Y. Malevergne
D. Sornette

Extreme Financial Risks

From Dependence to Risk Management

Springer

Extreme Financial Risks

Yannick Malevergne Didier Sornette

Extreme Financial Risks

From Dependence to Risk Management

Springer

Yannick Malevergne
Institut de Science Financière et d'Assurances
Université Claude Bernard Lyon 1
50 Avenue Tony Garnier
69366 Lyon Cedex 07
France

and

EM Lyon Business School
23 Avenue Guy de Collongue
69134 Ecully Cedex
France
E-mail: yannick.malevergne@univ-lyon1.fr

Didier Sornette
Institute of Geophysics and Planetary Physics
and Department of Earth and Space Science
University of California, Los Angeles
California 90095
USA

and

Laboratoire de Physique de la Matière
Condensée, CNRS UMR6622
and Université des Sciences
Parc Valrose
06108 Nice Cedex 2
France
E-mail: sornette@moho.ess.ucla.edu

Library of Congress Control Number: 2005930885

ISBN-10 3-540-27264-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-27264-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Typesetting: by the authors and TechBooks using a Springer LaTeX macro package
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 10939901 54/TechBooks 5 4 3 2 1 0

An error does not become truth by reason of multiplied propagation, nor does truth become error because nobody sees it.

M.K. Gandhi

Preface: Idiosyncratic and Collective Extreme Risks

Modern western societies have a paradoxical relationship with risks. On the one hand, there is the utopian quest for a zero-risk society [120]. On the other hand, human activities may increase risks of all kinds, from collaterals of new technologies to global impacts on the planet. The characteristic multiplication of major risks in modern society and its reflexive impact on its development is at the core of the concept of the "Risk Society" [47]. Correlatively, our perception of risk has evolved so that catastrophic events (earthquakes, floods, droughts, storms, hurricanes, volcanic eruptions, and so on) are no more systematically perceived as unfair outcomes of an implacable destiny. Catastrophes may also result from our own technological developments whose complexity may engender major industrial disasters such as Bhopal, Chernobyl, AZT, as well as irreversible global changes such as global warming leading to climatic disruptions or epidemics from new bacterial and viral mutations. The proliferation of new sources of risks imposes new responsibilities concerning their determination, understanding, and management. Government organizations as well as private institutions such as industrial companies, insurance companies, and banks which have to face such risks, in their role of regulators or of risk bearers, must ensure that the consequences of extreme risks are supportable without endangering the institutions in charge of bearing these risks.

In the financial sector, crashes probably represent the most striking events among all possible extreme phenomena, with an impact and frequency that has been increasing in the last two decades [450]. Consider the worldwide crash in October 1987 which evaporated more than one thousand billion dollars in a few days or the more recent collapse of the internet bubble in which more than one-third of the world capitalization of 1999 disappeared after March 2000. Finance and stock markets are based on the fluid convertibility of stocks into money and vice versa. Thus, to work well, money is requested to be a reliable standard of value, that is, an effective store of value, hence the concerns with the negative impacts of inflation. Similarly, investors look at the various financial assets as carriers of value, like money, but with additional return potentials (accompanied with downturn risks). But for this view to hold so as to promote economic development, fluctuations in values need to be tamed to minimize the risk of losing a lifetime of savings, or to avoid the risks of losing the investment potential of companies, or even to prevent economic and social recessions in whole countries (consider the situation of California after 2002 with a budget gap representing more than one-fourth of the entire State budget resulting essentially from the losses of financial and tax incomes following the collapse of the internet bubble). It is thus highly desirable to have the tools for monitoring, understanding, and limiting the extreme risks of financial markets. Fully aware of these problems, the worldwide banking organizations have promoted a series of advices and norms, known as the recommendations of the Basle committee [41, 42]. The Basle committee has proposed models for the internal management of risks and the imposition of minimum margin requirements commensurate with the risk exposures. However, some criticisms [117, 467] have found these recommendations to be ill-adapted or even destabilizing. This controversy underlines the importance of a better understanding of extreme risks, of their consequences and ways to prevent or at least minimize them.

In our opinion, tackling this challenging problem requires to decompose it into two main parts. First, it is essential to be able to accurately quantify extreme risks. This calls for the development of novel statistical tools going significantly beyond the Gaussian paradigm which underpins the standard framework of classical financial theory inherited from Bachelier [26], Markowitz [347], and Black and Scholes [60] among others. Second, the existence of extreme risks must be considered in the context of the practice of risk management itself, which leads to ask whether extreme risks can be diversified away similarly to standard risks according to the mean-variance approach. If the answer to this question is negative as can be surmized for numerous concrete empirical evidences, it is necessary to develop new concepts and tools for the construction of portfolios with minimum (but unavoidable) exposition of extreme risks. One can think of mixing equities and derivatives, as long as derivatives themselves do not add an extreme risk component and can really provide an insurance against extreme moves, which has been far from true in recent dramatic instances such as the crash of October 1987. Another approach could involve mutualism as in insurance.

Risk management, and to the same extent portfolio management, thus requires a precise and rigorous analysis of the distribution of the returns of the portfolio of risks. Taking into account the moderate sizes of standard portfolios (from tens to thousands of assets typically) and the non-Gaussian nature of the distributions of the returns of assets constituting the portfolios, the distributions of the returns of typical portfolios are far from Gaussian, in contradiction with the expectation from a naive use of the central limit theorem (see for instance Chap. 2 of [451] and other chapters for a discussion of the deviations from the central limit theorem). This breakdown of universality then requires a careful estimation of the specific case-dependent distribution of the returns of a given portfolio. This can be done directly using the time series of the returns of the portfolio for a given capital allocation. A more constructive approach consists in estimating the joint distribution of the returns of all assets constituting the portfolio. The first approach is much simpler and rapid to implement since it requires solely the estimation of a monovariate distribution. However, it lacks generality and power by neglecting the observable information available from the basket of all returns of the assets. Only the multivariate distribution of the returns of the assets embodies the general information of all risk components and their dependence across assets. However, the two approaches become equivalent in the following sense: the knowledge of the distribution of the returns for all possible portfolios for all possible allocations of capital between assets is equivalent to the knowledge of the multivariate distributions of the asset returns. All things considered, the second approach appears preferable on a general basis and is the method mobilizing the largest efforts both in academia and in the private sector.

However, the frontal attack aiming at the determination of the multivariate distribution of the asset returns is a challenging task and, in our opinion, much less instructive and useful than the separate studies of the marginal distributions of the asset returns on the one hand and the dependence structure of these assets on the other hand. In this book, we emphasize this second approach, with the objective of characterizing as faithfully as possible the diverse origins of risks: the risks stemming from each individual asset and the risks having a collective origin. This requires to determine (i) the distributions of returns at different time scales, or more generally, the stochastic process underlying the asset price dynamics, and (ii) the nature and properties of dependences between the different assets.

The present book offers an original and systematic treatment of these two domains, focusing mainly on the concepts and tools that remain valid for large and extreme price moves. Its originality lies in detailed and thorough presentations of the state of the art on (i) the different distributions of financial returns for various applications (VaR, stress testing), and (ii) the most important and useful measures of dependences, both unconditional and conditional and a study of the impact of conditioning on the size of large moves on the measure of extreme dependences. A large emphasis is thus put on the theory of copulas, their empirical testing and calibration, as they offer intrinsic and complete measures of dependences. Many of the results presented here are novel and have not been published or have been recently obtained by the authors or their colleagues. We would like to acknowledge, in particular, the fruitful and inspiring discussions and collaborations with J.V. Andersen, U. Frisch, J.-P. Laurent, J.-F. Muzy, and V.F. Pisarenko.

Chapter 1 describes a general framework to develop “coherent measures” of risks. It also addresses the origins of risks and of dependence between assets in financial markets, from the CAPM (capital asset pricing model) generalized to the non-Gaussian case with heterogeneous agents, the APT (arbitrage pricing theory), the factor models to the complex system view suggesting an emergent nature for the risk-return trade-off.

Chapter 2 addresses the problem of the precise estimation of the probability of extreme events, based on a description of the distribution of asset returns endowed with heavy tails. The challenge is thus to specify accurately these heavy tails, which are characterized by poor sampling (large events are rare). A major difficulty is to neither underestimate (Gaussian error) or overestimate (heavy tail hubris) the extreme events. The quest for a precise quantification opens the door to model errors, which can be partially circumvented by using several families of distributions whose detailed comparisons allow one to discern the sources of uncertainty and errors. Chapter 2 thus discusses several classes of heavy tailed distributions: regularly varying distributions (*i.e.*, with asymptotic power law tails), stretched-exponential distributions (also known as Weibull or subexponentials) as well as log-Weibull distributions which extrapolate smoothly between these different families.

The second element of the construction of multivariate distributions of asset returns, addressed in Chaps. 3–6, is to quantify the dependence structure of the asset returns. Indeed, large risks are not due solely to the heavy tails of the distribution of returns of individual assets but may result from a collective behavior. This collective behavior can be completely described by mathematical objects called *copulas*, introduced in Chap. 3, which fully embody the dependence between asset returns.

Chapter 4 describes synthetic measures of dependences, contrasting and linking them with the concept of copulas. It also presents an original estimation method of the coefficient of tail dependence, defined, roughly speaking, as the probability for an asset to lose a large amount knowing that another asset or the market has also dropped significantly. This tail dependence is of great interest because it addresses in a straightforward way the fundamental question whether extreme risks can be diversified away or not by aggregation in portfolios. Either the tail dependence coefficient is zero and the extreme losses occur asymptotically independently, which opens the possibility of diversifying them away. Alternatively, the tail dependence coefficient is non-zero and extreme losses are fundamentally dependent and it is impossible to completely remove extreme risks. The only remaining strategy is to develop portfolios that minimize the collective extreme risks, thus generalizing the mean-variance to a mean-extreme theory [332, 336, 333].

Chapter 5 presents the main methods for estimating copulas of financial assets. It shows that the empirical determination of a copula is quite delicate with significant risks of model errors, especially for extreme events. Specific studies of the extreme dependence are thus required.

Chapter 6 presents a general and thorough discussion of different measures of conditional dependences (where the condition can be on the size(s) of one or both returns for two assets). Chapter 6 thus sheds new light on the variations of the strength of dependence between assets as a function of the sizes of the analyzed events. As a startling concrete application of conditional dependences, the phenomenon of contagion during financial crises is discussed in detail.

Chapter 7 presents a synthesis of the six previous chapters and then offers suggestions for future work on dependence and risk analysis, including time-varying measures of extreme events, endogeneity versus exogeneity, regime switching, time-varying lagged dependence and so on.

This book has been written with the ambition to be useful to (a) the student looking for a general and in-depth introduction to the field, (b) financial engineers, economists, econometricians, actuarial professionals and researchers, and mathematicians looking for a synoptic view comparing the pros and cons of different modeling strategies, and (c) quantitative practitioners for the insights offered on the subtleties and many dimensional components of both risk and dependence. The content of this book will also be useful to the broader scientific community in the natural sciences, interested in quantifying the complexity of many physical, geophysical, biophysical etc. processes, with a mounting emphasis on the role and importance of extreme phenomena and their non-standard dependences.

Lyon, Nice and Los Angeles
August 2005

Yannick Malevergne
Didier Sornette

Contents

1

On the Origin of Risks and Extremes

1.1 The Multidimensional Nature of Risk and Dependence

In finance, the fundamental variable is the return that an investor accrues from his investment in a basket of assets over a certain time period. In general, an investor is interested in maximizing his gains while minimizing uncertainties ("risks") on the expected value of the returns on his investment, at possibly multiple time scales – depending upon the frequency with which the manager monitors the portfolio – and time periods – depending upon the investment horizon. From a general standpoint, the return-risk pair is the unavoidable duality underlying all human activities. The relationship between return and risk constitutes one of the most important unresolved questions in finance. This question permeates practically all financial engineering applications, and in particular the selection of investment portfolios. There is a general consensus among academic researchers that risk and return should be related, but the exact quantitative specification is still beyond our comprehension [414].

Uncertainties come in several forms, which we cite in the order of increasing aversion for most human beings:

(i) stochastic occurrences of events quantified by known probabilities;
(ii) stochastic occurrences of events with poorly quantified or unknown probabilities;
(iii) random events that are "surprises," *i.e.*, that were previously thought to be impossible or unthinkable until they happened and revealed their existence.

Here we address the first form, using the mathematical tools of probability theory.

Within this class of uncertainties, one must still distinguish several branches. In the simplest traditional theory exemplified by Markowitz [347], the uncertainties underlying a given set of positions (portfolio) result from the interplay of two components: risk and dependence.

(a) Risk is embedded in the amplitude of the fluctuations of the returns. its simplest traditional measure is the standard deviation (square-root of the variance).
(b) The dependence between the different assets of a portfolio of positions is traditionally quantified by the correlations between the returns of all pairs of assets.

Thus, in their most basic incarnations, both risk and dependence are thought of, respectively, as one-dimensional quantities: the standard deviation of the distribution of returns of a given asset and the correlation coefficient of these returns with those of another asset of reference (the "market" for instance). The standard deviation (or volatility) of portfolio returns provides the simplest way to quantify its fluctuations and is at the basis of Markowitz's portfolio selection theory [347]. However, the standard deviation of a portfolio offers only a limited quantification of incurred risks (seen as the statistical fluctuations of the realized return around its expected – or anticipated – value). This is because the empirical distributions of returns have "fat tails" (see Chap. 2 and references therein), a phenomenon associated with the occurrence of non-typical realizations of the returns. In addition, the dependences between assets are only imperfectly accounted for by the covariance matrix [309].

The last few decades have seen two important extensions.

- First, it has become clear, as synthesized in Chap. 2, that the standard deviation offers only a reductive view of the genuine full set of risks embedded in the distribution of returns of a given asset. As distributions of returns are in general far from Gaussian laws, one needs more than one centered moment (the variance) to characterize them. In principle, an infinite set of centered moments is required to faithfully characterize the potential for small all the way to extreme risks because, in general, large risks cannot be predicted from the knowledge of small risks quantified by the standard deviation. Alternatively, the full space of risks needs to be characterized by the full distribution function. It may also be that the distributions are so heavy-tailed that moments do not exist beyond a finite order, which is the realm of asymptotic power law tails, of which the stable Lévy laws constitute an extreme class. The Value-at-Risk (VaR) [257] and many other measures of risks [19, 20, 73, 447, 453] have been developed to account for the larger moves allowed by non-Gaussian distributions and non-linear correlations.
- Second and more recently, the correlation coefficient (and its associated covariance) has been shown to only be a partial measure of the full dependence structure between assets. Similarly to risks, a full understanding of the dependence between two or more assets requires, in principle, an infinite number of quantifiers or a complete dependence function such as the copulas, defined in Chap. 3.

These two fundamental extensions from one-dimensional measures of risk and dependence to infinitely dimensional measures of risk and dependence constitute the core of this book. Chapter 2 reviews our present knowledge and the open challenges in the characterization of distribution of returns. Chapter 3 introduces the notion of copulas which are applied later in Chap. 5 to financial dependences. Chapter 4 describes the main properties of the most important and varied measures of dependence, and underlines their connections with copulas. Finally, Chap. 6 expands on the best methods to capture the dependence between extreme returns.

Understanding the risks of a portfolio of N assets involves the characterization of both the marginal distributions of asset returns and their dependence. In principle, this requires the knowledge of the full (time-dependent) multivariate distribution of returns, which is the joint probability of any given realization of the N asset returns at a given time. This remark entails the two major problems of portfolio theory: (1) to determine the multivariate distribution function of asset returns; (2) to derive from it useful measures of portfolio risks and use them to analyze and optimize the performance of the portfolios. There is a large literature on multivariate distributions and multivariate statistical analysis [363, 468, 282]. This literature includes:

- the use of the multivariate normal distribution on density estimation [428];
- the corresponding random vectors treated with matrix algebra, and thus on matrix methods and multivariate statistical analysis [173, 371];
- the robust determination of multivariate means and covariances [297, 298];
- the use of multivariate linear regression and factor models [160, 161];
- principal component analysis, with excursions in clustering and classification techniques [276, 254];
- methods for data analysis in cases with missing observations [133, 310];
- detecting outliers [249, 250];
- bootstrap methods and handling of multicollinearity [461];
- methods of estimation using the plug-in principles and maximum likelihood [144];
- hypothesis testing using likelihood ratio tests and permutation tests [398];
- discrete multivariate distributions [253];
- computer-aided geometric design, geometric modeling, geodesic applications, and image analysis [464, 105, 426];
- radial basis functions [86], scattered data on spheres, and shift-invariant spaces [139, 433];
- non-uniform spline wavelets [139];
- scalable algorithms in computer graphics [76];
- reverse engineering [139], and so on.

The growing literature on (1) non-stationary processes [85, 210, 222, 361] and (2) regime-switching [172, 180, 215, 269] is not covered here. Nor do we address the more complex issues of embedding financial modeling within economics and social sciences. We do not cover either the consequences for risk assessment coming from the important emerging field of behavioral finance, with its exploration of the impact on decision-making of imperfect bounded subjective probability perceptions [36, 206, 437, 439, 474]. Our book thus uses objective probabilities which can be estimated (with quantifiable errors) from suitable analysis of available data.

1.2 How to Rank Risks Coherently?

The question on how to rank risks, so as to make optimal decisions, is recurrent in finance (and in many other fields) but has not yet received a general solution.

Since the middle of the twentieth century, several paths have been explored. The pioneering work by Von Neuman and Morgenstern [482] has given birth to the mathematical definition of the expected utility function, which provides interesting insights on the behavior of a rational economic agent and has formalized the concept of risk aversion. Based upon the properties of the utility function, Rothschild and Stiglitz [419, 420] have attempted to define the notion of increasing risks. But, as revealed by Allais [4, 5], empirical investigations have proven that the postulates chosen by Von Neuman and Morgenstern are actually often violated by humans. Many generalizations have been proposed for curing the so-called Allais' Paradox, but until now, no generally accepted procedure has been found.

Recently, a theory due to Artzner *et al.* [19, 20] and its generalization by Föllmer and Schied [174, 175] have appeared. Based on a series of postulates that are quite natural, this theory allows one to build *coherent* (resp., convex) measures of risks that provide tools to compare and rank risks [383]. In fact, if this theory seems well-adapted to the assessment of the needed economic capital, that is, of the fraction of capital a company must keep as risk-free assets in order to face its commitments and thus avoid ruin, it seems less natural for the purpose of quantifying the *fluctuations* of the asset returns or equivalently the deviation from a predetermined objective. In fact, as will be exposed in this section, it turns out that the two approaches consisting in assessing the risk in terms of economic capital on the one hand, and in terms of deviations from an objective on the other hand, are actually the two sides of the same coin as recently shown in [407, 408].

1.2.1 Coherent Measures of Risks

According to Artzner *et al.* [19, 20], the risk involved in the variations of the values of a market position is measured by the amount of capital invested in a risk-free asset, such that the market position can be prolonged in the future. In other words, the potential losses should not endanger the future actions of the fund manager of the company, or more generally, of the person or structure which underwrites the position. In this sense, a risk measure constitutes for Artzner *et al.* a measure of economic capital. The risk measure ρ can be either positive, if the risk-free capital must be increased to guarantee the risky position, or negative, if the risk-free capital can be reduced without invalidating it.

A risk measure is said to be *coherent* in the sense of Artzner *et al.* [19, 20] if it obeys the four properties or axioms that we now list. Let us call $\mathcal{G}$ the space of risks. If the space Ω of all possible states of nature is finite, $\mathcal{G}$ is isomorphic to $\mathbb{R}^{\mathrm{N}}$ and a risky position X is nothing but a vector in $\mathbb{R}^{\mathrm{N}}$. A risk measure ρ is then a map from $\mathbb{R}^{\mathrm{N}}$ onto $\mathbb{R}$. A generalization to other spaces $\mathcal{G}$ of risk has been proposed by Delbaen [123].

Let us consider a risky position with terminal value X and a capital α invested in the risk-free asset at the beginning of the time period. At the end of the time period, α becomes $\alpha \cdot (1 + \mu_0)$, where μ_0 is the risk-free interest rate. Then,

Axiom 1 (Translational Invariance)

$$\forall X \in \mathcal{G} \quad \text{and} \quad \forall \alpha \in \mathbb{R}, \quad \rho(X + \alpha \cdot (1 + \mu_0)) = \rho(X) - \alpha \ . \tag{1.1}$$

This simply means that an investment of amount α in the risk-free asset decreases the risk by the same amount α. In particular, for any risky position X, $\rho(X+\rho(X)\cdot(1+r)) = 0$, which expresses that investing an amount $\rho(X)$ in the risk-free asset enables one to exactly make up for the risk of the position X.

Let us now consider two risky investments X_1 and X_2, corresponding to the positions of two traders of an investment house. It is important for the supervisor that the aggregated risk of all traders be less than the sum of risks incurred by all traders. In particular, the risk associated with the position $(X_1 + X_2)$ should be smaller than or equal to the sum of the separated risks associated with the two positions X_1 and X_2.

Axiom 2 (Sub-additivity)

$$\forall (X_1, X_2) \in \mathcal{G} \times \mathcal{G}, \qquad \rho(X_1 + X_2) \le \rho(X_1) + \rho(X_2) \ . \tag{1.2}$$

The condition of sub-additivity encourages a portfolio managers to aggregate her different positions by diversification to minimize her overall risk. This axiom is probably the most debated among the four axioms underlying the theory of coherent measures of risk (see [131] and references therein). As an example, the VaR is well known to lack sub-additivity. At the same time, VaR is comonotonically additive, which means that the VaR of two comonotonic assets equals the sum of the VaR of each individual asset. But, since the comonotonicity represents the strongest kind of dependence (see Chap. 3), it is particularly disturbing to imagine that one can find situations where a portfolio made of two comonotonic assets is less risky than a portfolio with assets whose marginal risks are the same as in the previous situation but with a weaker dependence. Here is the rub with sub-additivity.

Axiom 3 (Positive Homogeneity)

$$\forall X \in \mathcal{G} \quad \text{and} \quad \forall \lambda \geq 0, \qquad \rho(\lambda \cdot X) = \lambda \cdot \rho(X) \ . \tag{1.3}$$

This third axiom stresses the importance of homogeneity. Indeed, it means that the risk associated with a given position increases with its size, here proportionally with it. Again, this axiom is controversial. Obviously, one can assert that the risk associated with the position $2 \cdot X$ is naturally twice as large as the risk of X. This is true as long as we can consider that a large position can be cleared as easily as a smaller one. However, it is not realistic because of the limited liquidity of real markets; a large position in a given asset is more risky than the sum of the risks associated with the many smaller positions which add up to the large position.

Eventually, if it is true that, for all possible states of nature, the risk of X leads to a loss larger than that of Y (*i.e.*, all components of the vector X in $\mathbb{R}^{\mathrm{N}}$ are always less than or equal to those of the vector Y), the risk measure $\rho(X)$ must be larger than or equal to $\rho(Y)$:

Axiom 4 (Monotony)

$$\forall X, Y \in \mathcal{G} \quad \text{such that} \quad X \leq Y, \qquad \rho(X) \geq \rho(Y) \ . \tag{1.4}$$

These four axioms define the coherent measures of risks, which admit the following general representation:

$$\rho(X) = \sup_{\mathbb{P} \in \mathcal{P}} \mathrm{E}_{\mathbb{P}} \left[\frac{-X}{1 + \mu_0} \right] , \tag{1.5}$$

where $\mathcal{P}$ denotes a family of probability measures. Thus, any coherent measure of risk appears as the expectation of the maximum loss over a given set of scenarios (the different probability measures $\mathbb{P} \in \mathcal{P}$). It is then obvious that the larger the set of scenarios, the larger the value of $\rho(X)$ and thus, the more conservative the risk measure.

It is particularly interesting that expression (1.5) is very similar to the result obtained in the theory of utility with non-additive probabilities [202, 203]. Indeed, in such a case, the utility of position X is given by

$$U(X) = \inf_{\mathbb{P} \in \mathcal{P}} \mathrm{E}_{\mathbb{P}} \left[u(X) \right] , \tag{1.6}$$

where $u(\cdot)$ is a usual utility function.

When the coherent risk measure is invariant in law and comonotonically additive, an alternative representation holds in terms of the spectral measure of risk [285, 471]

$$\rho(X) = p \int_0^1 \mathrm{VaR}_\alpha(X) \ dF(\alpha) + (1-p) \mathrm{VaR}_1(X) \ , \tag{1.7}$$

where F is a continuous convex distribution function on $[0,1]$, p is any real in $[0,1]$ and VaR_α is the Value-at-Risk defined in (3.85) page 125. Therefore, most coherent measures of risk appear as a convex sum of VaR_α (a non-coherent risk measure) at different probability levels. The weighting function F can be interpreted as a distortion of the objective probabilities, as underlined in the non-expected utility context [431, 495].

Coherent measures of risk can be generalized to define the so-called *convex* measures of risk by replacing the controversial axioms 2–3, by a single axiom of convexity of the risk measure [174, 175]. In the case where the risk measure is still positively homogeneous, this requirement is equivalent to the sub-additivity, but it becomes less restrictive when Axiom 3 is discarded. Then, one obtains the following representation of the convex risk measures:

$$\rho(X) = \sup_{\mathbb{P}\in\mathcal{M}} \mathrm{E}_{\mathbb{P}}\left[\frac{-X}{1+\mu_0} - \alpha\left(\mathbb{P}\right)\right], \tag{1.8}$$

where $\mathcal{M}$ is the set of all probability measures on $(\Omega, \mathcal{F})$, $\mathcal{F}$ denotes a σ-algebra on the state space Ω. More generally, $\mathcal{M}$ is the set of all finitely additive and non-negative set functions $\mathbb{P}$ on $\mathcal{F}$ satisfying $\mathbb{P}(\Omega) = 1$ and the functional

$$\alpha\left(\mathbb{P}\right) = \sup_{X\in\mathcal{G}|\rho(X)\le 0} \mathrm{E}_{\mathbb{P}}\left[\frac{-X}{1+\mu_0}\right] \tag{1.9}$$

is a penalty function that fully characterizes the convex measure of risk. In the case of a coherent risk measure, the set $\mathcal{P}$ (in (1.5)) is in fact the class of set functions $\mathbb{P}$ in $\mathcal{M}$ such that the penalty function vanishes: $\alpha(\mathbb{P}) = 0$.

Another alternative leads one to replace Axiom 4 by the following:

Axiom 5 (Expectation-Boundedness)

$$\forall X \in \mathcal{G} \quad \rho(X) \ge \frac{\mathrm{E}\left[-X\right]}{1+\mu_0}, \tag{1.10}$$

where the equality holds if and only if X is certain.[1] Then, together with axioms 1–3, it allows one to define the *expectation-bounded* risk measures [407]. They are particularly interesting insofar as they enable one to capture the inherent relationship existing between the assessment of risk in terms of economic capital and the measure of risk in terms of deviations from a target objective, as we shall see hereafter.

1.2.2 Consistent Measures of Risks and Deviation Measures

We now present a slightly different approach, which we think offers a suitable complement to coherent (and/or convex) risk measures for financial investments, and in particular for portfolio risk assessments. These measures are

[1] We say that X is *certain* if $X(\omega) = a$, for some $a \in \mathbb{R}$, for all $\omega \in \Omega$, such that $\mathbb{P}(\omega) \neq 0$, where $\mathbb{P}$ denotes a probability measure on $(\Omega, \mathcal{F})$ and $\mathcal{F}$ is a σ-algebra so that $(\Omega, \mathcal{F}, \mathbb{P})$ is a usual probability space.

called "*consistent* measures of risks" in [333] and "*general deviation measures*" in [407]. As before, we consider the future value of a risky position denoted by X, and we call $\mathcal{G}$ the space of risks.

Let us first require that the risk measure $\tilde{\rho}(\cdot)$, which is a functional on $\mathcal{G}$, always remains positive:

Axiom 6 (Positivity)

$$\forall X \in \mathcal{G} , \qquad \tilde{\rho}(X) \geq 0 , \tag{1.11}$$

where the equality holds if and only if X is certain. Let us now add to this position a given amount α invested in the risk-free asset whose return is μ_0 (with therefore no randomness in its price trajectory) and define the future wealth of the new position $Y = X + \alpha(1 + \mu_0)$. Since μ_0 is non-random, the fluctuations of X and Y are the same. Thus, it is desirable that $\tilde{\rho}$ enjoys a property of *translational invariance*, whatever X and the non-random coefficient α may be:

$$\forall X \in \mathcal{G},\ \forall \alpha \in \mathbb{R} , \qquad \tilde{\rho}(X + (1 + \mu_0) \cdot \alpha) = \tilde{\rho}(X) . \tag{1.12}$$

This relation is obviously true for all μ_0 and α; therefore, we set

Axiom 7 (Translational Invariance)

$$\forall X \in \mathcal{G},\ \forall \kappa \in \mathbb{R} , \qquad \tilde{\rho}(X + \kappa) = \tilde{\rho}(X) . \tag{1.13}$$

We also require that the risk measure increases with the quantity of assets held in the portfolio. This assumption reads

$$\forall X \in \mathcal{G},\ \forall \lambda \in \mathbb{R}_+ , \qquad \tilde{\rho}(\lambda \cdot X) = f(\lambda) \cdot \tilde{\rho}(X) , \tag{1.14}$$

where the function $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is increasing and convex to account for liquidity risk, as previously discussed. In fact, it is straightforward to show[2] that the only functions satisfying this requirement are the functions $f_\zeta(\lambda) = \lambda^\zeta$ with $\zeta \geq 1$, so that Axiom 3 can be reformulated in terms of positive homogeneity of degree ζ:

Axiom 8 (Positive Homogeneity)

$$\forall X \in \mathcal{G},\ \forall \lambda \in \mathbb{R}_+, \qquad \tilde{\rho}(\lambda \cdot X) = \lambda^\zeta \cdot \tilde{\rho}(X). \tag{1.15}$$

Note that the case of liquid markets is recovered by $\zeta = 1$ for which the risk is directly proportional to the size of the position, as in the case of the *coherent* risk measures.

These axioms, which define the so-called consistent measures of risk [333] can easily be extended to the risk measures associated with the return on the

[2] Using the trick $\tilde{\rho}(\lambda_1 \lambda_2 \cdot X) = f(\lambda_1) \cdot \tilde{\rho}(\lambda_2 \cdot X) = f(\lambda_1) \cdot f(\lambda_2) \cdot \tilde{\rho}(X) = f(\lambda_1 \cdot \lambda_2) \cdot \tilde{\rho}(X)$ leading to $f(\lambda_1 \cdot \lambda_2) = f(\lambda_1) \cdot f(\lambda_2)$. The unique increasing convex solution of this functional equation is $f_\zeta(\lambda) = \lambda^\zeta$ with $\zeta \geq 1$.

risky position. Indeed, a one-period return is nothing but the variation of the value of the position divided by its initial value X_0. One can thus easily check that the risk defined on the risky position is $[X_0]^\zeta$ times the risk defined on the return distribution. In the following, we will only consider the risk defined on the return distribution and, to simplify the notations, the symbol X will be used to denote both the asset price and its return in their respective context without ambiguity.

Now, restricting to the case of a perfectly liquid market ($\zeta = 1$) and adding a sub-additivity assumption

Axiom 9 (Sub-additivity)

$$\forall (X, Y) \in \mathcal{G} \times \mathcal{G} , \qquad \tilde{\rho}(X + Y) \le \tilde{\rho}(X) + \tilde{\rho}(X) , \tag{1.16}$$

one obtains the so-called general deviation measures [407]. Again, this axiom is open to controversy and its main *raison d'être* is to ensure the well-posedness of optimization problems (such as minimizing portfolio risks). It could be weakened along the lines used previously to derive the convex measures of risk from the coherent measures of risk.

One can easily check that the deviation measures defined in (1.16) correspond one-to-one to the expectation-bounded measures of risk defined in (1.10) through the relation

$$\rho(X) = \tilde{\rho}(X) + \frac{\mathrm{E}\left[-X\right]}{1 + \mu_0} \Longleftrightarrow \tilde{\rho}(X) = \rho\left(X + \mathrm{E}\left[-X\right]\right) . \tag{1.17}$$

It follows straightforwardly that minimizing the risk of a portfolio (measured either by ρ or by $\tilde{\rho}$) under constraints on the expected return is equivalent, as long as the constraints on the expected return are active. Indeed, in such a case, searching for the minimum of $\tilde{\rho}$ or of $\tilde{\rho}(X) + \frac{\mathrm{E}[-X]}{1+\mu_0}$ is the same problem since the value of the expected return is fixed by the constraints.

Additionally, it can be shown that the expectation-bounded measure of risk ρ defined by (1.17) is coherent if (and only if) the deviation measure $\tilde{\rho}$ satisfies [407]

$$\forall X \in \mathcal{G} , \qquad \tilde{\rho}(X) \le \mathrm{E}\left[X\right] - \inf X . \tag{1.18}$$

The general representation of the deviation measures satisfying this restriction can be easily derived from the representation of coherent risk measures. When such a requirement is not fulfilled, one can still have the following representation:[3]

[3] Strictly speaking, this representation only holds for lower semicontinous deviation measures, *i.e.*, deviation measures such that the sets $\{X | \tilde{\rho}(X) \le \epsilon\}$ are closed in $\mathcal{L}^2(\Omega)$, for all $\epsilon > 0$. This condition is fulfilled by most of the deviation measures of common use: the standard deviation, the semi-standard deviation, the absolute deviation, and so on.

$$\tilde{\rho}(X) = \sup_{Y \in \mathcal{Y}} \mathrm{E}\left[Y \cdot (\mathrm{E}[X] - X)\right] = \sup_{Y \in \mathcal{Y}} \mathrm{Cov}(-X, Y) , \tag{1.19}$$

where $\mathcal{Y}$ is a closed and convex subset of $\mathcal{L}^2(\Omega)$ such that

1. $1 \in \mathcal{Y}$,
2. $\forall Y \in \mathcal{Y}, \quad \mathrm{E}[Y] = 1$,
3. $\forall X \in \mathcal{L}^2(\Omega),\ \exists Y \in \mathcal{Y}$, such that $\mathrm{E}[Y \cdot X] < \mathrm{E}[X]$.

When the random variables in $\mathcal{Y}$ are all positive, they can be interpreted as density functions relative to some reference probability measure $\mathbb{P}_0$ on $(\Omega, \mathcal{F})$ (the objective probability measure). Thus, the term $\mathrm{E}[Y \cdot X]$ is nothing but the expectation of X under the probability measure $\mathbb{P}$, such that its Radon density $\frac{d\mathbb{P}}{d\mathbb{P}_0} = Y$. Therefore, one obtains a deviation measure associated with a coherent measure of risk.

These derivations show that deviation measures of risk on the one hand and coherent (or convex/expectation-bounded) measures of risk on the other hand are inextricably entangled. In fact, they are the two sides of the same coin, as mentioned in the introduction to this section. The various representation theorems show that, in most cases, these risk measures can be interpreted as worst-case scenarios, which rationalizes the use of *stress-testing* procedures as a sound practice for risk management.

In the more general case when the exponent ζ defined in Axiom 8 is no more equal to 1, and more precisely, when we only require that Axioms 6–8 hold, there is no general representation for the consistent risk measures to the best of our knowledge. The risk measures $\tilde{\rho}$ obeying Axioms 7 and 8 are known as the *semi-invariants* of the distribution of returns of X (see [465, pp. 86–87]). Among the large family of semi-invariants, we can cite the well-known centered moments and cumulants of X (including the usual variance). They are interesting cases that we discuss further below.

1.2.3 Examples of Consistent Measures of Risk

The set of risk measures obeying Axioms 7–8 is huge since it includes all the homogeneous functionals of $(X - \mathrm{E}[X])$, for instance. The centered moments (or moments about the mean) and the cumulants are two well-known classes of semi-invariants. Then, a given value of ζ can be seen as nothing but a specific choice of the order n of the centered moments or of the cumulants.[4] In this case, the risk measure defined via these semi-invariants fulfills the two following conditions:

$$\tilde{\rho}(X + \mu) = \tilde{\rho}(X) , \tag{1.20}$$
$$\tilde{\rho}(\lambda \cdot X) = \lambda^n \cdot \tilde{\rho}(X) . \tag{1.21}$$

[4] The relevance of the moments of high order for the assessment of large risks is discussed in Appendix 1.A.

In order to satisfy the positivity condition (Axiom 6), one needs to restrict the set of values taken by n. By construction, the centered moments of even order are always positive while the odd order centered moments can be negative. In addition, a vanishing value of an odd order moment does not mean that the random variable, or risk, $X \in \mathcal{G}$ is certain in the sense of footnote 1, since for instance any symmetric random variable has vanishing odd order moments. Thus, only the even-order centered moments seem acceptable risk measures. However, this restrictive constraint can be relaxed by first recalling that, given any homogeneous function $f(\cdot)$ of order p, the function $f(\cdot)^q$ is also homogeneous of order $p \cdot q$. This allows one to decouple the order of the moments to consider, which quantifies the impact of the large fluctuations, from the influence of the size of the positions held, measured by the degree of homogeneity of the measure $\tilde{\rho}$. Thus, considering any even-order centered moments, we can build a risk measure $\tilde{\rho}(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^{2n}\right]^{\zeta/2n}$, which accounts for the fluctuations measured by the centered moment of order $2n$ but with a degree of homogeneity equal to ζ.

A further generalization is possible for odd-order moments. Indeed, the *absolute* centered moments satisfy the three Axioms 6–8 for any odd or even order. So, we can even go one step further and use non-integer order absolute centered moments, and define the more general risk measure

$$\tilde{\rho}(X) = \mathrm{E}\left[|X - \mathrm{E}[X]|^{\gamma}\right]^{\zeta/\gamma} , \tag{1.22}$$

where γ denotes any positive real number.

Due to the Minkowski inequality, these risk measures are convex for any ζ and γ larger than 1 (and for $0 \le u \le 1$) :

$$\tilde{\rho}(u \cdot X + (1-u) \cdot Y) \le u \cdot \tilde{\rho}(X) + (1-u) \cdot \tilde{\rho}(Y) , \tag{1.23}$$

which ensures that aggregating two risky assets diversifies their risk. In fact, in the special case $\gamma = 1$, these measures enjoy the stronger sub-additivity property, and therefore belong to the class of general deviation measures.

More generally, any discrete or continuous (positive) sum of these risk measures with the same degree of homogeneity is again a risk measure. This allows us to define "spectral measures of fluctuations" in the spirit of Acerbi [2]:

$$\tilde{\rho}(X) = \int d\gamma \; \phi(\gamma) \; \mathrm{E}\left[|X - \mathrm{E}[X]|^{\gamma}\right]^{\zeta/\gamma} , \tag{1.24}$$

where ϕ is a positive real-valued function defined on any subinterval of $[1, \infty)$, such that the integral in (1.24) remains finite. It is sufficient to restrict the construction of $\tilde{\rho}(X)$ to normalized functions ϕ, such that $\int d\gamma \; \phi(\gamma) = 1$, since the risk measures are defined up to a global normalization factor. Then, $\phi(\gamma)$ represents the relative weight of the fluctuations measured by a given moment order and can be considered as a measure of the risk aversion of the risk manager with respect to large fluctuations.

The situation is not so clear for the cumulants, since the even-order cumulants, as well as the odd-order ones, can be negative (even if, for a large class of distributions, even-order cumulants remain positive, especially for fat-tailed distributions – even though there are simple but somewhat artificial counter-examples). In addition, cumulants suffer from another problem with respect to the positivity axiom. As for the odd-order centered moments, they can vanish even when the random variable is not certain. Just think of the cumulants of the Gaussian law. All but the first two (which represent the mean and the variance) are equal to zero. Thus, the strict formulation of the positivity axiom cannot be fulfilled by the cumulants. Should we thus reject them as useful measures of risks? It is important to emphasize that the cumulants enjoy a property which can be considered as a natural requirement for a risk measure. It can be desirable that the risk associated with a portfolio made of independent assets is exactly the sum of the risk associated with each individual asset. Thus, given N independent assets $\{X_1, \ldots, X_N\}$, and the portfolio $S_N = X_1 + \cdots + X_N$, we would like to have

$$\tilde{\rho}(S_N) = \tilde{\rho}(X_1) + \cdots + \tilde{\rho}(X_N) \ . \tag{1.25}$$

This property is verified for all cumulants, while it does not hold for centered moments excepted the variance. In addition, as seen from their definition in terms of the characteristic function

$$\mathrm{E}\left[e^{ik\cdot X}\right] = \exp\left(\sum_{n=1}^{+\infty} \frac{(ik)^n}{n!} C_n\right) \ , \tag{1.26}$$

cumulants C_n of order larger than 2 quantify deviations from the Gaussian law and therefore measure large risks beyond the variance (equal to the second-order cumulant).

What are the implications of using the cumulants as *almost* consistent measures of risks? In particular, what are the implications on the preferences of the agents employing such measures? To address this question, it is informative to express the cumulants as a function of the centered moments. For instance, let us consider the fourth-order cumulant:

$$C_4 = \mu_4 - 3 \cdot {\mu_2}^2 = \mu_4 - 3 \cdot {C_2}^2 \ , \tag{1.27}$$

where μ_n is the centered moment of order n. An agent assessing the fluctuations of an asset with respect to C_4 exhibits an aversion for the fluctuations quantified by the fourth central moment μ_4 – since C_4 increases with μ_4 – but is attracted by the fluctuations measured by the variance – since C_4 decreases with μ_2. This behavior is not irrational because it remains globally risk-averse. Indeed, it depicts an agent which tries to avoid the larger risks but is ready to accept the smallest ones. This kind of behavior is characteristic of any agent using the cumulants as risk measures. In such a case, having $C_4 = 0$ does not mean that the agent considers that the position is not risky (in the sense that the position is certain) but that the agent is indifferent between the large risks of this position measured by μ_4 and the small risks quantified by μ_2.

To summarize, centered moments of even orders possess all the minimal properties required for a suitable portfolio risk measure. Cumulants only partially fulfill these requirements, but have an additional advantage compared with the centered moments, that is, they fulfill the condition (1.25). For these reasons, we think it is interesting to consider both the centered moments and the cumulants in risk analysis and decision making. Finally let us stress that the variance, originally used in Markowitz's portfolio theory [347], is nothing but the second centered moment, also equal to the second-order cumulant (the three first cumulants and centered moments are equal). Therefore, a portfolio theory based on the centered moments or on the cumulants automatically contains Markowitz's theory as a special case, and thus offers a natural generalization encompassing large risks of this masterpiece of financial science. It also embodies several other generalizations where homogeneous measures of risks are considered, as for instance in [241].

We should also mention the measure of attractiveness for risky investments, the gain–loss ratio, introduced by Bernardo and Ledoit [50]. The gain (loss) of a portfolio is the expectation, under a benchmark risk-adjusted probability measure, of the positive (negative) part of the portfolio's excess payoff. The gain–loss ratio constitutes an improvement over the widely used Sharpe ratio (average return over volatility). The advantage of the gain–loss ratio is that it penalizes only downside risk (losses) and rewards all upside potential (gains). The gain–loss ratio has been show to yield useful bounds for asset pricing in incomplete markets that gives the modeler the flexibility to control the trade-off between the precision of equilibrium models and the credibility of no-arbitrage methods. The gain–loss approach is valuable in applications where the security returns are not normally distributed. Bernardo and Ledoit [50] cite the following domains of application: (i) valuing real options on non-traded assets; (ii) valuing executive stock options when the executive cannot trade the options or the underlying due to insider trading restrictions; (iii) evaluating the performance of portfolio managers who invest in derivatives; (iv) pricing options on a security whose price follows a jump-diffusion or a fat-tailed Pareto–Levy diffusion process; and (v) pricing fixed income derivatives in the presence of default risk.

1.3 Origin of Risk and Dependence

1.3.1 The CAPM View

Our purpose is not to review the huge literature on the origin of risks and their underlying mechanisms, but to suggest guidelines for further understanding. For enticing introductions and synopses, we refer to the very readable books of Bernstein [51, 52]. In [51], Bernstein reviews the history, since ancient times, of those thinkers who showed how to quantify risk:

> The capacity to manage risk, and with it the appetite to take risk and make forward-looking choices, are key elements [...] that drive the economic system forward.

The concept of risks in economics and finance is elaborated in [52], starting with the origins of the Cowles foundation as the consequence of Cowles's personal interest in the question: Are stock prices predictable? In the words of J.L. McCauley (see his customer review on www.amazon.com),

> this book is all about heroes and heroic ideas, and Bernstein's heroes are Adam Smith, Bachelier, Cowles, Markowitz (and Roy), Sharpe, Arrow and Debreu, Samuelson, Fama, Tobin, Samuelson, Markowitz, Miller and Modigliani, Treynor, Samuelson, Osborne, Wells-Fargo Bank (McQuown, Vertin, Fouse and the origin of index funds), Ross, Black, Scholes, and Merton. The final heroes (see Chap. 14, The Ultimate Invention) are the inventors of (synthetic) portfolio insurance (replication/synthetic options).

One of these achievements is the capital asset pricing model (CAPM), which is probably still the most widely used approach to relative asset valuation, although its empirical roots have been found to be weaker in recent years [59, 160, 223, 287, 306, 401]. Its major idea was that priced risk cannot be diversified and cannot be eliminated through portfolio aggregation. This asset valuation model describing the relationship between expected risk and expected return for marketable assets is strongly entangled with the Mean-Variance Portfolio Model of Markowitz. Indeed, both of them fundamentally rely on the description of the probability distribution function (pdf) of asset returns in terms of Gaussian functions. The mean-variance description is thus at the basis of the Markowitz portfolio theory and of the CAPM and its inter-temporal generalization (see for instance [359]).

The CAPM is based on the concept of economic equilibrium between rational expectation agents. Economic equilibrium is itself the self-organized result of complex nonlinear feedback processes between competitive interacting agents. Thus, while not describing the specific dynamics of how self-organization makes the economy converge to a stable regime [10, 18, 280], the concept of economic equilibrium describes the resulting state of this dynamic self-organization and embodies all the hidden and complex interactions between agents with infinite loops of recurrence. This provides a reference base for understanding risks.

We put some emphasis on the CAPM and its generalized versions because the CAPM is a remarkable starting point for answering the question on the origin of risks and returns: in economic equilibrium theory, the two are conceived as intrinsically entangled. In the following, we expand on this class of explanation before exploring briefly other directions.

Let us now show how an equilibrium model generalizing the original CAPM [308, 364, 429] can be formulated on the basis of the coherence measures adapted to large risks. This provides an "explanation" for risks from the point of view of the non-separable interplay between agents' preferences and their collective organization. We should stress that many generalizations have already been proposed to account for the fat-tailness of the assets return distributions, which led to the multimoments CAPM. For instance, Rubinstein [421], Krauss and Litzenberger [278], Lim [306] and Harvey and Siddique [223] have underlined and tested the role of the asymmetry in the risk premium by accounting for the skewness of the distribution of returns. More recently, Fang and Lai [162] and Hwang and Satchell [241] have introduced a four-moments CAPM to take into account the leptokurtic behavior of the assets return distributions. Many other extensions have been presented such as the VaR-CAPM [3] or the Distributional-CAPM [389]. All these generalizations become more complicated but unfortunately do not necessarily provide more accurate predictions of the expected returns.

Let us assume that the relevant risk measure is given by any measure of fluctuations previously presented that obey the Axioms 6–8 of Sect. 1.2.2. We will also relax the usual assumption of a homogeneous market to give to the economic agents the choice of their own risk measure: some of them may choose a risk measure which puts the emphasis on the small fluctuations, while others may prefer those which account for the larger ones. In such an heterogeneous market, we will recall how an equilibrium can still be reached and why the excess returns of individual stocks remain proportional to the market excess return, which is the fundamental tenet of CAPM.

For this, we need the following assumptions about the market:

- H1: We consider a one-period market, such that all the positions held at the beginning of a period are cleared at the end of the same period.
- H2: The market is perfect, *i.e.*, there are no transaction costs or taxes, the market is efficient and the investors can lend and borrow at the same risk-free rate μ_0.

Of course, these standard assumptions are to be taken with a grain of salt and are made only with the goal of obtaining a normative reference theory. We will now add another assumption that specifies the behavior of the agents acting on the market, which will lead us to make the distinction between homogeneous and heterogeneous markets.

Equilibrium in a Homogeneous Market

The market is said to be homogeneous if all the agents acting on this market aim at fulfilling the same objective. This means that:

- H3-1: All the agents want to maximize the expected return of their portfolio at the end of the period under a given constraint of measured risk, using the same measure of risks ρ_ζ for all of them (the subscript ζ refers to the degree of homogeneity of the risk measure, see Sect. 1.2).

In the special case where ρ_ζ denotes the variance, all the agents follow a Markowitz's optimization procedure, which leads to the CAPM equilibrium, as proved by Sharpe [429]. When ρ_ζ represents the centered moments, this leads to the market equilibrium described in [421]. Thus, this approach allows for a generalization of the most popular asset pricing in equilibrium market models.

When all the agents have the same risk function ρ_ζ, whatever ζ may be, we can assert that they have all a fraction of their capital invested in the same portfolio Π (see, for instance [333] for the derivation of the composition of the portfolio), and the remaining in the risk-free asset. The amount of capital invested in the risky fund only depends on their risk aversion and/or on the legal margin requirement they have to fulfill.

Let us now assume that the market is at equilibrium, *i.e.*, supply equals demand. In such a case, since the optimal portfolios can be any linear combinations of the risk-free asset and of the risky portfolio Π, it is straightforward to show that the market portfolio, made of all traded assets in proportion of their market capitalization, is nothing but the risky portfolio Π. Thus, as shown in [333], we can state that, whatever the risk measure ρ_ζ chosen by the agents to perform their optimization, the excess return of any asset i over the risk-free interest rate $(\mu(i) - \mu_0)$ is proportional to the excess return of the market portfolio Π over the risk-free interest rate:

$$\mu(i) - \mu_0 = \beta_\zeta^i \cdot (\mu_\Pi - \mu_0), \tag{1.28}$$

where

$$\beta_\zeta^i = \left. \frac{\partial \ln \left({\rho_\zeta}^{\frac{1}{\zeta}} \right)}{\partial w_i} \right|_{w_1^*, \cdots, w_N^*}, \tag{1.29}$$

where $w_1^*, \ldots, w_N^*$ are the optimal allocations of the assets in the following sense:

$$\begin{cases} \inf_{w_i \in [0,1]} \rho_\zeta(\{w_i\}) \\ \sum_{i \geq 0} w_i = 1 \\ \sum_{i \geq 0} w_i \mu(i) = \mu \ , \end{cases} \tag{1.30}$$

In other words, the set of normalized weights w_i^* define the portfolio with minimum risk as measured by any convex[5] measure ρ_ζ of risk obeying Axioms 6–8 of Sect. 1.2.2 for a given amount of expected return μ.

When ρ_ζ denotes the variance, we recover the usual β^i given by the mean-variance approach:

$$\beta^i = \frac{\mathrm{Cov}(X_i, \Pi)}{\mathrm{Var}(\Pi)} \ . \tag{1.31}$$

[5] Convexity is necessary to ensure the existence and the unicity of a minimum.

Thus, the relations (1.28) and (1.29) generalize the usual CAPM formula, showing that the specific choice of the risk measure is not very important, as long as it follows the Axioms 6–8 characterizing the fluctuations of the distribution of asset returns.

Equilibrium in a Heterogeneous Market

Does this result hold in the more realistic situation of an heterogeneous market? A market will be said to be heterogeneous if the agents seek to fulfill different objectives. We thus consider the following assumption:

- H3-2: There exist N agents. Each agent n is characterized by her choice of a risk measure $\rho_\zeta(n)$ so that she invests only in the mean-$\rho_\zeta(n)$ efficient portfolios.

According to this hypothesis, an agent n invests a fraction of her wealth in the risk-free asset and the remaining in Π_n, the mean-$\rho_\zeta(n)$ efficient portfolio, only made of risky assets. Again, the fraction of wealth invested in the risky fund depends on the risk aversion of each agent, which may vary from one agent to another.

The composition of the market portfolio Π for such a heterogeneous market is found to be nothing but the weighted sum of the mean-$\rho_\zeta(n)$ optimal portfolio Π_n [333]:

$$\Pi = \sum_{n=1}^{N} \gamma_n \Pi_n \ , \tag{1.32}$$

where γ_n is the fraction of the total wealth invested in the fund Π_n by the n^{th} agent.

Moreover, for every asset i and for any mean-$\rho_\zeta(n)$ efficient portfolio Π_n, for all n, the following equation holds

$$\mu(i) - \mu_0 = \beta_n^i \cdot (\mu_{\Pi_n} - \mu_0) \ , \tag{1.33}$$

where β_n^i is defined in (1.29). Multiplying these equations by γ_n/β_n^i, we get

$$\frac{\gamma_n}{\beta_n^i} \cdot (\mu(i) - \mu_0) = \gamma_n \cdot (\mu_{\Pi_n} - \mu_0) \ , \tag{1.34}$$

for all n, and summing over the different agents, we obtain

$$\left(\sum_n \frac{\gamma_n}{\beta_n^i} \right) \cdot (\mu(i) - \mu_0) = \left(\sum_n \gamma_n \cdot \mu_{\Pi_n} \right) - \mu_0 \ , \tag{1.35}$$

so that

$$\mu(i) - \mu_0 = \beta^i \cdot (\mu_\Pi - \mu_0) \ , \tag{1.36}$$

with

$$\beta^i = \left(\sum_n \frac{\gamma_n}{\beta^i_n} \right)^{-1} . \tag{1.37}$$

This allows us to conclude that, even in a heterogeneous market, the expected excess return of each individual stock is directly proportional to the expected excess return of the market portfolio, showing that the homogeneity of the market is not required for observing a linear relationship between individual excess asset returns and the market excess return.

The above calculations miss the possibility stressed by Rockafellar et al. [408] that two kinds of efficient portfolios Π_n may exist in a heterogeneous market: long optimal portfolios which correspond to a net long position, and short optimal portfolios which correspond to a net short position. If the existence of the second kind of portfolio is not compatible with an equilibrium in a homogeneous market,[6] their existence is not precluded in a heterogeneous market. Indeed, the net short positions of a certain class of agents can be compensated by the net long position of another class of agents. Thus, as long as a market portfolio Π corresponding to an overall long position exists, an equilibrium can be reached, and the results derived in this section still hold.

1.3.2 The Arbitrage Pricing Theory (APT) and the Fama–French Factor Model

The CAPM proposed a solution for what Roll [414] called

> perhaps the most important unresolved problem in finance, because it influences so many other problems, (which) is the relation between risk and return. Almost everyone agrees that there should be some relation, but its precise quantification has proven to be a conundrum that has haunted us for years, embarrassed us in print, and caused business practitioners to look askance at our scientific squabbling and question our relevance.

Indeed, past and recent tests cast strong doubts on the validity of the CAPM. The recent Fama–French analysis [160] shows basically no support for the CAPM's central result of a positive relation between expected return and global market risk (quantified by the so-called beta parameter). In contrast, other variables, such as market capitalization and the book-to-market ratio,[7] present some weak explanatory power.

[6] An equilibrium cannot be reached if all investors want to sell stocks.

[7] Ratio of the book value of a firm to its market value. Typically, the book-to-market is used to identify undervalued companies. If the book-to-market is less than one the stock is overvalued, while it is undervalued otherwise.

The Arbitrage Pricing Theory (APT)

The empirical inadequacy of the CAPM has led to the development of more general models of risk and return, such as Ross's Arbitrage Pricing Theory (APT) [418]. Quoting Sargent [427],

> Ross posited a particular statistical process for asset returns, then derived the restrictions on the process that are implied by the hypothesis that there exist no arbitrage possibilities.

Like the CAPM, the APT assumes that only non-diversifiable risk is priced. But it differs from the CAPM by accounting for multiple causes of such risks and by assuming a sufficiently large number of such factors so that almost riskless portfolios can be constructed. Reisman recently presented a generalization of the APT showing that, under the assumption that there exists no asymptotic arbitrage (*i.e.*, in the limit of a large number of factors, the market risk can be decreased to almost zero), there exists an approximate multi-beta pricing relationship relative to any admissible proxy of dimension equal to the number of factors [402]. Unlike the CAPM which specifies returns as a linear function of only systematic risk, the APT is based on the well-known observations that multiple factors affect the observed time series of returns, such as industry factors, interest rates, exchange rates, real output, the money supply, aggregate consumption, investor confidence, oil prices, and many other variables [414]. However, while observed asset prices respond to a wide variety of factors, there is much weaker evidence that equities with larger sensitivity to some factors give higher returns, as the APT requires.

The Fama–French Three Factor Model

This empirical weakness in the APT has led to further generalizations of factor models, such as the Fama–French three-factor model [160], which does not use an arbitrage condition anymore. Fama and French started with the observation that two classes of stocks show better returns than the average market: (1) stocks with small market capitalization ("small caps") and (2) stocks with a high book-value-to-price ratio (often "value" stocks as opposed to "growth" stocks). They added the overall market return to obtain the three factors: (i) the overall market return (Rm), (ii) the performance of small stocks relative to big stocks (SMB, small minus big), and (iii) the performance of value stocks relative to growth stocks (HML, high minus low). See the website of Professor K.R. French[8] which updates every quarter the benchmark factors and also presents the performance of several benchmark portfolios using different combinations of weights on the three factors. An important observation must be made concerning Fama and French's approach to risk in

[8] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`