Module 5: Factor Models

Daniel Barrera and Edgar Alcántara

August 21, 2025

Summary: In this Module, we explore empirical models that extend the Capital Asset Pricing Model (CAPM). It introduces the Fama and French Three-Factor Model, which incorporates the SMB (Small minus Big) and HML (High minus Low) risk premiums to capture the effects of size and value anomalies on stock returns. Subsequently, the Carhart Four-Factor Model is presented, adding the WML (Winners Minus Losers) risk premium to explain the momentum anomaly, which is based on the persistence of past returns. The document details how these factors are constructed by classifying stocks into different portfolios.

Contents

1 Fama & French: Three Factor Model

After understanding the theoretical outcomes of factor models related to Markowitz Portfolio Theory, such as the Capital Asset Pricing Model and its variations, we will explore extensions of these models with a greater focus on empirical approaches rather than theoretical derivations. Eugene Fama and Kenneth French identified empirical inconsistencies in the results of CAPM models. In response, they proposed a more comprehensive model that better explains certain market anomalies, which we will examine in this chapter.

1.1 Size and Value Anomalies

in Fama and French, 1993, the authors identified at least three anomalies in stock valuation when using the standard Capital Asset Pricing Model. These anomalies are commonly associated with the size of companies, the beta of the stock, and the book-to-market value. In this chapter, we will carefully explain these anomalies and demonstrate why Fama and French considered it important to incorporate them into the CAPM. Since we reviewed the Beta Anomaly while studying the CAPM in the previous chapter, we could omit it explanation in this section.

1.1.1 The Size Anomaly

While analyzing the results of the original Capital Asset Pricing Model (CAPM), Fama and French observed that companies with larger market capitalizations (calculated as the total number of shares outstanding multiplied by the stock price), or in simpler terms, larger companies, tend to have, on average, lower expected returns compared to companies with smaller market capitalizations. Therefore, they created a hypothesis proposing that the size of companies also affects their returns in addition to just the systematic risk explained by beta. If this hypothesis is true, then the returns of an asset could be explained by the following linear equation:

$$\mathrm{E}[R_i] = \alpha_i + \lambda_i \ln(\mathrm{MktCap}_i) + \varepsilon_i \tag{1}$$

If the coefficient λ_i is statistically significant (and negative) for a sufficiently large and representative sample of the capital market, we could conclude that Fama and French's hypothesis holds. Indeed, they demonstrated that this factor was significant in explaining the expected returns on stocks in the market, making it a relevant covariate to include in the CAPM. Nevertheless, Fama and French do not solely rely on the logarithmic transformation of market capitalization to understand the size anomaly. This is because such transformations can introduce endogeneity and multicollinearity problems since:

$$\ln(\mathrm{MktCap}_i) = \ln(P_i) + \ln(N_i) \tag{2}$$

Here, $\ln(P_i)$ represents the logarithmic transformation of the price of an asset, and $\ln(N_i)$ represents the logarithmic transformation of the number of shares outstanding. Since the returns of an asset inherently depend on its price, using this specification in a linear regression model could lead to biased results. To address this, Fama and French proposed a size risk premium to capture the idea that, given the size of the company, returns would differ. This premium is referred to as SMB (Small minus Big), and the model is adjusted as follows:

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_i \mathrm{SMB} + \varepsilon_i \tag{3}$$

Fama and French's hypothesis suggests that the SMB premium favors smaller companies. To explore this further, stocks can be categorized based on market capitalization. Suppose we rank stocks by market capitalization, resulting in:

$$\{\mathrm{MktCap}\} = \{\mathrm{MktCap}_1, \mathrm{MktCap}_2, ..., \mathrm{MktCap}_M, ..., \mathrm{MktCap}_{n-1}, \mathrm{MktCap}_n\}$$

Here, MktCap_M represents the median market capitalization. Stocks are then categorized as:

$$\text{Size}_i = \begin{cases} \text{Small}, & \text{if } \text{MktCap}_i \leq \text{MktCap}_M \\ \text{Big}, & \text{if } \text{MktCap}_i > \text{MktCap}_M \end{cases}$$

Technically, all companies on the left of the median market capitalization distribution will be considered small companies, while those to the right of the median will be classified as large companies. Fama and French's hypothesis suggests that the SMB (Small Minus Big) premium tends to favor smaller companies. Some interesting exercise to prove this hypothesis is to formulate a linear regression model with dummy variables with the following expression:

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_i \text{Small}_i + \varepsilon_i \tag{4}$$

Where:

$$\text{Small}_i = \begin{cases} 1, & \text{if } \text{MktCap}_i \leq \text{MktCap}_M \\ 0, & \text{otherwise} \end{cases}$$

If λ_i is significant, we could conclude that, on average, there is a difference in returns between small and large companies. Furthermore, we could combine the dummy variable with the market risk premium to assess whether there are significant beta differences between the two types of companies.

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_{0i} \text{Small}_i + \lambda_{1i} \left[\text{Small}_i * (\mu_M - r_f)\right] + \varepsilon_i \tag{5}$$

Now, if λ_{1i} this also would say that companies are more or less sensible to market shocks depending on their size. This because the 'real' beta coefficient of the stock would be $\beta_i + \lambda_{1i}$, adding an extra coefficient that represents the change of sensibility captured by the size anomaly. Nevertheless, we are going to see further how Fama and French measure the Small Minus Big premium.

1.1.2 The Value Anomaly

The value anomaly is somewhat harder to understand. It is related to valuation ratios, such as the Book-to-Market Ratio (B/M) and the Price-to-Earnings Ratio (P/E). Fama and French observed that value stocks tend to have different returns from growth stocks, uncovering another anomaly that cannot be explained by the market risk premium alone. The key idea is that value-stocks may be undervalued in the market, compared to growth-stocks, which gives them an additional premium, resulting in higher returns over the long term. First, we have to understand the **Price-to-Earnings Ratio (P/E)**. This ratio is commonly used to assess whether stock is cheap or expensive, comparing the market price of the asset with the earnings offered per share. The ratio is calculated with the next form:

$$\text{P/E}_i = \frac{P_i}{E_i} \tag{6}$$

A high P/E ratio is typically associated with growth-stocks. Investors are willing to pay a premium for these stocks because they are expected to have high growth rates in the future, potentially leading to higher dividends in subsequent periods. Conversely, a low P/E ratio is generally associated with value-stocks, where the market price of the stock is relatively low compared to the earnings per share. This suggests that the market may not expect significant growth in the future. For example, if a stock's market price is \$50 and the earnings per share (or dividends) are \$5, we would conclude that an investor is willing to pay 10 times the earnings (or dividends) for a single share of stock.

Then, there would be the **Book-to-Market Ratio (B/M)**, which is another metric that compares the book value of a company (commonly assets less liabilities) to the market capitalization (market value) of the company. This is another method to understand if a stock is a value-stock or a growth-stock:

$$\mathrm{B/M}_i = \frac{\mathrm{BookValue}_i}{\mathrm{MktCap}_i} = \frac{\mathrm{BookValue}_i}{N_i} * \frac{1}{P_i} \tag{7}$$

For example, if a stock has a Book-to-Market Ratio greater than one, we would conclude that the stock is undervalued because the market price does not fully reflect the intrinsic value of the company's assets. Conversely, if the Book-to-Market Ratio is less than one, the company would be considered overvalued, as the market price exceeds the value recorded in the books. This method provides a more intuitive way to classify stocks compared to using the P/E Ratio. First, we need to calculate the percentiles for the sample of companies based on their B/M ratios. Typically, the 30th and 70th percentiles are used for categorization. Companies with B/M ratios below the 30th percentile are classified as growth stocks, while those with B/M ratios above the 70th percentile are considered value stocks. Companies with B/M ratios falling between these two thresholds are categorized as neutral stocks. Then, the next expression represents the previous idea:

$$\mathrm{Value}_i = \begin{cases} \text{Growth}, & \text{if } \mathrm{B/M}_i < \mathrm{B/M}_{P30} \\ \text{Neutral}, & \text{if } \mathrm{B/M}_{P30} \leq \mathrm{B/M}_i \leq \mathrm{B/M}_{P70} \\ \text{Value}, & \text{if } \mathrm{B/M}_i > \mathrm{B/M}_{P70} \end{cases}$$

We can then propose linear regression models with dummy variables (excluding one category to avoid perfect collinearity) to test whether the hypothesis of the value risk premium holds true. For example, if we use the 'neutral' category as the reference group, the model would interpret the coefficients as the difference in returns relative to neutral stocks. The next expression is the representation of this model:

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_{0i}\,\mathrm{Growth}_i + \lambda_{1i}\,\mathrm{Value}_i + \varepsilon_i \tag{8}$$

If $\lambda_{0i} > 0$ and the coefficient is statistically significant, we can conclude that, on average, growth stocks offer better returns than neutral stocks. Conversely, if $\lambda_{1i} < 0$ and the coefficient is statistically significant, it would indicate that value stocks tend to provide smaller returns compared to neutral stocks.

1.2 Calculating the SMB and HML premiums

With this classification, we can compare companies to identify which ones are likely to provide better returns and which may not be the best investment options. However, if we get creative, we might also explore questions like how much better small value companies perform compared to large neutral companies. By combining dummy variables with interaction terms, we can create new subcategories from the existing ones and build increasingly complex models to analyze these relationships.

Instead of over-specifying a model, we aim to use variables that combine all these effects into a single framework. First, we need to identify the six categories to be used in the analysis and determine which companies belong to each group. The purpose is to create six portfolios of assets and obtain the returns for each of them:

- **Small-Growth (SG)**: Companies with market capitalizations below the sample median and Book-to-Market ratios below the 30th percentile.
- **Small-Neutral (SN)**: Companies with market capitalizations below the sample median and Book-to-Market ratios between the 30th and the 70th percentile.
- **Small-Value (SV)**: Companies with market capitalizations below the sample median and Book-to-Market ratios above the 70th percentile.

- **Big-Growth (BG)**: Companies with market capitalizations above the sample median and Book-to-Market ratios below the 30th percentile.
- **Big-Neutral (BN)**: Companies with market capitalizations above the sample median and Book-to-Market ratios between the 30th and the 70th percentile.
- **Big-Value (BV)**: Companies with market capitalizations above the sample median and Book-to-Market ratios above the 70th percentile.

The next step is to calculate the returns for these portfolios, not by using the classical equal-weighting approach but by implementing market capitalization weighting. This approach assigns larger weights to companies with higher market capitalization and smaller weights to those with lower capitalizations.

$$R_P = \sum_{i=1}^{n} \omega_i R_i \tag{9}$$

$$\omega_i = \frac{\text{MktCap}_i}{\sum_{i=1}^{n} \text{MktCap}_i} \tag{10}$$

After calculating the returns for the six portfolios, we can compute the Small-Minus-Big (SMB) risk premium by taking the difference between the arithmetic mean of the returns of the small company portfolios and the arithmetic mean of the returns of the large company portfolios. Naturally, the larger the sample, the better the approximation to the true value of this premium. Finally, the next mathematical expression captures the previously described method for the SMB premium:

$$\text{SMB} = \frac{1}{3}(R_{SV} + R_{SN} + R_{SG}) - \frac{1}{3}(R_{BV} + R_{BN} + R_{BG}) \tag{11}$$

Now, the next premium to calculate is the High-Minus-Low risk premium, which is just the difference between the average of the returns of the value portfolios and the average of the returns of the growth portfolios. Here, we are excluding the so-called neutral portfolios (portfolios composed of neutral stocks) since they would not capture the effect of the value anomaly. Then, mathematically, the premium will be:

$$\text{HML} = \frac{1}{2}(R_{SV} + R_{BV}) - \frac{1}{2}(R_{SG} + R_{BG}) \tag{12}$$

According to Fama and French, the creation of this risk-premiums will capture all in one, the effects of the categorization of the stocks and avoid multicollinearity problems, since we are reducing dimensions of analysis by creating a single factor to include the formulation of the two hypotheses. Now, we can use the Fama-MacBeth regression to formulate the Three Factor Model:

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_i^S \,\text{SMB} + \lambda_i^V \,\text{HML} + \varepsilon_i \tag{13}$$

Now, what we are trying to identify is whether the estimated coefficients of the regression are statistically significant to prove that the size hypothesis and the value hypothesis are true. As always, the coefficient β_i will capture the effects of the systematic risk that the stock is exposed to, λ_i^S will capture the size effect of the company, and λ_i^V will capture the value effect. As usual, we include the α_i coefficient to account for the additional effects on the expected returns of an asset that are explained by variables not included in the model. Thanks to these models, investors can determine the impact that non-systematic risks have on their portfolios. The SMB and HML risk premiums capture the effects of company characteristics on expected returns, which can be very useful for creating new strategies based on specific characteristics preferred by investors when building their portfolios.

2 Carhart: Four Factor Model

Some may consider the inclusion of the Carhart Four-Factor Model in the same chapter of this guide that explains the Fama and French factor model to be somewhat adventurous, given the controversies surrounding the results of the classical Fama and French Three-Factor Model and those obtained by Carhart. Carhart's predecessors commonly assumed that markets were generally efficient, meaning that stock prices accurately reflected all factors affecting them. However, Carhart was not a staunch proponent of market efficiency, and many other researchers, including Carhart himself, began analyzing the existence of the momentum anomaly.

2.1 The Momentum Anomaly

In Carhart, 1997, the author proposed that investors do not immediately react to exogenous shocks in the market; instead, it takes some time for them to adjust their investment strategies. As a result, today's prices often reflect 'lags' of the market movements from previous days. Investors may exhibit overconfidence in high-performing stocks, believing they can 'catch' the rallies even after some delay. Alternatively, this behavior could also be explained by investors underestimating certain stocks, waiting for further confirmation before making significant adjustments to their strategies.

Momentum can be defined as the persistence of past returns. Naturally, one would expect that, after positive returns, asset prices would rise as more investors are attracted to these stocks. However, this tendency would eventually correct itself when some investors start selling to realize profits, causing returns to decline. Yet, in reality, market corrections do not always occur perfectly; investors' expectations can lead to prolonged stock rallies.

2.1.1 Measuring Momentum and Classifying Stocks

In financial literature, there is no consensus on how to measure the momentum of stock returns or financial assets. Depending on the context, the measurement of momentum can vary, and statistical differences may lead to diverse empirical observations. One of these diverse ways to measure the same phenomena, and commonly used to measure momentum if the building of factor models, is the difference of the 12-month sum of the logarithmic returns of an asset minus the 1-month sum of the logarithmic returns. Mathematically, let us define the logarithmic returns as r_t:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{14}$$

Then, the 12-month sum, is going to be the daily sum of the last 252 days of history for the stock's returns, while the 1-month sum is just the daily sum of the last 21 days of history:

$$R_{12M,T} = \sum_{t=T-252}^{T-1} r_t, \tag{15}$$

$$R_{1M,T} = \sum_{t=T-21}^{T-1} r_t \tag{16}$$

After differentiating both series, we obtain a new series that captures the momentum of a stock based on its historical performance. Note that $R_{12M,T}$ and $R_{1M,T}$ represent the sums of logarithmic returns at time T. We can compute the same measure for every time point in the past, up to the first 252 observations.

$$p_t = R_{12M,t} - R_{1M,t} \tag{17}$$

We use p_t to represent momentum just as physicists do, but do not confuse it with P_t, that we use for prices. Some experts might argue that smoothing the series using moving averages could be valuable, and they may be right. However, we will explore that option in a later section of this chapter. For now, this is the simplest and most direct way to measure momentum. Next, let's examine how we can classify stocks as Winners or Losers based on their momentum.

$$\text{Momentum}_i = \begin{cases} \text{Losers}, & \text{if } p_i < p_{P30} \\ \text{Middle}, & \text{if } p_{P30} \leq p_i \leq p_{P70} \\ \text{Winners}, & \text{if } p_i > p_{P70} \end{cases}$$

After the classification, to test whether momentum is a relevant phenomenon in explaining the returns of an asset, we can propose a dummy linear regression model and assess whether the coefficients of the classification variables are statistically significant. Mathematically this regression model has the following form:

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_{0i}\,\text{Losers}_i + \lambda_{1i}\,\text{Winners}_i + \varepsilon_i \tag{18}$$

If Carhart's hypothesis holds, we would expect, for example, that the coefficient λ_{1i} is significant and greater than zero, while λ_{0i} is also significant but smaller than zero.

2.2 Calculating the WML premium

The calculation of the Momentum Factor, or the Winners Minus Losers (WML) Factor, depends on the construction of two portfolios (see Fama and French, 2012 where you can find a common derivation): the winners' portfolio and the losers' portfolio. Unlike the factors in the Fama and French models, these portfolios do not need to be market capitalization-weighted and can instead be equally weighted. The Momentum Factor can then be estimated using the following expression:

$$WML = R_W - R_L \tag{19}$$

Note that, unlike the Fama and French models, we are not calculating the average returns separated by size. We do not need to create separate portfolios for Winners-Big and Winners-Small and then calculate their average returns. Carhart proposes a more direct method for obtaining the Momentum Factor. Once we have determined the risk premium based on momentum, we can introduce it into our factor models.

$$\mu_i - r_f = \alpha_i + \beta_i(\mu_M - r_f) + \lambda_i^S\,\text{SMB} + \lambda_i^V\,\text{HML} + \lambda_i^M\,\text{WML} + \varepsilon_i \tag{20}$$

The direction of the premium will also influence the interpretation of the coefficient λ_i^M. For example, if the premium is negative ($WML < 0$) we can say that there is a negative premium favoring the loser stocks rather than the winner stocks. In this case, the coefficient will capture how sensitive the stocks are to this negative premium. The statistical significance of the coefficient will then determine whether Carhart's hypothesis is supported or rejected.

Thanks to this factor, we can capture delays that might occur in the market due to maladjustment in investors' actions. The Momentum Factor serves as a useful approximation for the positive rallies or negative streaks that stocks can experience, which are often caused by incomplete information or delayed reactions in the markets, phenomena extensively studied in fields like behavioral economics.

3 The Fama-MacBeth Regression

Commonly, in the investment world, these factors (previously reviewed) are used to understand risk. Naturally, our intuition would say that assets offering higher returns do so because of higher risks. The risk premium attributed to each asset can then be explained by different risk factors. Size, Value, Momentum, and other variables must be understood as risk factors, and our objective is to understand how we can estimate these risk determinants. Continuing along this path, we are going to study the most important methodology used in factor models to assess these risk factors.

3.1 Two-Step Estimation

In Fama and MacBeth, 1973, the authors propose an estimation method that follows two main steps. The first step is the estimation of the β coefficients associated with the risk factors under study. In this case, what we are estimating is the sensitivity of the returns of the selected stocks to the variables chosen for analysis. Assuming a K number of factors, the first regression will follow the next expression:

$$\mu_{i,t} = \alpha_i + \beta_{i,1}F_{1,t} + \beta_{i,2}F_{2,t} + \cdots + \beta_{i,K}F_{K,t} + \varepsilon_{i,t} \tag{21}$$

For a universe of n stocks, we then need to estimate a set of n different regressions, yielding a total of $N \times (K+1)$ coefficients, since we are also estimating the α coefficient.

$$\begin{aligned}
\mu_{1,t} &= \alpha_1 + \sum_{k=1}^{K} \beta_{1,k}F_{k,t} + \varepsilon_{1,t} \\
\mu_{2,t} &= \alpha_2 + \sum_{k=1}^{K} \beta_{2,k}F_{k,t} + \varepsilon_{2,t} \\
&\vdots \\
\mu_{N,t} &= \alpha_N + \sum_{k=1}^{K} \beta_{N,k}F_{k,t} + \varepsilon_{N,t}
\end{aligned}$$

The main output of this set of regressions is the estimated coefficients, which the authors propose to use in order to calculate the actual risk premiums attributed to each factor under study. How? Fama and MacBeth suggest employing these coefficients in a new cross-sectional regression. Note that the time-series regressions in the first step yield a single beta for the entire time period, and it is precisely these estimated betas that serve as the inputs (our explanatory variables) for the cross-sectional regression.

$$\mu_{i,t} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{i,1} + \gamma_{2,t}\hat{\beta}_{i,2} + \cdots + \gamma_{K,t}\hat{\beta}_{i,K} + \upsilon_{i,t} \tag{22}$$

The cross-sectional regression is estimated for each date available in our dataset. Thus, if we have a time-series of length T periods, we will calculate T regressions and obtain a total of $T \times (K+1)$ coefficients. Naturally, we expect the number of periods to be greater than the number of assets ($n < T$). Therefore, the output of the second step will be n time series, each with T periods.

$$\begin{aligned}
\mu_{i,1} &= \gamma_{1,0} + \sum_{k=1}^{K} \gamma_{1,k}\hat{\beta}_{i,k} + \upsilon_{i,1} \\
\mu_{i,2} &= \gamma_{2,0} + \sum_{k=1}^{K} \gamma_{2,k}\hat{\beta}_{i,k} + \upsilon_{i,2} \\
&\vdots \\
\mu_{i,T} &= \gamma_{T,0} + \sum_{k=1}^{K} \gamma_{T,k}\hat{\beta}_{i,k} + \upsilon_{i,T}
\end{aligned}$$

Then, we calculate the risk premium for each period of the series. All the γ coefficients represent the excess returns attributed to each of the risk factors under study. In other words, we obtain a time series for each of the risk factors considered. Finally, after running the regressions, the estimated risk premiums are obtained by computing the time average of these coefficients.

$$\bar{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T} \gamma_{t,k} \tag{23}$$

The relevance of this regression lies in assessing whether the time average of the risk factor coefficients is statistically significant. If it is, we have identified a relevant variable for explaining the excess returns of a risky asset. If no significance is observed, the factor does not contribute to explaining superior returns. In more advanced methodologies, analysts often use time-varying β coefficients in the cross-sectional regression. In this case, they calculate rolling β coefficients in the time-series step, resulting in a unique coefficient for each date in the series.

References

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, *105*(3), 457–472.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, *81*(3), 607–636.