A Statistical Explanation of the Capital Asset Pricing Model

Ronak Mehta

---

1 Introduction

The Capital Asset Pricing Model (CAPM) model is a widely adopted model for risk behavior of financial securities. This note presents the components of the model in the language of statistics - specifically linear regression. The first three sections provide a very brief probability and statistics recap, and can be skipped for anyone with the background. The fourth and fifth section present the model both mathematically and visually.

2 Random Variables

In order to denote ideas such as risk, we use random variables from Probability Theory. The following is a short introduction - you should consult the project chairs or the **Further Reading and Materials** section below for more resources to study. First, we loosely define an *experiment* and a *sample space.*

We will define **experiment** as a procedure that will randomly produce an outcome, and can be repeated. Similarly, we define the **sample space** Ω of an experiment to be the set of all possible outcomes of that experiment. (You should be comfortable with the notion of a *set* in mathematics.) For example, in the experiment conducted by rolling a 4-sided die, $\Omega = \{1, 2, 3, 4\}$, as these are the four numbers that could appear on the face of the die. In the experiment conducted by rolling two dice, however, $\Omega = \{(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)\}$, as which die rolled which number is taken into account. (Check: What is the sample space of flipping 4 coins?)

Moving on, a **random variable** X, is a function (you should be comfortable with those too) that maps from an element of the sample space (i.e. an outcome) to a real number in $\mathbb{R}$. In the case of the dice roll, a natural random variable would be the number on the face of the die, which is just a direct map from the outcome to the real number - for instance, if a 3 is rolled, the value of the RV is 3. In the two dice case, we could have $X =$ the sum of both dice values. Then, the outcome (1,2) from Ω corresponds to $X = 3$ in $\mathbb{R}$. (Note that the outcome (2,1) maps to 3 as well.) For the four coin flip experiment, we could have the random variable $X =$ the number of heads. Then, the outcomes HHTT, HTHT, HTTH, THHT, THTH, and TTHH all map to $X = 2$. (Check: How many outcomes map to $X = n$ in the four coin flip experiment?)

3 Properties of Random Variables

While the random variable function itself is deterministic (given an outcome, there is exactly on answer for the real number described by X), the outcome of the experiment is random. Therefore, before a given experiment is conducted, we cannot know exactly what the value of the random variable will be. However, we can still understand the nature of the numbers that will come out -

whether they are centered around a certain number, how widely the possible numbers are spread, what numbers are more likely that others, etc. This is where probability comes into play. The **probability function** P of an experiment is a function that maps from outcomes in the sample space to a real number in the interval [0,1], with a few properties.

(a) Let ω_i be a particular outcome of the experiment. Then it holds that $\sum_{i=1}^{|\Omega|} \mathrm{P}(\omega_i) = 1$. That is, the sum of the probabilities of each outcome in the experiment is 1.

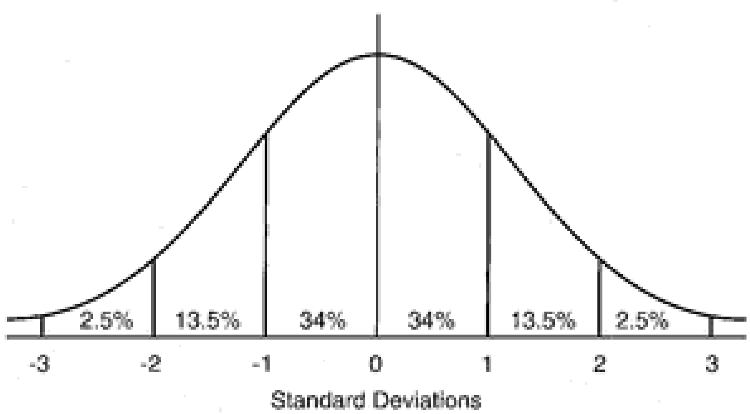
(b) Let A be a subset of the sample space, or $A = \{\omega_1, \omega_2, ..., \omega_n\}$, with $n \leq |\Omega|$. We can also call A an event. If this is the case, then $\mathrm{P}(A) = \sum_{j=1}^{n} \mathrm{P}(\omega_j)$. In other words, the probability of an event is the sum of all of the outcomes that could constitute that event (at least in the discrete case).

Using the probability function, we can assign probabilities to values of random variables. For example, in the two dice roll case, we have $\mathrm{P}(X = 3) = \mathrm{P}((1,2)) + \mathrm{P}((2,1)) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18}$. Here, the event is $X = 3$, and it contains all the outcomes in which X, the sum of the rolls, would equal 3.

Now, we look at the centrality of a random variable. For this, we define the **expected value** or **mean** function E. For the RV X, we have:

$$\mathrm{E}[X] = \mu = \sum_x x \cdot \mathrm{P}(X = x)$$

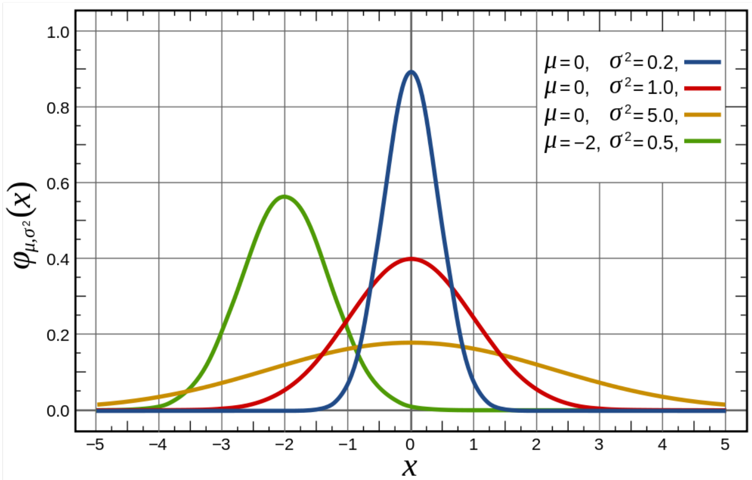
While this may not be clear at first, this is just the average of the values that the RV can take on, weighted by the probability that it takes on those values. Brief note: random variables can be continuous as well - think of the experiment recording the pH level of a subject's blood. Disregarding measurement precision, this could take on decimal values, therefore infinite an infinite number of outcomes. Here, we have a probability function that is continuous as well, where individual outcomes do not have probability, but intervals of RV values do. Consider a bell curve for a random variable:



Here, the RV is centered at 0, with regions of probability (listed in percent). With continuous function for probability $f(x)$, our rules for probability change slightly, namely $\sum_{i=1}^{|\Omega|} \mathrm{P}(\omega_i) = 1$ becomes $\int_x f(x)dx = 1$. Additionally, the expected value function becomes:

$$\mathrm{E}[X] = \mu = \int_x x \cdot f(x)dx$$

We know now that the expected value describes the average or central value of a random variable - but how far does it spread from the the central value? Take a look at the following bell curves: they are centered at the same spot, but spread differently.



Above, μ represents the expected value of X, while σ^2 represents our soon-to-be-defined measure of spread. The **variance** of a random variable X is:

$$\mathrm{Var}[X] = \sigma^2 = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

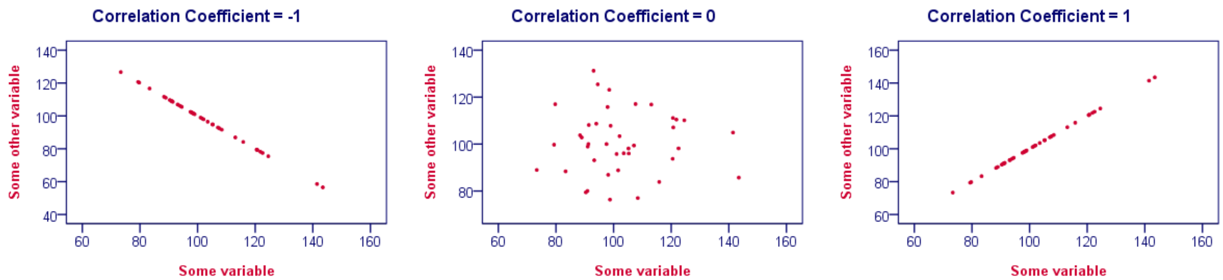
$$\mathrm{Var}[X] = \sigma^2 = \int_x (x - \mathrm{E}[X])^2 \cdot f(x)dx$$

Intuitively, the variance is the average of the squared deviations between the values of the random variable and its mean. RVs with high variance will have a large spread and be harder to predict. Similarly, those with small variance will center closely around their mean. The square-root of the variance is known as the **standard deviation**. The last two concepts we will define are **covariance** and **correlation** between random variables. The covariance describes linear dependence or linear proportionality between RVs. That is, if you plotted values of both RVs together, how much sense would it make to draw a line between them? RVs that are highly covariate or correlated will have some effect on each other - they will either grow proportionally or inversely. More on this in the next section, but understand the definitions below.

$$\mathrm{Cov}[X, Y] = \gamma = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

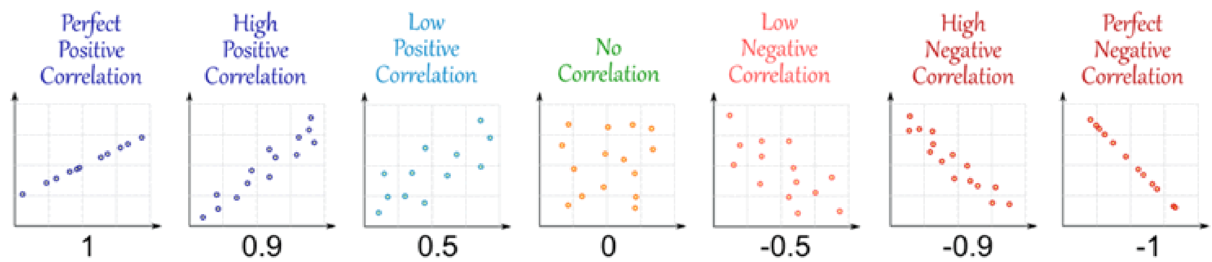
$$\mathrm{Corr}[X, Y] = \rho = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}}$$

The correlation, ρ, is a scaled version of the covariance, and will always be between 1 and -1. For a ρ value close to 1, the variables are strongly positively correlated, for -1 they are negatively correlated, and for 0 they are not correlated. See example below:

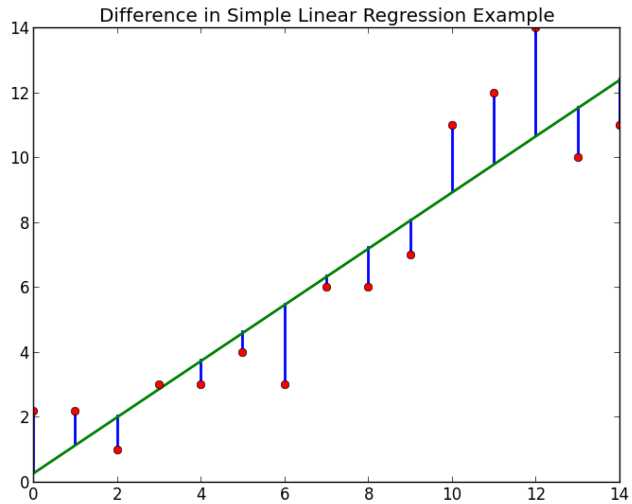


4 Linear Regression

Using these tools, we can tackle a very common problem - given a dataset of observations from two random variables, which we see to be strongly correlated, how can we draw a line of best fit through their plot? Can we use observations from one RV to predict the other RV? Observe another set of scatterplots below.



For the edge plots, we might want to find a linear function of the x-axis variable and use it to predict the y-axis. Let us use the function $y_i = mx_i + b + \epsilon_i$, where x_i and y_i are a particular predictor and response data point, m and b are constants, and ϵ_i is the observation of a random variable. Pay particular attention to the random variable ϵ - this RV has mean-zero, or $\mathrm{E}[\epsilon] = 0$. Note that because y is a function of a random variable, it also is a random variable. Adding the random term to the model accounts for the fact that some observations will be above the line, and some will be below. The difference in the y-direction between the line and the data point is the realized value for ϵ for data point i, or written compactly as ϵ_i. These deviations are made explicit in the diagram below.



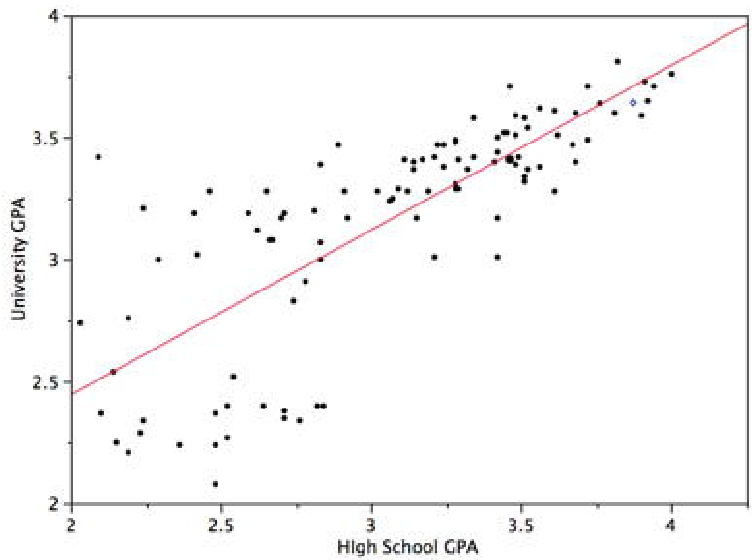
Suppose we are given a new x and are asked to predict its corresponding y value. Naturally, we would use the expected value of y. Because expected value affects only random variables, not constants, we can pull them out of the function (you can check this property with the integral above). The "|" sign means "given" in probability notation.

$$\begin{aligned} \mathrm{E}[y|x] &= \mathrm{E}[mx + b + \epsilon] \\ &= mx + b + \mathrm{E}[\epsilon] \\ &= mx + b \end{aligned}$$

This works, because the mean of ϵ is zero. Now, we need values for m and b. We provide both formulas below, but you will be responsible for just that of m (any of these formulas work).

$$\begin{aligned} m &= \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[X]} \\ m &= \mathrm{Corr}[X,Y] \cdot \sqrt{\frac{\mathrm{Var}[Y]}{\mathrm{Var}[X]}} \\ m &= \mathrm{Corr}[X,Y] \cdot \frac{\mathrm{St.Dev}[Y]}{\mathrm{St.Dev}[X]} \\ b &= \mathrm{E}[Y] - m \cdot \mathrm{E}[X] \end{aligned}$$

Using these values, we can capture the proportionality between observations of two RVs, and plot a line of best fit, as seen below:



5 Beta and the Capital Asset Pricing Model (CAPM)

We will now present two different, but closely related equations. Understanding the subtleties between them will be the key to understanding CAPM.

$$r_{i,t} - r_{\text{rf}} = \beta_i(r_{\text{market},t} - r_{\text{rf}}) + \alpha_i + \epsilon_{i,t} \tag{1}$$
$$r_i = \beta_i(r_{\text{market}} - r_{\text{rf}}) + r_{\text{rf}} + \alpha_i \tag{2}$$

With the following labels:

$$\begin{aligned}
r_{i,t} &= \text{the return on asset } i \text{ at time } t \\
r_i &= \text{the average return on asset } i \text{ across time} \\
\epsilon_{i,t} &= \text{the unsystematic risk of asset } i \text{ at time } t \\
r_{\text{rf}} &= \text{the risk-free rate} \\
r_{\text{market,t}} &= \text{the return on the market at time } t \\
r_{\text{market}} &= \text{the average return on the market across time} \\
\alpha_i &= \text{the abnormal return on asset } i \\
\beta_i &= \text{the systematic risk of asset } i
\end{aligned}$$

The first equation is the model used to calculate β, whereas the second is the statement of CAPM. They are very similar, but notice the differences. In the first equation, the independent variable is $(r_{\text{market},t} - r_{\text{rf}})$. Note there there is no i in this expression - the subscripts only contain t. This means that different observations of the independent variable will correspond to different observations in time. In the second equation, the independent variable is β_i. This expression (and the entire equation) does not have a t subscript, only i. This means that different observations of the independent variable here will refer to different companies (or assets). We will graph both of these equations in the next section. In the first graph, different data points will represent different points in time, whereas in the second, different data points will represent different companies. Pay attention to the notation and be crystal clear about this! Other differences include the risk-free rate being moved to the left-hand side in the first equation, and the existence of the $\epsilon_{i,t}$ term above. The other extremely important subtlety between the equations is the α_i term. In the first equation, because different data points depend on t, α_i is a constant, and NOT a random variable ($\epsilon_{i,t}$ is the RV). However, in the second equation, different observations *do* depend on i, so α_i is a random variable!

As you can imagine, we will use linear regression on both of the equations to calculate certain descriptive statistics about the the market and particular assets. Don't worry about the financial meaning of each of these terms yet - that will be in the next section. Just try to understand their statistical context: both equations are linear models in the form $y = mx + b + \epsilon$, ϵ being the RV. Identify which term is which in the model. For the first equation:

$$\begin{aligned}
r_{i,t} - r_{\text{rf}} &= \beta_i(r_{\text{market},t} - r_{\text{rf}}) + \alpha_i + \epsilon_{i,t} \\
y_1 &= m_1 x_1 + b_1 + \epsilon_1 \\
y_1 &= r_{i,t} - r_{\text{rf}} \\
x_1 &= r_{\text{market},t} - r_{\text{rf}} \\
m_1 &= \beta_i \\
b_1 &= \alpha_i \\
\epsilon_1 &= \epsilon_{i,t}
\end{aligned}$$

For the second equation:

$$\begin{aligned}
r_i &= \beta_i(r_{\text{market}} - r_{\text{rf}}) + r_{\text{rf}} + \alpha_i \\
y_2 &= m_2 x_2 + b_2 + \epsilon_2 \\
y_2 &= r_i \\
x_2 &= \beta_i \\
m_2 &= r_{\text{market}} - r_{\text{rf}} \\
b_2 &= r_{\text{rf}} \\
\epsilon_2 &= \alpha_i
\end{aligned}$$

Now, assume that you had a dataset of the risk-adjusted return on an asset i and return on the market for different points in time. How could you calculate the β_i of that asset? As seen from the first equation, you can use linear regression on those data points, and the slope will be β_i! Using the formulas from the **Linear Regression** section, we now have a formula for β_i:

$$\begin{aligned}
\beta_i &= \frac{\text{Cov}[r_{i,t} - r_{\text{rf}}, r_{\text{market},t} - r_{\text{rf}}]}{\text{Var}[r_{\text{market},t} - r_{\text{rf}}]} \\
&= \frac{\text{Cov}[r_{i,t}, r_{\text{market},t}]}{\text{Var}[r_{\text{market},t}]}
\end{aligned}$$

As seen above, we can actually take out the risk-free rate term from the covariance function, as it is a constant (moving a random variable by a constant would affect its mean, but not its spread). Similarly, given the β values and their expected annual returns for many different assets, we can estimate the return on the market using linear regression on the second equation.

$$r_{\text{market}} - r_{\text{rf}} = \frac{\text{Cov}[r_i, \beta_i]}{\text{Var}[\beta_i]}$$
$$r_{\text{market}} = \frac{\text{Cov}[r_i, \beta_i]}{\text{Var}[\beta_i]} + r_{\text{rf}}$$

In the next section, you will find these values plotted, and understand the financial implications each of these terms.

6 Graphical and Qualitative Interpretation

Now that you understand the linear model, we will go into the meaning of each of the above terms. We already know return to represent the percent change of an asset or index over a period of time. We have $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$, where p_t is the price at time t. Assume that we are always talking about daily returns, such that $t = 4$ represent the fourth day.

It is not unreasonable to think that there is a connection between risk and return - in many cases we speak of them interchangeably. A certain amount of risk needs to be incurred in order to expect a certain amount of return. When we adjust for the risk-free rate, risk and return essentially become the same quantity. So how can we measure and relate these values? Let us first dissect the first equation from the previous section:

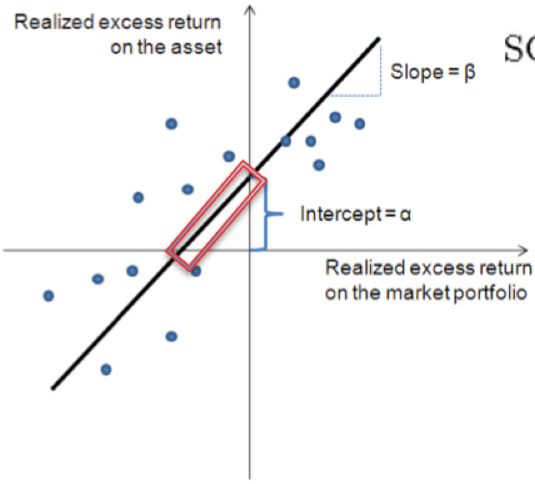
$$r_{i,t} - r_{\text{rf}} = \beta_i(r_{\text{market},t} - r_{\text{rf}}) + \alpha_i + \epsilon_{i,t}$$

Once again, this equation holds for a particular asset i. For that particular asset, there is a risk associated with the asset that is dependent on the performance of the market, and a component which is independent of the market. That which is dependent on the market is called **systematic risk** and is denoted by β (β_i meaning the systematic risk value for asset i). We multiply the adjusted return on the market by β to get the return on the asset. Naturally, we can use the β value as a *measure* of risk in comparison to the market.

$$\beta < 0 \implies \text{When the market does well, people buy less of this (i.e. gold).}$$
$$0 \leq \beta \leq 1 \implies \text{The asset has little correlation to market performance (i.e. alcohol).}$$
$$\beta > 1 \implies \text{Asset return depends heavily on performance of market.}$$

Similarly, there is an unsystematic risk component which is is attributed to random fluctuations in time. This is $\epsilon_{i,t}$, or the unsystematic risk of asset i at time t. We assume that $\text{E}_t[\epsilon_{i,t}] = 0$ and $\text{E}_i[\epsilon_{i,t}] = 0$. Note that here we took the expected value first with respect to time, meaning that the average is across time, and then with respect to companies. The first assumption justifies our using this model to calculate β with linear regression. The second assumption, that the average unsystematic risk at a particular time t averaged across companies is 0, meaning that we can be offset unsystematic risk by diversifying a portfolio with different companies. (While this may be intuitive, this gives mathematical justification for diversification of a portfolio to reduce risk.) The last value in the equation is the α_i term which refers to the abnormal rate of return for an asset. Note that this does not depend on time - it is just an inherent quality of a company, its status, its management, etc., that give it more or less than 0 extra return, other than that from the economy. Here, a similar assumption holds - that $\text{E}_i[\alpha_i] = 0$. You should keep in mind that $\text{E}_t[\alpha_i] = \alpha_i$

- within a reasonable period of time, α_i is just a constant. When using the above equation, we assume everything is known besides β_i and α_i, which can be calculated with linear regression. We now plot the equation:



$$\text{SCL} : R_{i,t} - R_f = \alpha_i + \beta_i \left(R_{M,t} - R_f\right) + \epsilon_{i,t}$$

Figure 1

α_i is the asset's alpha (abnormal return)
β_i - volatility compared to & correlated with market's
$\beta_i(R_{M,t} - R_f)$ - systematic risk
$\varepsilon_{i,t}$ - non-systematic risk
$R_{M,t}$ - market risk
R_f - risk-free rate

The name of this plot is the **securities charactaristics line**, or SCL. Commit to memory ALL parts of this plot. First, remember that this is a plot for a particular asset i at different times t. Additionally, when return is labeled as "excess" return, it means it is the return minus the risk-free rate. For any regression plot, there are six things you need to memorize:

(a) The common name (if any).

(b) The x-axis or independent variable.

(c) The y-axis or dependent variable.

(d) The slope.

(e) The y-intercept.

(f) The deviations between the data points and the regression line.

In the case of the SCL, the values are:

(a) Name: Security Characteristics Line

(b) x-axis/independent variable: Excess return on the market, or $r_{\text{market},t} - r_{\text{rf}}$.

(c) y-axis/dependent variable: Excess return on asset i, or $r_{i,t} - r_{\text{rf}}$.

(d) Slope: Systematic risk, or β_i.

(e) y-intercept: Abnormal rate of return, or α_i.

(f) Deviations: Unsystematic risk, or $\epsilon_{i,t}$.

This should be very similar to the information presented in the last section, just with a graphical component to the values. Now, we discuss the second equation and its corresponding graph.

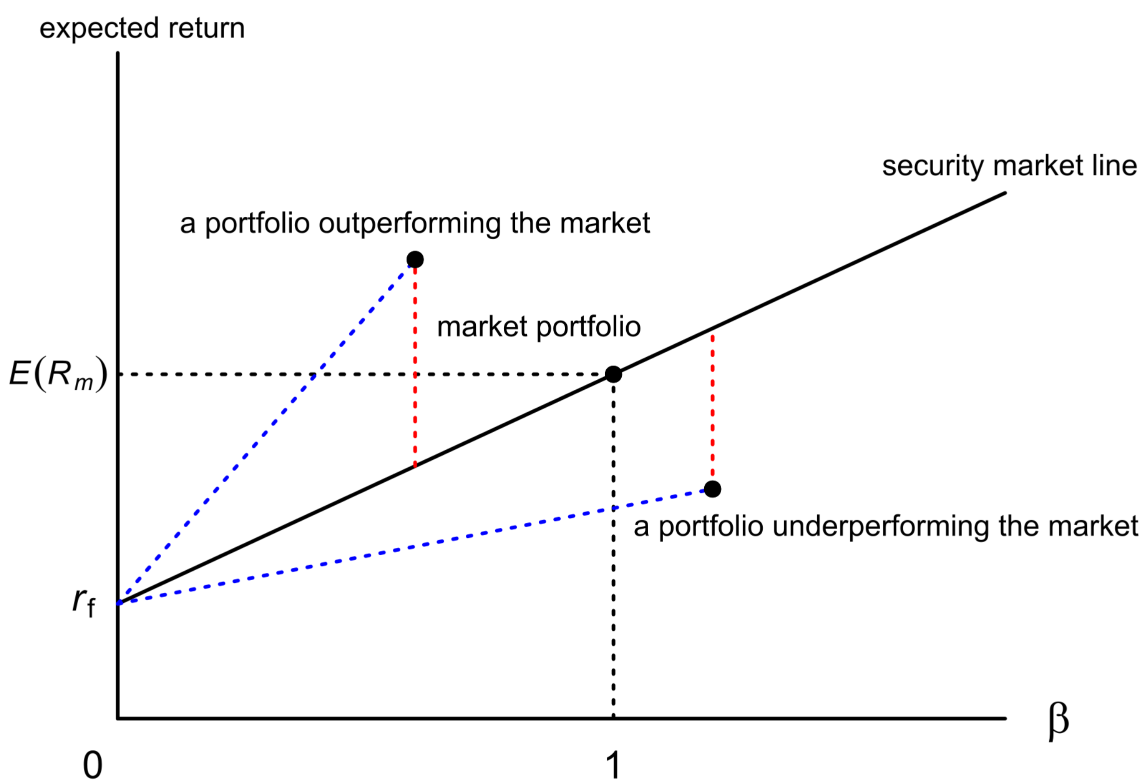
The second equation describes a market of *different* assets i in a market - assume we already took time averages to get the known values.

$$r_i = \beta_i(r_{\text{market}} - r_{\text{rf}}) + r_{\text{rf}} + \alpha_i$$

Here, r_i is the return on an asset not at a particular time, but the expected return across time. We can also write this equation equivalently as:

$$\text{E}_t[r_{i,t}] = \beta_i(\text{E}_t[r_{\text{market},t}] - r_{\text{rf}}) + r_{\text{rf}} + \alpha_i$$

Observe the plot below:



Notive the $E(R_m)$ label in the graph, which corresponds to $\text{E}_t[r_{\text{market},t}]$ in our notation. This is a plot of the expected return on different assets as a function of the risk (β) values of those assets. This line/plot is called the **security market line**. The slope of the line is $(r_{\text{market}} - r_{\text{rf}})$, which is also called the **market risk premium**. We label the parts of the plot for this one as well:

(a) Name: Security Market Line

(b) x-axis/independent variable: β values of different assets.

(c) y-axis/dependent variable: Expected return of those assets.

(d) Slope: Market Risk Premium, or $(r_{\text{market}} - r_{\text{rf}})$.

(e) y-intercept: Risk-free rate, or r_{rf}.

(f) Deviations: Abnormal rate of return, or α_i.

Finally, we look at characteristics between both of the plots. The first important characteristic in regression plots is that the deviation from the line is treated as a zero-mean random variable. This way, data points follow the line on average. The second aspect of the plots to consider is what is known and what is unknown. It is always the case that we have a set of data, and we want to use linear regression to calculate the slope and the y-intercept, which may have been originally unknown. In the security characteristic line plot, we use the data from an asset to calculate β_i. In the security market line plot, we use many β values to calculate the market risk premium. Be comfortable with the notation, the statistical interpretation, and the graphical interpretation of each of these market characteristics.