Exercise 4

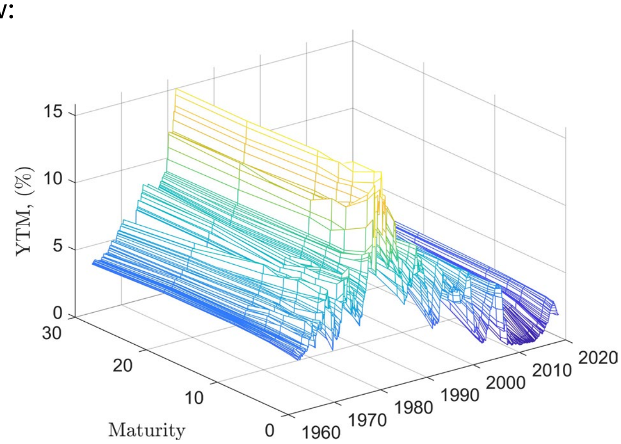
1. We are going to use the data files "FF3factors.txt" and "USTreturns.txt" Download them from Canvas (they are included in Exercise 4) and open them in a text editor (Notepad on Windows, TextEdit on Mac). **What do you notice about the data files? What is the file format – specifically what separates columns?**
2. Import the data into Python. FF3factors.txt contains data on special portfolio returns. These special portfolio returns help us understand how returns are distributed among all the risky assets out there. In order from left to right, the data columns are Year, Month, Market *Excess* Return ($r_m - r_f$), Small Minus Big (smb) and High Minus Low (hml). Smb is the return of the big portfolio we saw last time subtracted from the small return. Hml is the low book-to-market (Growth) subtracted from the high book-to-market portfolio return (Value).

   The USTreturns.txt contains data on US Treasury Securities. The first two columns are Year and Month, then it is 7 different *realized* bond returns (actual change in value from one month observation to the next). The next 7 columns are the yield to maturity corresponding to each bond (first return matches with first yield, second return with second yield, etc). The last 7 columns are the time to maturity (in years) or each bond (first return corresponding to the first time to maturity, etc.).

   What units are each data file's returns in? Do you need to transform one or the other (or both) at all?

3. Filter the data so you have observations between July 1963 and December 2019. **What are the averages maturities for the bonds? What about the standard deviation of maturities for each bond? Which bonds have more variable maturities? What original maturity likely corresponds to each bond? (ie when it was issues, what was the time to maturity?). Create a vector corresponding to these original maturities.**

4. Create a "surfboard" plot of the "yield curve." That is, the relationship between time to maturity and yield to maturity, but showing this yield curve across time. It should look something like below:



What happened to the general level of interest rates between the early 1980s and 2020? Since the early 1980s, there have been several recessions: July 1990 to March 1991, March 2001 to November 2001, December 2007 to June 2009. Examine the general relationship between short- term and long-term yields pre- and post-recession. What do you see?

5. Now let's switch gears to the FF3factors.txt data for a bit. Filter the data by date as we did with the bond returns. **What is the average *annualized* excess return for the market factor (the first column)? What is the *annualized* standard deviation (multiply monthly std by $\sqrt{12}$)?** If you divide the mean by the standard deviation from above, you get the Sharpe Ratio, which tells you *how much return you are getting per unit of risk*. **What is the Sharpe Ratio for the market?**

   Now do the same for the bonds from above. Make sure to use the bond *returns*. **Describe the relationship between maturity and these 3 values. Which maturity gives you the highest returns per unit of risk?**

6. Create a correlation table with market excess returns and bond returns. How does maturity affect correlation with market returns? Which bonds have the highest correlation with each other?

7. Now let's get into conditional distributions. Create a 5x5 bivariate histogram (which is an *empirical* distribution) for the 30 year bond and the market. Meaning, you will create 5 "buckets" for each variable, and then count the number of observations where each return is in each bucket. So if the market excess return is 0.1 and the bond return is -0.1, then that would count as one observation for the joint bucket (first tuple is for the market) [(0.08,0.11), (-0.11,-0.08)]. You should get counts that look something like this:

0	2	0	3	1
0	7	63	83	2
1	15	181	255	11
2	2	16	25	3
0	2	2	2	0

   If you divide this count table by the total number of observations, you will get a frequency table that will give you the *bivariate distribution* of 30 year bond returns and market returns.

   Using the relationship $PDF(Y|X=x) = \frac{PDF(Y=y,X=x)}{PDF(X=x)}$ find the conditional distributions of the 30 bond returns given returns of the market. Then find the conditional expectations of the 30 bond returns given returns on the market portfolio. **Explain what these numbers mean. When the stock market is doing well, how is the 30 bond performing? And when the market is doing poorly how does the 30 year bond perform? Is the bond a natural hedge for the stock portfolio?**

8. Now calculate a and b for the equation $y = a + bx + u$ where y is the 30 year bond return and x is the market excess return. Plot this line on the interval (-0.3, 0.3) and *on the same*

axes plot the conditional expectations of the 30 year bond return given market returns. **Describe what you see in the figure. How can we think about the regression line in this figure?**

9. Now calculate the regression line directly from the data using the formulas for the sample estimates of a and b that we saw in the slides. Plot this line on a scatter plot of market returns on the x-axis and corresponding 30 year bond returns on the y axis. **What is the R^2 of this model? Is the market return a good "explainer" of the bond return? Does this sample-driven regression line match the one we got from the conditional distribution? How does it differ and why?**