Chapter 4

An Introduction to Multi-Factor Models

- *What will you learn here:* What are multi-factor risk models; how many different types of models are there; and how are they estimated? Most importantly, how do they allow you to separate two worlds: the world of factors from that of company-specific returns?
- *Why does this matter:* Because it is the skeleton that supports the whole thing. Even if you do not see it, it is there and it is essential.
- *When will you need this:* Always and never. Always, because you will consume its byproduct daily; (almost) never, because you won't have to think about "factor portfolios", or cross-sectional regressions unless you really miss them.

4.1 From One Factor to Many

We have seen that alphas do not come from straightforward regressions. What about betas? Is the market beta the beta to end all betas? It turns out that this is just the beginning of the story. The single-beta factor model was proposed independently by Lintner and Mossin, and Sharpe, who gave it the now-standard name Capital Asset Pricing Model (CAPM), in the mid-1960s. It was immediately put to the test. Initial empirical studies confirmed the model [Black et al., 1972; Fama and MacBeth, 1973]. In the mid-1970s, three separate contributions by young researchers set the stage for a second revolution. First, Stephen Ross, then an assistant professor at Yale, extended the CAPM. His starting point is to assume that there is a small number of factors, compared to the number of assets, which act on the returns, not unlike the CAPM. Compare Figure 4.1 to Figure 3.2.

$$r = \alpha + \beta^1 \times f_1 + \ldots + \beta^m \times f_m + \epsilon \tag{4.1}$$

The betas in the Eq. (4.1) are termed *loadings* of a specific stock to a factor; this term is borrowed from the statistics literature on factor

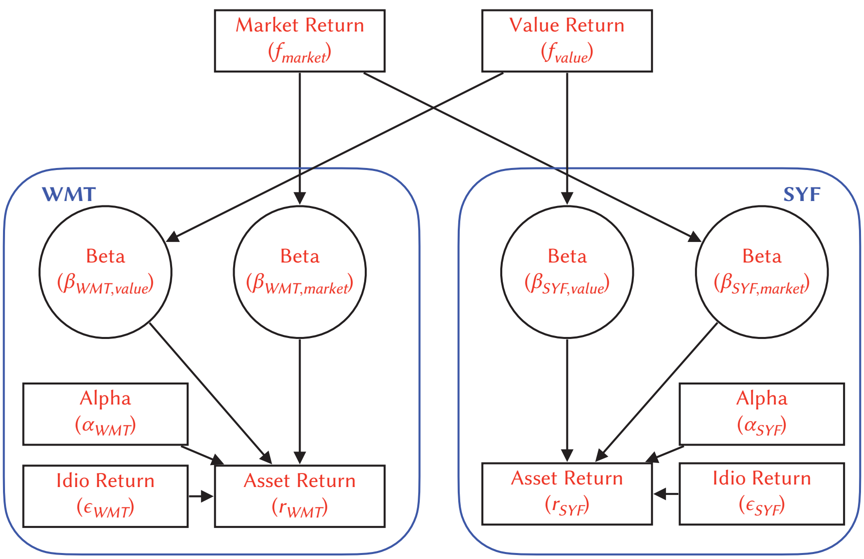

Figure 4.1 Analogous to Figure 3.2 but for two factors.

models. That same year, Barr Rosenberg (then an associate professor at Berkeley) and Vinay Marathe [Rosenberg and Marathe, 1976] proposed that loadings be *characteristics* of a stock. For example, we could create a financial ratio like cash flow to price (a profitability measure) for each company, and use it as a factor loading. The interpretation of the "cash flow to price" return then is similar to that of the market: when this return is positive, companies with high ratios have, everything else being equal, higher expected returns than companies with low ratios. There is one important difference, though: in the case of the CAPM, we could observe the market factor. Not so for the profitability factor. It is not a macroeconomic time series, nor a portfolio. No one "sees" this factor. However, there is a way around this problem. The basic idea is simple and powerful. We can build a "profitability" portfolio, i.e., a portfolio whose returns are very similar to the true profitability return. We will revisit this idea in this book, because it has applications for fundamental investors, too.

The last leg of the multi-factor stool came from Chicago. A Swiss assistant professor, Rolf Banz, then just starting at Northwestern, published part of his thesis [Banz, 1981]. This time, the challenge to the CAPM was empirical: there was evidence that small-cap stocks outperformed large-cap stocks; i.e., stocks exhibited different expected returns based on an attribute different than the market beta: market capitalization. Instead of one factor, we have at least two: a market-based one and a market-cap-based one. The formula giving the return of a stock now looks like this,

$$r = \alpha + \beta^M r^M + \beta^{BMS} r^{BMS} + \epsilon$$

where "M" stands for "Market" and "BMS" stands for "Big Minus Small". The study was exceptionally influential.[1] The papers of Ross and Banz together have more academic citations than there are seats in an average soccer stadium. Barr's papers are not nearly as cited, but that is because Rosenberg soon left academia to start a number of successful companies: Barra (now MSCI), Barclays Global Investors (now part of Blackrock), and AXA Rosenberg. Ross and

[1] In an interesting development, this original challenge was itself challenged! Today, the original "size effect" is not considered a real effect. A survey is [Alquist et al., 2018].

Banz had successful investment careers as well. The commercial factor models sold by Barra introduced conceptual innovations, but also broke a technological barrier. In 1980, there were approximately 5,000 listed companies in the US; a covariance matrix of these assets has approximately $(5000 \times 5000)/2$ distinct estimates.[2] That is 12.5M numbers, or about 50MB of data after compression. Even if the matrix had been generated once a month (daily models are an invention of the new millennium), that would have been a lot of data to transfer at a rate of a few KB/sec. Alternatively, you would have had to send a large pizza tape, or *a lot* of floppy disks to your clients. But with a reasonable model, say of 50 factors, the data size of the model would have been 50x smaller: by far the large piece of data would have been the stocks' characteristics, many of which are 0/1 numbers. Commercialization became possible.

Since 1981, there has been an explosion of factors explaining equity returns. Moreover, the factor approach has been applied successfully to other asset classes like government bonds and corporate credit. How are the factors discovered? And where do the betas come from? There are broadly three ways to attack this problem and we already saw two of them:

1. *The time-series approach.* We are surrounded by time-series data. Macroeconomic indicators, inflation, interest rates, economic growth, employment data, business cycle indicators. Economic sentiment data, like consumer confidence. Then, there are detailed economic activity data: same-store sales; purchasing manager indices; or international vs. domestic activity. Commodity prices (esp. oil) also affect economic activity. All of these data are factor candidates. You estimate the sensitivity (or beta) of a stock to these factors by performing a time-series regression of the stock return to the time series. The factor returns (the time series) are known; the betas are derived using asset and factor returns.
2. *The fundamental approach.* This is the method pioneered by Barr Rosenberg. Each stock has characteristics. These characteristics describe the company and its returns. We have to identify the relevant characteristics, which serve as the betas of the model. We

[2] The matrix is symmetric, so you only need the top-right half of the matrix.

then need to populate these fields for every stock, every characteristic, and every period. In this model, the betas are the primitive information, the factor returns are derived: given the betas and the asset returns, you can estimate the factor returns.

3. *The statistical approach.* In this method, you are only given the time series of asset returns. Both the beta and the factor returns are estimated using the asset returns alone. This may seem magical. Don't we need one or the other? The approach side-steps the problem by choosing betas and returns so as to describe most of the variation in the asset returns.

Each of these approaches has its merits and drawbacks, so that none of them dominates all of the others. In fact, risk models often use a mix of two or even three of them. The characteristic model has the benefit of being interpretable by the managers. The industry to which a company belongs is more like a simple yes/no flag than a time-series beta to an "industry return". These models can be extended with new characteristics and perform quite well in practical applications. Because of these two decisive advantages the fundamental (or characteristic) method is by far the most used model by fundamental managers. Yet, the time series model is necessary whenever the data are in the form of time series. In this case, we estimate time series betas. Time-series-based models are more popular in the academic literature than in the real world, both because of interpretability and performance concern. Time-series betas are however used as additional characteristics in fundamental models. Finally, statistical models have the least data requirement of all; and they have comparable performance for volatility prediction. However, beyond the first factor (or, at most, the second one), they are difficult to interpret. In equity models, factors have hardly any meaning, aside from the first one, which can be viewed as a market factor. And yet, statistical models occasionally find applications for fundamental managers, too. They are used as "second opinions", and also to check whether the base model (which is almost always a fundamental one) is missing some systematic source of risk. We summarize the pros and cons of the three approaches in Table 4.1. In the remainder of this chapter, I will focus on fundamental models, since they matter to you the most. And of their many aspects (estimation, performance, and

Table 4.1 Comparison of different approaches to modeling risk.

Model	Data Needs	Performance	Interpretability
Time Series	Medium	Low	Medium-High
Fundamental	High	High	High
Statistical	Low	High	Low

historical development) I will cover just two, but, in my opinion, the most important for you:

- First, I provide a list of the most important factors, their interpretation, and why they matter to you. On each of these factors, tens of papers have been written, with explanations, variants and tests. I will skim the surface, and point to a few books that contain a more thorough discussion.
- Second, I believe it's useful to peek inside the black box of model estimation. As a fundamental investor, you will never estimate a model in your life; as a human, you will never paint a Van Gogh, or design a turbine, either, but you will appreciate them more by understanding the design principles that guided their creation. You already know that the inputs to the model are the stock characteristics (for each stock) and the asset returns. The output are many: factor returns, idiosyncratic returns, idiosyncratic volatilities, factor covariances…what are these things, and where do they come from is the subject of the next section.

4.2 ★Frequently Asked Questions About Risk

This section collects several common questions and "how-tos" for risk models. These questions really do come up all the time. You can skip it on a first reading, but whenever you have a practical question or doubt, please check this section first.

Q: *How can two models with different versions of the factors (e.g., z-scored vs. raw) be different for specific factors and equivalent in the aggregate?*

A: On an intuitive level, the characteristics in the two models contain the same information, so it makes sense that operationally they are equivalent. On a more rigorous level, try the following experiment in Excel: perform a univariate linear regression (i.e., one dependent variable, one independent variable, and the intercept). Take note of the R-squared of the regression. Now, replace the first variable by its centered version and repeat the regression. You will see that the regression coefficients have changed, but the R-squared hasn't. The predictive power of the regression doesn't change if you mix and match the variables. The same principle is at work in a factor model.

Q: *What is the unit of measure of the loadings? Of the exposures? Of the portfolio volatility?*

A: Loadings are pure, dimensionless numbers. Exposures are numbers times positions in local currency, so they are in currency units. Volatility is in currency *or* in percentages. A \$10M annualized volatility means that the annual range of variation of the portfolio is \$10M; this number is independent of the Gross Market Value of the portfolio. An annualized volatility of 10% for a portfolio with a gross market value of \$1B means that the annual range of variation is 10% × \$1B, or \$100M.

Q: *How do I convert the annualized volatility of a portfolio into a weekly or a monthly volatility?*

A: Recall that the variances of independent random variable sum up. The annual variance of a portfolio is the sum of the weekly variances: variance(yearly) = variance(week 1) + variance(week 2) + … + variance(week 52) = 52 × variance(weekly). The volatility is the square root of the variance, therefore

$$(\textit{yearly volatility}) = \sqrt{52} \times (\textit{weekly volatility})$$

The weekly volatility is the annual volatility divided by $\sqrt{52}$. There are approximately 252 trading days in a year; by a similar reasoning, the daily volatility is the annual volatility divided by $\sqrt{252}$. Formulas for monthly, quarterly or multi-year vol are derived similarly.

Q: *Is the daily volatility equal to the average absolute daily return?*

A: No. If we assume that a return r is normally distributed with standard deviation σ and mean zero, the absolute return follows the distribution of a *folded normal distribution*. The expected value is given by [Leone et al., 1961] by

$$\begin{aligned}(\textit{average absolute daily return}) &= \sqrt{\frac{2}{\pi}} \times (\textit{daily volatility}) \\ &\simeq 0.8 \times (\textit{daily volatility})\end{aligned}$$

so it is slightly smaller than the daily volatility. In practice, since we saw that both factor returns and idiosyncratic returns are heavy-tailed, the average absolute daily return is smaller than 80% of the daily volatility.

Q: *Why can't I estimate expected returns from historical returns using some law of large numbers? If I use hourly returns instead of daily returns to increase the size of my dataset, can I get a more accurate estimate?*

A: You can't hope to increase the accuracy of your historical estimates from historical data. Say that the returns of a stock are normally distributed, have expected annualized returns α and annualized volatility equal to σ. If you split one year in n intervals (252 for daily returns; 1640 for hourly returns; 9800 for 10-minute returns), then the expected return in one period is α/n and the volatility is $\sigma/\sqrt{n}$ (as we have seen from the previous question about volatility scaling). When you take the average of the n returns, the expected value of the per-period return is α/n; the standard error of the estimate is equal to the standard error of the return around the mean (i.e., $\sigma/\sqrt{n}$) divided by the square root of the number of observations, as you may recall from introductory statistics. Hence the estimation error is

$$(\sigma/\sqrt{n})/\sqrt{n} = \sigma/n$$

So our 95% confidence interval for the expected return is

$$\frac{\alpha}{n} \pm 1.96\frac{\sigma}{n} = \frac{1}{n}(\alpha \pm 1.96\sigma)$$

The greater the value of n, the smaller the value, as expected. But the relative size of the error doesn't depend on n! The signal-to-noise ratio, defined as the ratio of the expected value of the estimate to its standard error, is α/σ; and it is reasonable to expect α to be much smaller than σ. Even if we have a very large number of observations, the accuracy does not improve. The problem is that, as we increase the number of observations n, we have to estimate a fainter signal: not α, but α/n.

Q: *Your model predicts a volatility that is 50% my current realized volatility. But I have been given by my hedge fund a risk budget based on predicted volatility. It is very difficult for me to make money!*

A: Usually this is occurring in the aftermath of a large volatility event: think September 2008, November 2016, or March 2020. Volatility models still rely on recent history; and if history shows a big spike in returns, it takes at least three to six months for a model suited for fundamental investors to return to normal levels. If the underestimation occurs in "normal" times, then there could be issues with the model specification.

Q: *Your model predicts the wrong market beta.*

A: There could be several reasons for this. First, the predicted beta might not be to your regional benchmark, but to a custom benchmark based on the estimation universe of the risk model. Sometimes this is not made sufficiently clear. My preference is to enable the PM with her favorite benchmark, provided that it satisfies a few requirements: it has to be sufficiently broad and representative of the market; and it has to be sufficiently liquid, via an ETF or a future, or both. Secondly, the beta is estimated with respect to a risk model, which in turn may not be adequate for the purpose. For example, using a global model (i.e., representing all tradable world assets) to estimate the beta of a European portfolio to the STOXX50 is not advisable. Better choose a European model.

Q: *Your model predicts a beta that is different from the Bloomberg beta.*

A: The Bloomberg beta (*BETA* ⟨*GO*⟩) is a historical beta. It is estimated by performing a linear univariate regression of an asset's returns against the benchmark returns. All the observations are equal weighted. The model beta is estimated based on the model's factor covariance matrix and is a *predicted beta*. The factor covariance matrix and the idiosyncratic volatilities are estimated using more sophisticated methods than equal weighting. The historical beta is a less accurate measure of forward-looking sensitivity for several reasons. First, while the risk model is also based on historical data, it weights past observations using a weighting scheme that discounts observations further in the past, and is optimal for the forecast horizon of interest (e.g.) two months out. Historical betas are equal-weighted, with no horizon optimization. Second, both factor and idio volatilities are estimated separately, with different weighting schemes. Third, modelers apply several advanced adjustments for the estimates that capture time-series dependencies and changes in volatility regimes. Taken together, these methods make a difference.

Q: *In Section* 3.4.2 *(Table* 3.6*) you say that the idiosyncratic volatility of the benchmark is essentially zero. But why?*

A: Let us consider the easy part first: the beta of the benchmark is by definition, one, because the beta is computed with respect to the benchmark itself. Now, for the idiosyncratic volatility. The sum of the weights of the constituents is one. By Pythagoras's Theorem (again, in Section 3.4.2), the squared volatility of the portfolio is the sum of the squared volatilities of its constituents:

$$(\textit{weight stock } 1)^2 \times (\textit{idio vol stock } 1)^2 + \dots \\ + (\textit{weight stock } n)^2 \times (\textit{idio vol stock } n)^2$$

Now, take the largest idiosyncratic volatility among the constituents, and denote it $\sigma_{\max}$. The sum above is smaller than

$$[(\textit{weight stock } 1)^2 + \dots + (\textit{weight stock } n)^2]\sigma^2_{\max}$$

If sum of the weights is one, then the sum of the squared weights is much smaller than one. For example, if the weights are all equal to $1/n$, then the idiosyncratic variance is equal to $(1/n^2 + \dots + 1/n^2)\sigma^2_{\max} = \sigma^2_{\max}/n$. The idiosyncratic volatility is $\sigma_{\max}/\sqrt{n}$. Since n is usually at least equal to 50, and can be as large as several thousands, then the idiosyncratic vol is small. In practice, on one side the weights are not all identical, and this makes diversification worse; but on the other, stocks with the largest weights are the large-cap companies, with lower idiosyncratic risk, and therefore their contribution to the sum is damped.

Q: *I know linear algebra and statistics and want to learn more about the details of the estimation process. Can you recommend a reference?*

A: The short survey paper [Connor and Korajczyk, 2010] (also available online) is a good starting point. The book by the same authors et al. [Connor et al., 2010] is suitable for self-study and as a reference. The collections [Jurczenco, 2015; Litterman, 2003] are also useful.

4.3 ★The Machinery of Risk Models

As we mentioned, the distinctive feature of characteristic models is that they use stock features (or characteristics) as primitives. Characteristics are numerical values: each stock, at each date, and for each characteristic type (the "factor"), has one specific value. For each date, each factor has an associated return. For example, consider the market beta of a stock as its characteristic. The associated factor is the "market", proxied by a weighted portfolio of the traded stocks; and the market return is just the return of that portfolio. For the size factor, the characteristic is the order of magnitude of the market capitalization (e.g., the logarithm of the market cap) of a stock, and so on. Characteristics are intuitive, factor models less so. The goal of this section is to give an intuitive explanation of how factor models are constructed and estimated.

Perhaps the easiest way to think about a factor model is as a *superposition* of effects on the stocks. A wave breaking on the shore does not have a perfect, sinusoidal shape. It is made of a large wave, and then of a few smaller waves riding on it, and then many ripples on top of the smaller waves. These effects *sum up*. Similarly, stock returns are the effect of a large shock (the market), then a few smaller ones (sectors, the larger style factors), then a few even smaller ones. And then there are shocks that are specific to each stock only and that need to be estimated separately. Like waves, many of these movements have similar amplitude over time. In technical terms, we say that their volatility is persistent. This is a blessing, because it allows us to estimate future volatilities from past ones. In some cases, however, the volatility of a factor is not very persistent: in the case of the short interest, hedge fund holdings or of momentum, the factors experience large returns; you can think of these large returns as the tsunami that will wreak havoc on your portfolio if you are not prepared for it. We will cover them in detail in the next chapter.

There are many steps in the generation of a fundamental risk model. The major tasks involved are:

1. On each date, in advance of market open, an up-to-date list of characteristics of the stocks are received from data vendors, checked for data integrity, and preprocessed; for example, certain loadings are scaled and centered, so that average of the loadings over a certain investment universe is zero, and their standard deviation is one. These characteristics are arranged into a *loadings matrix* $\mathbf{B}$ with a number of rows equal to the number of assets, and a number of columns equal to the number of characteristics.
2. Factor returns and idiosyncratic returns for the previous day are estimated using a cross-sectional linear regression. This means that the stock returns are the dependent variable (the "y" in a regression equation) and the matrix of stock characteristics are the independent variables. The estimates – the regression coefficients – are the factor returns; the list of factor returns is more or less universally denoted by $\mathbf{f}$. The residuals of the regression are the idiosyncratic returns (or residual returns) of the stock; and the vector of residuals, like the symbol for residuals in undergraduate statistics courses, is denoted by the symbol ϵ.
3. Once the latest factor and idiosyncratic returns are estimated, we enrich the existing time series of these returns and, using the entire time series, estimate the factor covariance matrix and the individual idiosyncratic volatilities.

As a byproduct of the second step, the estimation procedure also generates portfolios, one for each factor. These portfolios have returns equal to the factor returns; hence the term *factor-mimicking portfolios* (FMPs). They play an important role for performance attribution and hedging. The process is exemplified in Figure 4.2. These steps are repeated every day.

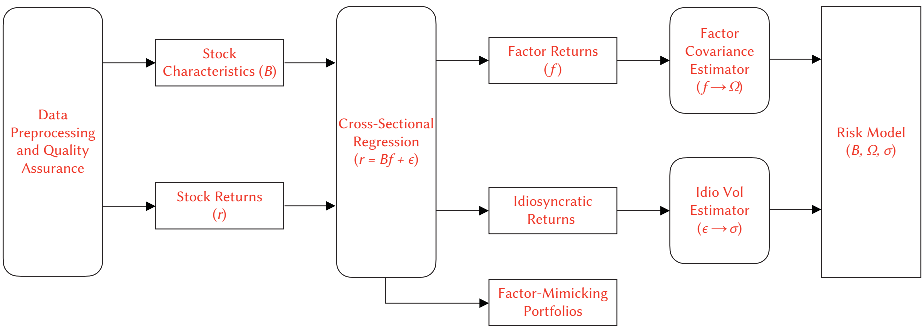

Figure 4.2 Steps needed to generate a risk model.

4.4 Takeaway Messages

There are three types of factor models:

1. *Fundamental;*
2. *Time series;*
3. *Statistical.*

Fundamental models rely on characteristics of the individual stocks and are commonly used by fundamental investors. They describe two features of returns:

1. Their *expected returns;*
2. Their *risk.*