An Introduction to Linear Factor Pricing Models and their Use Cases

Jay Sivadas, Ronak Datta, and Clara Ee

Introduction to Regression Analysis

Simple and Multiple Linear Regressions

A linear regression is defined as the statistical method that allows us to summarize and study relationships between continuous, quantitative variables. The relationship is expressed in the form of an equation or a model connecting the response or dependent variable and one or more explanatory or predictor variables. We denote the response variable by Y and the set of predictor variables by X_1, X_2, '', X_p , where p denotes the number of predictor variables. The true relationship between Y and X_1, X_2, ..., X_p can be approximated by the regression model

$$Y = f\,(X_1 + X_2 + \ldots + X_p) + \boldsymbol{\epsilon}$$

where ϵ denotes a random error representing the discrepancy in approximation and the function describes the relationship between Y and X_1, X_2 ... X_p

An example of the linear regression model is below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \boldsymbol{\epsilon}$$

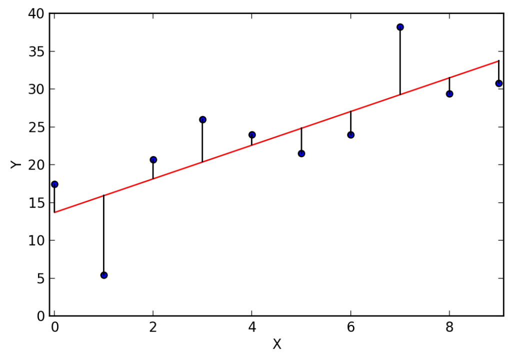
Where the betas (β) are called the regression parameters or coefficients and to be determined from the data. In the context of X being the regressor and Y being the regressand, the betas represent regressand variation related to the associated regressor. In this particular case, β_0 can also be denoted by alpha (α) and represents the y-intercept of the line described by the model. It is helpful to think of α as representing how much on average cannot be attributed to the predictor variables alone.

For the intent of this paper, we will not focus on nonlinear regression although as the name suggests, it can be employed when analyzing nonlinear relationships between the explanatory variable(s) and response variable. However, do note that it is possible to represent nonlinear functions as linear functions through transformations. Moreover, doing so makes the interpretation of the regression coefficients difficult.

The method most commonly used to choose the best parameters is by minimizing the sum of the squared errors. In other words, we select values of regression coefficients that minimize the following equation:

$$\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \cdots - \beta_k x_{k,t})^2.$$

Visually, we can think of this technique known as Ordinary Least Squares as minimizing the sum of the squares of the distance between the regression line (i.e. "line of best fit") and the observed data values. The discrepancy, or distance, between the line and observed values, is referred to as the residual and is represented by the black lines connecting the points to the red line in the image below:



The process of finding the appropriate regression coefficients is referred to as "fitting" the model, with the estimated values found through the process of fitting being denoted by a hat (^) as in the following estimated regression equation:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t}.$$

Nowe we must make the distinction between simple linear regression and multiple linear regression. The former can best be thought of as the simplest case of linear regression - where we have one predictor variable X_1 and only two regression coefficients, β_0 and β_1 as described below.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Alternatively, we can instead opt to use α instead of β_0

$$Y = \alpha + \beta_1 X + \epsilon$$

The equation generated closely resembles the slope-intercept form of a line (y = mx + b), making interpretations of the regression coefficients much more intuitive as β_1 represents the change in the response variable Y, for every 1 unit change in the explanatory variable X.

Multiple linear regression, on the other hand, is an extension of simple linear regression to include multiple explanatory variables while still maintaining one response variable. This is not to be confused with multivariate regression, where we have two or more response variables as opposed to the traditional univariate regression discussed. In the context of multiple linear regression, the regression coefficients(β) can be thought of as measuring the effect of each predictor variable after taking into account the effects of all the other predictors in the model. Therefore, the coefficients essentially measure the marginal effect of each predictor variable on the response value.

We can summarize how well a linear regression model fits the data through the coefficient of determination (denoted by R^2). This value can be calculated by dividing the square of the correlation between observed response values and predicted response values ($\hat{y}$). More commonly, the following equation is used:

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2},$$

where y_t represents the actual value of the response value, $\hat{y}_t$ represents the predicted value of the response variable, and $\bar{y}$ represents

Essentially, the R^2 statistic tells us the proportion of variation in the response variable that is explained by the regression model. Keep in mind that in simple linear regressions, R^2 is equal to the square of the correlation between the response variable and predictor variable. Moreover, R^2 varies between 0 and 1, with a value closer to 1 indicating that the predictions are close to the actual values recorded and thus the model explains a significant portion of the variation in the response variable.

In practice and in the context of multiple linear regression, we instead opt for adjusted R^2, which is defined by the formula below:

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where R^2 denotes R-squared, N equals the total number of samples, and p represents the number of predictor variables

In most contexts, adjusted R^2 is a more useful tool as it essentially penalizes for adding independent variables(X_k) that do not fit the model in an effort to protect against overfitting and making the model misleading. While R^2 itself assumes that every single variable explains the variation in the dependent variable, adjusted R^2 tells us the proportion of variation explained by only the independent variables that actually affect the dependent variable. Note that adjusted R^2 will always be less than equal to regular R^2.

The presented regression model allows us to predict the response value for predictor variables not contained in the original data used to fit the parameters. However, the predictor value must be contained in the original range of data used to estimate the regression model as out-of-sample figures produce invalid projections when plugged into the model. This is primarily because the relationship may no longer be linear, or adhere to the original model parameters even if it is linear, thus rendering the model ineffective.

Ordinary Least Squares Assumptions

The following are core assumptions made as part of the Ordinary Least Squares regression technique discussed earlier.

(i) Linearity
The model must be linear in parameters (i.e.coefficients such as $\boldsymbol{\beta}$)

(ii) Random Sample
The sample used to fit correlation coefficients must be drawn randomly from the population, the number of observations taken in the sample should be greater than the number of estimated parameters, and explanatory variables should impact the response variable. For the purpose of this paper, this assumption is not critical in the context of factor pricing models.

(iii) No Collinearity
There can be no perfect relationship between the individual explanatory variables ($\boldsymbol{X}_i$). This also implies that there should be sufficient variation in the explanatory variables ($\boldsymbol{X}_i$) in order to draw better conclusions. Multicollinearity can be detected using Variable Inflation Factors (VIF), which is defined as the strength of correlation between the independent variables in multiple linear regression. It is important to note that a VIF exists for each predictor in a multiple regression model and simply represents the factor by which the variance of a regression coefficient bj is inflated by the existence of correlation among the predictor variables in the model. It can be defined by the following equation:

$$VIF_j = 1/(1-R_j^2)$$

where is the R^2-value obtained by regressing the j^{th} predictor on the remaining predictors.

To address multicollinearity we may simply remove predictor variables with high VIFs above 10 as this suggests significant multicollinearity and re-run the regression on the remaining predictor variables and subsequently re-calculate VIF. Ultimately the goal is to remove high VIFs until the only ones left are those below 4.

(iv) Zero Conditional Mean
This states that predictor variable X provides no information about the expected value of the error epsilon and thus must now demonstrate a systematic pattern and have a mean of 0. Mathematically this is denoted as

$$E[e|X] = E[e] = 0$$

(v) Normality of Errors
Errors are independent of predictor variables and are independently and identically distributed (iid) according to the following distribution:

$$e \sim N(0, \sigma)$$

where N denotes a normal distribution with a mean of 0 and standard deviation of σ.

Hypothesis Testing

For each regression coefficient (β), we can run the hypothesis test listed below to understand how precise our coefficient measurement is as well as calculate their corresponding p-values to measure their statistical significance.

$$H_0: \beta_1 = 0 \text{ (the slope is equal to zero)}$$
$$H_A: \beta_1 \neq 0 \text{ (the slope is not equal to zero)}$$

We can calculate the test statistic *t* using the following equation:

$$t = \beta_1 / s_1$$

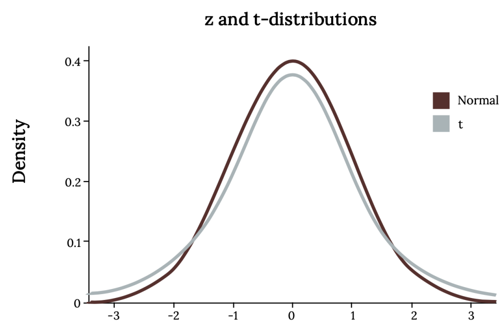
Where β_1 represents coefficient estimate and the s_1 represents the standard error of coefficient β_1.

The standard error of a regression coefficient can be thought of as a measure of how precise the model estimates the coefficient's unknown value. It can be calculated through the following equation:

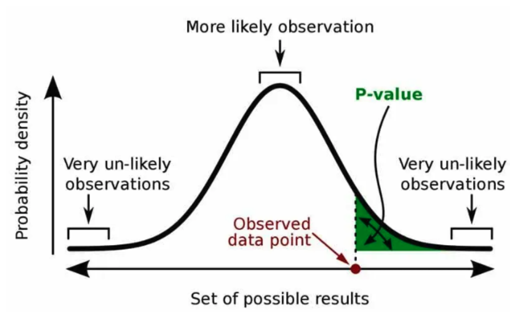
$$s(b_1) = \sqrt{\frac{1}{n-2} * \frac{\sum (y_i - \hat{y}_i)^2}{\sum (x_i - \bar{x})^2}}$$

where n represents the total sample size, y_i represents the actual value of the response value, $\hat{y}_i$ represents the predicted value of the response variable, x_i represents the actual value of the predictor variable, and $\bar{x}$ represents mean value of the predictor variable

The t-statistic generated enables us to understand how confident we are that a regression coefficient is not 0 and thus significant in terms of the relationship between the associated predictor variable and response variable. Its associated p-value tells us the probability of an observed or more extreme result assuming that the null hypothesis is true. We can find the p-value using the Student t-distribution pictured before:



Visually, we can visually imagine the p-value to represent the area under the curve of observations equal to or more extreme than the observed data point as indicated in the shaded region in the diagram below.



If the p-value produced is less than or equal to the significance level used, we conclude that we have sufficient evidence to reject the null hypothesis (i.e. conclude that the alternative hypothesis is true). Moreover, this suggests we are (1-α)% confident that the measured regression coefficient (β_1 in the example used) is not equal to 0. Here, α represents the significance levels chosen, with some popular figures in practice being 0.10, 0.05, and 0.01.

A Brief Introduction to Modern Portfolio Theory

This section will cover a brief introduction to Modern Portfolio Theory and mean-variance analysis as it is a basis for understanding linear factor pricing models. The section is by no means a comprehensive explanation of Modern Portfolio Theory, but it should give enough of an overview so you can understand linear factor pricing models, assuming no background in modern portfolio theory.

Mean-variance analysis was invented by Nobel prize winner Harry Markowitz in his development of Modern Portfolio Theory. The goal of the analysis is to provide a set of portfolio allocation weightings given any combination of assets that optimizes the Sharpe ratio, or mean return to volatility, of that portfolio. This invention is one of the most widely used formulas for portfolio allocation and is implemented by both retail and institutional investors. The mean-variance theorem implies that given any combination of *n* assets, an investor could optimize the risk-return profile of the portfolio by selecting a composition of the *n* assets that maximize returns given a certain level of risk.

The Sharpe ratio is a ratio of the return of an investment to its risk or volatility. The mean excess returns of a portfolio could be used to measure return, while risk or volatility could be derived from the mean-variance. As such, the Sharpe ratio = $\frac{\mu^p - r^f}{\sigma^p}$, where μ^p, σ^p and r^f represent the average returns, standard deviation of returns and the risk-free rate respectively.

Consider a portfolio of n > 1 assets, each with a portfolio weighting of W_i, where the sum of the constituent weights, $W_1 + W_2 + \ldots + W_n = 1$. Adjusting the combination of individual asset weightings in the portfolio results in different risk-return characteristics for the portfolio.

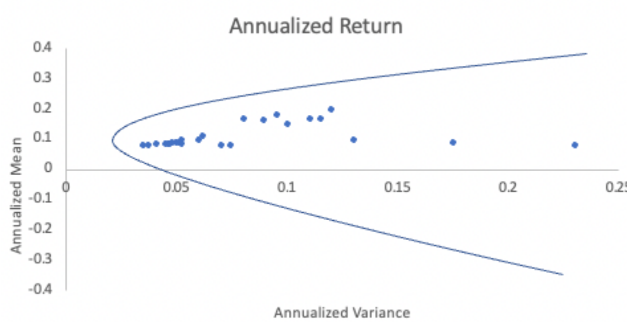


Figure 1: Mean-Variance Curve

Figure 1 (above) illustrates a representation of the risk-return profiles of different portfolios, each comprising the same *n* assets, but with different weights allocated to each asset. The set of all possible compositions of portfolios forms a convex set in the mean-variance space, and the boundary of the set delineates a hyperbola. This boundary, within which every possible combination of risk-return profiles generated from n assets is contained, is known as the mean-variance frontier.

Notice on the boundary that there are two return values (y) given each variance (x), with each value of y representing the largest and smallest possible returns respectively. An investor who constantly looks to maximize his returns will select a portfolio on the upper bound. This upper half of the hyperbola is denoted as the *efficient frontier*, which consists of portfolios yielding the highest possible rates of return for a given level of risk. The objective of adjusting the weights of a portfolio's constituent assets is thus to maximize the return of the portfolio such that the portfolio lies on the efficient frontier.

<u>Tangency Portfolio</u>

The tangency portfolio is the point on the efficient frontier with the maximum Sharpe ratio any portfolio could possibly attain given n assets. The tangency portfolio assumes the existence of a risk-free asset. From this assumption, the tangency portfolio is derived by drawing a line from the point of intercept of the risk-free asset–which has zero variance and a given return–that is tangent to the mean-variance frontier. The point of tangency is the portfolio that has the maximal attainable Sharpe ratio from any combination of those n assets, as is called the tangency portfolio. This line, on the other hand, is known as the capital market line or the capital allocation line.

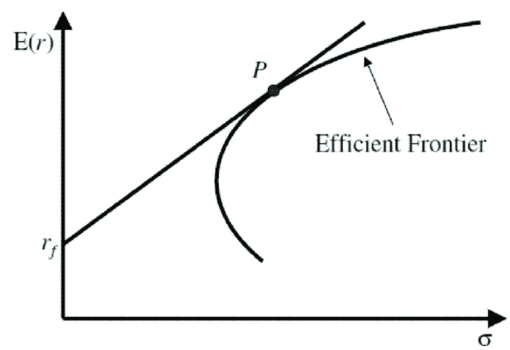


Figure 2: Tangency portfolio (Research Gate)

The capital market line represents a set of portfolios that optimize the risk-return relationship. The slope of the capital market line is equivalent to the Sharpe ratio of the tangency portfolio, meaning that any portfolio constructed on the capital market line holds the same Sharpe ratio as the tangency portfolio. Rational investors can choose to invest in a portfolio anywhere along the capital market line, including the tangency portfolio. Different points on the capital market line differ in that their levels of return and risk vary from each other. Investors can adjust their holdings on the capital market line by holding different combinations of risk-free and risky assets to achieve targeted levels of risk and return.

The tangency portfolio assumes 100% holdings in risky assets. Holding a larger percentage of risk-free assets moves you left along the line–to a portfolio that has lower return and lower risk–while selling the risk-free asset (i.e. levering your holdings) moves you to the right of the capital market line that has higher risk and higher reward.

Linear Factor Pricing Models

Linear regression is the mathematical basis for linear factor pricing models. Linear Factor Pricing Models are a subset of modern finance popularized in the mid-1900s by William Sharpe, Eugene Fama, and Kenneth French. As the derivations of Linear Factor Pricing Models are an extension of modern portfolio theory, it is important that one retains a basic grasp of its underlying concepts before delving into factor pricing models.

Linear factor pricing models in theory predict the excess returns of any asset by taking the relationship between the asset and a theoretical tangency portfolio. This relationship is known as the beta, which we will derive in different examples in subsequent sections to predict the expected excess return of the asset. Recall from modern portfolio theory that the tangency portfolio lies on the efficient frontier with a maximum Sharpe ratio attainable by n assets.

There are many different linear factor pricing models. Where they differ primarily is their assertion on the tangency portfolio and its composition. They are, in essence, different combinations of the relationship between an asset and the tangency portfolio it belongs to and provides predictive pricing for the asset. This tangency portfolio can be constructed from one or more portfolios, each of which is referred to as a *factor*.

Factors are “attributes” driving the return of assets. One can think of a factor as a portfolio with its own return characteristics, very much like that of a stock. Factors are oftentimes constructed as a portfolio of different assets through the application of desired parameters.

Theoretically, the risk and return of a stock or asset can be predicted based on its relationship to a factor or a group of factors. Factor investing thus adopts regression models to analyze this relationship, deriving a resulting value that provides insight into an asset's returns. In the regression model, the factor is simply the independent variable or regressor.

The use cases of factor analyses include performance breakdowns, portfolio tracking, or hedging.

Capital Asset Pricing Model

The capital asset pricing model (CAPM) is one of the most famous linear factor pricing models. Recall that the discounted cash flow valuation method derives the cost of equity of an asset using the CAPM. The CAPM discounts equity cash flows based on the expected return rate of return given a certain level of risk. The CAPM thus exhibits qualities characteristic of linear factor pricing models, which determine the expected return given a relationship to a factor and its constituent risk, or beta.

The CAPM is thus a linear factor pricing model that adopts the value-weighted market portfolio as the tangency portfolio in the pricing model.

The market portfolio is constructed as a value-weighted portfolio of all available publicly traded assets. In a value-weighted market portfolio, each asset is allocated a weight that is determined by its value or size—as measured by market capitalization—with larger assets being allocated a larger weight in the portfolio. In practice, a broad equity index such as the SPY is generally used as a proxy for the market portfolio.

The CAPM can be derived from the following equation, where the expected excess return of the asset is given by the beta between the asset and the market multiplied by the excess return of the market:

$$\beta_i (ER_m - R_f)$$

The excess returns of the market is more conventionally referred to as the market risk premium, or the extra return an investor expects to receive by taking on market risk in excess of the risk-free rate. Further, the CAPM is a component of the equation for the cost of equity, which is derived by simply adding the risk-free rate of return:

$$E_R = R_f + \beta_i (ER_m - R_f)$$

The two components of return on equity change. The risk-free rate of return varies with the 10-Year Treasury Bonds rate while market return varies with time, typically as a consequence of various macroeconomic factors. The beta, β_i, is determined by running a regression. The excess return of an asset over a historical period is regressed onto the excess return of the market over the same time period. The resulting value is the beta, which can be used to extrapolate the market return into the future to estimate the expected return of an asset over a period of time. The beta could also be given by the following equation:

$$\beta^{i,m} \equiv \frac{cov(r^i, r^m)}{var(r^m)}$$

where r^i and r^m represent the return of an asset, i and the market return respectively

The CAPM first asserts that all returns have a joint normal distribution based on the mean and variance of their returns. Assuming the existence of a tangency portfolio, investors will aim to optimize the mean-variance of their portfolio by investing in a portfolio that lies on the mean-variance frontier. Each investor holds a combination of the tangency portfolio and the risk-free portfolio. While every investor holds different assets, the aggregate of all tangency portfolios of every investor investing on the mean-variance frontier will be equivalent to the tangency portfolio containing every possible publicly traded asset. This is known as the market portfolio.

In the CAPM, the market portfolio is the tangency portfolio containing every possibly publicly traded asset. The CAPM is developed from the underlying concept that the aggregate of the tangency portfolios of all investors would be equivalent to the value-weighted market portfolio. This is intuitively sound because a stock will have a higher market capitalization (or value) if more investors own it.

The CAPM assumes that the market beta is the only risk investors will demand a higher return for. Recall from modern portfolio theory the discussion on systematic and idiosyncratic risk. Because idiosyncratic risk can be diversified easily and cheaply by holding a large number of assets in a portfolio, investors will not demand a return from the portfolios' aggregate idiosyncratic risk. As such, they demand only a higher return from systematic risk or the market beta.

An important point to note is that the CAPM assumes that the market portfolio perfectly predicts all asset returns. A regression of an asset's return onto the market portfolio should thus imply zero alpha as no return is unexplained by the market.

<u>Limitation of the CAPM</u>

However, in practice, a regression results in a risk-reward trade-off modeled after the red line as shown below. If the CAPM were to be an accurate representation of reality, the black line below would be the best-fit line, as all asset returns should be a scalar adjustment from the market portfolio return based on the beta. Furthermore, unlike the CAPM, a regression of real market data would produce values that imply some level of alpha.

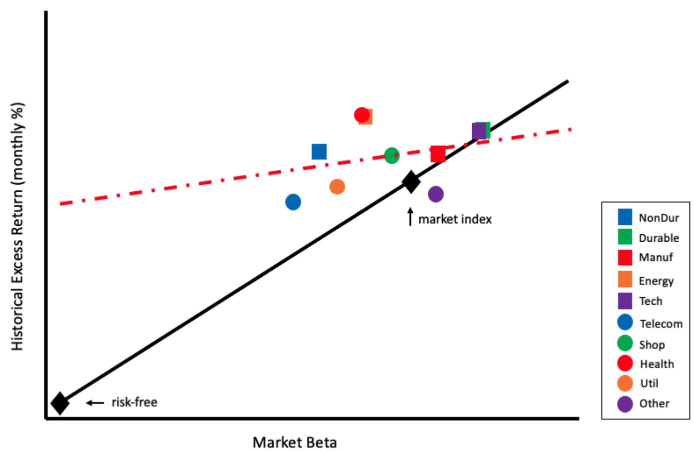


Figure 3: Linear regression of industry portfolios

Multi-Factor models

Linear factor pricing models can have one or more factors. Thus far, we've covered single-factor pricing models, with the most famous of them being the capital asset pricing model (CAPM). There also exist linear factor pricing models that base their predictions of expected return on multiple factors.

Multi-factor models are simply an extension of linear factor models. However, they assert that the tangency portfolio is a linear combination of multiple factors, unlike single-factor models that assume the single factor—in the case of CAPM, the market portfolio—to be the tangency. This means that the tangency portfolio is now a combination of multiple factors as well. The excess return of the tangency portfolio is thus given by the weighted average return of the individual factors and their constituent weights. The equation for the excess return of the tangency portfolio is given by the following:

$$r^{t} = w' r^{z}$$

Multi-factor models could simply be understood as multiple linear regression models, each regressing the return of an asset onto multiple independent variables or the factors. Each independent variable has a relationship to the asset very much like the beta in CAPM, and the aggregate of these variables determines the rate of return of the asset. Very much like the single-factor model, only the systematic risk is being considered. The equation for the expected return of a portfolio is thus given by the following:

$$E(r^{p}) = (\beta^{p,z})' E(r^{z})$$

Fama French

The Fama-french factor models are well-known examples of multi-factor models. They were developed by Kenneth French and Eugene Fama in the late 1900s, both of whom became Nobel prize laureates. The Fama-French 3-factor model assumes that the tangency portfolio is a linear combination of 3-factors: the market portfolio, size portfolio (SMB), and value portfolio (HML). The expected return of a 3-factor portfolio is thus given by the following:

$$E_R = R_f + \beta_i (ER_m - R_f) + \beta_S * SMB + \beta_V * HML$$

Notice that the market portfolio is simply the value-weighted portfolio of all liquid stocks in the market.

To calculate the expected return of an asset, you simply multiply the betas between the asset and each factor by the expected return of that factor and take a summation of the resulting values. Like how one would approach a single-factor pricing model, betas are first calculated by running a regression—this time a multi-variable regression—on historical return data. The return on each factor is then extrapolated to predict the return of the asset.

The size portfolio (SMB) assumes that smaller market capitalization stocks have higher return potential than larger market capitalization stocks. The size portfolio is hence constructed by going long smaller market capitalization stocks and short-selling larger market capitalization stocks.

The value portfolio (HML) assumes that assets trading at low multiples relative to peers will have a higher return than assets trading at high multiples. Low-multiple and high-multiple assets are coined “value” and “growth” stocks respectively, and the value portfolio is constructed by going long-value stocks and short-selling growth stocks.

However, there are infinite ways in which one can construct a factor portfolio. For instance, different metrics could be used to construct a value factor portfolio. Significant research goes into constructing each factor, with different models built to test the efficacy of each factor portfolio construction parameter on actual asset return. Here are some examples of how some factor portfolios could be constructed:

$(ER_m - R_f)$	Excess market return = Value weighted return on NYSE, AMEX & NASDAQ stocks – one month Treasury Bill Rate
SMB	SMB = average return on 3 small portfolios – average return on 3 big portfolios = ⅓ (Small Value + Small Neutral + Small Growth) – ⅓ (Big Value + Big Neutral + Small Growth)
HML	HML = average return on 2 value portfolios – average return on 2 growth portfolios = ½ (Small Value + Big Value) – ½ (Small Growth + Big Growth)

Source: French, K.R. *Description of Fama/French 5 Factors.*

The Fama-French 3-factor model has also been expanded to include 2 additional factors (hence the Fama-French 5-factor model):

RMW	Robust-Minus-Weak = ½ (Small Robust + Big Robust) = ½ (Small Weak + Big Weak)
CMA	Conservative-Minus-Aggressive = ½ (Small Conservative + Big Conservative) - ½ (Small Aggressive + Big Aggressive)

Source: French, K.R. *Description of Fama/French 5 Factors.*

In the Fama-French 5-factor model, the SMB calculation would look like the following:

SMB	$SMB_{(B/M)}$ = ⅓ (Small Value + Small Neutral + Small Growth) - ⅓ (Big Value + Big Neutral + Big Growth) $SMB_{(OP)}$ = ⅓ (Small Robust + Small Neutral + Small Weak) - ⅓ (Big Robust + Big Neutral + Big Weak) $SMB_{(INV)}$ = ⅓ (Small Conservative + Small Neutral + Small Aggressive) - ⅓ (Big Conservative + Big Neutral + Big Aggressive) SMB = average return on 9 small portfolios – average return on 9 big portfolios = ⅓ ($SMB_{(B/M)}$ + $SMB_{(OP)}$ + $SMB_{(INV)}$)

Source: French, K.R. *Description of Fama/French 5 Factors.*

As noted above, there are different ways of identifying a characteristic—such as value—and weighing each asset to construct a factor portfolio. The Fama-French model adopts the book-to-market multiple as a measure of trading value in its construction of the value factor. The book-to-market multiple simply measures the ratio of the book value of equity to the market value of equity. The book value of equity is the value of shareholder's equity on the balance sheet of a company. This includes any equity raised, retained earnings from net incomes and dividends, and other extraneous factors. A low book-to-market multiple signals the potential "overvaluation" of a stock, where the market pays a significant premium over its true value. As such, value stocks have high book-to-market multiples. Figure 4 (below) shows the annualized return of US stocks sorted by book-to-market. Notice that there is a positive correlation between excess return and value.

Figure 4: Annualized Return of U.S. Stocks Sorted by Book-to-Market (1951 - 2016)

Other popular accounting metrics used to construct the value factor portfolio include the earnings-to-price, EBITDA-to-price, and dividend-to-price ratios. Depending on the company or industry, it is also possible to use non-accounting metrics such as subscribers-to-price and scientists-to-price that could represent a company's future earnings potential.

Value, growth, and market are just some examples of factors one might consider. There are two main categories of factors—style and macro. Style factors describe the characteristics of assets and include those discussed above such as value, size, and momentum. Macro factors capture the broad macroeconomic risk influencing asset return. This can include inflation, Gross Domestic Product (GDP) growth, and interest rates, just to name a few.

Investment vehicles that adopt factor analysis invest generously to develop perfect factor models to improve their investment portfolios.

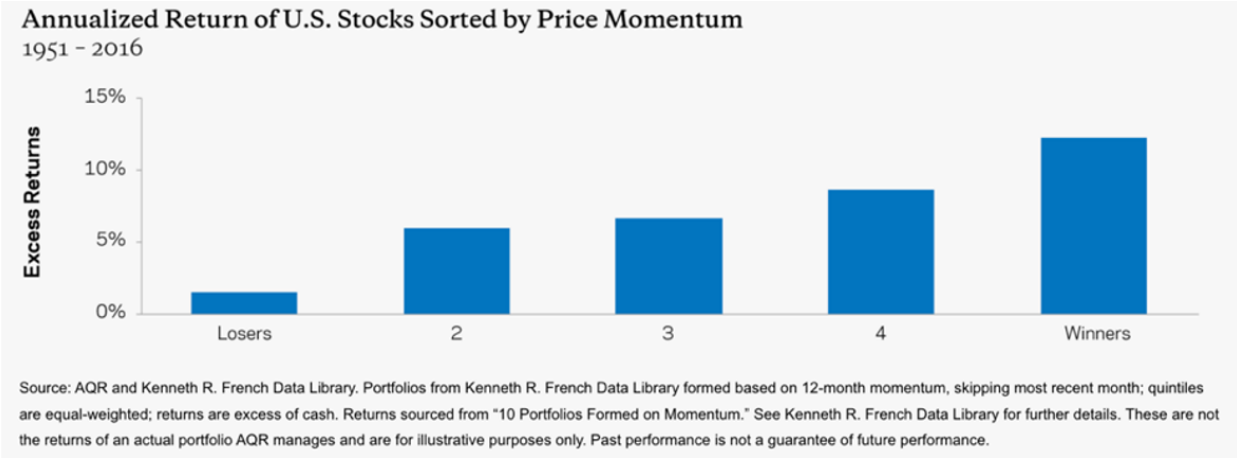


Figure 5: Annualized Return of U.S. Stocks Sorted by Price Momentum (1951 - 2016)

Factor Investing in Practice

The two most popular methods of factor investing are "smart beta" and systematic factor portfolio construction.

Smart beta is a broad term encompassing any investing strategy utilizing factor analysis to allocate asset weights or select stocks. Developed a few decades ago, it is now mostly a marketing strategy. The most popular application of smart beta strategies is the construction of a Smart Beta Exchange-Traded Fund (ETF). An ETF is a portfolio of assets managed by firms such as Blackrock and Vanguard, which can simply be bought and sold like stocks using a brokerage account. ETF investors benefit from the return of the portfolio while firms make a small fee. Traditional ETFs are weighted by market capitalization. In the S&P 500, for instance, larger market capitalization stocks such as Apple are allocated larger weights relative to smaller market capitalization stocks. The weighting is automatically adjusted and typically favors overpriced assets, as such assets typically have larger market capitalizations.

Smart Beta ETFs typically adopt a factor tilt to adjust weightings. By applying some weighting schema to the market capitalization weight of an asset, factor exposure could be calibrated. While there are many variations, one example of a single factor tilt is provided below:

$$\widehat{W}_i = \frac{S_i * W_i}{\sum_{j \in U} S_j * W_j}$$

for an underlying index universe U, with underlying index weights W_i and the standard cumulative normal distribution function of the cross-sectional Z-Scores for a given factor be represented by S_i

Systematic factor portfolio construction, on the other hand, is a way of selecting assets from a portfolio through factor regression analysis. We first apply an initial set of criteria—such as a minimum market capitalization of 1 billion—to a universe of equities, and run a regression on all the resulting equities. After which, certain parameters are applied to the regression results to select a predetermined number of assets. Investment management firms such as AQR capital management construct portfolios systematically using this method and adjust their holdings with changing market environments and factor model updates. The PNG investment strategy aims to construct a portfolio using a similar idea, except that the final holdings are selected from the portfolio on a discretionary basis.

Constructing Multi-Factor Portfolios

Two ways of applying parameters in systematic factor portfolio construction are the heuristic construction and optimized construction methods.

The heuristic construction method assigns each equity a factor score using a factor regression analysis. The assets are then ranked by factor score and the top *n* assets are selected to construct a portfolio. A sample factor score is given by the following equation:

$$\alpha_i = 0.2 * F_{1,i} + 0.2 * F_{2,i} + 0.2 * F_{3,i} + 0.2 * F_{4,i} + 0.2 * F_{5,i}$$

Source: NASDAQ

However, other complex scoring systems such as the information and Treynor ratios could also be adopted to construct the portfolio.

The optimized construction method applies a variety of security and sector-level constraints to construct the desired portfolio. An example would be the adoption of a risk framework to adjust the expected portfolio risk and one minimum risk optimization function could be given by the following:

$$\min_{W} \bar{\sigma}_p = W^T \Sigma W$$

Subject to

$$0 \leq w_i \leq j \text{ for every stock } i$$

$$sum(w_i * f_{k,i}) = F_k \text{ for every factor } k$$

$$sum(w_i * s_{l,i}) = S_l \text{ for every sector } l$$

$$sum(w_i) = 1$$

Source: NASDAQ

Performance Analysis

Factor models are also frequently used to measure fund performance. Oftentimes, portfolio managers aim to outperform a certain benchmark, and this excess return over the benchmark is referred to as *alpha*. Factor models can be used to derive the value of alpha to evaluate portfolio performance and provide insight into return drivers.

For instance, fund managers can achieve impressive returns. However, such outperformance may not necessarily be attributable to the fund manager's capabilities if the portfolio has a high, positive correlation with the market and market returns are favorable over the assessment period. A high, positive correlation with the market implies that the portfolio exhibits similar return characteristics to the market, thus allowing it to capture a rate of return that mirrors that of the market.

Benchmark selection for performance analysis is of equal importance, as the benchmark used should be of a similar risk profile as the actual portfolio. For example, if a technology-only investment fund generated 20% on an annual basis from 2012 to 2020, benchmarking the investment fund against the S&P 500 would yield an access return of 7% on an annual basis. However, using the S&P 500 as a benchmark does not accurately capture the risk of a technology-only portfolio, given the greater volatility of technology stocks relative to the S&P 500. As such, a different benchmark such as the NASDAQ, which leans heavily towards technology stocks, would be a more appropriate choice.

(i) Linear Factor Decomposition

The return of an asset can be decomposed by running a regression of the asset onto a factor. Consider a portfolio invested across all industries. One can obtain the historical return statistics of the portfolio and regress those returns onto a factor to understand how the factor might explain the return. Regressing the portfolio onto the S&P 500 would allow one to determine the alpha.

The values obtained from a linear factor decomposition that we are most concerned with are the alpha, beta and R^2.

Alpha is the return that is not explained by the benchmark, or the overall market in this instance. Active managers seek to generate alpha to provide clients with returns beyond investing passively in the S&P 500.

The beta is variation or risk relating to the benchmark. A beta of 1 indicates perfect correlation; a beta greater than 1 indicates greater risk relative to the benchmark; a beta of less than 0 indicates inverse correlation; a beta of 0 indicates zero correlation.

R^2 is a measure of how well a factor explains the variation of return and allows an evaluation of whether a benchmark accurately reflects the risk of a portfolio. A higher R^2 value reflects a greater ability of the benchmark to capture the variation of a portfolio.

A high alpha could mean that a fund manager is generating high returns that are not attributable to the market alone. However, if an improper benchmark is used, as indicated by a low resulting R^2, the value of the alpha becomes a misrepresentation.

(ii) Performance Metrics

Performance metrics analyze the risk-to-return ratio of the asset, as opposed to the upside, as is the case of alpha.

Recall from modern portfolio theory that the Sharpe ratio is the mean excess return over the standard deviation of the portfolio return. Mean-variance analysis is built upon this performance ratio. However, there are also other common performance ratios that utilize different metrics of risk and return.

One example is the information ratio, which analyzes the characteristics of the regression that are unexplained by a factor in measuring performance. It uses alpha—the return unexplained by the factor—as a measure of return and the standard deviation of the residuals—the variation unexplained by the factor—as a measure of risk. A portfolio manager seeks to maximize alpha while retaining a high R^2 value. In other words, the portfolio manager would aim to generate excess return while retaining a risk profile similar to the benchmark.

Just like the Sharpe ratio, the Treynor ratio adopts the mean excess return of a portfolio as a measure of return. The difference lies in the use of systematic risk—captured in the beta—as a measure of the volatility of the portfolio. Formulas for the Sharpe ratio, information ratio, and Treynor ratio are given by the following:

$$\text{Sharpe Ratio} = \frac{\tilde{\mu}^p}{\sigma^p} \qquad \text{IR} = \frac{\alpha}{\sigma_\epsilon} \qquad \text{Treynor Ratio} = \frac{\mathbb{E}[\tilde{r}^i]}{\beta^{i,m}}$$

Hedging

The regression model could also be used as a risk-management tool to hedge against the risk of exposure to a certain market. Hedging is the process of buying and selling assets with similar return behavior or buying assets with inversely correlated return behavior to minimize the potential downside risk of an investment. While the process of hedging reduces return, the risk-return profile of the investment improves in the process.

The process of hedging involves selecting an asset and identifying a characteristic or risk to hedge against. For example, one can choose to hold Apple shares and hedge against exposure to the overall market. To do so, we will first run a regression of Apple's return onto the overall market's return. We then short-sell a market portfolio by beta dollars for every dollar we put into Apple shares. Through this process, systematic risk relating to the market is eliminated, leaving us with the alpha and residuals, which are the return and variation unexplained by the factor.

Tracking

One other use case for regression models is the construction of tracking portfolios, where investors aim to generate a portfolio that replicates the return and variation of an index or target portfolio without knowledge of its specific holdings.

The process of tracking involves running a multiple regression of the target portfolio's return data onto all of the investor's holdings. For instance, if the investor owns a list of stocks such as Apple, Google, Facebook, and Walmart, these individual stocks would be factors used in the regression analysis.

Given a set of regression results, the investor then invests in each holding an amount equal to its corresponding beta multiplied by the total value of the portfolio. This method provides a set of portfolio allocation weights that would theoretically replicate the return and variation of the target portfolio.

Here, the investor aims to maximize the R^2 value of the tracking portfolio, as it explains how well it replicates the performance characteristics of the original portfolio.

Conclusion

We began the paper with a brief introduction of the concept of linear regression, which enables us to model the relationship between a scalar response and one or more predictor variables. The resulting equation produced through the ordinary least squares technique, along with subsequent statistical calculations detailed earlier in the paper, allow us to understand both the strength and nature of the relationship between the response and explanatory variables.

The understanding of regression analysis leads us into the theory behind linear factor pricing models, which derives its analysis from the regression results of factors onto assets. The theory behind linear factor pricing models is built off modern portfolio theory and the tangency portfolio, as pricing models are simply different assertions about the tangency portfolio and the covariances between an asset and the tangency portfolio.

The application of regression analysis extends beyond asset pricing in finance, to topics including portfolio construction, performance breakdown, hedging, and tracking. An understanding of regression analysis and factor pricing models provides an investor optionality and flexibility in the analysis and construction of their portfolios.

References

Chatterjee, S., Hadi, A.S. (2013). *Regression Analysis by Example*. (ed. 5). Hoboken, NJ: Wiley

French, K.R. (No date). *Description of Fama/French 5 Factors* [Article]. Retrieved from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html#:~:text=The%20Fama%2FFrench%205%20factors,formed%20on%20size%20and%20investment

Lin, R. (No date). *A Practitioner's Guide to Multi-Factor Portfolio Construction* [Article]. Retrieved from https://www.nasdaq.com/docs/2020/09/01/A-Practitioner%E2%80%99s-Guide-to-Multi-Factor-Portfolio-Construction.pdf

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91.