Chapter 1

Factor Models and General Definition

COPYRIGHTED MATERIAL

1.1. Introduction

Today, the linear factor model is a benchmark in portfolio management theory [MAR 52, LIN 65], and arbitrage pricing theory (APT) [ROS 76, ROL 80]. In practice, factor models are also used widely to understand the cross-section dispersion of asset returns, whatever the asset class is. Even in the hedge fund industry, where returns feature high nonlinearity, this approach is largely implemented, in general with some improvement such as nonlinear factors or time-varying parameters. This chapter introduces not only the common version of linear factor models but also discusses its limits and the developments described in the following chapters of this book.

In section 1.2, we introduce the different notations and discuss the model and its structure. We list in section 1.3 the reasons why factor models are generally used in finance, and discuss the limits of this approach. Section 1.4 describes the different steps in the building of factor models, i.e. factor selection and parameter estimation. This section is a direct introduction to Chapter 2 for the factor selection step, and Chapters 3 and 4 for the parameter estimation step. Finally, section 1.5 concludes the chapter by giving a historical perspective on the use of factor models in finance.

1.2. What are factor models?

1.2.1. *Notations*

We first consider a set of N risky assets, indexed $i = 1, \ldots, N$. We denote by $s_{i,t}$ the price of the asset i at time t, and $r_{i,t}$ the corresponding return for the period $(t-1, t)$. This return is defined by:

$$r_{i,t} = \frac{s_{i,t}}{s_{i,t-1}} - 1. \quad [1.1]$$

If prices are observed on a daily (respectively, weekly, monthly, etc.) basis, then $r_{i,t}$ represents daily (respectively, weekly, monthly, etc.) returns. We then denote by μ_i the expected return of asset i, and σ_i^2 its variance:

$$\mu_i = \mathbb{E}(r_{i,t}) \text{ and } \sigma_i^2 = V(r_{i,t}). \quad [1.2]$$

The variance σ_i^2 (or the volatility σ_i) is often used to measure the risk relative to the asset i. The greater the volatility, the greater is the risk. The covariance $\sigma_{i,j} = Cov(r_{i,t}, r_{j,t})$ between assets i and j will be useful to compute the risk of a portfolio of several assets.

As opposed to risky assets, we also consider a risk-free asset that gives a return r_f, called the risk-free rate (supposed to be constant). In practice, short-term government securities such as US Treasury Bills are used as a proxy for the risk-free asset; this results in a non-constant r_f.

The expected excess return (or risk premium) of asset i is the return we can expect from asset i in excess of r_f, that is

$$\mathbb{E}(r_{i,t}) - r_f, \quad [1.3]$$

and is the premium relative to the risk taken when investing in asset i.

According to Markowitz's [MAR 52] theory of *mean-variance efficiency*, investors (should) therefore require a higher expected return for holding a more risky asset and want to earn the highest possible return for a level of risk that they are willing to take. A portfolio is said to be mean-variance efficient if we cannot create another portfolio with a greater expected return and the same variance of returns (or another portfolio with a lower variance and the same expected return).

Figure 1.1 illustrates the time series evolution of three risky assets in comparison to an asset that is a proxy for the risk-free asset. Their mean-variance trade-offs are given in Table 1.1.

Name	$\mu(\%)$	$\sigma(\%)$	$(\mu - r_f)/\sigma$
T-Bill 1Y (r_f)	2.93	1.19	-
HFRI ED Index	8.91	7.1	0.84
S&P500	4.28	14.69	0.09
APPLE	58.3	41.05	1.35

Table 1.1. *Annualized return (μ), annualized volatility (σ) and return-to-risk ratio (Sharpe Ratio) calculated between Feburary 2003 and November 2009 for the four financial instruments shown in Figure 1.1 and ordered by increasing level of risk*

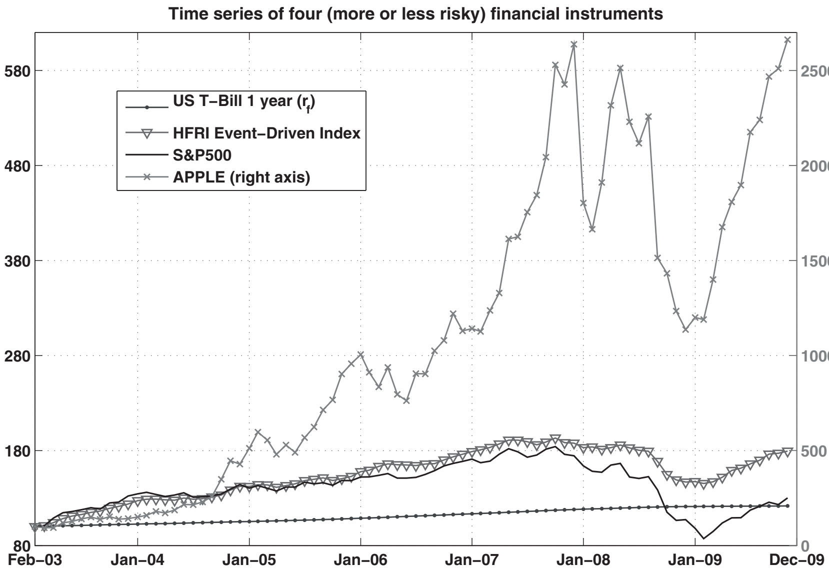

Figure 1.1. *Monthly evolution of four time series (in US dollars and based at $100 at the end of January 2003) representing: (1) an investment in the US T-Bill 1 year, a 1 year maturity US Treasury Bill, considered as a proxy for the risk-free asset, (2) the NAV of HFRI ED index, an index of event-driven (ED) hedge funds, (3) the price of the S&P500, a US market index and (4) the price of the equity APPLE (on the right axis)*

The net asset value (NAV) gives the evolution over time of $100 invested in each asset. We can easily compute the NAV from the arithmetic returns with:

$$NAV_{i,t} = NAV_{i,t-1}\left(1 + r_{i,t}\right), \forall i = 1, \cdots, N \text{ and } \forall t = 1, \cdots, T$$

with $NAV_{i,0} = 100$ (end of January 2003).

1.2.2. *Factor representation*

A factor model is a multivariate regression linking the returns of a set of risky assets to several factors. We focus our attention on linear factor models where the relationship between factors and returns is linear.

Factors represent fundamental data, statistical factors or specific portfolios. Fundamental data are specified using economic theory and the knowledge of financial markets and include macroeconomic variables such as the inflation rate, the unemployment rate and the gross national product. Typically, macroeconomic variables are correlated.

To select uncorrelated factors, empirical dimension reduction techniques, such as factor analysis (FA) or principal component analysis (PCA), are performed on the covariance matrix of the returns of the risky assets. It gives rise to *eigenfactors* arising from an eigenvalue decomposition of the covariance matrix.

However, factors can be substituted by specific portfolios, especially if they represent different strategies that an investor can pursue at a low cost. The factors represent the various sources of risk present in the market to which an investor is exposed.

We now introduce the following linear factor model specification. The returns of a set of N risky assets indexed by $i = 1, \cdots, N$ are assumed to be wide-sense stationary, and can be expressed, for $t = 1, \cdots, T$, by:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\,\mathbf{f}_t + \boldsymbol{\epsilon}_t, \qquad [1.4]$$

where:

– $\mathbf{r}_t = [r_{1,t}, \cdots, r_{N,t}]'$ is the N-dimensional vector of the risky asset returns at time t;

– $\mathbf{f}_t = [f_{1,t}, \cdots, f_{K,t}]'$ is the K-dimensional vector of values of *common risk factors* at t whose covariance matrix is $\boldsymbol{\Sigma}_{\mathbf{f}}$;

– $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_K]$ is a $N \times K$ matrix where each element $b_{i,k} \in \mathbb{R}$ defines the exposure (or sensitivity) of the asset i to risk factor k. The sensitivities of asset i to the K factors are the K-dimensional row-vector $\mathbf{b}_i' = [b_{i,1}, \cdots, b_{i,K}]$. It is also referred to as the *beta*;

– $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_N]'$ denotes the N-dimensional vector of intercepts and is called *alpha*; and

– $\boldsymbol{\epsilon}_t = [\epsilon_{1,t}, \cdots, \epsilon_{N,t}]'$ is the N-dimensional vector of the zero-mean asset-specific residual returns whose covariance matrix is $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

In [1.4], the unknown $N \times (K+1)$ matrix of parameters is $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_N]' = \left[\boldsymbol{\alpha}\ \mathbf{B}\right]$ and the inputs of the model are $\mathbf{r}_t$ and $\mathbf{f}_t$. Let $\boldsymbol{\epsilon}_i = [\epsilon_{i,1}, \cdots, \epsilon_{i,T}]'$ and $\mathbf{f}_k = [f_{k,1}, \cdots, f_{k,T}]'$ denote the T-dimensional vectors of, respectively, the residual returns of asset i and the values of factor k.

Additional assumptions are made for [1.4]:

A1) The residual returns are uncorrelated with each of the factors: $\mathbb{E}(\boldsymbol{\epsilon}_i\, \mathbf{f}_k') = \mathbf{0}_T$, $i = 1, \cdots, N$, $k = 1, \cdots, K$.

A2) The residual returns are temporally uncorrelated: $\mathbb{E}(\boldsymbol{\epsilon}_{t_1} \boldsymbol{\epsilon}_{t_2}') = \mathbf{0}_N$ for $t_1 \neq t_2$.

A3) The residual returns are uncorrelated, that is $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is a diagonal matrix.

Assumption A3 means that the only sources of correlation among asset returns are those that arise from their exposures to the factors and the covariances among the factors. Residuals of asset returns are assumed to be unrelated to each other and hence totally *specific* to each asset. In other words, the risk associated with the residual return is *idiosyncratic* to the asset in question.

Note that the covariance matrix of the asset returns specified by [1.4] is:

$$\mathbf{\Sigma_r} = \mathbf{B}\,\mathbf{\Sigma_f}\,\mathbf{B}' + \mathbf{\Sigma}_\epsilon. \qquad [1.5]$$

The right-hand side of [1.5] consists of two distinct terms: $\mathbf{B}\,\mathbf{\Sigma_f}\,\mathbf{B}'$ is called the *systematic risk*, that is the risk explained by the K common factors and also known as *non-diversifiable risk*, *beta risk* or *market risk*.

The second term, $\mathbf{\Sigma}_\epsilon$, is called the *idiosyncratic risk*, also known as *diversifiable risk* or *asset-specific risk* and is totally specific to the assets.

As opposed to the *non-diversifiable systematic risk*, an investor has the possibility to reduce the amount of the *asset-specific risk* by properly diversifying his (or her) investments. Note also that $\mathbf{\Sigma_f}$ is diagonal when *eigenfactors* are selected.

When N is large, using [1.5] helps to narrow down the dimensionality to estimate the covariance matrix $\mathbf{\Sigma_r}$. On the one hand, the number of terms that must be estimated reduces significantly and on the other hand, $K < T$ usually meets the requirement to obtain an invertible estimated covariance matrix.

1.3. Why factor models in finance?

1.3.1. *Style analysis*

Let us consider a portfolio of n risky assets whose returns are $r_{i,t}, i = 1, ..., N$. The portfolio returns are denoted by $r_{p,t}$ and are obtained from $r_{i,t}, i = 1, ..., N$ by the following formula:

$$r_{p,t} = \sum_{i=1}^{N} a_i r_{i,t} = \mathbf{a}'\mathbf{r}_t,$$

where $\mathbf{r}_t = [r_{1,t}, \cdots, r_{N,t}]'$ and $\mathbf{a} = [a_1, \cdots, a_N]'$ is the vector of (fixed) weights of assets in the portfolio. The distribution of $r_{p,t}$ depends directly on the joint distribution of the vector $\mathbf{r}_t$, which can be considered as the factors explaining the portfolio performance. However, these factors can be highly correlated and/or unobserved (when we do not know the portfolio manager investment universe). A more parsimonious and tractable representation is then obtained using a small number of observed factors correlated with the portfolio performance. However, we must consider in this case the potential error made in explaining portfolio returns with a set of "common" factors. This explains why we introduce error terms in the linear factor representation.

A given portfolio allocation basically reflects the portfolio manager's bets. If we assume that these bets remain unchanged over the whole observation period, then the approach described above is relevant. However, from a practical standpoint, these bets are in general time-varying. The portfolio manager reallocates his/her portfolio on a continuous basis, and only the average exposure to factors is obtained through this classical return-based style analysis (see [SHA 92] for greater details on this approach). The linear combination of factors exposures and their respective performance gives then the strategic or long-term portfolio benchmark. The value added by the portfolio manager (or market timing ability) is then defined as the return difference between this portfolio returns and the strategic long-term benchmark (see [DAR 12]).

As a portfolio manager's bets are made on a continuous basis, it could also be interesting to track their impacts on the portfolio performance in the short term. Tactical portfolio allocation decisions rely on short-term portfolio manager's forecasts of risk premia (bets on factors) and can, as a result, also be captured using a linear factor model, but with time-varying exposures. Since information arrives randomly, and tactical bets are assumed to be responses to new information, we expect the exposure to risk factors to evolve randomly over time. Equation [1.4] in this case becomes:

$$\mathbf{r}_t = \boldsymbol{\alpha}_t + \mathbf{B}_t \mathbf{f}_t + \boldsymbol{\epsilon}_t,$$

where $\mathbf{B}_t$ (respectively, $\boldsymbol{\alpha}_t$) denotes the time-varying exposure of assets to factors (respectively, alpha). Risk factor exposures are not directly observed and must be filtered from $\mathbf{r}_t$.

Let us consider the case of a portfolio of hedge funds (or fund of hedge funds). Hedge funds are used in the following chapters to give empirical illustrations. Although the trend is very clearly toward more transparency, investors do not systematically have access to the full composition of hedge funds, and their evolution over time. Fund of hedge fund managers themselves do not always have a complete view of the risk factor exposures of their underlying investments, and, as a result, of the bets they implicitly make. This is all the more true when the trading frequency of the underlying funds is significantly higher than their reporting frequency (i.e. embedded risks can be dramatically different from those shown at a specific date), or when the number and the diversity of positions make it difficult to come up with accurate aggregated factor exposures. Tactical bets explicitly (at the portfolio level) and implicitly (at the underlying level) made by the fund of hedge fund manager can alternatively add up or cancel each other. Using a return-based style analysis therefore allows us to mitigate one of the main shortcomings of holding-based approaches, by capturing and assessing both effects concomitantly.

This example shows that this return-based style analysis must be done with time-varying parameter to filter from portfolio returns both long-term and short-term bets.

1.3.2. *Optimal portfolio allocation*

Let us consider a portfolio of N risky assets whose returns are $r_{i,t}, i = 1, \ldots, N$, with N large. According to Markowitz's [MAR 52] theory, optimal diversified portfolios are obtained by inverting the covariance matrix of asset returns $\mathbf{\Sigma_r}$. This matrix is unknown, and we have to estimate it from observations $r_{i,t}, i = 1, \ldots, N, t = 1, \ldots, T$ by using, for example, the empirical covariance matrix:

$$\hat{\mathbf{\Sigma}}_{\mathbf{r},\mathbf{T}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_t - \mathbb{E}(\mathbf{r}_t))(\mathbf{r}_t - \mathbb{E}(\mathbf{r}_t))',$$

where $\mathbf{r}_t = [r_{1,t}, \cdots, r_{N,t}]'$ is the vector of asset returns. If the dimension T is lower than N, then $\hat{\mathbf{\Sigma}}_{\mathbf{r},\mathbf{T}}$ is not invertible, and the optimal Markovitz's portfolios cannot be computed. Moreover, even in the case where N is lower than T, this empirical covariance matrix can have a determinant close to zero, and we can encounter numerical problems when trying to invert it with the usual algorithms. This is, in particular, the case when correlations between assets are high. Factor models can be used to manage these high correlations and therefore give an alternative to the direct numerical inversion approach.

Let us develop these computations in the very simple case [1.4]. We assume in the following that the risk-free rate is set to zero and excess risky asset returns $\mathbf{r}_t$ satisfy the following linear single risk factor model:

$$r_{i,t} = b_i f_t + \epsilon_{i,t}, i = 1, \ldots, N,$$

where the unobserved single factor f_t is assumed to be Gaussian $N(m_f, \sigma_f^2)$ and $\epsilon_{i,t}, i = 1, \ldots, N$ have Gaussian distribution $N(0, \sigma_\epsilon^2)$. The expected excess return is $\mathbb{E}(r_{i,t}) = b_i m_f$ and the idiosyncratic risk σ_ϵ^2. The systematic source of risk (i.e. the unobserved single factor) creates an additional individual risk equal to $b_i^2 \sigma_f^2$, but also a covariance between excess returns of two different assets:

$$Cov(r_{i,t}, r_{j,t}) = b_i b_j \sigma_f^2,$$

for $i \neq j$. This covariance term can lead to correlation close to 1 when the systematic risk is high relative to the idiosyncratic risk. However, we can compute explicitly the mean-variance optimal allocation. The vector of efficient allocations in the N risky assets is indeed proportional to:

$$\mathbf{a}^* = V(\mathbf{r}_t)^{-1} \mathbb{E}(\mathbf{r}_t),$$

where $\mathbf{r}_t = [r_{1,t}, \cdots, r_{N,t}]'$. Using the factor structure, we can explicit the two first moments of $\mathbf{r}_t$ and get the following closed-form formula:

$$\mathbf{a}^* = (\sigma_\epsilon^2 \mathbf{I} + \sigma_f^2 \mathbf{B}\mathbf{B}')^{-1} m_f \mathbf{B} = \frac{m_f}{\sigma_\epsilon^2 + \sigma_f^2 \mathbf{B}'\mathbf{B}} \mathbf{B},$$

where $\mathbf{B} = [b_1, \cdots, b_N]'$. We then obtain the efficient allocation by estimating the parameters involved in the previous formula, without having to numerically invert the covariance matrix.

In this example, the single factor is unobserved and then must be filtered from asset returns, or replaced by a proxy that is able to explain all the correlation structure observed between risky assets.

1.4. How to build factor models?

1.4.1. *Factor selection*

The factor selection problem is not new in the financial literature. The main issue is related to the delicate balance between using too many or too few factors. On the one hand, adding too many factors lowers the regressors efficiency when we estimate factor exposures using equation [1.4]. On the other hand, working with too few factors also has an important risk of missing the correlation structure observed between asset returns. This raises the question whether it is possible to build a factor selection methodology allowing us to consider only the appropriate factors.

The first solution consists of using a predefined set of observable factors, already documented in the financial literature for their ability to explain the cross-section of asset returns. These factors are used to build long- and short-term benchmarks in the style analysis of section 1.3.1. They are also used to reduce the dimension of the covariance matrix, when optimal portfolio allocation must be computed for a large set of risky assets in section 1.3.2. If we miss some risk factors in the style analysis, both long- and short-term benchmarks are misspecified, and the portfolio manager's added value is not correctly calculated. In particular, we can interpret positive risk-adjusted returns as a manager's skills and omit an important risk exposure. In the portfolio allocation example, omitted factors imply residual correlations between idiosyncratic risk terms. As these correlations are not taken into account in the calculation of the optimal portfolio, we underestimate the portfolio risk or, in other words, we overestimate the diversification effect.

A second solution is to use statistical approaches to filter from the asset returns distribution (and, in particular, the covariance matrix) unobservable factors that are able to explain the cross-section of asset returns. If this approach seems appealing to a statistical point of view, it has also many drawbacks from a financial point of view. First, it is, in general, difficult to give an economic interpretation of these statistical factors. Second, the factor representation is not unique, and any linear combination of a given set of factors defines an equivalent factor model. Third, the factor decomposition is, in general, time-varying and then difficult to interpret. Both approaches are described in detail in Chapter 2.

1.4.2. *Parameters estimation*

Once a set of factors is chosen, parameters estimation consists of computing numerical values of factors exposures. When parameters are not time-varying, least squares (LS) approaches can be used to compute these estimators, depending on the residual properties. However, both in the style analysis of section 1.3.1 and the optimal portfolio allocation problem of section 1.3.2, parameters can be time-varying and more complex statistical filtering approaches must be used (see e.g. [BOL 09, PAT 13]) such as the Flexible Least Squares [KAL 89, MAR 04], the Kalman filter (KF) [RON 08a, RON 08b, RAC 10] or Markov switching regimes [BIL 12].

The LS estimator can always be computed on rolling windows, updated each time we get new data. This provides a time series of estimators that have then a time-varying behavior. However, this approach, if useful when changes in parameters over time are smooth, can be misleading when big changes occur. A style analysis computed using rolling window LS can miss a sudden style rotation decided by the portfolio manager. On the contrary, this estimation method provides very good results when style drifts are small and/or implemented step by step.

The KF approach is by definition more reactive and can capture in theory quick style rotations. However, this approach works well only when the relevant factors are used in equation [1.4]. Indeed, assumptions made up of the statistical behavior of the residual returns are more constraining, and any failure can have a huge impact of the filtered time-varying exposures. It is, for example, the case when residual returns do not follow a Gaussian distribution. Commonalities and discrepancies between LS and KF are discussed in Chapter 3, and an enhanced version of KF is discussed in Chapter 4.

1.5. Historical perspective

Factor models have been the focus of numerous studies in empirical finance since Treynor [TRE 62], Sharpe [SHA 64], Lintner [LIN 65] and Mossin [MOS 66] developed the capital asset pricing model (CAPM) in the 1960s.

1.5.1. *CAPM and Sharpe's market model*

CAPM lays the foundation for all the existing factor models. It gives the theoretical *equilibrium* relationship that should occur between the returns and the risks of individual assets with regard to the market returns.

Its conception is based on two fundamental financial concepts: market equilibrium that occurs if the amount of demand is balanced by the amount of supply (represented by the market portfolio), and, as a result, the mean-variance efficiency of the market portfolio.

Given the risk-free rate r_f and the risk σ_m of the market portfolio m, the CAPM stipulates that the returns we expect from individual assets in excess of r_f are given in proportion to the excess returns expected from the market, as follows:

$$\mathbb{E}(r_{i,t}) - r_f = \beta_i \left(\mathbb{E}(r_{m,t}) - r_f\right), \qquad [1.6]$$

where $\beta_i = \gamma_{im}/\sigma_m^2$ is related to the amount of risk given by i, with γ_{im} being the covariance between r_i and r_m. Larger values of β_i correspond to larger expected return and larger risk for asset i. The term $(\mathbb{E}(r_{m,t}) - r_f)$ is called market risk premium.

In practice, the unknown parameter β_i is estimated through the following univariate regression using historical data for the asset returns, the market portfolio and the risk-free rate:

$$r_{i,t} - r_f = \alpha_i + \beta_i \left(r_{m,t} - r_f\right) + \epsilon_{i,t}, \qquad [1.7]$$

where $\epsilon_{i,t}$ satisfies $\mathbb{E}(\epsilon_{i,t}) = 0$. Additional parameter α_i is the asset *alpha*. The strict form of [1.6] specifies that *alpha* must be zero and that deviation from zero is the result of temporary disequilibrium. With $\alpha_i = 0$ in [1.7], this model coincides with Sharpe's market model that is an output of [1.6]. The market portfolio is usually replaced by a major standard equity index (such as the S&P500) since Black [BLA 72] has shown that in market equilibrium, such market-value weighted indices are always efficient. The parameter α_i can be used to perform fitted performance analysis (see [DAR 10]).

CAPM results can be represented by the so-called security market line (SML) as shown in Figure 1.2. The SML is the plot of the expected return of any asset i as a function of its *beta*, as given by [1.6]. It is obtained for a fixed period of time through [1.6] where r_f is the mean of the risk-free asset returns, $\mathbb{E}(r_{m,t})$ is the mean of the returns for the proxy of the market portfolio and where some theoretical values for the β_is are chosen. For each risky asset i, α_i and β_i are estimated through [1.7] and $\hat{\mu}_i$ is the estimated mean of its returns. According to CAPM, given $\hat{\beta}_i$, if $\hat{\mu}_i$ does not lie on the theoretical SML, then the asset is mispriced.

In practice, this model has received several criticisms (e.g. see Roll's critique [ROL 77]). First, the market portfolio is unobservable. Standard equity indices usually substitute for it but do not reflect all the wealth in the economy as the market portfolio does. Second, the mean-variance approach depends only on the first two moments of the asset returns, which is too restrictive.

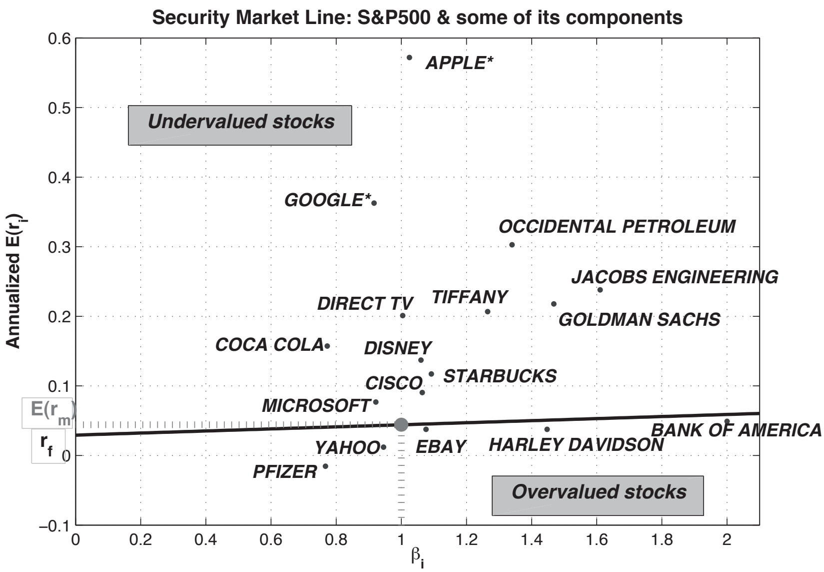

Figure 1.2. *The security market line (SML, the black line) shows the expected return of asset i as a function of its beta, as given by* [1.6]*, and given in annualized values over the period August 2004 / October 2010:* $r_f = 2.93\%$, $\mathbb{E}(r_{m,t}) = 4.41\%$, *and* $\sigma_m = 22.81\%$. *Here,* r_f *is the average rate of the US T-Bill 1 year over the whole period,* m *is the S&P500 and the risky assets are some of its components. Each (blue) point has for coordinates* $(\hat{\beta}_i, \hat{\mu}_i)$ *– the estimated values, for asset* i, *of* β_i *and* $\mathbb{E}(r_{i,t})$. *If* $\hat{\mu}_i$ *is lower (respectively, higher) than* $\mathbb{E}(r_{i,t})$, *obtained from* [1.6] *given* r_f, $\mathbb{E}(r_{m,t})$ *and* $\hat{\beta}_i$, *then the asset is said to be under (respectively, over) valued. Stocks with an * have significant (at 99.5%) non-zero* α_i

Finally, empirical studies often invalidate the CAPM because of its strong assumptions and show that more than one factor is necessary to identify the market risks.

1.5.2. *APT for arbitrage pricing theory*

Ross [ROS 76] and Roll and Ross [ROL 77, ROL 80] developed an alternative *arbitrage* model to the CAPM, called APT model. APT follows from two basic postulates:

P1) the risky assets follow a K*-factor structure* like [1.4] with three additional assumptions:

A4) $K << N$;

A5) $\mathbb{E}(f_k) = 0$, $k = 1, \cdots, K$ that leads to $\mathbb{E}(\mathbf{r}_t) = \boldsymbol{\alpha}$;

A6) $\exists\, s < \infty$ such that $\sigma^2_{\epsilon_i} = \mathbb{E}(\epsilon^2_{i,t}) \leq s^2$.

P2) pure *arbitrage* profits are impossible.

Pure arbitrage profits are risk-free profits at zero cost: an investor can earn a positive return on any combination of assets without undertaking risk and without making some net investment of funds. With P1 and P2, APT stipulates that the risk premiums of the risky assets are given through a linear combination of the factor *risk prices* weighted by the factor sensitivities $\mathbf{b}_i$ of [1.4]:

$$\mathbb{E}(r_{i,t}) = r_f + \sum_{k=1}^{K} b_{i,k}\,\lambda_k, \qquad [1.8]$$

where λ_k defines the *price of risk* for factor k. The full APT expression is then obtained by replacing α_i, $i = 1, \cdots, N$ in [1.4] by [1.8]. Jointly estimating the K-dimensional vectors $\mathbf{b}_i$ and $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_K]'$ is not trivial. In practice, a two-pass methodology can be used: first estimate $\mathbf{b}_i$ through [1.4] (giving rise to $\hat{\mathbf{b}}_i$), then use $\hat{\mathbf{b}}_i$ as input to estimate $\boldsymbol{\lambda}$ through:

$$r_{i,t} = r_f + \sum_{k=1}^{K} \hat{b}_{i,k}\,\lambda_k + u_{i,t},$$

where $u_{i,t}$ represents zero-mean residual errors with usual assumptions. Unlike the CAPM, APT does not provide any information about the nature of the K factors.

1.6. Glossary

Volatility

The volatility σ of an asset a refers to the standard deviation of the continuously compounded returns of a within a specific time horizon. Volatility is usually expressed in annualized terms: if the volatility σ is computed using daily (respectively, weekly) returns and if we consider that a year is made us of 250 business days (respectively, 52 weeks), then the annualized volatility will be equal to $\sqrt{250}\,\sigma$ (respectively, $\sqrt{52}\,\sigma$).

Risk

The risk of holding an asset a may be quantified by the volatility of a.

Risk-free rate

The theoretical rate of return of an investment with zero risk; the risk-free rate, commonly denoted r_f, represents the minimum return an investor expects for any investment when taking no risk. This is often considered as a lower bound of what should at least give any riskier investment since bearing any risk should be more remunerating than a risk-free investment.

Expected (excess) return

The expected return of an asset a is the return on a expected in the future and is computed by $\mathbb{E}(r_a)$. The expected excess return is the expected return on a in excess of the return given by the risk-free rate (or by any other market measure): $\mathbb{E}(r_a) - r_f$.

Risk premium

The risk premium of an asset a is the expected return on a in excess of the risk-free rate, that is $\mathbb{E}(r_a) - r_f$.

Zero-cost portfolio

A zero-cost portfolio is a portfolio for which the weights add up to zero. For example, the excess return of an asset a, that is $\mathbb{E}(r_a) - r_f$ is a zero-cost portfolio. It invests 100% in a and -100% in the risk-free rate. The amount borrowed at r_f is invested in a. As the investor will have to pay the interest for borrowing money at r_f, he/she would therefore expect to receive more than r_f in return for his/her investment in a.

Tradable portfolio

A portfolio is said to be tradable if it is a zero-cost portfolio.

Portfolio diversification

Diversifying a portfolio is the act of adding more investments to one's portfolio to reduce the risk inherent in any one investment. It increases the possibility of making a profit, or at least avoiding a loss. In general, the broader the diversification, the less is the risk and the return. Adding more investments to the portfolio for diversification involves subdividing the portfolio among many smaller investments. If the portfolio's size increases instead of remaining constant, the portfolio's risk may not decrease especially if the assets added in the portfolio are uncorrelated. Here is a very simple example of diversification benefits on the variance of the portfolio. If the N assets in the portfolio are mutually uncorrelated and have identical variances σ^2, portfolio variance is minimized by holding all assets in the equal proportions $1/N$. Then, the portfolio variance equals σ^2/N that is monotonically decreasing in N. So, even if the added assets are uncorrelated, the portfolio variance decreases. Benefits of diversification amplify when adding negatively correlated assets in the portfolio. The modern understanding of diversification dates back to the work of Harry Markowitz [MAR 52] in the 1950s.

Market portfolio

A market portfolio is a perfectly well-diversified portfolio and represents the evolution of the market as a whole. In the factor model framework, a perfectly diversified portfolio admits a pure factor structure, that is a factor structure where there is no additional idiosyncratic risk.

Market efficiency

The market is said to be efficient if the price of the assets in the market reflects all information available.

Market equilibrium

The market equilibrium occurs if the amount of demand is balanced by the amount of supply. In this condition, the prices should tend not to change unless demand or supply change.

Arbitrage

Arbitrage is the possibility of a risk-free profit at zero cost. For example, if the same asset does not trade at the same price on all markets, buying and selling this asset simultaneously and instantaneously on two different markets will take advantage of the price difference.

Market capitalization

Market capitalization, also known as Market Cap. or cap. (MC), is the number of outstanding shares of a firm times the price of its share in the market. MC measures the size of the firm. It changes every day because of the quoted firm price (P). For example, in October 24, 2010, the six largest MC in the world were (T$ is for trillions US dollars and B$ for billions US dollars):

(1) Telecom Brasil with MC = T$6.57 and P = $5.99,

(2) Exxon Mobil with MC = B$337.69 and P = $66.32,

(3) Apple with MC = B$282.77 and P = $309.52,

(4) PetroChina with MC = B$229.20 and P = $125.23,

(5) BHP Billiton with MC = B$224.83 and P = $80.81, and

(6) Microsoft with MC = B$219.97 and P = $25.42.