamlined design, this tool empowers users to follow their own hypotheses and gain insights into the factors influencing academic job markets.