Project #2 Time series analysis for interest rate hedging

The first three principal components of the US Treasury yield curve explain about 95% of the variation in the yield curve over time. The first principal component is usually referred to as a parallel shift (even though the eigenvectors are not all the same value) which increases or decreases the entire yield curve. The second principal component represents a steepening or flattening of the yield curve and the third principal component represents a change in curvature with the middle of the yield curve moving in a different direction than the short and long end of the curve.

I would like to do a time series analysis on the changes in the treasury yield curve based on this principal component analysis looking at how the US treasury yield curve evolves as a function of current and past shocks to the US treasury curve, changes in foreign exchange rates and bond yields of other large countries and Federal Reserve Board member speeches. I believe using vector autoregressions and impulse response functions could lead to an interesting analysis of what is noise and what represents a trended change in yield curves. I would also like to look into a spectral analysis to see whether Fed speeches that are different have a longer wavelength on the yield curve than other sources of signal.

This is an important topic for corporates that need to issue long-term debt and are uncertain about the timing of hedging their debt issuance. For example, suppose a hypothetical local tech company has $3.25B in debt maturing over the next 18 months and needs to refinance this debt. The company can hedge the expected future issuance at a cost, but the cost of hedging swings wildly over time. Is there some signal in the changes to foreign exchange rates, US yields or foreign yields that would indicate the variance of treasury yields is increasing or decreasing that help make the decision of when to hedge or to remain exposed to chances in the yield curve.

DATA PIPELINE:

Collection:

Historical Treasury rates and foreign exchange rates will come in from Quandl or several other online data sources. I will need exchange rates with a London time stamp (before the open of the US market, and I believe getting this data will not be a problem.)



Federal reserve speeches since 2010 are available online at the St. Louis Fed’s web site (<https://www.stlouisfed.org/fomcspeak/viewbydate>) in text format. Some of the speeches are video format, but I believe I can use the speech recognition library on pypi to translate the speaches.)

Cleaning:

The majority of the cleaning will be focused on getting the Federal Reserve speech`s into a vectorized format. This will include converting the audio into text before we can turn it into vectors.

I will need to adjust the historical foreign exchange and interest rate data for differences in holidays.

Modeling:

* Performing principal component analysis on the interest rates will be pretty straight forward in python
* Determining if the new speech information or exchange rate information has an impact on the principal components of the term structure will require regression techniques
* I would like to look at spectral analysis on the eigenvalues to determine the cycle length of shocks. The yield curve tends to go on long (several week) moves in one direction before settling down into a pattern that looks more like random noise. I would like to see whether these longer
* Vector Auto Correlations and variance decomposition

Evaluation:

* Out of sample forecasts of potential impact of certain events

Challenges:

* Spectral analysis
* Federal Reserve speeches have historically been very similar. The governors themselves have begun using their speeches to sway fed policy, but the main press releases have been very similar. Will the differences in their speeches be enough difference to find something very interesting?

End:

* I hope to have a good understanding of the potential longer-term impact of shocks to the Treasury yield curve
  + How long will it take for a substantially different fed speech to dissipate into Treasury markets
  + What percentage of the variance of the yield curve is due to new information (monetary and fiscal policy) versus other noise.
  + Do parallel shifts in the yield curve (the first principal component) have a lagged impact on the slope of the yield curve in the future
  + A web page that demonstrates the impact of the latest fed speech, currency move and what my model is forecasting for interest rates.

Packages:

Numpy, pandas, scikit learn, nlp, speech recognition and Flask