Capital Asset Pricing Models

Oliver Bandosz

Abstract:

An introduction to Linear Regression application in finance. The Capital Asset Pricing Model (CAPM) is use to assess the relationship between the market performance and the asset. Using linear regression, we evaluate the sensitivity of stocks to market movement, β, and analyse the impact of a macroeconomic factors, such as oil price fluctuations, by extending the CAPM model to include oil returns.

Contents

1 Introduction

Captial Asset Pricing Models (CAPM) is a widely used financial model that estimates expected return on an asset based on the risk involved relative to the broader market. CAPM provides a framework for understanding how assets should be priced in equilibrium given their risks profiles, and it was developed by Sharpe and Lintner.

CAPM is based on the assumption that the expected return of an asset is proportional to its systematic risk, which cannot be diversified away. Systematic risk is the portion of risk that is inherited by the entire market, this risk cannot be eliminated by diversification, unlike unsystematic risk. CAPM focuses only on systematic risk and assumes investors should be compensated for this type of risk alone.

In such a market, the CAPM is based on the following equation:

$$\mathbb{E}(R_i) = R_f + \beta_i \cdot (\mathbb{E}(R_m) - R_f), \tag{1}$$

where, $\mathbb{E}(R_i)$ is the expected return on asset i, R_f is the risk-free rate, and β_i is the asset's sensitivity to market movements. $(\mathbb{E}(R_i) - R_f)$ is the market risk premium, or the excess return expected for investing in the asset rather than risk-free assets. CAPM builds the foundations asset valuation in portfolio theory. β represents the systematic risk in the model, such that when β is greater than 1 the asset is considered more volatile than the market, compared to an asset with a β less than 1.

The CAPM model relies on several assumptions. Although these assumptions simplify real-world dynamics, they are unrealistic, and are a limitation to the model. The first of which is that the CAPM model assumes that investors hold a well-diversified portfolios, meaning that they are only affected by systematic risk. Additionally, the CAPM model assumes all investors have the same standardised time horizon, typically one year. This standardisation allows for comparing the relationship of returns across different investments.

CAPM models also assume perfect capital markets, where there are no taxes or transaction costs, all investors are rational, information is freely available to all investors, as well as liquid markets ensuring efficient pricing. Finally, CAPM assumes that an investor can borrow and lend at the risk-free rate which is not realistic. This discrepancy often results in a shallower slope for the Security Market Line (SML) in practice, compared to a steep theoretical slope assumed by CAPM.

2 Linear Regression and CAPM

2.1 Linear Regression

Linear regression is a technique to model and extract relationships between one or more independent variables (predictors) and dependent variables (response). The formula for a simple linear regression is:

$$y = \beta_0 + \beta_1 x + \epsilon,$$

where β_0 is the intercept and β_1 is the slope, representing the relationship between x and y. x is the independent variable and y is the dependent variable, with ϵ as the residual. Later, when we want to extend our CAPM model to include oil prices we will conduct a multiple linear regression. The formula for a multiple linear regression is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon,$$

where y is the dependent variable, $x_1, ..., x_k$ are the independent variables, $\beta_0, ..., \beta_k$ are the coefficients associated with each independent variable. ϵ is the residuals.In the context of CAPM we use simple linear regression to estimate the relationship between a stock's excess returns and the market's excess return, and apply multiple linear regression when including additional factors, such as macro-economic factors.

2.2 Ordinary Least Squares (OLS) in Linear Regression

The Ordinary Least Squares (OLS) method is used for estimating the coefficients associated with each independent variable. The objective of OLS is to find the line (or hyperplane in multiple regression) that minimises the sum of the squared residuals. The least squares criterion: for $\underline{\boldsymbol{\beta}} = (\beta_1, \beta_2, ..., \beta_p)^T$, define

$$\mathcal{Q}(\underline{\boldsymbol{\beta}}) = \sum_{i=1}^{N} [y_i - \hat{y}_i]^2.$$

Ordinary Least Squares estimate, $\hat{\beta}$, minimises $\mathcal{Q}(\underline{\boldsymbol{\beta}})$. By taking $\underline{\hat{\boldsymbol{y}}} = (\hat{y_1}, ..., \hat{y_n})^T = \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}}$, where

$$\underline{\boldsymbol{X}} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{pmatrix},$$

then we get the following:

$$\begin{aligned} \mathcal{Q}(\underline{\boldsymbol{\beta}}) &= \sum_{i=1}^{N} [y_i - \hat{y}_i]^2 = (\underline{\boldsymbol{y}} - \underline{\hat{\boldsymbol{y}}})^T (\underline{\boldsymbol{y}} - \underline{\hat{\boldsymbol{y}}}), \\ &= (\underline{\boldsymbol{y}} - \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}})^T (\underline{\boldsymbol{y}} - \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}}), \end{aligned}$$

To minimise $\mathcal{Q}$, we need to solve for $\hat{\beta}$ the following equation: $\frac{\partial \mathcal{Q}}{\partial \beta_j} = 0$, for $j = 1, 2, ..., p$.

$$\frac{\partial \mathcal{Q}(\beta)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \left(y_i - \sum_{k=1}^{p} x_{i,k}\beta_k \right)^2,$$

$$\frac{\partial \mathcal{Q}(\beta)}{\partial \beta_j} = -2 \sum_{i=1}^{n} x_{i,j} \left(y_i - \sum_{k=1}^{p} x_{i,k}\beta_k \right),$$

$$\frac{\partial \mathcal{Q}(\underline{\boldsymbol{\beta}})}{\partial \beta_j} = -2\underline{\boldsymbol{X}}_{[j]}^T (y - \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}}),$$

where $\underline{\boldsymbol{X}}_{[j]}$ is the j$^{\text{th}}$ column of $\underline{\boldsymbol{X}}$. Therefore,

$$\frac{\partial \mathcal{Q}(\underline{\boldsymbol{\beta}})}{\partial \underline{\boldsymbol{\beta}}} = -2\underline{\boldsymbol{X}}^T(y - \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}}).$$

So the OLS estimate must satisfy the normal equations:

$$\underline{\boldsymbol{X}}^T(\underline{\boldsymbol{y}} - \underline{\boldsymbol{X}}\underline{\boldsymbol{\beta}}) = 0,$$

$$\iff \underline{\boldsymbol{X}}^T\underline{\boldsymbol{X}}\underline{\hat{\boldsymbol{\beta}}} = \underline{\boldsymbol{X}}^T\underline{\boldsymbol{y}}.$$

The OLS estimate is:

$$\underline{\hat{\boldsymbol{\beta}}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = (\underline{\boldsymbol{X}}^T\underline{\boldsymbol{X}})^{-1}\underline{\boldsymbol{X}}^T\underline{\boldsymbol{y}}.$$

The fitted values $\underline{\hat{\boldsymbol{y}}}$, which are the predicted values based on the OLS estimate $\underline{\hat{\boldsymbol{\beta}}}$, are given by:

$$\underline{\hat{\boldsymbol{y}}} = \underline{\boldsymbol{X}}\underline{\hat{\boldsymbol{\beta}}} = \underline{\boldsymbol{X}}(\underline{\boldsymbol{X}}^T\underline{\boldsymbol{X}})^{-1}\underline{\boldsymbol{X}}^T\underline{\boldsymbol{y}} = \underline{\boldsymbol{H}}\underline{\boldsymbol{y}},$$

where the hat matrix is a projection matrix. The hat matrix projects $\Re^n$ onto the column space of $\underline{\boldsymbol{X}}$
. The residuals are,

$$\underline{\hat{\boldsymbol{\epsilon}}} = \begin{pmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_2 \\ \vdots \\ \hat{\epsilon}_p \end{pmatrix} = \underline{\boldsymbol{y}} - \underline{\hat{\boldsymbol{y}}} = (\boldsymbol{I} - \underline{\boldsymbol{H}})\underline{\boldsymbol{y}}$$

<u>Note:</u> The least-squares residuals vector $\underline{\hat{\boldsymbol{\epsilon}}}$ is orthogonal to the column space of $\underline{\boldsymbol{X}}$.

<u>Note:</u> For $\underline{\hat{\boldsymbol{\beta}}}$ to exist uniquely, $(\underline{\boldsymbol{X}}^T\underline{\boldsymbol{X}})$ must be invertible $\iff$ $\underline{\boldsymbol{X}}$ has a full column rank.

2.3 Linear Regression for CAPM

CAPM can be expressed as a simple linear regression, where the dependent variable is the excess return of a stock over the risk-free rate, and the independent variable is the market's excess return over the risk-free rate. The CAPM regression equation is:

$$R_i - R_f = \alpha + \beta(R_m - R_f) + \epsilon,$$

where α is the intercept of regression, in this case α represents the out-performance relative to the market (or adbnoraml returns).

3 CAPM Modelling

In this section, we will use Python to model the relationship between a stock's returns and the broader market using capital asset pricing model. The goal is to estimate the beta's of two stocks, American Electric Power (AEP) and Bank of America (BAC), and assess how sensitive these stocks are to market movements. To start, we will analyse AEP using the basic CAPM model, before extending the analyse by including oil price fluctuations. Finally, we will compare the results for AEP and BAC to better understand how the two stocks behave under different market conditions. We will be using statsmodels to program the ordinary least squares, and will be using scikit-learn to program the linear regression model. We will model the risk-free rate using the 3 month Treasury bills (data used from FRED). Additionally we will model the market as the closing price of the S&P 500 index (data used from yfinance). All data used will be from 01/01/2010 to 01/01/2020.

3.1 CAPM for American Electric Power (AEP)

The regression model for CAPM for American Electric Power (AEP) is,

$$R_{AEP} - R_f = \alpha + \beta(R_{SP500} - R_f) + \epsilon.$$

From the set time frame, we get the $\beta = 0.6815$ and an $\alpha = -0.0016$ (4 s.f.). The beta indicates that AEP is less volatile than the market and move less to changes in the market.

3.2 Extending CAPM to Include Macro-Economic Factors

As mentioned, we will implement crude oil prices over the time frame into the model. The new regression model is defined as,

$$R_{AEP} - R_f = \alpha + \beta_1(R_{SP500} - R_f) + \beta_2(\text{Oil Price Change}) + \epsilon.$$

The extended CAPM results for AEP revealed two important findings. First, the beta for the market remained positive at a 0.7095, confirming that AEP's returns are positively correlated with the market. Secondly, the coefficient for oil price returns was -0.0678, indicating a negative relationship between oil prices and AEP's returns. Consequently, this suggests that increasing oil prices negatively affect AEP stocks. This is common for energy firm, as rising oil prices increases operational costs which squeezes profits if unable to push onto the consumer.

In addition to estimating the coefficients, we can conduct regression diagnostics on the model to check the assumptions of ordinary least squares are not violated. This helps confirm the validity and reliability of the regression results. We can check for patterns that may indicate problems such as heteroscedasticity (non-linearity), or influential points by analysing the residuals and leverages.

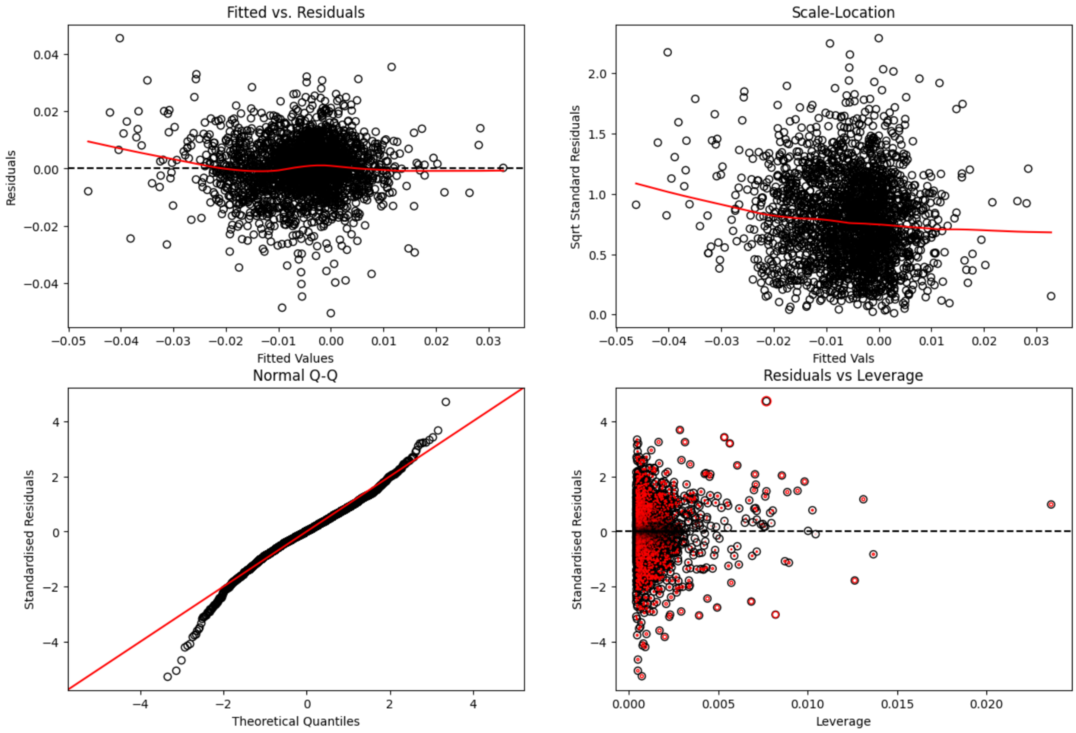

Figure 1: Regression Diagnostics

Starting with the fitted vs residuals plot. This shows the residuals against the fitted values. In this plot, the LOWESS smoothing line stays close to the horizontal asymptote at 0, although in the negative fitted values, we see a slight spread from this horizontal line which indicated some non-linearity behavour.

Next, the scale-location plot analyses the spread of the residuals. The line slopes slightly in the negative fitted values like the fitted vs residuals plot. This indicates that the slight spread is caused from heteroscedasticity. Although, this does not cause major concerns in the model, but a relatively flat line would confirm that the residuals are spread evenly across the fitted values.

The Normal Q-Q plot further assess the distribution of residuals. The residuals mostly stay close to the 45-degree line, except for some deviation in the tails, where the residuals breakoff from the line. This suggests that a majority of the residuals follow a normal distribution, although there are some outliers at both ends of the distribution, which could affect the interpretation of the model.

Finally, the residuals vs leverage plot helps identify any influential points. Several high leverage points were detected by Cook's distance, which shows points that have a disproportionate impact on the regression coefficients. This suggests data points are exerting more influence on the model than others.

Overall, we can conclude that the model meets the necessary assumptions of OLS, with minor concerns about normality of residuals at the tails and presence of high-leverage points.

3.3 Comparing the BAC's and AEP's CAPM Models

By implementing the same methods used for AEP's CAPM model, we can create another model for Bank of America (BAC). After analysing the data, we observe significant differences in how AEP and BAC respond to market conditions. AEP is a more defensive stock with a lower beta, making it less sensitive to market volatility. However, it is negatively impacted by rising oil prices, which is typical for the utility sector since this increases operational costs. BAC, on the other hand exhibits higher market sensitivity with a beta 1.316 (4 s.f.), which means for 1% moves in the market BAC will respond with a 1.316% move. Additionally, BAC also benefits from rising oil prices. Finally, the R-Squared on AEP was lower than BAC's R-Squared. This indicates that the market movements and oil returns have a stronger predictive power for BAC compared to AEP.