Yaowen Mei [1177855]

ymei@uofguelph.ca

FIN6000 fall21 MIDTERM PROJECT

Prediction of Asset Pricing with the CAPM Model

Table of Contents

[Table of Contents 1](#_Toc86271460)

[Chapter 1 Introduction to the purpose of the model 2](#_Toc86271461)

[1.1 To get the historical data 2](#_Toc86271462)

[1.2 To build the model and use it to predict the future 2](#_Toc86271463)

[1.3 To estimate the efficiency/accuracy of this model 2](#_Toc86271464)

[Chapter 2 Challenges encountered in the modeling process 3](#_Toc86271465)

[Chapter 3 Application and the limitations of the model 4](#_Toc86271466)

[Chapter 4 Modeling techniques learned in this assignment 5](#_Toc86271467)

[Chapter 5 Financial concepts explored in this assignment 6](#_Toc86271468)

[Chapter 6 Conclusion 7](#_Toc86271469)

[6.1 The stock market’s CAPM predicted excess return is different than the CRSP theoretical value 7](#_Toc86271470)

[6.2 The excess return of the US Fema-French Portfolios are well explained by the CAPM model comparing to the stocks. 7](#_Toc86271471)

[6.3 The percentage of variation on the excess return rate on the previous 5 year’ CAPM beta is relatively a constant regardless of asset types, and benchmark types. 9](#_Toc86271472)

[Bibliography 10](#_Toc86271473)

# Introduction to the purpose of the model

The purpose of this assignment is to 1) build a static Capital Asset Pricing Model (CAPM) based on historical data, 2) use this model to predict assets’ the expected monthly return for the following month, 3) then determine the efficiency/accuracy of this model. This whole project is based on Chapter 10 of Craig W.Holden’s book and excel template[1].

## To get the historical data

In order to build this CAPM model from the historical data, the monthly return rate data for 6 stocks (ABX, IBM, KEP, SI, TV, and YPF), 6 Fama-French US portfolios (Small-Growth, Small-Neutral, Small-Value, Big-Growth, Big-Neutral and Big-Value), and 6 Country Exchange-Trade Funds (Australia, Canada, Germany, Malaysia, Mexico, and Singapore) are collected and benchmarked to S&P500 Index (SPY), CRSP Index (CRSPTM1) and Dow Jones Global Index (W1DOW) for the time period between Oct, 2011 to Sept 2021.

## To build the model and use it to predict the future

In order to build the Static CAPM model, I followed the standard Fama-MacBeth two-pass regression method. The first stepis to conduct a timeseries regression (over each 5 years window) by fitting each assets’ risk premium to market index’s risk premium and calculate the averaged CAPM betas’ for each of the assets. The second step is to sub-group the assets by types, i.e. stocks, portfolios, ETFs, then do a cross-sectional regression by fitting the each individual’s risk premium values to the CAPM betas calculated from the 1st step. The average slope of this second pass regression is called CAPM Risk Premium, while the average intersection is called CAPM Risk Intercept.

Finally, putting the average first pass beta into the second regression function will give people the expected return rate for the next month.

## To estimate the efficiency/accuracy of this model

By calculating the regression coefficient, R2, One can estimate the efficiency/accuracy of this Static CAMP model. In order to do this, I first calculate the R2 for the first pass time series regression, and take average value over all the 5 year windows. Then, I calculate the R2 for the second pass cross-section regression, and take average over all the months. .

# Challenges encountered in the modeling process

There are several challenges I faced while I was doing this assignment.

Firstly, I have hard time to download the historical data for Barrick Gold (ABX), this used to be a NYSE stock, but changed to Toronto. Therefore, the historical data I got from Yahoo finance does not match what I have in the excel sheet. Eventually, my classmate Uday found this data from a website called [Inversting.com](https://in.investing.com/equities/barrick-gold-corp.-historical-data?end_date=1356978600&interval_sec=daily&st_date=1325356200&interval_sec=monthly), and this data is ONLY available for the India version of this website (not for USA version, Canada version, and not for Chinese version).

Secondly, I have hard time to get the Dow Jones World Index (W1DOW) from Yahoo finance, as there are only 107 months’ data available there, but my model requires 120 data points. Again, I found that the India version of Inversting.com website has the full data, so I eventually get the Dow Jones Index data from the India website. Without this India website, I really cannot do this assignment with just Yahoo Finance and Google Finance.

Thirdly, I have hard time to figure out the monthly return is calculated from the closing price between two contiguous months. Intuitively, my first thought was that monthly return should be calculated between the open price of the first trading day of a month and the closing price of the last trading day of a month, but I found my result does not match what I already have in the excel sheet. It took me a lot of time and effort to do the trail-and-error thing to figure out the proper time window to correctly calculate monthly return as well as how to properly add dividends into the monthly return. I am proud that I figure out the equation for monthly return rate independently, see Equation (3‑1), and I will remember this equation forever.

|  |  |
| --- | --- |
|  | **(3‑1)** |

Finally, It is time consuming to calculate all the monthly return data and add dividends into them, so I have created a Python Web Crawler with selenium library that can automatically download any stock’s historical price data and dividends data from Yahoo Finance website and covert them to a monthly return CSV file [[Github Link](https://github.com/y5mei/YahooFinanceCrawler)]. I hope this Yahoo Finance Crawler will help me in the future and make the final exam easier for me.

# Application and the limitations of the model

The application of this model is to predict each asset’s expected monthly return for the next month (Oct 2021). This is also called the Cost of Equity Capital. As can be seen in the table below (also in the line 393 of the attached excel sheet), with the past 10 years historical data, I expect the all the stocks are going to lose money by 0.3% except for the gold company, Barrick Gold, which is predicted to only reduce by about 0.2%.

Table 1 The expected monthly return rate for Oct-2021 based on three different market benchmarks

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **OCT-2021 Exp Return** | Barrick Gold (ABX) | IBM (IBM) | Korea Electric (KEP) | Siemens (SI) | Grupo Televisa (TV) | YPF  (YPF) |
| S & P 500 | -0.20% | -0.31% | -0.27% | -0.33% | -0.33% | -0.37% |
| CRSP | -0.25% | -0.31% | -0.28% | -0.31% | -0.31% | -0.33% |
| DJ Global | 0.07% | -0.20% | -0.10% | -0.28% | -0.31% | -0.52% |

This model can also be used to evaluate the percentage of the variation in each asset’s risk premium that can be described by the benchmark market index (in the line 478 of the attached excel sheet). This is done with the first pass timeseries regression. Finally, this model is also used to estimate how much monthly return variation in each subgroup of asset (i.e. stocks, portfolios, or ETFs) in the next month that can be described by the CAPM model’s beta. This is calculated from the time period of the previous 60 months (5 years).

In the contrast, there are also several limitations for this model. Firstly, this model sometimes will produce a negative average CAPM risk premium, and an average intercept value which is far away from zero (the theoretical intercept should always be zero), this big intercept and negative risk premium will lead to the fact that the model does not make any sense in terms of economics. If this happened, even though, this model produces some expected return rate with R2, one should note that this model has been already lost the ability to explain any economic behavior in these situations. The second limitation is the choice of benchmark index. I have only included 3 benchmark index, namely to be S&P500, CRSP, and Dow Jones Global Index, these benchmarks are good index to describe North America’s economic activities, but these might not correlate well with developing countries stock market (i.e. China, India, and Vietnam). These developing countries has been playing a role as raw material providers and manufactures for North America, so there might be a time lag between developing countries market index with developed countries market index. This model will not provide accurate estimations for the stock market for countries like China and India.

# Modeling techniques learned in this assignment

I have learned the simple linear regression, and subgroup samples to conduct ANOVA test from previous physics and engineering study, but this is the first time I have ever learned the timeseries linear regression and subgroup samples to do cross-section linear regression. By sliding the starting point of the 60-month-window, a lot of extra information can be obtained from the time series data than the simple linear regression method.

Theoretically, there are two steps involved in **Fama-MacBeth regression**. In the first pass, we try to fit each of the asset returns to the market index over every 60 months’ period via sliding window technique. For example, we have 6 stocks, the regression will be a group of Equation (5‑1) over all the available time slots. For each asset, , the beta exposure is determined by taking average of over all available ; the percentage of explained fluctuation in asset’s risk premium that can be described by the benchmark market index is determined by taking average of all the over all available t.

|  |  |
| --- | --- |
|  | (5‑1) |

In the second pass, by subgrouping all the stocks, portfolios, and ETFs into subgroups, we try to fit all the monthly excess returns to their estimated betas (which are calculated from the previous pass) for each available month. This will determine the averaged CAPM risk premium and its intersection. For example, with these 6 stocks, we will conduct regression with subgroup of stock based on Equation (5‑2) for all the available months. Eventually, the CAPM risk premium is the average value of over all the available t, and the CAPM intercept is the average value of all the over all the available t. The percentage of fluctuation in the excess monthly return rate that can be described by the previous 5 years’ CAPM betas is the determined by over all the available t.

|  |  |
| --- | --- |
|  | (5‑2) |

Finally, the expected following month return for stock, i, can be calculated by Equation (5‑3):

|  |  |
| --- | --- |
|  | (5‑3) |

# Financial concepts explored in this assignment

In this assignment, I have explored several very important financial concepts.

Firstly, I have obtained a deeper understanding of the formula of monthly return rate. As shown in Equation (3‑1), I have derived the equation for monthly return rate (with dividends) by myself. But still, I need to read more papers to understand how to calculate monthly return rate incase of stock split happens.

Secondly, I have built a solid understanding of CAPM model. The basic concept for CAPM model is simple, the individual stock’s excess return should be linear correlated with the market index’s excess return, and if a portfolio is diversified enough, all the intersection terms for each of the individual stocks will cancel each other (non-systematic risk can be hedged), and the correlation between the performance of the portfolio and the market index is the beta exposure of the portfolio (only systematic risk will be left). However, it is interesting to see how Fama and MacBeth went beyond this simple CAPM model, and presenting such rich information from the time series data by firstly subgroup the data into a series of 60 months’ time frames, and secondly subgroup the data by asset category.

Last but not the least, during the process of obtaining the last 10 years historical data, I have explored different financial websites, such as Yahoo Finance, Google Finance, and Investing. All these websites provide users slightly different information, and in order to get a full set of accurate historical data, one need to try his best to take advantage of all these websites. Additionally, I have explored different Dow Jone’s index, namely, to be the Dow Jones Global Index (W1DOW) and The Global Dow (GDOW), I found that the previous one has a very strong correlation with the US S&P 500 and CRSP, but the latter one behaviors slightly different than US S&P 500. In the future, I want to explore more about the definition of each of the most famous market index. As the market index is on the right hand side of the CAPM equation, it must be very critical for me to choose a proper benchmark index when doing any prediction work with the CAPM model.

# Conclusion

There are several interesting conclusions can be drawn from this project.

## The stock market’s CAPM predicted excess return is different than the CRSP theoretical value

I have found that using the previous 10 years data, and using the S&P500 index as market benchmark. The second pass Static CAPM model is providing an opposite prediction on the future month’s excess return on all the stocks except for the gold company (ABX) and YPF. The data are show in the Table 2 below, for ABX and YPF, both the CAPM and the CRSP theoretical value are predicting a negative excess return rate for month of Oct 2021; while, for the rest of the stocks, the CAPM model is predicting a negative excess return rate, while the CRSP theory value are all positive. The deviation between empirical prediction and theory prediction might raising from the huge volatility induced by COVID19, under situation like this, each stock as individual does not always follow the theorical model, while as I will explain in the next sub-section, by grouping the companies by size and book-to-market value (Fema French method), CAPM will have a very accurate prediction consistent with the CRSP theory value.

Table 2 The Expected Excess Return Rate of CAPM vs. CRSP theoretical value

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Oct 2021*  *Return* | *Barrick Gold*  *(ABX)* | *IBM*  *(IBM)* | *Korea Electric*  *(KEP)* | *Siemens*  *(SI)* | *Grupo Televisa*  *(TV)* | *YPF*  *(YPF)* |
| *CAPM* | *-0.25%* | *-0.31%* | *-0.28%* | *-0.32%* | *-0.31%* | *-0.34%* |
| *Theory* | *-0.20%* | *1.26%* | *0.59%* | *1.41%* | *1.37%* | *1.98%* |

## The excess return of the US Fema-French Portfolios are well explained by the CAPM model comparing to the stocks.

It is interesting to find out that CAPM prediction for the next month’ excess return rate on these 6 FF-Ports are very consistent with the CRSP theoretical value. The reason behind this fact is that by subgrouping all the companies in the market to the companies size and book to market value, people are capturing the underlying factors which are not included in the traditional CAMP model. This is also part of the reason why Fema is famous and got the Nobel Prize. As can be seen from Table 3, all these 6 F-F portfolios, both the CRSP and the CAPM models are predicting a excess return rate about 1.1%.

Table 3 The Expected Excess Return Rate of US-FF-Ports vs. CRSP Theoretical Values

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Oct 2021*  *Return* | *Small-Growth* | *Small-Neutral* | *Small-Value* | *Big- Growth* | *Big-Neutral* | *Big-Value* |
| *CAPM* | *1.28%* | *1.24%* | *1.24%* | *1.15%* | *1.16%* | *1.23%* |
| *Theory* | *1.53%* | *1.40%* | *1.39%* | *1.12%* | *1.14%`* | *1.37%* |

Chart, bar chart

Description automatically generated

Figure 1. The Empirical CAPM prediction of the Risk Prem and Intercept vs the CRSP theorical value. Note that a positive slop and a close to zero intercept makes this mathematical model make sense in the scope of economics..

## The percentage of variation on the excess return rate on the previous 5 year’ CAPM beta is relatively a constant regardless of asset types, and benchmark types.

Another interesting fact I found from this project is that regardless of how I choose the market portfolio benchmarks and the asset types, the percentage of fluctuation of the next month’s excess return explained by the previous 5 years’ CAPM betas is relatively a constant (cell B353 in the attached excel sheet). The value varies between 22.9% (use Dow Jones as benchmark to predict US FF Port) to 30.1% ( use CRSP as benchmark to predict US FF Port). This reflects the fact that generally, CAPM can explain about 30% of the excess rate, and economics need to work harder to in-rich the CAPM model by adding more factors inside and do more round of sub-grouping and regressions on the historical data to make CAPM to explain more of the future’s asset excess return.

Bibliography

[1] Holden, C. W., & Womack, K. L. (2000). Spreadsheet modeling in finance and investment courses. *Available at SSRN 241708*.