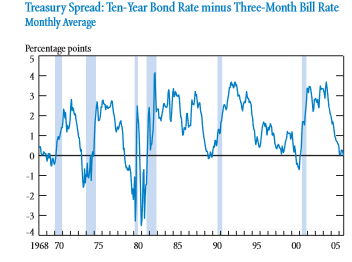
**ETL Project Report   
Suhaib Kiani, Jared Madris**

**What is the Yield Curve?**

Since the 1980s, economists have argued that the slope of the yield curve—the spread between long- and short-term interest rates—is a good predictor of future economic activity. The chart below highlights the spread between long- and short-term interest rates, which is believed to be a good predictor of future economic activity as before each of the last six recessions, this spread became negative. In October, 2019 (and briefly at the end of January 2020) the spread became negative sparking speculation about a possible economic recession in the near future.



**Project Overview**

Our ETL Project seeks to lay the ground for further study in the use of the interest rate yield curve as a real time forecasting tool for future economic activity particularly for forecasting economic recessions. We used datasets for historical interest rates to calculate the yield spread and combine it with GDP data to highlight the relationship between yield curves and past recessions.

**ETL Summary : *Extract***

We downloaded three historical datasets in CSV format from the Federal Reserve Bank website :

- Monthly Federal Reserve Funds rate (short term interest rates)

- Monthly 10 year traded treasury bills (long term interest rates)

- Quarterly GDP data with GDP growth percentages recession markers

All three datasets dated back to 1954 and were current up to December 2019. The datasets were read into a jupyter notebook for the project.

**ETL Summary : *Transform***

The three datasets were saved into Pandas dataframes. Data cleaning was performed which involved changing column names and changing format of "Date" column to datetime. The three dataframes were merged into a singledataframe in two steps. An outer merge was performed to address discrepancy between the monthly datasets (interest rates) and quarterly dataset (GDP). The difference between short and long term rates (spread denoting the yield curve) was calculated and saved as a new column.

**ETL Summary : *Load***

The combined dataframe was saved as an SQL database to PostgreSQL using an SQLAlchemy engine. We chose to use PostgreSQL as the interest rate data is arranged by date of market closure and well suited for highly structured PostgreSQL relational database format.

**Potential Next Steps**

The database can be used to create a real time forecasting tool for future economic activity particularly for predicting economic recessions. A regression analysis can be performed to understand the historical relationship between GDP growth/recession and yield curve profiles . This could lead to real time tool for calculating the probability and/or magnitude of an economic downturn in the future based on current yield curve analysis.