Project Report

Objective

The aim was to model the Market Yield on U.S. Treasury Securities at different Constant Maturity Time Periods, specifically, 1-Year, 2-Year, 3-Year, 5-Year, 7-Year, and 10-Year returns. The additional task was to try and reduce dimensions of the combined data set, while still explaining most of the variation.

Introduction

United States Department of the Treasury issues United States Treasury securities, often known as Treasuries or Treasuries, as a form of tax-free financing for the federal government. The Bureau of the Fiscal Service, which replaced the Bureau of the Public Debt, has overseen managing U.S. government debt since 2012. This historical data was collected by FRED (Federal Reserve Economic Data) from where we will query and retrieve our data from. The prediction of each of the US Treasury Securities requires data from returns of other treasury securities and their past values. The implications of this are problematic because it leads to a curse of dimensionality. This project dives into the historical data provided by FRED and uses Lasso to reduce dimensions to provide an easily interpretable model for the market yield of every different U.S. Treasury Securities for the next month. As a follow-up, this project uses ARIMA for roll-over predictions – moving beyond next month predictions.

Data

The first step was to query and retrieve the data series from FRED. To do this, Full FRED toolkit was used. Data was collected for all the 6 different U.S. Treasury Securities, and we removed any data prior to '1976-06-01' to make all the different time series data of equal sizes. This was done as not all the returns data started at the same date. The first 70% of the data was used for training, the next 10% for validation and the last 20% for testing.

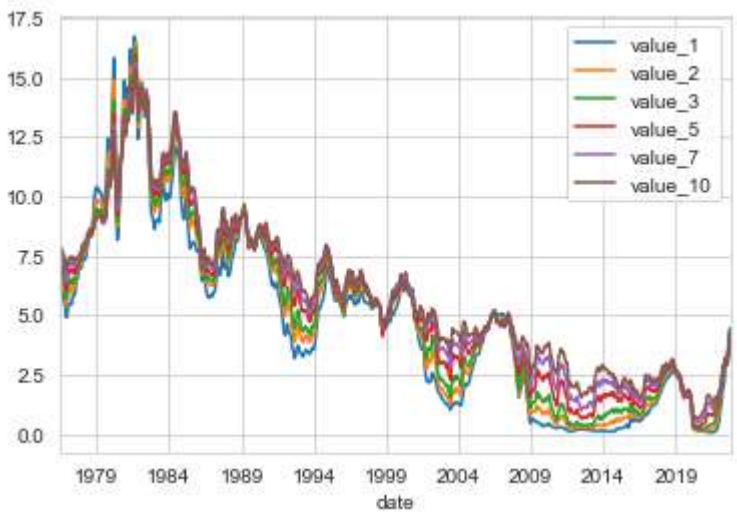

Figure 1 - U.S. Treasury Securities

Analysis & Modelling

Since the range of the data was small, there was no need to scale. We could directly move to other pre-processing steps and modelling. This section consists of the pre-processing steps and dimensionality reduction by using a r2-norm regularized regression algorithm called Lasso.

Sliding Window

Using a sliding window of 2 year i.e., 24 entries allowed us to change the problem into a supervised time series learning problem where each row had 24 independent variables and a dependent variable. All the 6 types of treasury returns were transformed in the same way.

Baseline vs Linear Regression on Sliding Window

The baseline model is taking the average of the previous 2 years to predict the next month's values for each Series. This is compared to using Linear Regression with the previous 2 years data of all 6 series to predict the next month's values. Refer to Model 1 in the table at the end.

Lasso

Since I used the past data of all 6 series to predict each of the next series, the dimension of the data was $6 * 24$ for each series. I used Lasso Regression to do variable selection. I found that for each variable, I only needed the last month's data of 2 or 3 of the other variables, reducing the dimension from 144 to 2 or 3, depending on the variable. Below are the tables:

1-Year

Variable	Coeff
value_1_23	0.908028
value_1_0	0
value_7_12	0
value_5_21	0
value_5_22	0
value_5_23	0
value_7_0	0
value_7_1	0
value_7_2	0
value_7_3	0

2-Year

Variable	Coeff
value_2_23	0.709465
value_1_23	0.190547
value_1_0	0
value_5_20	0
value_5_22	0
value_5_23	0
value_7_0	0
value_7_1	0
value_7_2	0
value_7_3	0

3-Year

Variable	Coeff
value_2_23	0.704379
value_3_23	0.172347
value_1_0	0
value_7_2	0
value_5_21	0
value_5_22	0
value_5_23	0
value_7_0	0
value_7_1	0
value_7_3	0

5-Year

Variable	Coeff
value_3_23	0.630898
value_5_23	0.228335
value_5_1	0
value_5_20	0
value_5_21	0
value_5_22	0
value_7_0	0
value_7_1	0
value_7_2	0
value_7_3	0

7-Year

Variable	Coeff
value_5_23	0.638484
value_3_23	0.199838
value_1_0	0.006164
value_1_2	0.001779
value_7_3	0
value_5_21	0
value_5_22	0
value_7_0	0
value_7_1	0
value_7_2	0

10-Year

Variable	Coeff
value_5_23	0.584846
value_7_23	0.233095
value_1_0	0.010859
value_7_11	0
value_5_20	0
value_5_21	0
value_5_22	0
value_7_0	0
value_7_1	0
value_7_2	0

Testing

I did Linear Regression on the variables selected using Lasso. This gave me results on all 6 series over the last 2 years. Refer to Model 2 in the table below.

Post-hoc Analysis with ARIMA

The fact that I needed only last month's data made me realize that I needed to try a different approach as I wasn't using the sliding window to its full extent. My second approach was to use ARIMA model on each of the data series.

	Model1	Model2	Model1-Model2	Arima	Model2-Arima
Value_1	-45.2194	0.826133	-46.045538	0.960259	-0.134126
Value_2	-5.7009	0.882889	-6.583788	0.966988	-0.084098
Value_3	-1.688	0.903075	-2.591073	0.953809	-0.050733
Value_5	0.041119	0.914334	-0.873214	0.922633	-0.008299
Value_7	0.409791	0.891898	-0.482106	0.93157	-0.039673
Value_10	0.478536	0.915231	-0.436694	0.923049	-0.007818