Correlation between Equities and Bonds:

ECON 493 QFE Lab Data Visualization

Larry Li

April 30, 2020

1 Introduction

This project discovers and visualizes the relationship between bonds and equities. Specifically, it visualizes the correlation between different kinds of US Treasury bonds, with maturity from 1 month to 30 years and 5 different equity portfolios by industries. Time horizon is from 2006 to 2020. I used Microsoft Power BI to do the visualization backed by Matlab for early-stage data processing. There are two parts for this project. The first part is matrix for unconditional correlation between the above-mentioned two groups of data. The second part is time-variant conditional correlation between the two groups. Users will be able to choose from a fixed rolling window size. The initial idea is from Prof. Aguilar and we extended it to this working application.

2 Motivation

2.1 Learning Objective

There are mainly two learning objective from this application. First is to have a deeper understanding about correlation, which measures the dependency between two random variables or two series of data. For a fixed amount of time-series data, one can get a unconditional correlation, but also able to get a time-variant conditional correlation, using a rolling window strategy.

The second learning objective is to discover the relationship between equities and bonds. In general, an important reason for discovering the relationship is that when bond prices rise, bond yields goes down. And as bonds yield goes down, equities become more appealing. Vice versa. Usually, equity prices have an inverse relationship with bond yields. The relationship is useful for investors to make decisions. The SP500 and 10-year Treasury yield curves reflect this very well. However, we are not sure whether it's still the case when looking at US Treasuries with different maturities or subset of SP500 (for example, equities in different industries).

Another important motivation is that although we are able to find tons of financial data on websites and databases, there is no place that investors/researchers/students can find detailed correlation information between these two groups, especially considering the specific subset that we are focusing. Thus, I think this application is useful and beneficial for a various groups of people who are involved in these areas and industries.

2.2 Targeting Audience

The target audiences for this application is not limited. It targets a wide range of ECON major students. Students learning Econometric will find it useful for understanding the concept and application of correlation and covariance. Students currently in ECON 420/425 will benefit from it in understanding the relationship between bonds and equities. Students currently in ECON 525 might find the data and visualization useful to support their potential research project.

It may also have impact for audiences outside of class. Like briefly mentioned above, since there are no compiled visualization available online, investors and researchers may also find it interesting.

3 Design

The overall design focuses on the capability of extension (including different data categories, time horizon, rolling window size, etc.) and duplication. Students trained from ECON 525 or with programming background should be able to follow the code and duplicate the application. In this section, I will present the way that I choose data for this project, the structure for this project, a ”user manual” for this project, and the limitation that it has.

3.1 Data

There are two categories of data: bonds and equities. It is intriguing to discover the relationship between bonds that have different maturities and equities that are from different industries. The maturities for the bonds are 1month, 3month, 6month, 1yr, 2yr, 3yr, 5yr, 7yr, 10yr, 20yr, 30yr, 11 maturities in total. Equities are from Ken French data library 5 industry portfolios, including consumer(cnsmr), manufacture(manuf), High-Tech(HiTec), Health(Hlth), and others. Time span is from 2006 to 2020. The reason is that all bonds data from the 11 different maturities are available after 2006. To maintain a large group, I decide to start from 2006. However, it is also possible to get rid of some bond data and start earlier.

The application is available in both monthly data and daily data for the unconditional correlation. I downloaded daily bond rate data from US Department of The Treasury website and monthly data from FRED. (daily data are also available from FRED). The 5 industry portfolio data are from Ken French data library for both of the data frequencies.

3.2 Structure

The overall structure is using Matlab to do the data processing and use Power BI for visualization. For daily data, a difficulty to deal with is synchronize the dates from two different sources. I use Power BI to do the initial round of process but I think Matlab should also be able to do that.

3.2.1 Matlab

For this particular project, it's very heavy-weight, so using Excel for processing data is not applicable. There are 11 bonds series and 5 equities series, and each pair-up will have it's own correlation series in the conditional series. Thus, we need to calculate the conditional correlation using rolling window for 55 pairs of data series. We pick 3 different rolling period, 20 days, 60 days, and 100 days. Thus, for each of the 55 pairs, we need to calculate correlation for 3 times. This scale shows that Excel will be inefficient. After the calculation, I also use Matlab script to range the data into the long format that Power BI requires.

3.2.2 Microsoft Power BI

The first page of power BI shows the unconditional correlation for these assets. The top graph is in daily frequency while the bottom one uses monthly data. Users should be able to sort the rates and sort the maturities by clicking the headers on first rows. The background of the cells also show the relative relationship between the size of numbers.

The second page is the conditional correlation series. Top graph is categorized by rolling window so it can show maximum 3 series with 3 different rolling window. Middle graph is categorized by different combinations from equities and bonds, so it can show a maximum of 55 lines. The bottom series graph will allow users to drag and change the time that is shown in the graph. The slicers on the bottom left are used to choose what pairs of bonds and equities as well as the rolling period. Default is selecting all rolling periods and all pairs. Ideally, if users pick more than one options from one box, they should only pick one from the other box. For example, if I pick cnsmr10yr and cnsmr20yr from the left slicer, I should only pick 1 rolling window period from the other slicer. The reason is that if users pick more than 1 from both of the slicers, the calculation on the graphs will be the average of the data for the other category. If the users pick one from both of the slicers, the two graphs will be the same. (Note: using control+left click for multi-selection)

3.3 Assessment of Feasibility

This application follows functional programming architecture in the Matlab code so it can be easily extended and updated. Details in the comment of Matlab code. All the data are available downloading directly online as well as pulling by API.

4 Potential Extension

4.1 Limitations

The first limitation for this application is that I, when I was developing and deploying the project, was downloading the data from websites directly, which can be time consuming, depending on the wide-range of the datasets. One way to improve is pulling by API in Matlab (I didn't use this method since when I started working on the project and downloaded all the data, I thought I can use Excel so I didn't use Matlab at the first place). Another way is to using Power BI's built-in data import feature, which is still a new realm to explore, and I'm not sure whether it's applicable, regarding the opportunity cost of using Matlab, although I had some hands-on trying.

Another limitation is mentioned in the Power BI subsection that when you choose two or more categories from both of the slicers, it will show the average of correlation, which is not meaningful. However, users can still make use of this application based on the instruction above. This limitation should be able to be fixed using another design of the visualization, but I haven't figured out a better way than the current model to handle such a large amount of data.

4.2 Advantage

Like mentioned above, this application should be easily extended and duplicate based on the instruction that I discussed above and the Matlab code that I commented. The advantage of using Power BI is that all the data will be automatically updated once you update the source file. Thus, ideally, by running the Matlab code manually, the source file will be updated, and the Power BI will be automated accordingly.

A more ideal way is to using Matlab's compatibility with Power BI. There is a Matlab Production Server Interface for Microsoft Power BI that enables to send data from Power BI to Matlab engine. By using Matlab, Matlab Compiler SDK and Matlab Production Server, and connect to Power BI using RESTful API, the process will be fully automated. However, this service requires subscription and staff from Matlab haven't gotten back to me on the free trial.

5 Appendix

Database reference:

Ken French Data Library:

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

FRED:

https://fred.stlouisfed.org/

US Department of The Treasury:

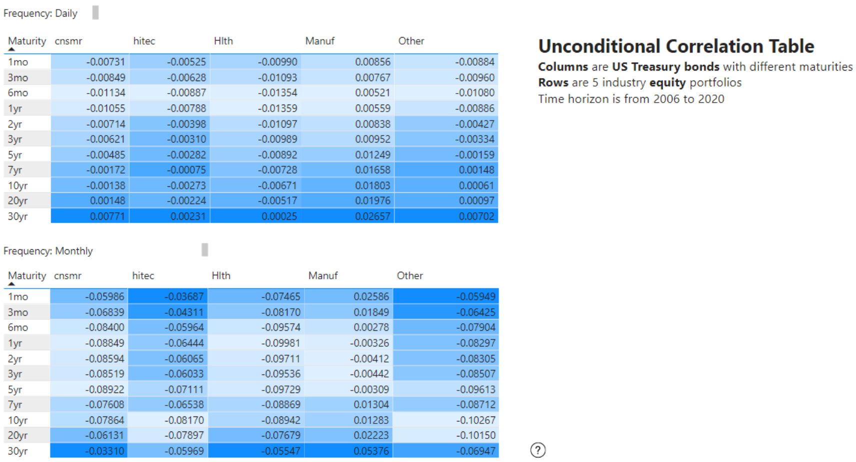
https://www.treasury.gov/resource-center/data-chart-center/interest-rates/pages/textview.aspx?data=yie

Code reference:

https://github.com/liruochen1998/DataViz-2020

Screenshots of application:

Figure 1: Unconditional Correlation Page



Frequency: Daily

Maturity	cnsmr	hitec	Hlth	Manuf	Other
1mo	-0.00731	-0.00525	-0.00990	0.00856	-0.00884
3mo	-0.00849	-0.00628	-0.01093	0.00767	-0.00960
6mo	-0.01134	-0.00887	-0.01354	0.00521	-0.01080
1yr	-0.01055	-0.00788	-0.01359	0.00559	-0.00886
2yr	-0.00714	-0.00398	-0.01097	0.00838	-0.00427
3yr	-0.00621	-0.00310	-0.00989	0.00952	-0.00334
5yr	-0.00485	-0.00282	-0.00892	0.01249	-0.00159
7yr	-0.00172	-0.00075	-0.00728	0.01658	0.00148
10yr	-0.00138	-0.00273	-0.00671	0.01803	0.00061
20yr	0.00148	-0.00224	-0.00517	0.01976	0.00097
30yr	0.00771	0.00231	0.00025	0.02657	0.00702

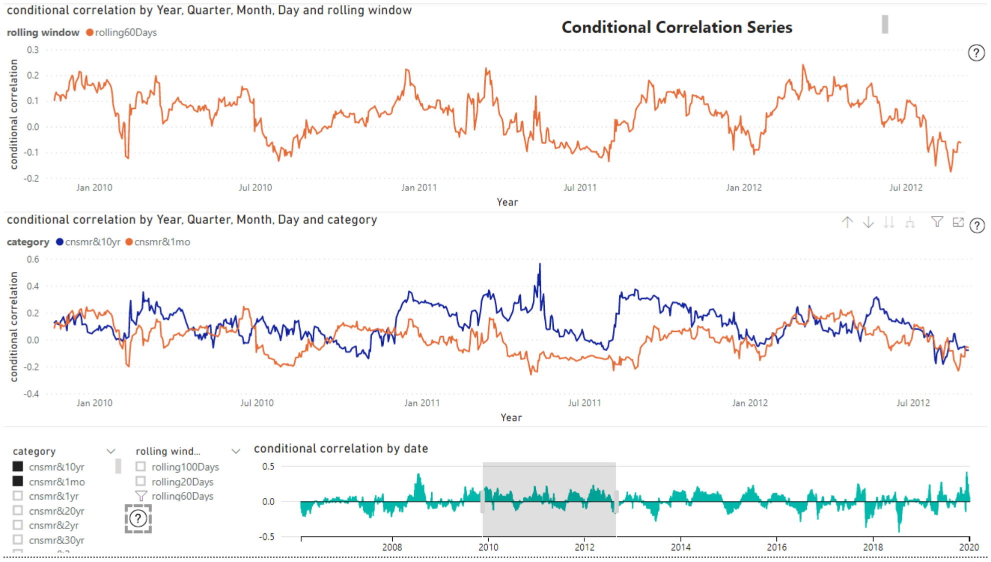
Frequency: Monthly

Maturity	cnsmr	hitec	Hlth	Manuf	Other
1mo	-0.05986	-0.03687	-0.07465	0.02586	-0.05949
3mo	-0.06839	-0.04311	-0.08170	0.01849	-0.06425
6mo	-0.08400	-0.05964	-0.09574	0.00278	-0.07904
1yr	-0.08849	-0.06444	-0.09981	-0.00326	-0.08297
2yr	-0.08594	-0.06065	-0.09711	-0.00412	-0.08305
3yr	-0.08519	-0.06033	-0.09536	-0.00442	-0.08507
5yr	-0.08922	-0.07111	-0.09729	-0.00309	-0.09613
7yr	-0.07608	-0.06538	-0.08869	0.01304	-0.08712
10yr	-0.07864	-0.08170	-0.08942	0.01283	-0.10267
20yr	-0.06131	-0.07897	-0.07679	0.02223	-0.10150
30yr	-0.03310	-0.05969	-0.05547	0.05376	-0.06947

Unconditional Correlation Table

Columns are **US Treasury bonds** with different maturities
Rows are 5 industry **equity** portfolios
Time horizon is from 2006 to 2020

Figure 2: Conditional Correlation Page



category
■ cnsmr&10yr
■ cnsmr&1mo
□ cnsmr&1yr
□ cnsmr&20yr
□ cnsmr&2yr
□ cnsmr&30yr

rolling wind...
□ rolling100Days
□ rolling20Days
▽ rolling60Days