International Journal of Theoretical and Applied Finance
Vol. 1, No. 1 (1998) 25–41

TAMING LARGE EVENTS: OPTIMAL PORTFOLIO THEORY FOR STRONGLY FLUCTUATING ASSETS

J. P. BOUCHAUD

*Service de Physique de l'Etat Condensé**
CEA, Orme des Merisiers, 91 191 Gif sur Yvette CEDEX

D. SORNETTE

Laboratoire de Physique de la Matière Condensée
Université de Nice-Sophia Antipolis B. P. 70, Parc Valrose, 06108 Nice CEDEX 2

C. WALTER

Direction de la Gestion et des Relations Investisseurs- BC 11152 Crédit Lyonnais
19, Boulevard des Italiens, 75002 Paris

J. P. AGUILAR

Capital Futures Management, 109-111 rue Victor Hugo, 92932 Levallois CEDEX

Received 26 August 1997

We propose a method of optimization of asset allocation in the case where the stock price variations are supposed to have "fat" tails represented by power laws. Generalizing over previous works using stable Lévy distributions, we distinguish three distinct components of risk described by three different parts of the distributions of price variations: unexpected gains (to be kept), harmless noise inherent to financial activity, and unpleasant losses, which is the only component one would like to minimize. The independent treatment of the tails of distributions for positive and negative variations and the generalization to large events of the notion of covariance of two random variables provide explicit formulae for the optimal portfolio. The use of the probability of loss (or equivalently the Value-at-Risk), as the key quantity to study and minimize, provides a simple solution to the problem of optimization of asset allocations in the general case where the characteristic exponents are different for each asset.

1. Introduction

Despite quite a number of early insightful studies [35, 24, 25, 18, 33], the fact that many natural phenomena must be described by power law statistics has only been fully accepted in the past ten years. Correspondingly, an intense activity has developed in order to understand both the physical content of the mathematical tools devised by P. Lévy [22] and others [19], and the origin of these ubiquitous

*Also at: Science et Finance, 109-111 rue V. Hugo, 92 532 Levallois, France.

power law tails. This has led to several interesting ideas [43], in particular, the seminal concept of "self-organized criticality" [4].

As a matter of fact, it is in economy and finance that these power law distribution were first noted by Pareto [35] and, in the early sixties, by Mandelbrot and Fama [24, 25, 18]. Their ideas were however discarded and did not have a large practical impact: their work appeared when the "standard" model of efficient markets was rapidly developing [30, 39] with notable successes, such as the CAPM [31] (Capital Asset Pricing Model) or the Black–Scholes option pricing theory [8]. The efficient market hypothesis implies that market prices are not predictable, prices change randomly and no consistent positive profit can be obtained from speculation. The paradigm of stock price variations is thus the multivariate normal distribution [12, 3]. The strength of the efficient market hypothesis is its conceptual and mathematical simplicity: based on the elementary notions of mean and variance, the algebraic manipulations are eased by the hypothesis of Gaussian statistics, for which a score of mathematical results are available. The "power-law" hypothesis, on the contrary, may force one to abandon the notion of mean and variance (although this is not always the case, see below) and find different, less elementary objects as the building-blocks of the theory. The mathematical tool-box is also, in that case, much more involved [37] and somewhat poorer.

The "power-law" (or "paretian") hypothesis remained somewhat dormant until recently, notwithstanding the continuing confirmation that distributions of price variations and other commodities present a strong leptokurtosis (anomalously large ratio, sometimes much larger than 3 characterizing Gaussian statistics, of the fourth moment over second moment squared), indicating anomalously large fluctuations. The recent revival of these ideas in finance is partly due to the 1987 crash, and more recent smaller ones, which spotlights the crucial importance of large, catastrophic events and the limitation of Gaussian theories, which dramatically underestimates their probability of occurence. More recently, the notion of "value-at-risk" (VaR) has become a central quantity to assess risk. A correct determination of this VaR, and its necessary control again requires new theoretical tools.

A series of authors have thus explored, following [24, 25, 18], the possibility that the stable Lévy distributions could represent price fluctuations more accurately than the normal distribution (see [42, 26 and references therein]). These analysis of stock market fluctuations (see [42, 34, 27, 29, 11]) consistently show that a Lévy distribution (or more precisely a *truncated* Lévy distribution — see below) of returns is a better representation of the data.

An additional clue comes from the well-known observation that the standard "efficient portfolio" theory leads to several difficulties in practice: the weights of the different assets must be revised all the time and the optimal portfolio often retain a small fraction of all assets — an aspect that is felt unreasonable and impractical by many operators.

The aim of the present paper is to extend previous theoretical approaches [38, 2, 6] of asset allocation developed for stable Lévy laws to more general situations where the power laws hold only on the tail of the distribution, i.e. *on the most notable events* which have an immediate meaning for all operators on the financial markets, or are characterized by different exponents (possibly larger than 2), i.e. are not stable in the sense of Lévy. We also propose, loosely speaking, to replace the standard return/risk ratio by a return/(probability of large losses) ratio. Our basic idea being that "risk" should be decomposed into three distinct components: unexpected gains (to be kept), harmless noise inherent to financial activity, and unpleasant losses, which is the only component one would like to minimize.

The paper is constructed as follows: in the first section, we introduce notations and specify the class of probability distribution (or probability density) that we shall adopt for the description of real data. In Sec. 2, we shall propose a generalization of the notion of covariance of two random variables which is adapted to large events. In Sec. 3, important results on the addition of power-law distributed random variables will be recalled (heuristic proofs are given in the Appendix A), and used in Sec. 4 to establish our results on optimal portfolio in the extended sense mentioned above. Section 5 discusses the applicability of these ideas, and present a concrete example of portfolio optimization. This paper is not claiming mathematical rigour, but rather discusses these topics in an intuitive, and hopefully useful, manner.

2. Power-Law Distributions

Let us denote by v the (daily, weekly, monthly, ...) variation in the value of a given asset. Any progress on the problem of asset allocation depends on the determination of the mathematical form for the probability density $P(v)$ of v's, since it controls the existence or not of mean return and variance. As recalled above, the repeated observations of a strong anomalous leptokurtosis has led to the proposal that the "Gaussian paradigm" should be replaced by the "Lévy paradigm". Considering Lévy laws is a natural step, since they are characterized by power-law tails and are stable, as is the Gaussian distribution, with respect to the addition of variables. The first property gives hope to take into account the large observed fluctuations and corresponding leptokurtosis. The second one ensures that the mathematical description is invariant with respect to the chosen time step (daily, weekly, monthly, ...) and leads to the property of stationarity. However, we shall not take a dogmatic point of view but rather let us guide by empirical data, in particular allowing for deviations from a Lévy distribution.

Here, we will thus explore the more general situation where, for large enough positive and negative variations v of the asset price, the probability density $P(v)$ is a power law:

$$P(v) \simeq_{v \to \pm\infty} \frac{C_\pm}{|v|^{1+\mu_\pm}} \,. \tag{2.1}$$

C_+ (resp. C_-) is the scale factor for the positive (resp. negative) price variations. μ^+ (resp. μ^-) is the exponent of the tail of the probability density for positive (resp. negative) price variations. Notice that μ (when $\mu^+ = \mu^-$) can be identified with the index α for the particular case where $P(v)$ is a Lévy distribution L_α.

This form is compatible with recent empirical studies [42, 27, 32, 29, 10] which suggests that an appropriate description of price variations involve Lévy distributions. In fact, this might be too restrictive and a more flexible description might be needed, involving power tails with exponents larger than 2 (i.e. not stable) [13, 23] or rather cross-overs from a power law to an exponential behavior, leading to a "truncated Lévy distribution" [29, 28, 10, 11]. In this respect, it is interesting to note that the limit $\mu \to \infty$ formally corresponds to exponential tails [11].

The expression (2.1) provides an interesting and useful parametrization because robust statistical tools can be used to extract the exponent μ and scale parameter C on a relatively small data set [44, 17, 1, 20, 16, 15, 14, 13]. We would like to mention in particular the "rank ordering" technique pioneered first by [44] which puts the emphasis on the analysis of the tails of distributions (see [40] for a recent review). For the practical implementation of our proposed asset allocation strategy, we will use this method to retrieve the values of the exponents and scale parameters.

The scale coefficient $C_\pm$ will play a crucial role in the following: it reflects the scale of the fluctuations of the price variations v. More precisely, the order of magnitude of the largest event out of N drawn from the distribution (2.1) is given by the condition

$$N \int_{v_{\max}}^{+\infty} P(v) dv \simeq 1 \tag{2.2}$$

yielding

$$v_{\max}^{+}(N) \simeq C_{+}^{\frac{1}{\mu_+}} N^{\frac{1}{\mu_+}} \tag{2.3}$$

for the typical largest positive variation and

$$v_{\max}^{-}(N) \simeq -C_{-}^{\frac{1}{\mu_-}} N^{\frac{1}{\mu_-}} \tag{2.4}$$

for the largest negative variation. Another interpretation of $C_\pm$, which will be exploited in the sequel is the following: the total probability of observing events larger than a certain value is proportional to $C_\pm$. Indeed, the probability for a loss (resp. gain) larger than a certain value λ is given by

$$P_{\text{loss}}(|v| > \lambda) = \frac{C_-}{\mu \lambda^{\mu_-}}; \quad P_{\text{gain}}(v > \lambda) = \frac{C_+}{\mu \lambda^{\mu_+}}. \tag{2.5}$$

The power law (2.1) structure of the probability density $P(v)$ describes a *self-similar* process, i.e. one for which no intrinsic characteristic scale exists. Although no precise model has been proposed to explain this in finance (see however [4, 41, 21, 5]), this seems to be a rather reasonable assumption, at least in a restricted range of variations, by analogy with other collective systems. This assumption may however break down at large variations, pointing out the possible existence of characteristic scales. For example, the power-law tail is bound to break down ultimately beyond some characteristic value $v_{\max}$, where some extrinsic mechanism might operate such as quotation suspension or market closure, or some intrinsic feedback mechanism, leading to a faster decay (e.g. exponential) beyond $v_{\max}$.

It is easy to see (from its very definition) that the qth moment of P ceases to exist as soon as $q \geq \mu$. In particular, the mean of the distribution is formally infinite when $\mu \leq 1$. Similarly, the variance of the distribution is infinite when $\mu \leq 2$. In fact, because of the "cut-off" mechanism just described, the mean (resp. variance) will remained well defined even for $\mu \leq 1$ (resp. 2), but proportional to the cut-off $v_{\max}^{1-\mu}$ (resp. $v_{\max}^{2-\mu}$):[a] hence in these cases, the mean (resp. variance) is obviously not the interesting quantity to consider since it is unrelated to the *typical order of magnitude* of the variations (resp. their fluctuations). Note that even for $\mu > 1$ (resp. $\mu > 2$) but not too large, the mean (resp. variance) remains a tricky notion to use in applications, since its determination from a time series of length N converges only slowly towards the theoretical value. As already shown, the largest variations scale as $v_{\max} \sim N^{\frac{1}{\mu}}$. Consequently, for $1 < \mu < 1$ the relative error on the mean converges to zero as $N^{\frac{1}{\mu}-1}$. For $\mu > 2$, the standard $N^{-\frac{1}{2}}$ convergence of the mean is recovered. For $2 < \mu \leq 4$, the estimation of the variance converges slowly as $N^{\frac{2}{\mu}-1}$. For $\mu > 4$, one recovers the standard $N^{-\frac{1}{2}}$ convergence of the variance.

The important practical implication is that, if N is not very large, these basic parameters for the determination of an optimal portfolio within the usual Gaussian framework are changing with time — leading to rather strong instabilities in the optimal weights. Let us illustrate further the importance of the slow convergence of the mean (or of the variance) even in cases where their mathematical convergence is ensured, with applications to real data in mind. Suppose that a given stock can be described by a μ-variable[b] with $\mu = 1.5$, which is an oft-cited value [42, 24, 25, 26]. Since $\mu > 1$, $\langle v \rangle$ is *a priori* well defined. We have generated 1000 samples of 500 variable $v > 0$ sequences each distributed according to Eq. (2.1) with $\mu = 1.5$ and C such that the theoretical average value of v is $\langle v \rangle = 3/2$. We have then determined the empirical average value either directly, or by reconstructing the distribution using the rank ordering method [44, 40] and computing analytically this mean value using the empirical values of μ and C. The results are $\langle v \rangle_{\text{direct}} = 1.54 \pm 0.12$ and $\langle v \rangle_{\text{reconstructed}} = 1.50 \pm 0.07$, to be compared with $\langle v \rangle_{\text{exact}} = 1.50$. In the following, we shall thus use the rank ordering technique to extract from a given time series the parameters μ, C_+, C_- and do the calculations from the reconstructed distributions.

[a]We drop from now on the distinction between μ_+ and μ_-.
[b]We use the name μ-variable to denote a random variable distributed according to the power-law probability density (2.1) with exponent μ.

3. Tail Covariance. Generalized β Model and Multifactor Models

The diversification of extreme risks is only possible if large events are to some extent uncorrelated. An important step towards obtaining a useful portfolio theory is thus to generalize the usual notion of covariance to the "tail" of the distributions. Since any definition of covariance is bound to involve products and linear combinations of the price variations of different assets, we need to use the toolbox which is available for the determination of the statistical properties of sums and products of μ-variables.

The properties, that will be useful for our purpose are the following (a brief derivation is given in Appendix A):

1. If w_i and w_j are two *independent* μ-variables, characterized by $C_i^{\pm}$ and $C_j^{\pm}$, then $w_i + w_j$ is a $\mu-$variable with $C^{\pm}$ given by $C_i^{\pm} + C_j^{\pm}$.
2. If w is a μ-variable with a certain C then $p \times w$ is a μ-variable with a C equal to $p^{\mu} C$.
3. If w is a μ-variable, then w^q is a $\frac{\mu}{q}$-variable.
4. If w_i and w_j are two *independent* μ-variables, then the variable $x = w_i w_j$ is (up to logarithmic corrections) a μ-variable. Intuitively, this means that cases where x is large corresponds to cases where — say — w_i is large and w_j takes typical values in the central part of the distribution. Cases where both w_i and w_j are very large are negligible, since these are uncorrelated variables.

3.1. *Generalized "one factor" β-model*

In order to progress, consider a particular model of correlation between all assets, which is a generalization of the well-known β-model [31]. Following previous works [18, 38, 2, 6], we assume that v_i (the daily, weekly, monthly,... variation of asset i) has a part which reflects a common evolution of all assets, and a part which is intrinsic to each asset, i.e. using property (1) above:

$$v_i \simeq \beta_i^{\pm} w_0 + w_i \qquad (v_i \to \pm\infty)\,, \tag{3.6}$$

where w_0 and $\{w_i\}$ are *independent* μ-variables with $C_{\pm}(w_0) \equiv 1$ and $C_{\pm}(w_i) \equiv \gamma_i^{\pm}$. This model was in fact first proposed by Fama [18] in the present context of strongly fluctuating assets. The aim is thus to extract the set of parameters $\beta_i^{\pm}$ and $\gamma_i^{\pm}$ from the data. In this goal, let us study the product $v_i v_j$, whose probability distribution constitutes, in the power-law world, the natural generalization of the covariance.[c] Using properties (3) and (4), one finds the following result, expressed

[c]Other generalizations were proposed in the context of stable laws, in particular the "covariation" discussed in [37] and recently applied to construct a "Stable" CAPM model in [7]. Its intuitive meaning is however, at least to our eyes, less transparent than our own definition, which explicitly measures the correlations between large events only, while to "covariation" picks up contributions from the "core" of the distributions, and mixes positive and negative variations.

rather symbolically:

$$v_i v_j \simeq \beta_i \beta_j \frac{\mu}{2}\text{-variable} + \beta_i \mu\text{-variable} + \beta_j \mu\text{-variable} + \mu\text{-variable}\,, \qquad (3.7)$$

which shows that the tail of the distribution of $v_i v_j$ is *dominated by the first term*, and will directly be sensitive to the product $\beta_i \beta_j$. More precisely, a rank ordering analysis of $v_i v_j$ should give a slope of $-\frac{2}{\mu}$ and a $C_{ij}^{\pm}$ proportional to $(\beta_i^{\pm})^{\mu}(\beta_j^{\pm})^{\mu}$.

From the set of all $C_{ij}^{\pm}$, one then checks first whether or not the "one factor" model is sufficiently accurate, and, if it is, determines all the $\beta_i^{\pm}$. From the direct rank ordering analysis of the individual v_i, one also has the values of $C_i^{\pm}$, from which, using (1), one finally gets

$$\gamma_i^{\pm} \equiv C_i^{\pm} - (\beta_i^{\pm})^{\mu}\,. \qquad (3.8)$$

Note that for the number of equations to equal the number $2M$ of unknown, where M is the number of assets taken into account, one has to choose a particular asset i_0 as the reference, i.e. set $\gamma_{i_0}^{\pm} \equiv 0$.

3.2. *Generalized "multi-factor" model*

More generally, one can easily envisage the case where the assets v_i are linear combinations of M independent μ-variables w_α, $\alpha = 1, \ldots, M$, each characterized by a tail amplitude C_α^-. Since the sum of μ-variables is still a μ-variable (point 1 above), we can write, for $v_i \to -\infty$:

$$v_i = \sum_{\alpha=1}^{M} a_i^{\alpha} w_\alpha\,. \qquad (3.9)$$

The tail covariance of assets i, j is then given by $C_{ij}^- = \sum_{\alpha=1}^{M} (a_i^\alpha a_j^\alpha)^{\frac{\mu}{2}} C_\alpha^-$, which indeed gives back the usual definition of the covariance matrix in the "gaussian limit" $\mu = 2$ if one identifies C_2 with the variance (as it should). The empirical determination of the "tail covariance matrix" C_{ij} (using the method explained above) then enables one to obtain the matrix $A = a_i^\alpha$, since the matrix $A^{\bullet\mu/2}$ (where $\bullet\mu/2$ means that each element of the matrix is raised to the power $\mu/2$) is the orthogonal matrix allowing one to diagonalize C_{ij}^-; its eigenvalues being the C_α^-.

This procedure constitutes the natural generalization of the calculation of covariance in the presence of power-law distributions of asset price variations. With this in hand, we can now turn to the determination of the "extreme risk optimized portfolio".

4. Optimal Portfolio Theory for Strongly Fluctuation Assets

Consider a given portfolio $\mathcal{P}$ characterized by the set of $\{p_i\}_{i=1,M}$, $0 \le p_i \le 1$, giving the weights of M different assets, with $\sum_{i=1}^{M} p_i = 1$. Using a generalization of assertions (1,2) above, the value V of the portfolio, given by $V = \sum_{i=1}^{M} p_i v_i$, is a μ-variable with a scale parameter given by

$$C_{\mathcal{P}}^{\pm} = \left(\sum_{i=1}^{M} p_i \beta_i^{\pm} \right)^{\mu} + \sum_{i=1}^{M} p_i^{\mu} \gamma_i^{\pm} \,. \tag{4.10}$$

This result generalizes to arbitrary μ and to assymetric distributions a result first obtained by Fama [18]. By arbitrary μ, we mean, not only $\mu < 2$ characterizing stable Lévy distributions but, also $\mu \geq 2$; and in particular the limit case $\mu = \infty$ which formally corresponds to exponential tails [11], for which the central limit theorem holds and ensures well-behaved asymptotic convergences. We have already pointed out that, even in this case, the determination of the mean return and variance could be ill-conditionned (see below for an extension of this discussion). Note that our taking account of the assymetry of the distributions is automatically done by considering separately the tails for large positive and large negative variations v. We thus avoid the rather delicate problem of the determination of the assymetry parameter of stable Lévy distributions which fit best the asset time series. Another interest in calculating separately $C_{\mathcal{P}}^{+}$ and $C_{\mathcal{P}}^{-}$ will be obvious in the sequel in order to define quality ratios, generalizing the Sharpe ratio.

$C_{\mathcal{P}}^{-}$ is the amplitude of the large loss part of the distribution of returns for portfolio $\mathcal{P}$, and $C_{\mathcal{P}}^{+}$ the corresponding amplitude for the large gain side. More precisely, the probability for a loss (resp. gain) larger than a certain value λ is given by

$$P_{\text{loss}}(|V| > \lambda) = \frac{C_{\mathcal{P}}^{-}}{\mu \lambda^{\mu}} \,, \tag{4.11}$$

respectively

$$P_{\text{gain}}(V > \lambda) = \frac{C_{\mathcal{P}}^{+}}{\mu \lambda^{\mu}} \,. \tag{4.12}$$

What should be the strategy underlying the determination of the optimal portfolio? In the context of stable Lévy distributions with $\mu \geq 1$ for which the mean exists mathematically, Fama has first proposed to minimize $C_{\mathcal{P}}$ at fixed average return, the solution of which has been provided by Samuelson [38]. This idea is the most natural extension of the standard Markowitz portfolio model [30, 39, 31], which consists of minimizing the variance of expected returns at fixed average return. Indeed, as can be seen straightforwardly from the expression of the characteristic functions of stable Lévy distributions, the coefficient $C_{\mathcal{P}}$ of the Lévy distribution degenerates into the variance of the Gaussian law which corresponds to the special borderline stable law with $\mu = 2$. Arad [2] has reviewed a variety of alternatives always written in terms of two parameters (return-risk), involving for instance linear combinations of mean and variance, or considering the probability of return larger than a minimum value. Inspired by the intuitive meaning of the Sharpe ratio, Bawa *et al.* [6]

have proposed that the optimal portfolio would be that which maximizes the ratio of the average return (minus the return of a riskless asset) divided by the scale parameter $[C_{\mathcal{P}}]^{1/\mu}$ of the portfolio. This is again directly inspired by the standard strategy used for normally distributed returns which consists of maximizing the Sharpe ratio, since the coefficient $C_{\mathcal{P}}$ of the Lévy distribution maps to the variance of the Gaussian law for $\mu = 2$.

Following a similar vein, we propose that the optimal portfolio is characterized by the set of $\{p_i\}_{i=1,M}$, *such that* $C_{\mathcal{P}}^-$ *is minimized*. This minimization must be done with some other constraints, for instance, for a given value of the expected return $R = \sum_{i=1}^{M} p_i R_i$, where R_i is the expected return of the ith asset. Another possibility is to minimize $C_{\mathcal{P}}^-$ while maximizing the probability of large gains which is proportional to $C_{\mathcal{P}}^+$. This will give the generalization of the "efficient frontier" in the usual portfolio theory and can be called a "tail chiseling" technique: the frequency of very large, unpleasant losses is minimized for a certain level of return. Note that, if this looks quite similar to previous works, there is a fundamental difference: we do not attempt to minimize the global coefficient $C_{\mathcal{P}}$ of the stable Lévy distribution taken as the best representation of the whole distribution of price variations; this strategy, common to Fama, Samuelson, Arad and Bawa *et al.*, follows the idea that it is better to have less fluctuations (both in gains and in losses) than to let the possibility that large losses coexist with large gains. The strategy discussed here focuses specifically on the scale parameter $C_{\mathcal{P}}^-$ weighting the large *losses*. In other words, this amounts to minimizing the "Value-at-Risk" of the portfolio, which is the value of λ corresponding to a certain loss probability — say 1%. In this process, since we have separated the analysis of the tails for the positive and negative variations, we are not *a priori* impeding the potential for large gains.

This strategy is justified by the common observation that losses or gains stem from the behaviour of the portfolio over a small fraction of the total investment period. Take for instance the US S&P500 index, which showed an average return of 16.2% per year over 10 years from 1983 to 1992. Over these 2526 trading days, 80% of the index return stem from the 40 best days (defined as those days where the index rose the most), corresponding to less than 1.6% of the 2526 trading days! Defining a portfolio strategy based on the importance of large gains and large losses should thus aim correctly at the part of the market moves which are significant for the behaviour of the portfolio.

The practical implementation of the minimization of $C_{\mathcal{P}}^-$ can be performed using Lagrange multipliers, which leads to

$$\mu\beta_i^- \left(\sum_{i=1}^{M} p_i \beta_i^-\right)^{\mu-1} + \mu p_i^{\mu-1}\gamma_i^- = \alpha_0 + \alpha_1 R_i \tag{4.13}$$

with α_0 and α_1 chosen as to ensure normalization of the weights and that R has its chosen value. Following the remark at the end of Sec. 1, we suggest an alternative definition of the average return R_i of the ith asset, determined only by the tails of the distribution. It is easy to show that the corresponding "tail" return, which only retains contribution from large events, is given by

$$R_i^{\text{tail}} = \frac{1}{\mu - 1}\left[(C_i^+)^{\frac{1}{\mu}} - (C_i^-)^{\frac{1}{\mu}}\right] . \tag{4.14}$$

Instead of considering the average return given by Eq. (4.14), another interesting criterion to determine optimal portfolio could be to minimize $C_{\mathcal{P}}^-$ for a given value of $C_{\mathcal{P}}^+$. The generalization of Sharpe's ratio to this case is clearly the quality factor

$$\mathcal{Q} \equiv \frac{C_{\mathcal{P}}^+}{C_{\mathcal{P}}^-} . \tag{4.15}$$

Let us note that these minimization procedures only lead to non-trivial solutions if $\mu > 1$. If $\mu < 1$, it can be shown, for instance by studying the limit $\mu \to 1$ from above, that the optimal corresponds to choosing the particular asset with the lowest C_i^- with weight $p_i = 1$. In this case, as first foreseen by Fama, increasing diversification deteriorates the risk.

The above ideas can be extended to treat cases where assets are characterized by different μ_i's, which is probably important for applications. The idea is still to minimize the probability of losses greater than a certain value λ, obtained simply by generalizing Eq. (4.11):

$$P(|V| > \lambda) = \sum_{i=1}^{M} \frac{p_i^{\mu_i} C_i^-}{\mu_i \lambda^{\mu_i}} . \tag{4.16}$$

The difference with the case where all μ are equal lies in the fact that the value of λ is irrelevant in the latter case, but will matter in general. Different loss tolerance levels λ will therefore correspond to different asset allocations. This formula (4.16) quantifies, through the factor $\frac{1}{\lambda^{\mu_i}}$, the intuition that the allocation of the assets with the smallest μ_i's will be highly susceptible to the investor attitude towards risk (i.e. to the chosen λ).

Figure 1 presents an example of the efficiency frontier for a portfolio containing 18 different assets, all characterized by different μ_+ and μ_-. The representation we have used is to plot the average number of days between losses larger than some value (here 2%) as a function of the average return normalized per day. This plot generalizes the celebrated Markowitz–Sharpe variance-return diagram. The line gives the generalized efficient border, defined as the maximum time span without large losses which can be obtained (i.e. equivalently the minimum probability of large losses) for a given expected return.

Our approach has common roots with the "safety first" criterion introduced by Roy [36], which consists of minimizing the probability of a loss larger than some predetermined value, or equivalently of minimizing the admissible loss threshold (safety level or Value at Risk) given a given low probability of tolerance.

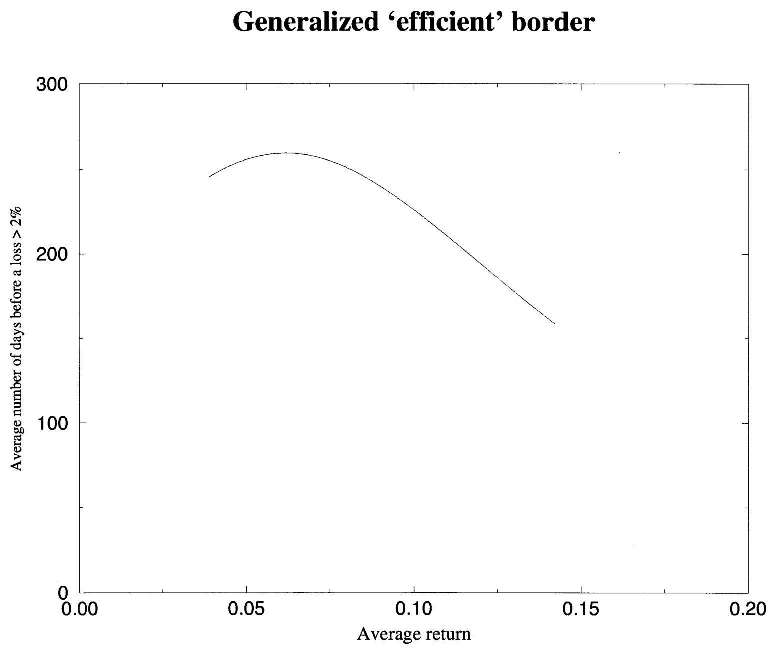

Fig. 1. Generalized efficient border for a portfolio containing 18 different assets, all characterized by different μ_+ and μ_-. The average number of days between losses larger than some value (here 2%) is shown as a function of the average return normalized per day.

Initially discussed in the framework of Gaussian statistics [36], this criterion has recently been generalized to power-law distributions [14] using extreme probability theory. However, these authors do not address the most general practical problem of finding the best weights p_i for constructing the best portfolio as we do here but rather compare a few bond and equity portfolios and propose to select the one which gives the lowest loss probability. No method is obtained to determinate the p_i's. Thus, their method is more a measure of risk than a practical portfolio optimization.

4.1. *The general (multi-factor) case*

Finally, let us give the formulae for the general case where the assets are decomposed in as Eq. (3.9): $v_i = \sum_{\alpha=1}^{M} a_i^\alpha w_\alpha$. The tail parameter of the portfolio is now given by

$$C_{\mathcal{P}}^- = \sum_{\alpha=1}^{M} \left(\sum_i a_i^\alpha p_i \right)^\mu C_\alpha^- \, . \tag{4.17}$$

The minimization of $C_{\mathcal{P}}^-$ at fixed R now leads to the equation (written in matrix form):

$$\mathcal{P} = [A^\dagger]^{-1} \left\{ (A\Delta)^{-1} (\alpha_0 + \alpha_1 \mathcal{R})^{\bullet \frac{1}{\mu - 1}} \right\} , \tag{4.18}$$

with $(\mathcal{P})_i = p_i$ and $(\mathcal{R})_i = R_i$, and α_0 and α_1 are Lagrange multipliers, and Δ the diagonal matrix made of the C_α^-'s. Again, the limit $\mu = 2$ gives back the usual Markowitz formulae. The limit $\mu \to \infty$ can also be given a precise meaning and corresponds to the case where the tail of the distributions are exponential, which is indeed the case for in the extreme fluctuation regime [11].

5. Tail Survival: The Central Limit Theorem and its Limitations

At this point, the alert reader might worry that what we have claimed is in plain contradiction with the central limit theorem, which in fact constitutes the strong argument usually used to justify the standard approach to portfolio optimization. Here again, the cases $\mu > 2$ and $\mu < 2$ must be distinguished. While it is certainly true that for $\mu > 2$, the sum $V = \sum_{i=1}^{M} p_i v_i$ converges towards a Gaussian variable in the limit $M \to \infty$ (provided the v_i are not too strongly correlated), in the case $\mu < 2$ (infinite variance), V is distributed, in the large M limit, according to a Lévy stable distribution [19, 9]. These Lévy distributions have themselves tails decaying as $V^{-1-\mu}$ for large values of V. Hence the above arguments are certainly applicable when $\mu < 2$. Our point is that if M is not very large, *our arguments are also relevant for* $\mu > 2$. In order to understand this, let us qualitatively explain how the Gaussian limit is approached. When M becomes large, the distribution of V exhibits two regimes: a *central* part, which is essentially Gaussian of relative width sharpening as $M^{-1/2}$, extends up to $V \simeq \pm V^{*\pm}$, and *tails* extending from V^* to ∞ with precisely the same power as the original variables v_i, i.e. as $V^{-1-\mu}$. Hence, schematically,

$$P(V) \simeq \frac{A_M}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{V^2}{2\sigma^2}\right) \tag{5.19}$$

for $V^{*-} < V < V^{*+}$ (we have assumed for simplicity that the mean value of V is zero) and

$$P(V) \simeq \frac{C_{\mathcal{P}}^{\pm}}{|V|^{1+\mu}}, \tag{5.20}$$

respectively for $V > V^{*+}$ and $V < -V^{*-}$, where A_M and $V^{*\pm}$ are determined by the overall normalization of $P(V)$, and by continuity of $P(V)$ around $V = \pm V^{*\pm}$. The important point is that V^* grows as $\sigma\sqrt{\log(M)}$ when M tends to infinity, i.e., the regime (5.20) disappears in that limit (leaving, as well known, the Gaussian as the limit distribution), but only very slowly. The "anomalous" weight contained in the tails (5.20) in fact decay as

$$\frac{C_{\mathcal{P}}^{\pm} M^{1-\frac{\mu}{2}}}{(\log M)^{\frac{\mu}{2}}}. \tag{5.21}$$

Note that for $\mu < 2$, this weight increases with M signaling the breakdown of the convergence to the Gaussian and the attraction to the μ-stable Lévy distribution.

However, even for $\mu > 2$ and when M is not very large, the optimization proposed here based on the idea of chiseling the tails of the distribution of returns is quite relevant. Similar arguments in fact apply when $\mu < 2$ in the presence of a finite cut-off $V_{\max}$: only for extremely large values of M the Gaussian will be recovered, while for intermediate M a description in terms of Lévy stable distributions will be adequate (see [28] for a recent discussion of this point).

6. Conclusion

Let us summarize our main points: many studies point towards the existence of power-law tails in return probability distributions, with a rather low value of the index μ, typically in the range $\mu = 1.2 - 1.8$, which is usually cut-off above a certain characteristic value $V_{\max}$ (beyond which the distribution might also be a power law with larger value for $\mu \simeq 3 - 4$ [13, 23], or more probably an exponential [29, 11]). This means that the usual concept of variance, although formally well defined, may not be relevant to the situation since it is dramatically sensitive to large events. Even the expected return is not very well defined: for a time series of length N, the relative uncertainty is of order $N^{\frac{1-\mu}{\mu}}$ ($\simeq 10\%$ for $N = 1000$ and $\mu = 1.5$: see the values quoted at the end of Sec. 1) not to speak about the intrinsic non stationarity of financial time series. We propose here to replace these quantities by their natural analogue in the case of power-law distributions, which are the tail amplitudes C_+, C_-. This generalization was in fact suggested on formal grounds and for $\mu < 2$ in Fama and Samuelson, but the link with the tail amplitudes, and a systematic method to extract the relevant parameters, was not explicited. We have suggested to use the method of rank ordering to extract the "tail covariance", and proposed explicit formulae for the optimal portfolio in an extended sense, based on the minimization of the probability of large losses. Our main point is to treat separately the tails for positive and negative variations, thus allowing an independent minimization of the probability of large losses, while keeping the potential for large gains. Similar ideas were proposed for option pricing in a power-law world in [10, 11]. Let us also stress that the idea that risk should be decomposed into three components with the probability of loss as the key quantity to study and minimize, provides a simple solution to the problem of optimization of asset allocations in the case where the exponents μ are different for each asset, or even for more general distributions. In general, the optimal portfolio in the sense of the Value at Risk is different for the Gaussian (Markowitz) optimal portfolio.

Appendix. Results on Sums and Products of Power-Law Variables

The central result about μ variables which is used to obtain properties (1–4) quoted in the text is the following. Suppose for simplicity that v is a positive variable, such that

$$P(v) \simeq_{v\to\infty} \frac{C}{|v|^{1+\mu}} \,. \tag{A.1}$$

A useful tool is to take the Laplace transform of $P(v)$: $\hat{P}(\beta) \equiv \int_0^\infty dv P(v) e^{-\beta v}$. $\hat{P}(\beta)$ then reads

$$\hat{P}(\beta) = \mu \int_1^\infty dw P(w) \frac{e^{-\beta w}}{w^{1+\mu}} = \mu \beta^\mu \int_\beta^\infty dx \frac{e^{-x}}{x^{1+\mu}} . \tag{A.2}$$

Let us define m as the integer equal to the integer part of μ, i.e. such that $m < \mu < m+1$. Integrating by part m times, we obtain (with $C = 1/\mu$)

$$\hat{P}(\beta) = e^{-\beta}\left(1 - \frac{\beta}{\mu - 1} + \cdots + \frac{(-1)^m \beta^m}{(\mu-1)(\mu-2)\ldots(\mu-m)}\right) \tag{A.3}$$

$$+ \frac{(-1)^m}{(\mu-1)(\mu-2)\ldots(\mu-m)} \int_\beta^\infty dx e^{-x} x^{m-\mu} . \tag{A.4}$$

This last integral is no more singular as $\beta \to 0$ and is given by

$$\int_\beta^\infty dx e^{-x} x^{m-\mu} = \Gamma(m+1-\mu)[\beta^\mu + \beta^{m+1}\gamma^*(m+1-\mu), \beta)] , \tag{A.5}$$

where Γ is the gamma function ($\Gamma(n+1) = n!$) and

$$\gamma^*(m+1-\mu, \beta) = e^{-\beta} \sum_{n=0}^{+\infty} \frac{\beta^n}{\Gamma(m+2-\mu+n)} \tag{A.6}$$

is the incomplete gamma function. One thus observes that $\hat{P}(\beta)$ has a regular Taylor expansion in β only up to order $m = [\mu]$, followed by a term of the form β^μ. For the general shape $P(w) = \frac{C}{w^{1+\mu}}$ with arbitrary C, we can thus write

$$\hat{P}(\beta) = 1 + c_1\beta + \cdots + c_m\beta^m + c_\mu\beta^\mu + \mathcal{O}(\beta^{m+1}) , \tag{A.7}$$

with $c_1 = -\langle v \rangle$, $c_2 = \frac{\langle v^2 \rangle}{2}, \ldots$ and c_μ is proportional to C. For small β, we can rewrite $\hat{P}(\beta)$ under the form

$$\hat{P}(\beta) = \exp\left[\sum_{k=1}^m d_k \beta^k + c_\mu \beta^\mu\right] , \tag{A.8}$$

where the coefficient d_k can be simply expressed as functions of c_k's. This expression generalizes, to arbitrary μ and the positive tail, the canonical form of the Fourier transform of stable Lévy distributions. We retrieve the Lévy distributions for $\mu < 2$ by replacing β by ik.

Now the sum of two μ-variables has a distribution whose Laplace transform is the product of the two individual Laplace transform by the convolution theorem. Hence the term proportional to β^μ is simply the sum of the corresponding c_μ, showing that the C's themselves add up, leading to (1).

Points (2,3) are trivial and simply come from a change of variable $v \to pv$ or $v \to v^q$.

Finally, if $z = v_i v_j$ where v_i and v_j are independent, then

$$P(z) \equiv \int_0^\infty \int_0^\infty dv_i dv_j P_i(v_i) P_j(v_j) \delta(z - v_i v_j) = \int_0^\infty \frac{dv_j}{v_j} P_j(v_j) P_i(\frac{z}{v_j}) \,. \quad \text{(A.9)}$$

For large values of z, one thus finds

$$P(z) \propto \frac{C_i}{z^{1+\mu}} \int_0^z dv_j v_j^\mu P_j(v_j) \simeq \frac{C_i C_j \log z}{z^{1+\mu}} \,, \quad \text{(A.10)}$$

which is the result announced in (4). Intuitively, (4) means that cases where x is large corresponds to cases where — say — w_i is large and w_j takes typical values in the central part of the distribution. Cases where both w_i and w_j are very large are negligible, since these are uncorrelated variables.

References

[1] K. Aki, *Maximum likelihood estimate of b in the formula log N = a − bm and its confidence limits*, Bull. Earthquake Res. Inst. Tokyo Univ. **43** (1965) 237–239.

[2] R. Arad Wiener, *The implications of a long-tailed distribution structure to portfolio selection and capital asset pricing*, PhD thesis, Princeton Univ. (1975).

[3] L. Bachelier, *Theory of Speculation* (translation of 1900 French edition), pp. 17–78, in Cootner, P.H. editor.

[4] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. Lett. **59** (1987) 381, Phys. Rev. **A38** (1988) 364. P. Bak, K. Chen, J. A. Scheinkman and M. Woodford, Ricerche Economiche **47** (1993) 3–30.

[5] P. Bak, M. Paczuski and S. Shubnik, preprint cond-mat (1996).

[6] V. Bawa, E. J. Elton and M. J. Gruber, J. Finance **34** (1979) 1041–1047.

[7] L. Belkacem, J. Lévy-Véhel and Ch. Walter, *Generalized market equilibrium: Stable CAPM*, Proc. AFFI conference (1995).

[8] F. Black and M. Scholes, J. Political Economy **3** (1973) 637.

[9] For a simple introduction, see: J.P. Bouchaud and A. Georges, Phys. Rep. **195** (1990) 127.

[10] J. P. Bouchaud, D. Sornette and M. Potters, *Option pricing in the presence of extreme fluctuations*, to appear in Proc. of the Newton Institute workshop on Financial Mathematics, eds. M. Dempster and S. Pliska, Cambridge Univ. Press (1996a).

[11] J. P. Bouchaud and M. Potters, *Theory of Financial Risk: Portfolios and Options*, ed. C. Godrèche, Aléa-Saclay (1997) (in French).

[12] P. H. Cootner, ed. *The Random Character of Stock Market Prices*, Cambridge, MIT Press (1964).

[13] M. M. Dacorogna, U. A. Muller, O. V. Pictet and C. G. de Vries, *The distribution of extremal exchange rate returns in extremaly large data sets*, Preprint of O&A Research Group (1995).

[14] L. de Haan, D. W. Jansen, K. Koedijk and C. G. de Vries, *Safety first portfolio selection, Extreme value theory and long run asset risks*, in *Extreme Value Theory and Applications*, eds. J. Galambos et al., 471–487, (Kluwer Academic Publ., The Netherlands (1994); D. W. Jansen and C. G. de Vries, *On the frequency of large stock returns: putting booms and busts into perspective*, Rev. of Economics and Statistics **18** (1991) 18–24.

[15] A. L. Dekkers and L. De Haan, *On the estimation of the extreme-value index and large quantile estimation*, Ann. Statistics **17**, (1989) 1795–1832.

[16] W. H. DuMouchel, *Estimating the stable index α in order to measure tail thickness: a critique*, Ann. Statistics **11** (1983) 1019–1031.
[17] B. Eistein and M. Sobel, J. Amer. Statistical Association **48** (1953) 486–502.
[18] E. Fama, Management Science **11** (1965) 404–419.
[19] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sum of independent random variables*, Addison Wesley, Reading, MA (1954).
[20] B. M. Hill, *A simple general approach to inference about the tail of a distribution*, Ann. Statistics **3**, (1975) 1163–1174. B. M. Hill, *Bayesian forecasting of extreme values in an exchangeable sequence*, J. Res. Natl. Inst. Stand. Technol. **99** (1994) 521–538.
[21] T. Hirabayashi, H. Takayasu, H. Miura and K. Hamada, Fractals **1** (1), (1993) 29–40.
[22] P. Lévy, *Théorie de l'addition des Variables Aléatoires*, Gauthier Villars, Paris (1937–1954).
[23] F. Longin, *The asymptotic distribution of extreme stock market returns*, J. Business, forthcoming (1995).
[24] B. B. Mandelbrot, J. Business **36** (1963) 394.
[25] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco (1983).
[26] B. B. Mandelbrot, *Selected works on Finance*, Springer (to appear, 1997).
[27] R. Mantegna, Physica A **179** (1991) 232.
[28] R. Mantegna and H. E. Stanley, Phys. Rev. Lett. **28** (1994) 2946–2949.
[29] R. Mantegna and H.E. Stanley, Nature **376** (6535) (1995) 46–49.
[30] H. Markovitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley and Sons, New York (1959).
[31] R. C. Merton, *Continuous-Time Finance*, Blackwell, Cambridge (1990).
[32] S. Mittnik and S. T. Rachev, *Reply to comments on "Modelling asset returns with alternative stable distributions" and some extensions*, Econometric Rev. **12** (1993) 347–389.
[33] E. Montroll and M. Shlesinger, *On the wonderful world of random walks*', in *Non-equilibrium Phenomena II, From Stochastic to Hydrodynamics, Studies in Statistical Mechanics XI*, eds. J. L. Lebowitz and E. W. Montroll) North Holland, Amsterdam (1984).
[34] U. A. Müller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz and C. Morgenegg, J. Banking and Finance **14** (1990) 1189–1208.
[35] V. Pareto, *Cours d'économie Politique*, reprinted as a volume of *Oeuvres Complètes*, Droz, Geneva (1896–1965).
[36] A. D. Roy, *Safety first and the holding of assets*, Econometrica (1952) 431–449.
[37] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York (1994).
[38] P. A. Samuelson, J. Financial and Quantitative Analysis **2** (1967) 107–122.
[39] W. F. Sharpe, Management Science (1963) 9–2, 277–293.
[40] D. Sornette, L. Knopoff, Y. Y. Kagan and C. Vanneste, *Rank-ordering statistics of extreme events: application to the distribution of large earthquakes*, J. Geophys. Res. **101** (1996).
[41] H. Takayasu, H. Miura, T. Hirabayashi and K. Hamada, Physica **A184** (1992) 127–134.
[42] Ch. Walter, Bulletin de l'Institut des actuaires français **349** (1990) 3–32 and **350** (1991) 4–23. Ch. Walter, *Lévy-stable distributions and fractal structure on the Paris market: an empirical examination*, Proc. of the first AFIR colloquium, Paris (April 1990), **3**, 241–259. Ch. Walter, PhD Thesis, Univ. of Paris, Inst. d'Etudes Politiques (1994), unpublished.

[43] G. Zaslavsky, U. Frisch and M. Shlesinger, eds. *Lévy flights and related topics in physics*, in *Lecture Notes in Physics* **450**.

[44] G. K. Zipf, *Human behavior and the Principle of Least-Effort*, Addison-Wesley, Cambridge (1949).