William Reed

Fama-French 3-Factor Model Extension

Executive Summary

One of the most important things to do when analyzing a company is understand how its returns are affected by various kinds of market risks. The relationship between a company's stock returns (in excess of the risk-free rate) and risks is the corner stone to coming up with a valuation for the company (e.g. discounted cash flows analysis). The prevailing method for calculating a stock's expected return is using the capital asset pricing model (CAPM), which only accounts for the market risk premium. More recently, the Fama-French 3-Factor model expanded on the CAPM by adding two other risk premiums observed in the market, the small-minus-big and high-minus-low premiums. My analysis here seeks to expand on the 3-Factor model by adding CPI inflation as a risk. To test my theory that inflation is a risk can affect stock returns, I will be analyzing Walmart's monthly returns from 2011 – 2024. These 168 months will be large enough to have substantially sized training and testing sets. This is the model to be fit:

$$\mathrm{Y} = \beta_0 + \beta_1 {*}(MRP) + \beta_2 {*}(SMB) + \beta_3 {*}(HML) + \beta_4 {*}(INF) + \varepsilon$$

In the above equation, the response variable is the excess market returns of a stock, and the predictors are the previously mentioned risk premiums. The model satisfied all of the regression assumptions when fit to the training set and there is no multicollinearity between the predictors. The model fit the training set well, given the highly volatile nature of stock returns. Better yet, the model had a lower RMSE on the testing set than on the training set, indicating that the model is not overfit.

In the way of statistical inference, the coefficient on INF is statistically significant at the 95% confidence level, meaning that the null hypothesis can be rejected. Inflation indeed has an adverse affect on WMT's returns in excess of the risk-free rate. Going forward, the 4-Factor model I created could be applied to other companies in other industries. Further, more macroeconomic risks could be investigated to see whether they affect excess stock returns.

Variable Selection

The CAPM seeks to understand the relationship between a stock's returns, and those of the market. To quantify this relationship, the stock's returns in excess of the risk-free rate are regressed on the market's returns (S&P 500) in excess of the risk-free rate (or "market risk premium"). The resulting beta coefficient quantifies the relationship that the stock has with the systematic risk of the market.

The Fama-French 3-Factor model expands on the CAPM by exploring the relationship that stock returns have with two other market risks: the "small-minus-big" (SMB) premium, which captures the riskier nature of small-cap stocks, and the “high-minus-low” (HML) premium, which captures the riskier nature of stocks with a high book-to-market ratio. The former of these is calculated as the spread between small-cap and large-cap stock returns, and the latter is calculated as the spread between stocks with high book to market ratios and those with low book to market ratios.

I want to expand on the 3-Factor model by investigating whether there are other systematic risks that affect stock returns. While there are countless risks that one could test, the most natural risk that could be added to the 3-Factor model is inflation (INF). The subject of my own 4-Factor model will be Walmart (WMT), as the company is characteristic of a very important industry (grocery stores) and exposed to many macroeconomic risks. While inflation is only one factor in my model, it is the one I am most interested in, as the other three have been widely accepted and used for quite some time. The 4-Factor model is as follows:

$$Y = \beta_0 + \beta_1*(MRP) + \beta_2*(SMB) + \beta_3*(HML) + \beta_4*(INF) + \varepsilon$$

The units of all variables are monthly percentage changes. Intuitively, inflation seems to be factor that would affect the stock returns of companies in many industries, especially grocery stores.

Fitting the Model

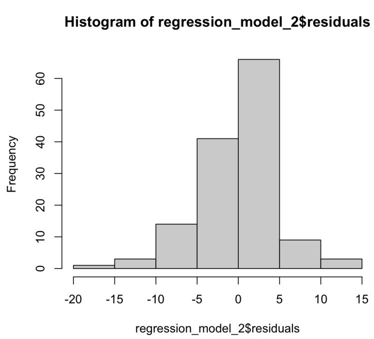
The model was fit to monthly data consisting of the market risk premium, small-minus-big premium, high-minus-low premium, CPI inflation data, and WMT returns in excess of the risk-free rate, from the years 2011 – 2024. The most recent 168 months will be enough to adequately train the model and even construct a testing set. The first three risk premiums were easily accessible on Eugene Fama’s website, the CPI inflation data was easily accessible from the Bureau of Labor Statistics, and WMT monthly returns were available on Stooq.com. I combined the data from these three sources using Excel.

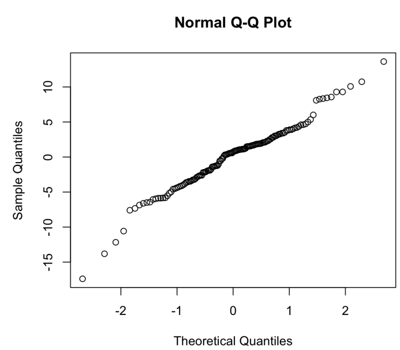
William Reed

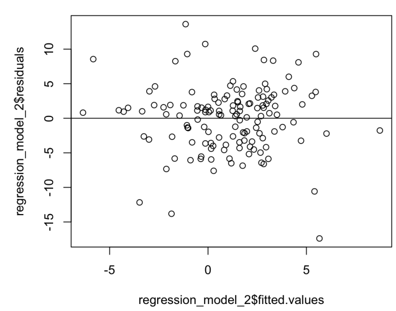
First, I split the data into the training and testing set. The 75% of the observations were chosen to form the training set, and the remaining 25% became the testing set. I then fit the 4-Factor model on the training set using ordinary least squares (OLS). Here are the results:

```
Call:
lm(formula = WMT_ExR ~ MRP + SMB + HML + INF, data = train)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.3851  -3.0744   0.7362   2.5642  13.6193 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.16556    0.48716   2.393  0.01814 *  
MRP          0.45744    0.09986   4.581 1.06e-05 ***
SMB         -0.44436    0.16244  -2.736  0.00709 ** 
HML         -0.24263    0.12443  -1.950  0.05331 .  
INF         -2.97877    1.19961  -2.483  0.01428 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.809 on 132 degrees of freedom
Multiple R-squared:  0.2155,	Adjusted R-squared:  0.1917 
F-statistic: 9.063 on 4 and 132 DF,  p-value: 1.687e-06
```

To assess the conditions for regression, I created a histogram of the residuals, a residual-fitted plot, and a normal Q-Q plot for the residuals.







In the histogram, the tails of the distribution taper rapidly, which is characteristic of the normal distribution. Looking at the Q-Q plot, the plotted line is nearly diagonal. The histogram and Q-Q plot suggest that the residuals are more-or-less normally distributed. Turning to the residual-fitted plot, the mean of the residuals is zero, and the variance is constant. Also, the residual-fitted plot does not reveal cone-like pattern but rather has all of the residuals lying in a relatively horizontal band, indicating that independence and linearity are satisfied. Thus, all of the regression conditions are satisfied (linearity, independence, and normally distributed residuals with mean of zero and constant variance).

Now to check for multicollinearity between the predictors. To do this, I created a correllation matrix of the four predictors:

```
             MRP         SMB         HML         INF
MRP  1.000000000  0.33264946 0.009884402 -0.11308636
SMB  0.332649461  1.00000000 0.119227367 -0.09073878
HML  0.009884402  0.11922737 1.000000000  0.05944898
INF -0.113086361 -0.09073878 0.059448985  1.00000000
```

None of the correllation coefficients deviate much from zero, so multicollinearity is not a concern with this model.

Assessing Model Fit

As can be seen in the summary print-out from earlier, the R-squared value is 0.2155, adjusted R-squared is 0.1917, and RMSE is 4.809%. While this would suggest a fairly good fit given the volatility of stock returns (standard deviation of WMT's excess returns is 5.35%), it is possible that the model is overfit to the data. So, to test for overfit, the fitted model was used to predict the response variable in the testing set, yielding a RMSE of 3.53%. The fitted model did an even better job predicting the testing set than the training set, indicating that it was not overfit.

Statisitcal Inference

Turning back to the original fitted model summary and the primary quesiton of if inflation affects WMT's returns in excess of the risk-free rate. The training set on which the model was fit is a randomly selected, representative sample of WMT's monthly returns, meaning that statistical inference to the population can be made. The coefficient for INF is -2.979, and the associated p-value is 0.0143, meaning that it is statisically significant at the 95% confidence level. So, we can reject the null hypothesis that inflation does not affect WMT's returns in excess of the risk-free rate.

Now to use the model to make a prediction. Say that one month, the market risk premium is 4.3%, the small-minus-big premium is -2.4%, the high-minus-low premium is 1.2%, and inflation is 0.7%. The model predicts that the expected excess returns of WMT in this month is 1.82%. The 95% confidence interval for such a month is [-0.065%, 3.71%], meaning that the expected excess returns of WMT will fall in this range with 95% certainty.

William Reed

Conclusion

As it turns out, adding monthly inflation to the long-standing 3-Factor model proved to be successful in helping predict Walmart's excess returns. As previously discussed, the coefficient for INF was statistically significant at the 95% confidence level. In practical application, this model could be used to calculate Walmert's cost of equity, which analysts could then use to value the firm.

Extensions of this project would be looking at other companies to see if inflation affects their monthly excess returns. If this analysis is completed with enough companies in enough induistries, one could determine which sectors in the stock market are most exposed to inflation risk. In addition, factors other than inflation could be considered, ultimately leading to more accurate models that better capture the types of that affect stock returns.

Code Appendix

```
library(car)

data <- read.csv('regression_data.csv')

data$WMT_ExR = data$WMT_R - data$RF

set.seed(110)


subset <- sample(c(TRUE, FALSE), nrow(data), replace=TRUE, prob=c(0.75, 0.25))

train <- data[subset, ]
test <- data[!subset, ]

regression_model_2 <- lm(WMT_ExR ~ MRP + SMB + HML + INF, data = train)

summary(regression_model_2)

plot(regression_model_2$fitted.values, regression_model_2$residuals)
abline(h=0)
qqnorm(regression_model_2$residuals)
hist(regression_model_2$residuals)

cor_matrix <- cor(train[, c("MRP", "SMB", "HML", "INF")])
cor_matrix

sd(train$WMT_ExR)

sqrt(mean((test$WMT_ExR - predict(regression_model_2, test))^2))

newdata <- data.frame(MRP = c(4.3), SMB = c(-2.4), HML = c(1.2), INF = c(0.7))

predict(regression_model_2, newdata = newdata, interval = "confidence")
```