Università degli Studi di Milano-Bicocca

Dipartimento di Statistica e Metodi Quantitativi

Dottorato di ricerca in matematica per l'analisi

dei mercati finanziari

XXVI Ciclo

MULTIVARIATE LÉVY MODELS: ESTIMATION AND ASSET ALLOCATION

Tesi di Dottorato di ANGELA LOREGIAN

matricola 072377

Relatori: Proff. Gianluca Fusai e Laura Ballotta

Abstract

Multidimensional asset models based on Lévy processes have been introduced to meet the necessity of capturing market shocks using more refined distribution assumptions compared to the standard Gaussian framework. In particular, along with accurately modeling marginal distributions of asset returns, capturing the dependence structure among them is of paramount importance, for example, to correctly price derivatives written on more than one underlying asset. Most of the literature on multivariate Lévy models focuses in fact on pricing multi-asset products, which is also the case of the model introduced in Ballotta and Bonfiglioli (2014). Believing that risk and portfolio management applications may benefit from a better description of the joint distribution of the returns as well, we choose to adopt Ballotta and Bonfiglioli (2014) model for asset allocation purposes and we empirically test its performances. We choose this model since, besides its flexibility and the ability to properly capture the dependence among assets, it is simple, relatively parsimonious and it has an immediate and intuitive interpretation, retaining a high degree of mathematical tractability. In particular we test two specifications of the general model, assuming respectively a pure jump process, more precisely the normal inverse Gaussian process, or a jump-diffusion process, precisely Merton's jump-diffusion process, for all the components involved in the model construction.

To estimate the model we propose a simple and easy-to-implement three-step procedure, which we assess via simulations, comparing the results with those obtained through a more computationally intensive one-step maximum likelihood estimation.

We empirically test portfolio construction based on multivariate Lévy models assuming a standard utility maximization framework; for the exponential utility function we get a closed form expression for the expected utility, while for other utility functions (we choose to test the power one) we resort to numerical approximations.

Among the benchmark strategies, we consider in our study what we call a 'non-parametric optimization approach', based on Gaussian kernel estimation of the portfolio return distribution, which to our knowledge has never been used.

A different approach to allocation decisions aims at minimizing portfolio riskiness requiring a minimum expected return. Following Rockafellar and Uryasev (2000), we describe how to solve this optimization problem in our multivariate Lévy framework, when risk is measured by CVaR. Moreover we present formulas and methods to compute, as efficiently as possible, some downside risk measures for portfolios made of assets following the multivariate Lévy model by Ballotta and Bonfiglioli (2014). More precisely, we consider traditional risk measures (VaR and CVaR), the corresponding marginal measures, which evaluate their sensibility to portfolio weights alterations, and intra-horizon risk measures, which take into account the magnitude of losses that can incur before the end of the investment horizon. Formulas for CVaR in monetary terms and marginal measures, together with our approach to evaluate intra-horizon risk, are among the original contributions of this work.

Contents

List of Figures

List of Tables

Introduction

Multidimensional asset models based on Lévy processes have been introduced to meet the necessity of capturing market shocks using more refined distribution assumptions compared to the standard Gaussian framework. In particular, along with accurately modeling marginal distributions of asset returns, capturing the dependence structure among them is of paramount importance, for example, to correctly price derivatives written on more than one underlying asset. Most of the literature on multivariate Lévy models focuses in fact on pricing multi-asset products, which is also the case of the model introduced in Ballotta and Bonfiglioli (2014). Believing that risk and portfolio management applications may benefit from a better description of the joint distribution of the returns as well, we choose to adopt Ballotta and Bonfiglioli (2014) model for asset allocation purposes and we empirically test its performances. We choose this model since, besides its flexibility and the ability to properly capture the dependence among assets, it is simple, relatively parsimonious and it has an immediate and intuitive interpretation, retaining a high degree of mathematical tractability. In particular we test two specifications of the general model, assuming respectively a pure jump process, more precisely the normal inverse Gaussian process, or a jump-diffusion process, precisely Merton's jump-diffusion process, for all the components involved in the model construction.

To estimate the model we propose two alternative approaches: a simple and easy-to-implement three-step procedure and a more computationally intensive one-step maximum likelihood estimation. Both of the approaches are assessed via simulations.
In our empirical test of portfolio construction based on multivariate Lévy models, we assume a standard utility maximization framework; for the exponential utility function we get a closed form expression for the expected utility, while for other utility functions (we choose to test the power one) we resort to numerical approximations.

Among the benchmark strategies, we consider in our study what we call a 'non-parametric optimization approach', which to our knowledge has never been used, despite being quite straightforward to implement. The approach is based on Gaussian kernel estimation of the portfolio return distribution and leads to a closed expression for the expected utility in the exponential case.

A different approach to allocation decisions aims at minimizing portfolio riskiness requiring a minimum expected return. Following Rockafellar and Uryasev (2000), we describe how to solve this optimization problem in our multivariate Lévy framework, when risk is measured by conditional value at risk. Moreover we present formulas and methods to compute, as efficiently as possible, some downside risk measures for portfolios made of assets following the multivariate Lévy model by Ballotta and Bonfiglioli (2014). More precisely, we consider traditional risk measures (value at risk and conditional value at risk), the corresponding marginal measures, which evaluate their sensibility to portfolio weights alterations, and intra-horizon risk measures, which take into account the magnitude of losses that can incur before the end of the investment horizon. Formulas for CVaR in monetary terms and marginal measures, together with our approach to evaluate intra-horizon risk, are among the original contributions of this work.

The outline of the thesis is as follows. We start with a brief review of the literature on multivariate Lévy processes, with particular attention to applications in the asset allocation area.[1] Chapter 1 recalls some basic notions needed for the best understanding of the present work. In Chapter 2 we introduce Ballotta and Bonfiglioli (2014) model and we propose two alternative estimation procedures, which we assess via simulations under two specifications of the general multivariate model. In Chapter 3 we show how to perform portfolio selection in the standard framework of expected utility maximization when assets follow the multivariate Lévy model by Ballotta and Bonfiglioli (2014), empirically testing the out-of-sample performance. Formulas and methods to efficiently compute downside risk measures for portfolios made of multivariate Lévy assets are developed and applied in Chapter 4, where we also deal with portfolio selection when the objective is to minimize CVaR. Finally, we conclude and give some directions for future work.

[1] References and detailed reviews of the contributions relative to more specific topics are presented within the respective chapters.

Literature review

Dependence modeling is of paramount importance across all in quantitative finance, both in the derivatives pricing branch, when dealing with multi-asset derivatives, and in the risk and portfolio management area.

Before the introduction of alternative models based on copulas (see Patton (2009) for an overview), multivariate time series and stochastic processes were mainly represented through elliptical models, like multivariate normal or t-distributions, whose popularity comes from their mathematical tractability but is questioned by empirical stylized fact observed in financial data (Cont (2001)).

Constructing multivariate Brownian motions or diffusion based processes, like log-normal processes, is the traditional way adopted to model dependence, but it reveals many limitations: besides its very restricted symmetric dependence structure, the marginal processes poorly reproduce the dynamics of single assets.

To incorporate the well-documented heavy tails of stock returns and the volatility skew effects observed in the option market, a vast literature on more sophisticated models emerged in the last decades, stochastic volatility models and Lévy based models above all. The transition to Lévy processes is natural since they preserve the statistical properties of Brownian motion's increments, relaxing the path continuity by allowing jump-alike discontinuities, which are consistent with the real evolution of stock prices through time and give rise to flexible distributions to describe financial asset returns. Notable parametric pure jump Lévy processes are the variance gamma (VG) process (Madan and Seneta (1990), Madan et al. (1998)), the normal inverse Gaussian (NIG) process (Barndorff-Nielsen (1997)) and the more general CGMY process (Carr et al. (2003)); examples of jump-diffusion Lévy processes are the Merton model (Merton (1976)), with normal distributed jump sizes, and the double exponential jump-diffusion model by Kou (2002).

While these models successfully explain the dynamics of a single price process, modeling a higher dimensional Lévy process is not as straightforward as in the case of multivariate Brownian motion. Recently, there has been an increasing interest in developing multivariate Lévy processes.

Madan and Seneta (1990) first introduced the multivariate symmetric variance gamma

process by subordinating a multivariate Brownian motion without drift by a common gamma process. Similarly, Barndorff-Nielsen (1998) studied a multivariate NIG process, using a common inverse Gaussian subordinator. Cont and Tankov (2004), Luciano and Schoutens (2006) developed the asymmetric case, studying multivariate Lévy processes with variance gamma components:

$$X^{(i)}(t) = \theta_i \Gamma(t) + \sigma_i W^{(i)}(\Gamma(t)), \quad \text{for } i = 1, \dots, n, \tag{1}$$

where Γ is a gamma process with unit mean rate and variance rate ν, and the Brownian motions $W^{(i)}$ and $W^{(j)}$ driving different components are correlated with correlation coefficient ρ_{ij}.

These models are easily tractable, but they do not accommodate independence, and linear correlation cannot be fitted once the marginals are fixed. Moreover, sharing the same subordinator on all the marginal components makes hard their joint calibration.

To build dependence among arbitrary marginal variance gamma processes, Semeraro (2008), Luciano and Semeraro (2010a) studied the multivariate subordination of multivariate Brownian motions. The marginal processes are modeled as

$$X^{(i)}(t) = \theta_i G^{(i)}(t) + \sigma_i W^{(i)}(G^{(i)}(t)), \quad \text{for } i = 1, \dots, n, \tag{2}$$

where

$$G^{(i)}(t) = Y^{(i)}(t) + a_i Z(t) \tag{3}$$

with $Y^{(i)}$ and Z independent gamma processes.

The same formulation has been immediately extended to other time changed Brownian motions, like NIG and CGMY processes.

In this way, the case of full independence is accommodated, when all the subordinators are independent. The correlation can be fitted by choosing the parameters of the common component of the subordinator. However, using this construction, the number of involved parameters increases with the square of the number of components, making the model cumbersome when dealing with many assets. Moreover, the closed-form joint characteristic function can only be written in the case of independent Brownian motions; in the latter case the dependence mainly comes from the drift part and is sometimes too weak

for financial modeling purposes (see Wang (2009)).

A theoretical motivation supporting this kind of constructions is that every Lévy process can be written as a subordinated Brownian motion (Monroe (1978)). However, this result has a practical utility only if the subordinator is identifiable. This prevents, for example, the tractability of jump-diffusion processes, which in some cases may represent the most appropriate choice to reproduce the empirical features of asset return distributions, typically in the medium run.

Recently, a different approach, based on Lévy copulas, has been introduced by Kallsen and Tankov (2006), and Tankov (2006). A Lévy copula can be seen as a function gluing the marginal Lévy measures together to form the Lévy measure of the joint process, which is guaranteed to be a Lévy process as well. Despite the theory being elegant, the application to financial data is still challenging since both estimation and simulation can be numerically demanding. For more details we refer to Cont and Tankov (2004) and to the works mentioned above.

A further way to build multivariate Lévy processes, which we choose to adopt throughout the present work, is via linear combination of independent Lévy processes; the closure of Lévy processes under linear transformations is exploited in this case. As we discuss in Chapter 2, a very simple construction, proposed by Ballotta and Bonfiglioli (2014), leads to a flexible model which can properly describe both the marginal distributions and the dependence structure among assets. Moreover the model is relatively parsimonious, the overall number of parameters growing linearly with the number of components.

Most of the multivariate Lévy models cited above, included that of Ballotta and Bonfiglioli (2014), were introduced with the main purpose of pricing derivatives. Calibration procedures and applications to multi-asset options, such as exchange options, spread options, basket options and cross-currency options, have been developed.

Although these models are still not widespread in risk and portfolio management applications, short term allocation decisions and their riskiness assessment may be improved by modeling more realistically asset returns and capturing their dependence structure.

To our knowledge, the main works focusing on asset allocation based on multivariate Lévy processes are Madan and Yen (2007), Bertini et al. (2007), Staino et al. (2007) and Hitaj

and Mercuri (2013).

Madan and Yen (2007) apply a signal processing technique known as independent component analysis (ICA) to a multivariate financial time series, decomposing it into a linear combination of statistically independent components (ICs). Assuming that the ICs follow a variance gamma process, they build a multivariate variance gamma portfolio model, which reduces to univariate problems of component investment, solved in closed form for the exponential utility. They compare in terms of Sharpe ratio, certain equivalent and gain-loss ratio the out-of-sample performance of portfolios built using their model with those obtained assuming Gaussian ICs, observing a substantial improvement.

Staino et al. (2007) model returns as a multidimensional time-changed Brownian motion where the subordinator follows either an inverse Gaussian process or a gamma process. They discuss static and dynamic portfolio selection models using Brownian motion as benchmark model to compare results. In particular, for the static case they consider allocations based on exponential utility and on the mean-value at risk ratio. In the former case, they describe directly the portfolio distribution by a univariate Lévy model, in order to obtain an analytical expression for expected utility. In the latter they consider multivariate Lévy models (VG and NIG) for asset log-returns, observing their better, yet not substantially different, performances with respect to those obtained using the Brownian motion benchmark. However, the proposed estimation procedure appears quite cumbersome with an increasing number of assets, since it requires two maximum likelihood estimations for each marginal and a maximum likelihood estimation for each pairwise correlation coefficient. To compute the value at risk of a portfolio made of multivariate Lévy assets, Staino et al. (2007) suggest to adopt a moment based approximation approach, as the Cornish-Fisher expansion or that proposed by Iaquinta et al. (2009).
Bertini et al. (2007) integrate the work of Staino et al. (2007) focusing specifically on the multi-period allocation.

Hitaj and Mercuri (2013) perform asset allocation under three specifications of the multivariate variance gamma model: Madan and Seneta (1990), Semeraro (2008) and Wang (2009). They empirically investigate the effect of using higher moments in portfolio allocation, comparing the performance obtained using a mean-variance approach, a

mean-variance-skewness-kurtosis approach based on the fourth-order Taylor expansion of expected utility and the maximization of analytical expected utility, which can be expressed in closed form for Lévy processes when the utility function is assumed to be exponential. The parametric computation of moments and comoments according to the three multivariate variance gamma models gives better results than the use of sample moments, and including skewness and kurtosis in the decision process appears to be useful. Among the three models they take into account, the one by Semeraro (2008) has the worst performances while the one by Wang (2009) has the best ones, most likely because of their different ability in capturing the dependence structure among assets.

Chapter 1

Essential notions

Let us start introducing some fundamental concepts needed for the best understanding of the main part of this work. We mostly follow Schoutens (2003), Cont and Tankov (2004), Wasserman (2003) and Meucci (2005).

In Section 1.1 we introduce some basic definition on stochastic processes and filtrations; Section 1.2 is about characteristic functions, moments and cumulants of probability distributions; in Section 1.3 we review the most fundamental concepts on Lévy processes, closely examining the normal inverse Gaussian process and the Merton's jump diffusion process, which we choose as reference examples to carry on throughout the present work. Section 1.4 describes the estimation methods adopted in the following and Section 1.5 recalls some useful notions to evaluate estimators.

1.1 Stochastic processes and filtrations

Let T be a fixed finite time horizon. We introduce a probability space $(\Omega, \mathcal{F}, P)$, where Ω is the set of all possible outcomes, $\mathcal{F}$ is a sigma-algebra (a family of subsets of Ω containing Ω itself and closed under the complement and countable unions), and P is a function assigning probabilities to the events in $\mathcal{F}$. We say that the probability space is *complete* if for each $B \subset A \in \mathcal{F}$ such that $P(A) = 0$, we have that $B \in \mathcal{F}$. We will always work with complete spaces, since every probability space $(\Omega, \bar{\mathcal{F}}, P)$ can be completed considering $\mathcal{F} = \sigma(\bar{\mathcal{F}} \cup \mathcal{N})$, where

$$\mathcal{N} = \left\{ B \subset \Omega : B \subset A \ \text{ for some } \ A \in \bar{\mathcal{F}}, \ \text{ with } P(A) = 0 \right\}$$

and $\sigma(\mathcal{C})$ is the smallest sigma-algebra on Ω containing $\mathcal{C}$. Moreover, we add a *filtration* $\{\mathcal{F}_t\}_{0\leq t\leq T}$ to our probability space. A filtration is a non decreasing family of sub-sigma-algebras of $\mathcal{F}$: $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}_T \subset \mathcal{F}$ for $0 \leq s < t \leq T$; $\mathcal{F}_t$ represents the information available at time t, while the filtration $\{\mathcal{F}_t\}_{0\leq t\leq T}$ represents the information flow evolving with time.

We assume that the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\leq t\leq T}, P)$ satisfies the so-called 'usual conditions':

(a) $\mathcal{F}$ is complete;

(b) $\mathcal{F}_0$ contains all P-null sets of Ω (intuitively, we know which events are possible and which are not);

(c) $\{\mathcal{F}_t\}_{0\leq t\leq T}$ is right continuous, i.e. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

We will always work with filtered probability spaces satisfying the usual conditions, since for every filtered probability space $(\Omega, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}_{0\leq t\leq T}, P)$ it is possible to build the so-called usual P-augmentation completing the sigma algebra $\bar{\mathcal{F}}$ as specified above and setting

$$\mathcal{F}_t = \bigcup_{s<t} \sigma(\bar{\mathcal{F}}_s \cup \mathcal{N}) \ \ \forall\, 0 \leq t \leq T.$$

A *stochastic process* $X = \{X_t, 0 \leq t \leq T\}$ is a family of random variables defined on a complete probability space $(\Omega, \mathcal{F}, P)$. The process X is said to be adapted to the filtration $\{\mathcal{F}_t\}$ if X_t is $\mathcal{F}_t$-measurable for each t ($X_t \in \mathcal{F}_t\ \forall t$), i.e. X_t is known at time t.

Given a process X on a complete probability space, the *natural filtration* of X, $\{\mathcal{F}_t^X\}$ is the P-augmentation of $\{\bar{\mathcal{F}}_t^X\}$ where, for each $0 \leq t \leq T$, $\bar{\mathcal{F}}_t^X$ is the smallest sigma-algebra such that X_t is $\bar{\mathcal{F}}_t^X$-measurable. The natural filtration is thus the 'smallest' filtration satisfying the usual conditions and containing all the information that can be achieved observing the evolution of X through time.

1.2 Characteristic function, moments and cumulants

Let us first introduce the Fourier transform. There are several common conventions for defining the Fourier transform $\mathcal{F}$ of an integrable function $f : \mathbb{R} \to \mathbb{C}$; throughout this

work we adopt the following definition:

$$\mathcal{F}[f](u) = \int_{-\infty}^{\infty} f(x)e^{-iux}dx. \tag{1.1}$$

The characteristic function ϕ of a random variable X is defined as

$$\phi_X(u) = E[e^{iuX}] = \int_{-\infty}^{\infty} e^{iuX}dF(x) \quad u \in \mathbb{R}, \tag{1.2}$$

where F indicates the cumulative distribution function $F_X(x) = P(X \leq x)$.

In particular, if X is a continuous random variable, the characteristic function is the Fourier transform of the probability density function f_X

$$\phi_X(u) = E[e^{iuX}] = \int_{-\infty}^{\infty} e^{iuX}f_X(x)dx \quad u \in \mathbb{R}, \tag{1.3}$$

being $f_X(x) = \frac{d}{dx}F_X(x)$.

The characteristic function is always defined and it is continuous; $\phi(0) = 1$ and $|\phi(u)| \leq 1$ for all $u \in \mathbb{R}$.

It is worth noting that the Fourier transform takes convolutions into multiplications: if X and Y are two independent random variables with characteristic functions ϕ_X and ϕ_Y respectively, then the characteristic function of $X + Y$ is given by $\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)$. Moreover, the characteristic function fully and uniquely determines the distribution function F_X.

From ϕ we can easily derive the moments of X; assume that there exists the k-th moment of X, i.e. suppose $E[|X|^k] < \infty$, then:

$$E[X^k] = i^{-k}\frac{d^k}{du^k}\phi(u)\bigg|_{u=0}. \tag{1.4}$$

The function $m(u) = \phi(-iu)$, if defined for all $u \in \mathbb{R}$, is called the moment generating function.

The function $k(u) = \log m(u)$ is called cumulant function of X. The cumulants c_r, $r = 1, 2, \ldots$ are the coefficients in the Taylor expansion of the cumulant generating function about the origin:

$$k(u) = \log m(u) = \sum_{r=1}^{\infty} c_r \frac{u^r}{r!}, \tag{1.5}$$

and they can be obtained as

$$c_r = \frac{d^r}{du^r} k(u)\Big|_{u=0}. \tag{1.6}$$

In the development of this work we will deal with the first four moments and cumulants, for which the following relations hold (Cornish and Fisher (1937))

$$c_1 = \hat{\mu}_1 \tag{1.7}$$

$$c_2 = \hat{\mu}_2 \tag{1.8}$$

$$c_3 = \hat{\mu}_3 \tag{1.9}$$

$$c_4 = \hat{\mu}_4 - 3\hat{\mu}_2^2, \tag{1.10}$$

where c_r is the r-th cumulant and $\hat{\mu}_r$ is the r-th central moment, defined as

$$\hat{\mu}_r = E[(X - E(X))^r].$$

1.3 Lévy processes in finance

In this section we introduce some fundamental concepts about Lévy processes; for a deeper introduction to Lévy processes refer to Cont and Tankov (2004), Schoutens (2003) or Sato (1999).

Lévy processes can be thought of as random walks in continuous time, i.e. they are stochastic processes with independent and stationary increments.

Brownian motion, the best known of all Lévy processes, was introduced as a model for stock prices in early 1900s by Bachelier. Osborne (1959) and Samuelson (1965) refined Bachelier's model by proposing the exponential of a Brownian motion (called exponential or geometric Brownian motion) as stock price model. It was Mandelbrot (1963) who studied the first non-normal exponential Lévy process and introduced the α-stable Lévy motion with index $\alpha < 2$. A few years later, an exponential Lévy process model with a non-stable distribution was proposed by Press (1967). His log-price process is a superposition of a Brownian motion and an independent compound Poisson process with normally distributed jumps and it represented the basis for the construction of jump-diffusion models (Merton (1976), Kou (2002)). Later, models based on more general pure jump Lévy processes were introduced, such as variance gamma (VG) by Madan and Seneta (1990),

normal inverse Gaussian (NIG) by Barndorff-Nielsen (1997) and CGMY by Carr et al. (2003).

The general structure of this class of processes has been gradually discovered by de Finetti, Kolmogorov, Lévy, Khintchine and Itô, and it is described by the celebrated Lévy-Khintchine formula.

For notational simplicity, here we consider only $\mathbb{R}$-valued processes, although the results hold also for $\mathbb{R}^n$-valued processes.

Assume as given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$.

Lévy process. An adapted, càdlàg process $X := \{X_t : t \geq 0\}$, with $X_0 = 0$ almost surely, is a Lévy process if

(i) X has increments independent of the past, i.e. the random variable $X_t - X_s$ is independent of $\mathcal{F}_s$ for any $0 \leq s < t < \infty$;

(ii) X has stationary increments, i.e. $X_t - X_s$ is distributed as X_{t-s} for any $0 \leq s < t < \infty$;

(iii) X is continuous in probability, i.e. $\lim_{s\to t} X_s = X_t$, where the limit is taken in probability.

The third condition does not imply that the sample paths are continuous; it only means that for a given time t, the probability of seeing a jump at t is null: discontinuities occur at random times.

Infinite divisibility. If X is a Lévy process, then for any $t > 0$, the distribution of X_t is infinitely divisible, i.e. for every integer $n > 2$ there exist n i.i.d. random variables $X^{(1)}, \ldots, X^{(n)}$ whose sum is equal in distribution to X_t. Conversely, if F is an infinitely divisible distribution, then there exists a Lévy process X such that the law of X_1 is given by F. Examples of infinitely divisible distributions are Gaussian, gamma, Poisson, log-normal, Pareto, Student-t and α-stable distributions.

By the infinitely divisibility, the characteristic function of Lévy processes can be

expressed in a simple form. Moreover it is possible to characterize all Lévy processes in terms of their characteristic functions, as showed by the Lévy-Khintchine formula.

Lévy-Khintchine representation. The inner structure of Lévy processes has been described in full details by Paul Lévy and A. Ya. Khintchine in the 1930's in terms of their characteristic functions. Consider a triple (a, σ, ν), where $a \in \mathbb{R}$, $\sigma \geq 0$ and ν is a positive measure on $\mathbb{R}_0 := \mathbb{R}\backslash\{0\}$ (i.e. the real line, possibly excluding zero) such that

$$\int_{\mathbb{R}_0} (1 \wedge |x|^2)\nu(dx) < \infty. \tag{1.11}$$

Let X be a Lévy process; then

$$\phi_X(u;t) = E(e^{iuX_t}) = e^{t\varphi(u)}, \tag{1.12}$$

where

$$\varphi(u) = iua - u^2\frac{\sigma^2}{2} + \int_{\mathbb{R}_0} (e^{iux} - 1 - iux1_{\{|x|<1\}})\nu(dx). \tag{1.13}$$

The measure ν is called the *Lévy measure*; the triple (a, σ, ν) is called the *characteristic triple* of the process X; the function φ is also known as *Lévy exponent*.

The Lévy triple fully determines the path properties of a Lévy process. For example, if $a = 0$ and $\nu = 0$, the process is a standard Brownian motion and has continuous random paths. If $\sigma = 0$, the Lévy process is a pure jump process, meaning that there is no diffusion component. If the Lévy measure also satisfies $\nu(dx) = \lambda\delta(1)$, where $\delta(1)$ is the Dirac delta computed at 1, then we obtain a Poisson process with rate parameter λ.
Pure jump Lévy processes can also be split into two categories depending on the arrival rate of jumps. A Lévy process is called of *finite activity* if $\int_{\mathbb{R}_0} \nu(dx) < \infty$. If the integral diverges instead, then the process has *infinite activity*, which means its arrival rate of jumps is infinity.

Let us examine closely a couple of examples of Lévy processes, which we will use as basic components for building the multivariate Lévy models adopted in the application part of this work: the normal inverse Gaussian (NIG) process and the jump-diffusion (MJD) model proposed by Merton (1976). In the following we will take as known some

very basic notions on the two most important continuous time stochastic processes, the Brownian motion (Bachelier (1900)) and the Poisson process (Cox (1955)), both of which happens to be Lévy.

1.3.1 The normal inverse Gaussian process (NIG)

The NIG process belongs to a class of models based on subordinated Brownian motion, and in particular to the so-called *normal tempered stable* processes. A subordinated Brownian motion X is a Lévy process obtained by observing an arithmetic Brownian motion on a time scale governed by an independent increasing and positive Lévy process G, called subordinator. Hence X_t can be written as

$$X_t = \theta G_t + \sigma W(G_t), \quad \theta \in \mathbb{R},\ \sigma > 0, \tag{1.14}$$

where W is a Brownian motion and it is independent of the subordinator G.

The characteristic function reads

$$\phi_X(u;t) = \exp\left(t\varphi_G\left(iu\theta - u^2\frac{\sigma^2}{2}\right)\right), \quad u \in \mathbb{R}, \tag{1.15}$$

where φ_G is the characteristic exponent of the subordinator.

Usually the parameters for the distribution of the subordinator are chosen in such a way that $E(G_t) = t$, to make the random time an unbiased reflection of calendar time (see, for example, Madan et al. (1998)). One class of subordinators quite popular in financial modeling due to its mathematical tractability is the family of tempered stable processes, which have characteristic exponent

$$\varphi_G(u) = \frac{\alpha - 1}{\alpha k}\left[\left(1 - \frac{iuk}{1-\alpha}\right)^{\alpha} - 1\right], \quad u \in \mathbb{R}, \tag{1.16}$$

where $k > 0$ is the variance rate of G_t and $\alpha \in [0, 1)$ is the index of stability. As particular cases, if $\alpha = 0$, expression (1.16) is to be understood in a limiting sense, G_t is a gamma process and X is called variance gamma process (Madan and Seneta (1990), Madan et al. (1998)); if $\alpha = 1/2$, the subordinator follows an inverse Gaussian process and X is the NIG process of Barndorff-Nielsen (1997). Lévy processes based on subordinated Brownian motion are widespread in financial modeling for many reasons: they are highly tractable

from a mathematical point of view, as under trade time the log-return distribution is Gaussian, and empirical evidence reveals this to be a realistic assumption (see for example Ane and Geman (2000)); moreover, the time change construction recognizes that stock prices are largely driven by news, and the time between news, as well as their impact, is random.

As we stated above, the NIG model is the normal tempered stable process with stability parameter $\alpha = 1/2$. Its characteristic function reads

$$\phi_X(u;t) = \exp\left(\frac{t}{k}(1 - \sqrt{1 - 2iu\theta k + u^2\sigma^2 k})\right), \quad u \in \mathbb{R}. \tag{1.17}$$

The NIG process is among those Lévy processes whose probability density function is available in (semi) closed analytical form; this allows for direct maximum likelihood estimation techniques.

According to the parametrization adopted in Cont and Tankov (2004), the probability density function is

$$f_t(x) = C\exp\left(Ax\frac{K_1(B\sqrt{x^2 + t^2\sigma^2/k})}{\sqrt{x^2 + t^2\sigma^2/k}}\right), \tag{1.18}$$

where

$$A = \frac{\theta}{\sigma^2} \tag{1.19}$$

$$B = \frac{\sqrt{\theta^2 + \sigma^2/k}}{\sigma^2} \tag{1.20}$$

$$C = \frac{t}{\pi}e^{t/k}\sqrt{\frac{\theta^2}{k\sigma^2} + \frac{1}{k^2}}, \tag{1.21}$$

and $K_v(x)$ is the Bessel function of the second kind with order v.

The Lévy measure reads

$$\nu(x) = \frac{C}{|x|}\exp\{Ax\}\, K_1(|B|x); \tag{1.22}$$

since the integral of (1.22) over the real line is infinite, the NIG process has infinite activity.

We report below the first four cumulants of X_t, needed to estimate the model by method

of moments:

$$c_1 = \theta t \tag{1.23}$$

$$c_2 = \sigma^2 t + \theta^2 k t \tag{1.24}$$

$$c_3 = 3\sigma^2 \theta k t + 3\theta^3 k^2 t \tag{1.25}$$

$$c_4 = 3\sigma^4 k t + 15\theta^4 k^3 t + 18\sigma^2 \theta^2 k^2. \tag{1.26}$$

From (1.26) we can observe that θ controls the mean and the skewness of X_t (the distribution is symmetric when $\theta = 0$), σ dictates the overall variability and k controls the kurtosis or tail heaviness of the distribution.

1.3.2 The Merton's jump-diffusion model (MJD)

A Lévy jump-diffusion process has the following form

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N_t} J_i, \tag{1.27}$$

where W is a standard Brownian motion, N is a Poisson process counting the jumps of X and J_i are i.i.d. variables describing the jump sizes. The first terms in (1.27) represent the *diffusion component*, which we indicate by D_t, while $\sum_{i=1}^{N_t} J_i$ is the *jump component*. All the random objects involved, W, N and J_i (for all i), are assumed to be mutually independent. In the Merton's jump-diffusion model (Merton (1976)) jump sizes are all normally distributed, $J_i \sim \mathrm{N}(\nu, \tau^2)$ for all i. This allows to obtain the probability density function of X_t as a quickly converging series; indeed, from

$$\mathrm{P}(X_t \in A) = \sum_{k=0}^{\infty} \mathrm{P}(X_t \in A | N_t = k) \mathrm{P}(N_t = k), \tag{1.28}$$

we easily get

$$f_{X_t}(x) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{\exp\left\{-\frac{x-\mu t - k\nu}{2(\sigma^2 t + k\tau^2)}\right\}}{\sqrt{2\pi(\sigma^2 t + k\tau^2)}} \frac{(\lambda t)^k}{k!}. \tag{1.29}$$

Thus X_t has an infinite Gaussian mixture distribution with mixing coefficients given by a Poisson distribution with parameter λ.

The Lévy measure is given by

$$\nu(x) = \frac{\lambda}{\tau\sqrt{2\pi}} \exp\left\{-\frac{(x-\nu)^2}{2\tau^2}\right\} \tag{1.30}$$

and it clearly integrates to $\lambda < \infty$, meaning that the jump component of MJD model is of finite activity. The economical interpretation is that in JD models, jump are considered rare events caused by the arrival of important new information which has more than a marginal effect on prices, as compared, for example, to temporary imbalances between demand and supply.

The first four cumulants of X_t are

$$c_1 = \mu t + \lambda \nu t \tag{1.31}$$

$$c_2 = t(\sigma^2 + \lambda(\nu^2 + \tau^2)) \tag{1.32}$$

$$c_3 = t\lambda(3\tau^2\nu + \nu^3) \tag{1.33}$$

$$c_4 = t\lambda(3\tau^4 + 6\tau^2\nu^2 + \nu^4). \tag{1.34}$$

We can observe how the parameters λ, ν and τ control the non-Gaussian properties of the density (1.29); in particular, ν controls skewness (the density function is symmetric when $\nu = 0$), and the density (1.29) approaches normality when either λ or $E(J_{i,t}^2) = \nu^2 + \tau^2$ converge to zero.

The characteristic function reads

$$\phi_{X_t}(u) = \exp\left\{iu\mu t - \frac{u^2\sigma^2}{2}t + \lambda t\left(e^{iu\nu - \frac{\tau^2 u^2}{2}} - 1\right)\right\}, \quad u \in \mathbb{R}. \tag{1.35}$$

A convenient reparametrization, suggested by Duncan et al. (2009) and exploited to simplify the solution of the EM algorithm they propose to estimate the model, is obtained by setting $\nu = \alpha\sigma$ and $\tau = \beta\sigma$ (with $\beta > 0$). In the following we will apply their estimation procedure, using the two parametrizations interchangeably (see Section 2.3.1.3 and Appendix B for more details).

1.4 Estimation methods

In this section we briefly review the parametric methods we will use in the estimation of univariate Lévy processes.

1.4.1 Method of moments (MoM)

The method of moments introduced by Pearson (1894) estimates the k model parameters in such a way that the first k theoretical moments match the first k sample moments. We will use this method to initialize our maximum likelihood estimation procedures.

1.4.2 Maximum likelihood (ML)

The maximum likelihood method, introduced by Fisher (1925), selects the set of parameters that maximizes the likelihood function, i.e. the estimates are built in such a way that the observed data become the most likely outcomes of the parametric distribution. Formally, the maximum likelihood estimates are defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\ f(x_1, \ldots, x_n|\boldsymbol{\theta}), \tag{1.36}$$

where Θ is the parameter space and $f(x_1, \ldots, x_n|\theta)$ is the likelihood function, i.e. the joint probability density function of the observations, given the set of parameters θ. In practice it is often more convenient to work with the logarithm of the likelihood function, called the log-likelihood. When dealing with i.i.d. observations, with density function f, the likelihood maximization becomes

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}), \tag{1.37}$$

or, equivalently, the log-likelihood maximization becomes

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \ln f(x_i|\boldsymbol{\theta}). \tag{1.38}$$

1.4.3 ML via Expectation-Maximization algorithm (EM)

The EM algorithm proposed by Dempster et al. (1977) is an efficient iterative procedure to compute the maximum likelihood estimates in presence of missing data or latent variables.

Given a set $\mathbf{y} = (\mathbf{x}; \mathbf{z})$ of complete data, where $\mathbf{x}$ are observed and $\mathbf{z}$ are latent or missing data, and a vector of unknown parameters $\boldsymbol{\theta}$, along with a complete likelihood function $L_c(\boldsymbol{\theta}; \mathbf{y}) = f_c(\mathbf{y}|\boldsymbol{\theta})$, each iteration of the EM algorithm consists of two steps: the E-step computes the conditional expectation of the complete log-likelihood, given the observed

data $\mathbf{x}$ and under the current parameters estimate $\boldsymbol{\theta}$; the M-step finds a new set of parameters maximizing the expected log-likelihood obtained in the E-step. More precisely, above we denoted by $f_c(\mathbf{y}|\boldsymbol{\theta})$ the probability density function of the random vector Y corresponding to the complete data vector $\mathbf{y}$, and by $L_c(\boldsymbol{\theta};\mathbf{y})$ the log likelihood function that could be specified for $\boldsymbol{\theta}$ if $\mathbf{y}$ were fully observable. As the complete log-likelihood is unobservable, it is replaced by its conditional expectation given the available data $\mathbf{x}$, using the current fit for $\boldsymbol{\theta}$. Therefore, on the $k+1$ iteration of the EM algorithm

- E-step: compute[1]
$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)}) := E_{\boldsymbol{\theta}^{(k)}}\left\{\ln(L_c(\boldsymbol{\theta};\mathbf{y}))|\mathbf{x}\right\} \tag{1.39}$$
- M-step: choose $\boldsymbol{\theta}^{(k+1)}$ to maximize $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)})$ with respect to $\boldsymbol{\theta}$.

The E-step and the M-step are alternatively repeated until convergence, which may be determined setting a suitable stopping rule based either on the distance among two subsequent parameter estimates or among the difference in the respective log-likelihoods. For further details, refer for example to Ng et al. (2012).

1.4.4 Spectral generalized method of moments (sGMM)

Spectral GMM estimators, or GMM-type estimators, aim at minimizing the distance among the theoretical and the empirical characteristic functions; formally they are based on the following condition

$$E[h(u,Y;\boldsymbol{\theta})] = 0 \quad \forall u \in \mathbb{R} \tag{1.40}$$

where

$$h(u,Y;\boldsymbol{\theta}) = e^{iuY} - \phi_{\boldsymbol{\theta}}(u), \tag{1.41}$$

and $\boldsymbol{\theta}$ is the vector of true parameters.

The idea is that, since there is a one to one correspondence between the characteristic function and the density function, they enclose the same information; this fact suggests that the estimation based on the empirical characteristic function should be in principle as efficient as the maximum likelihood estimation.

The most popular way to deal with (1.40) is the discrete approach proposed by Feuerverger

[1]The conditional expectation is denoted by Q since it is the best predictor of the complete log-likelihood according to the quadratic loss.

and McDunnough (1981), which consists in choosing a finite grid $\mathbf{u} = (u_1, \dots, u_k)'$ and using the $2k$ conditions:

$$E[\mathbf{h}_\theta(Y)] = \mathbf{0}, \tag{1.42}$$

where $\mathbf{h}_\theta(Y) = (\mathbf{h}_\theta{}^{\Re}(Y), \mathbf{h}_\theta{}^{\Im}(Y))'$. The estimator is then defined as

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ [\hat{\mathbf{h}}_n(\boldsymbol{\theta})]' \mathbf{W}_n [\hat{\mathbf{h}}_n(\boldsymbol{\theta})], \tag{1.43}$$

where $\hat{\mathbf{h}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{h}_\theta(y_i)$ and $\mathbf{W}_n$ is a proper weighting matrix. Despite being a simple and widely spread method (see for example Singleton (2001), Jiang and Knight (2002), Chacko and Viceira (2003), Yu (2004)) for the estimation of Lévy models, which are defined in term of characteristic function (1.12), the choice of the grid, i.e. how many and which moment conditions to impose, is a very delicate issue which can strongly affect the efficiency of the estimates.

Although we won't go through this path, it is worth knowing that a continuous approach to the problem (1.40), called *continuous generalized method of moments* (cGMM), was proposed by Carrasco and Florens (2000); the procedure matches empirical and theoretical characteristic functions continuously over an interval. Imposing a continuum of moment conditions, the information enclosed in the empirical characteristic function is better exploited, leading to a gain in efficiency.

1.5 Evaluating estimators

In this section we define the main measures we use to evaluate estimators, i.e. error, bias and inefficiency, following Wasserman (2003) and Meucci (2005). Let $x_1, \dots, x_T$ be T data points from a same distribution, meaning that the random variables $X_1, \dots, X_T$ are i.i.d.. A point estimator $\hat{\boldsymbol{\theta}}_T$ of a parameter (or more generally a vector of parameters) $\boldsymbol{\theta}$ is some function of the data

$$\hat{\boldsymbol{\theta}}_T = g(x_1, \dots, x_T). \tag{1.44}$$

A reasonable requirement for an estimator is that it should converge to the true parameter value as more and more data are collected; more precisely, if $\hat{\boldsymbol{\theta}}_T \xrightarrow{P} \boldsymbol{\theta}$, the estimator is said to be *consistent*. Being a function of the data, $\hat{\boldsymbol{\theta}}_T$ is clearly a random variable, whose distribution is called the sampling distribution. If the expected value of the sampling

distribution coincides with the true value of the parameter, the estimator is said to be *unbiased*. The *bias* of an estimator is indeed defined (in the general multivariate case) by

$$\text{bias}^2(\hat{\theta}_T) = \left\| E(\hat{\boldsymbol{\theta}}_T) - \boldsymbol{\theta} \right\|^2, \tag{1.45}$$

where $\|\cdot\|$ denotes the Euclidean norm.

In the univariate case, the standard deviation of the sampling distribution is called standard error; in the more general multivariate case, *inefficiency* is a measure of the dispersion of the estimator defined as

$$\text{ineff}^2(\hat{\boldsymbol{\theta}}_T) = E\left\{ \left\| \hat{\boldsymbol{\theta}}_T - E[\hat{\boldsymbol{\theta}}_T] \right\|^2 \right\}. \tag{1.46}$$

We define the *loss* of an estimator as the *quadratic loss*

$$\text{loss}(\hat{\boldsymbol{\theta}}_T) = \left\| \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|^2, \tag{1.47}$$

and the *error* as the square root of the expectation of the loss (since the loss is a squared distance), i.e.

$$\text{err}(\hat{\boldsymbol{\theta}}_T) = \sqrt{E\left\{ \left\| \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|^2 \right\}}. \tag{1.48}$$

In the univariate case, the error reduces to the so called *root mean squared error* (RMSE). The following relation holds for error, bias and inefficiency

$$\text{err}^2 = \text{bias}^2 + \text{ineff}^2. \tag{1.49}$$

To make the evaluation scale independent we can normalize the loss and the error by the length of the true value, if it is not zero. So we have the relative (or percentage) loss and relative (or percentage) error, defined as

$$\text{Ploss}(\hat{\boldsymbol{\theta}}_T) = \frac{\left\| \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|^2}{\|\boldsymbol{\theta}\|^2} \tag{1.50}$$

$$\text{Perr}(\hat{\boldsymbol{\theta}}_T) = \frac{\sqrt{E\left\{ \left\| \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|^2 \right\}}}{\|\boldsymbol{\theta}\|} \tag{1.51}$$

In the univariate case, the relative error is called *relative root mean squared error* (RRMSE).

Chapter 2

Multivariate Lévy models via linear transformation

There exist several methods to build multivariate Lévy processes.

The traditional way is to subordinate a Brownian motion through a univariate subordinator (see, e.g., Monroe (1978)). Following this approach, Luciano and Schoutens (2006) propose a common gamma subordinator and an uncorrelated multidimensional Brownian motion to build a multivariate variance gamma model. Cont and Tankov (2004) and Leoni and Schoutens (2007) extend this model to correlated Brownian motions, which allow for uncorrelated asset returns. One drawback of these models is that the time change is assumed to be the same for all assets, which is contrary to economic intuition. Another problem is that the achievable range of dependence is limited. For example, asset returns cannot be independent in this approach.

Recent research has focused on alleviating these problems by using multivariate subordinators (see for example Semeraro (2008), Wang (2009), Luciano and Semeraro (2010b), Luciano and Semeraro (2010a) for multivariate variance gamma and multivariate normal inverse Gaussian models, Prause (1999) for a more general model with a generalized inverse Gaussian subordinator). Yet, as pointed out by Wallmeier and Diethelm (2012), although these models offer more flexibility with respect to the marginal distributions, some of them do not improve modeling of the dependence structure, still resting on the assumption of independent Brownian motions. Semeraro (2008) and Luciano and Semeraro (2010b) improve the richness of the correlation structure introducing correlated Brownian motions, but this is achieved at the cost of increasing the complexity of the model, with a

number of parameters growing with the square of the number of assets.

Another way to build multivariate models is based on the so called Lévy copula, which instead of being defined on the cumulative distribution function, as is the standard copula, is defined on the Lévy measure $\nu(dx)$ itself (see Section 1.3). Lévy copulas can be seen as functions gluing the marginal Lévy measures together to form the Lévy measure of the whole process. For more details refer to Tankov (2006), Barndorff-Nielsen and Lindner (2004) and Kallsen and Tankov (2006).

A third method, which is the one we will adopt in the further discussion, is via linear combination of independent Lévy processes; in fact Lévy processes are closed under linear transformations. As we see in Sections 2.1 and 2.2, through a very simple construction, as that proposed by Ballotta and Bonfiglioli (2014), we get a flexible model which can properly describe both the marginal distributions and the dependence structure among assets. Moreover the model is relatively parsimonious, since the overall number of parameters involved grows linearly with the number of components. In Section 2.3 we discuss the estimation of the model, which is one of the main contributions of the present work.

2.1 Model construction

Let us start presenting the multivariate Lévy model introduced by Ballotta and Bonfiglioli (2014), which represents the general framework for the models adopted in the following.

Proposition 1

Let $Z, Y^{(j)}, j = 1, \ldots, n$ be independent Lévy processes on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with characteristic functions $\phi_Z(u;t)$ and $\phi_{Y_j}(u;t)$, for $j = 1, \ldots, n$, respectively.

Then, for $a_j \in \mathbb{R}, j = 1, \ldots, n$

$$\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(n)})' = (Y_t^{(1)} + a_1 Z_t, \ldots, Y_t^{(n)} + a_n Z_t)' \tag{2.1}$$

is a Lévy process on $\mathbb{R}^n$ with characteristic function

$$\phi_{\mathbf{X}}(\mathbf{u};t) = \phi_Z\left(\sum_{j=1}^n a_j u_j; t\right) \prod_{j=1}^n \phi_{Y_j}(u_j;t), \mathbf{u} \in \mathbb{R}^n. \tag{2.2}$$

Proof. The proof, given in Appendix A, follows from the properties of Lévy processes.

Corollary 1

Let $\mathbf{X}$ be the multivariate Lévy process introduced in Proposition 1.

(i) For $j = 1, \ldots, n$, the m^{th} cumulant c_m of the j^{th} component of $\mathbf{X}_t$ is

$$c_m(X_t^{(j)}) = t[c_m(Y_1^{(j)}) + a_j^m c_m(Z_1)]. \tag{2.3}$$

(ii) For any $j \neq l$

$$\text{Cov}(X_t^{(j)}, X_t^{(l)}) = a_j a_l \text{Var}(Z_1)t. \tag{2.4}$$

Proof. The proof is given in Appendix A.

2.2 Strengths of the model

Interpretation. The construction of the model offers a simple and intuitive interpretation: for each $X^{(j)}$, $j = 1 \ldots, n$, the process Z can be considered as the systematic part of the risk, while the process $Y^{(j)}$ can be seen as capturing the idiosyncratic shocks.

Dependence structure. Due to the presence of the common factor Z_t, the components of $\mathbf{X}_t$ are dependent and also may jump together. The joint distribution follows by construction from the distributions chosen for Z and $Y^{(j)}$, $j = 1 \ldots, n$; in other words, the copula is implicitly defined by the model.

Corollary 2

For each $t \geq 0$, $\mathbf{X}$ is positive associated if either $a_j \geq 0$ for $j = 1 \ldots n$ or $a_j \leq 0$ for $j = 1 \ldots n$, i.e.

$$\text{Cov}(f(\mathbf{X}_t), g(\mathbf{X}_t)) \geq 0 \tag{2.5}$$

for all non decreasing functions $f, g : \mathbb{R}^n \to \mathbb{R}$ for which the covariance is well-defined.

Proof. The proof is given in full detail in Ballotta and Bonfiglioli (2014).

In general, the coefficients a_j do not have the same sign for all $j = 1 \dots n$; in this case, the components $X^{(j)}$ and $X^{(l)}$ are pairwise positive (negative) quadrant dependent if $a_j a_l > 0$ ($a_j a_l < 0$). From the construction of $\mathbf{X}$, it follows that, conditioning to the common factor Z, the components are independent; further, if $\mathbf{Y}$ is degenerate, the component of $\mathbf{X}$ are perfectly (linearly) dependent; on the other hand, if Z is degenerate, the components of $\mathbf{X}$ are independent. The dependence between components is correctly described by the pairwise linear correlation coefficient

$$\rho^{\mathbf{X}}_{j,l} = \text{Corr}(X^{(j)}_t, X^{(l)}_t) = \frac{a_j a_l \text{Var}(Z_1)}{\sqrt{\text{Var}(X^{(j)}_1)}\sqrt{\text{Var}(X^{(l)}_1)}}. \tag{2.6}$$

In fact, for fixed $a_j, a_l \neq 0$, the correlation coefficient is null if and only if $\text{Var}(Z_1) = 0$, i.e. Z is degenerate and the components are independent. Moreover, $\left|\rho^{\mathbf{X}}_{j,l}\right| = 1$ if and only if $Y^{(j)}$ and $Y^{(l)}$ are degenerate, i.e. there is no idiosyncratic factor in the components $X^{(j)}$ and $X^{(l)}$. If we consider the correlation coefficient as a function of the loading parameters, a_j and a_l, we observe that for fixed $a_j = \bar{a} > 0$ ($a_j = \bar{a} < 0$), $\rho^{\mathbf{X}}_{j,l}$ is monotone increasing (decreasing) in a_l and can take any value from -1 to 1 (from 1 to -1). In particular, $\rho^{\mathbf{X}}_{j,l} = 0$ if at least one of the loading is null (i.e. at least one of the two components has no systematic risk), while $\left|\rho^{\mathbf{X}}_{j,l}\right| = 1$ as a limit case for $|\bar{a}| = +\infty$ and $|a_l| = +\infty$ (i.e. the idiosyncratic risk plays no role in the dynamic of the components $X^{(j)}$ and $X^{(l)}$). Finally, $\text{sign}(\rho^{\mathbf{X}}_{j,l}) = \text{sign}(a_j a_l)$, therefore both positive and negative correlations can be accommodated.

Moreover, since

$$\text{Corr}(X^{(j)}_t, Z_t) = a_j \sqrt{\frac{\text{Var}(Z_1)}{\text{Var}(X^{(j)}_1)}}, \tag{2.7}$$

the pairwise correlation coefficient $\rho^{\mathbf{X}}_{j,l}$ can be expressed in terms of the correlation between each component of $\mathbf{X}$ and the systematic component Z, as

$$\rho^{\mathbf{X}}_{j,l} = \text{Corr}(X^{(j)}_t, Z_t)\text{Corr}(X^{(l)}_t, Z_t). \tag{2.8}$$

The dependence structure presented above is one of the strength points of the model proposed in Ballotta and Bonfiglioli (2014) compared to the multivariate subordinator approach of Semeraro (2008) and Luciano and Semeraro (2010a), where the correlation coefficient can be null even though the processes are dependent, and the factor copula

approach of Baxter (2007) and Moosbrucker (2006), whose models can accommodate only strictly positive correlations.

Parsimony. The relative parsimony of the model is a further point in support of the construction presented in Proposition 1. We already remarked that a multivariate subordination of correlated Brownian motions implies that the number of parameters grows with the square of the number of assets, due to the presence of a correlation matrix for the Brownian motion part of the components. The construction proposed by Ballotta and Bonfiglioli (2014) manages to model the whole range of dependencies retaining a high degree of parsimony, the number of parameters growing linearly with the number of assets.

Flexibility. A further pro of the construction above is that we are allowed to specify any univariate Lévy process for $Y^{(j)}$, $j = 1 \dots n$, and Z. The joint distribution might not be known analytically, but it can be recovered numerically using the characteristic function given in Proposition 1. Conversely, for a chosen distribution for the process $X^{(j)}$, $j = 1, \dots, n$, it is possible to impose convolution conditions on the processes $Y^{(j)}$ and Z s.t. $Y_t^{(j)} + a_j Z_t \stackrel{d}{=} X_t^{(j)}$, $\forall j = 1 \dots n$, i.e.

$$\varphi_{X^{(j)}}(u) = \varphi_{Y^{(j)}}(u) + \varphi_Z(a_j u) \quad j = 1, \dots, n. \tag{2.9}$$

The latter feature is particularly convenient when the model is aimed at describing the information provided by traded derivative contracts, as argued in Ballotta and Bonfiglioli (2014).

2.3 Model estimation

Here the present work parts from the one carried out by Ballotta and Bonfiglioli (2014); in fact, our aim is to apply the model for asset allocation purposes, and we focalize on historical estimation rather than calibration.

In particular, we propose a three-step estimation procedure, based on the assumption that the common factor Z is observable, being well proxied by the returns on a broad-based in-

dex, and a more computationally intensive one-step maximum likelihood estimation.
To assess the effectiveness of the two approaches we test them through simulation studies in two particular specifications of the multivariate model (2.1): one assumes all the involved processes to be normal inverse Gaussian processes with drift ('all-NIG'); the other assumes all the involved processes to be Merton's jump-diffusion processes ('all-MJD').

2.3.1 A three-step estimation procedure

In this section we propose a three-step estimation procedure for the multivariate Lévy model described in Section 2.1.

Let us denote with S_t the price of a financial asset. In the class of exponential-Lévy models, the price S_t is represented as

$$S_t = S_0 \exp(L_t), \tag{2.10}$$

where L is a Lévy process, whose characteristic exponent (1.13) we denote by φ. Assuming that we observe the price process on an equally-spaced time grid $t = 1, 2, \ldots, n$, the log-returns or compounded returns, defined as

$$X_t = \log\left(\frac{S_t}{S_{t-1}}\right) = L_t - L_{t-1}, \tag{2.11}$$

are i.i.d. infinitely divisible random variables whose characteristic function reads

$$\phi_{\mathbf{X}}(u) = \phi_{\mathbf{L_1}}(u) = e^{\varphi(u)}. \tag{2.12}$$

We adopt the multivariate model presented in Section 2.1 to describe the joint process of compounded stock returns.

In a preliminary stage, we need to specify the nature of Lévy processes assumed for $Y^{(j)}$, $j = 1 \ldots n$ and Z. We are in principle free to choose any Lévy process for each of them. In our leading example we suppose them to be all of the same kind, precisely all NIG processes with a drift component or all Merton's jump-diffusion processes.
We do not impose any convolution condition on $\mathbf{X}$, whose nature is fully determined by its characteristic function (2.2) once the parameters of Z, $\mathbf{Y}$ and the loadings a have been estimated.

Step 1. Systematic component.

Our main assumption is that the past realizations of the systematic component are somehow observable, being well-proxied by the returns of a broad-based index. Therefore the first step of our procedure consists in the univariate estimation of the parameters leading the process Z given the time series of the index log-returns.

This univariate estimation can be performed by any technique, therefore we shall choose the one which best suits the particular model. The maximum likelihood method is preferable in terms of efficiency if the probability density function of the process assumed for Z is available in analytical form (as it is for the NIG process), or if it can be performed via EM algorithm (as it is for the Merton's jump-diffusion process), but this is often not the case; however, the return density can be recovered numerically inverting the analytical characteristic function (1.12), for example via FFT, as suggested by Carr et al. (2003). The spectral GMM estimation described in Section 1.4.4 can be also applied.

For the NIG model, we apply maximum likelihood directly, using the method of moments estimates as starting points for the numerical optimization. The spectral GMM, though less accurate, can be an adequate alternative. See Section (2.3.1.2) for more details.

For the MJD model we apply the EM algorithm proposed by Duncan et al. (2009); in fact, ML estimation for jump-diffusion processes is not straightforward and requires a careful numerical optimization (see for example Honoré (1998)). Refer to Section (2.3.1.3) for more details.

Step 2. Loading parameters.

In the second step we estimate the loading parameters $\mathbf{a} = [a_1, \ldots, a_n]'$. The loadings determine the dependence structure among the components of the process $\mathbf{X}$, therefore they must be recovered by extracting the information enclosed in the covariance (or correlation) matrix of the stock returns.

In particular, we estimate the vector $\mathbf{a}$ fitting the non diagonal entries of the sample covariance matrix to their theoretical counterparts predicted by our multivariate model.

From (2.6), we can express the covariance matrix of $\mathbf{X}$ as

$$\text{Cov}(\mathbf{X}) = \mathbf{a}\mathbf{a}'\text{Var}(Z_1) + \text{diag}([\text{Var}(Y^{(1)}), \dots, \text{Var}(Y^{(n)})]), \tag{2.13}$$

where the non-diagonal entries depend only on the loadings and on the variance of Z_1. We then simply minimize with respect to $\mathbf{a}$ the Frobenius distance between the sample covariance and expression (2.13), setting the diagonal entries to zero in both of them. In expression (2.13) we can plug-in the sample variance of the stock index returns or the parametric expression for the variance, using the estimates obtained in Step 1; in the former case, this step turns out to be independent of the specification of the Lévy processes involved in the multivariate model construction. For a reasonable initialization of the algorithm we suggest to perform a simple linear regression of the stock returns on the broad-based index returns.

Step 3. Idiosyncratic components.

We are left with the parameters leading the processes $Y^{(j)}, j = 1 \dots n$, to be estimated. From the very definition of our model (2.1), we can write each component $Y^{(j)}$ as

$$Y^{(j)} = X^{(j)} - a_j Z. \tag{2.14}$$

The nature of the process $Y^{(j)}$ is decided exogenously (e.g. each $Y^{(j)}$ is a NIG process or a MJD process as in our main examples); given the observed time series of assets returns $x^{(j)}$ and index returns z, and plugging in the estimate of the loadings $\hat{a}_j$ obtained in Step 2, we can then recover the parameters of the process $Y^{(j)}$ by means of a univariate estimation on the fictitious time series

$$y_t^{(j)} = x_t^{(j)} - \hat{a}_j z_t. \tag{2.15}$$

As in Step 1, the univariate estimation can be performed by any technique. To estimate the NIG model we apply direct maximum likelihood. The spectral GMM estimation can also be applied, but this could be hard when the number of asset n is high, since the estimation of each component $Y^{(j)}, j = 1 \dots n$ may require different moment conditions to produce reliable results, as we realized in our simulation study.

For the MJD model we apply the EM algorithm proposed by Duncan et al. (2009).

In the following sections we present the results of a simulation study aimed at assessing the estimation procedure described in Section 2.3.1. Since the estimation of loadings is model independent[1], we start presenting the evaluation of Step 2 (Section 2.3.1.1). Sections 2.3.1.2 and 2.3.1.3 deal with the estimation assessment of Step 1 and Step 3, when $Y^{(j)}$, $j = 1 \dots n$, and Z are all assumed to be NIG processes with drift or all Merton's JD processes respectively. The parameters used for the simulations were chosen applying the three-step routine to a real dataset of daily returns of 30 stocks in the S&P500 index, using the index returns as proxy for Z.

2.3.1.1 Simulation study: loadings

In the following we present part of the results of the estimation assessment relative to Step 2, i.e. the estimation of the loadings **a**. The proper estimation of loading parameters is of crucial importance, both to correctly model the dependence structure among asset returns and because their estimates are involved in the idiosyncratic component parameters estimation. We report in Table 2.1 the relative and absolute root mean square error (1.51, 1.48), bias (1.45) and inefficiency (1.46) of the first loading a_1 as the number of assets varies in $n = [5, 10, 15, 30]$ and the length of the simulated series for the estimation varies in $T = [250, 500, 750, 1000]$.[2] We can notice that the estimates improve both for increasing T, which is an indicator of consistency, and also with the number of assets n.
Figure 2.1 shows the distributions of the estimators (on the left) and of the quadratic loss (1.47) (on the right) in the particular case $n = 30$ and $T = 250$. On the left hand plot, the true value of a_1 and the mean of the estimator are highlighted respectively with red and green dots. On the right hand plots we highlight the mean square error with a blue bar and the squared bias with an orange bar, which in this case is almost not visible, reminding that the difference among them represents the squared inefficiency.[3]

[1] We choose to use the sample variance of Z_1 in expression 2.13, to compute the theoretical covariance matrix.
[2] Analogous results for loadings other than the first, omitted for the sake of brevity, are available upon request.
[3] For Figures 2.1, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.12, 2.13 and 2.14 we refer to Meucci (2014).

a_1= 0.8898	*n=5*	*n=10*	*n=15*	*n=30*
T=250				
RRMSE	8.42%	7.88%	7.87%	7.58%
RMSE	7.50E-02	7.01E-02	7.00E-02	6.74E-02
Bias	1.13E-03	7.80E-04	2.12E-03	2.55E-03
Inefficiency	7.50E-02	7.01E-02	7.00E-02	6.74E-02
T=500				
RRMSE	5.94%	5.40%	5.44%	5.36%
RMSE	5.29E-02	4.81E-02	4.84E-02	4.77E-02
Bias	1.49E-03	5.13E-04	6.24E-05	1.90E-03
Inefficiency	5.28E-02	4.81E-02	4.84E-02	4.76E-02
T=750				
RRMSE	4.80%	4.41%	4.43%	4.36%
RMSE	4.27E-02	3.93E-02	3.94E-02	3.88E-02
Bias	1.30E-04	3.85E-04	4.97E-04	2.82E-04
Inefficiency	4.27E-02	3.93E-02	3.94E-02	3.88E-02
T=1000				
RRMSE	4.14%	3.83%	3.85%	3.81%
RMSE	3.68E-02	3.41E-02	3.42E-02	3.39E-02
Bias	3.17E-04	4.34E-04	7.33E-04	5.84E-04
Inefficiency	3.68E-02	3.41E-02	3.42E-02	3.39E-02

Table 2.1: Estimation assessment results for the loading parameter a_1.

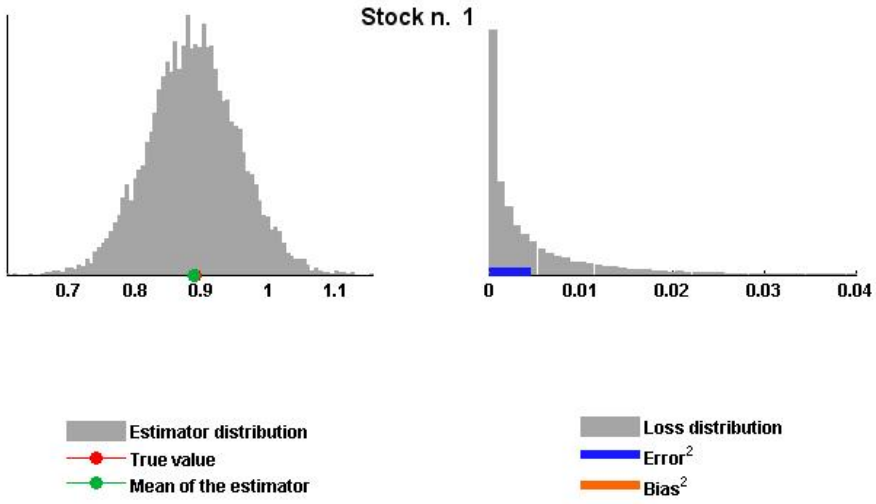

Figure 2.1: Distributions of the estimators and of the square loss for the loading parameter of the first stock (a_1). Number of simulations: 10000; length of the simulated time series: 250; number of assets: 30.

In order to analyze in more depth the behavior of the loadings estimators as the number of assets varies, we perform a further experiment, simulating datasets all made of series of a fixed length T, and with number of assets spanning the interval $[2, 60]$. To choose the parameters we first apply the three-step routine (2.3.1) to a real dataset made of 60 assets belonging to the FTSE100 index, using the index returns as proxy for the systematic factor Z. For each n we simulate a dataset and we estimate the loadings, repeating the simulation-estimation procedure 10000 times. We then compute the average error, average bias, average standard error and average interquartile range of the loadings in correspondence of each n, meaning that, given n, we compute these measures for all $a_j,\ j = 1, \dots, n$, and then we take the average.

The computations are repeated for simulated series of increasing length: $T = [250, 500, 750, 1000]$. Results are plotted in Figure 2.2. The estimates of the loadings appear to be consistent, since all the average error measures decrease when estimation is performed on longer time series.

We observe that for $n = 3, \dots, 5$ we get the highest, but still quite small errors (the average percentage root mean square error in the worst case, i.e. $n = 3$ and $T = 250$, is less than 8%), both due to higher inefficiency and bias, while for a higher number of assets, average error, bias and inefficiency decrease and almost stabilize to lower values.

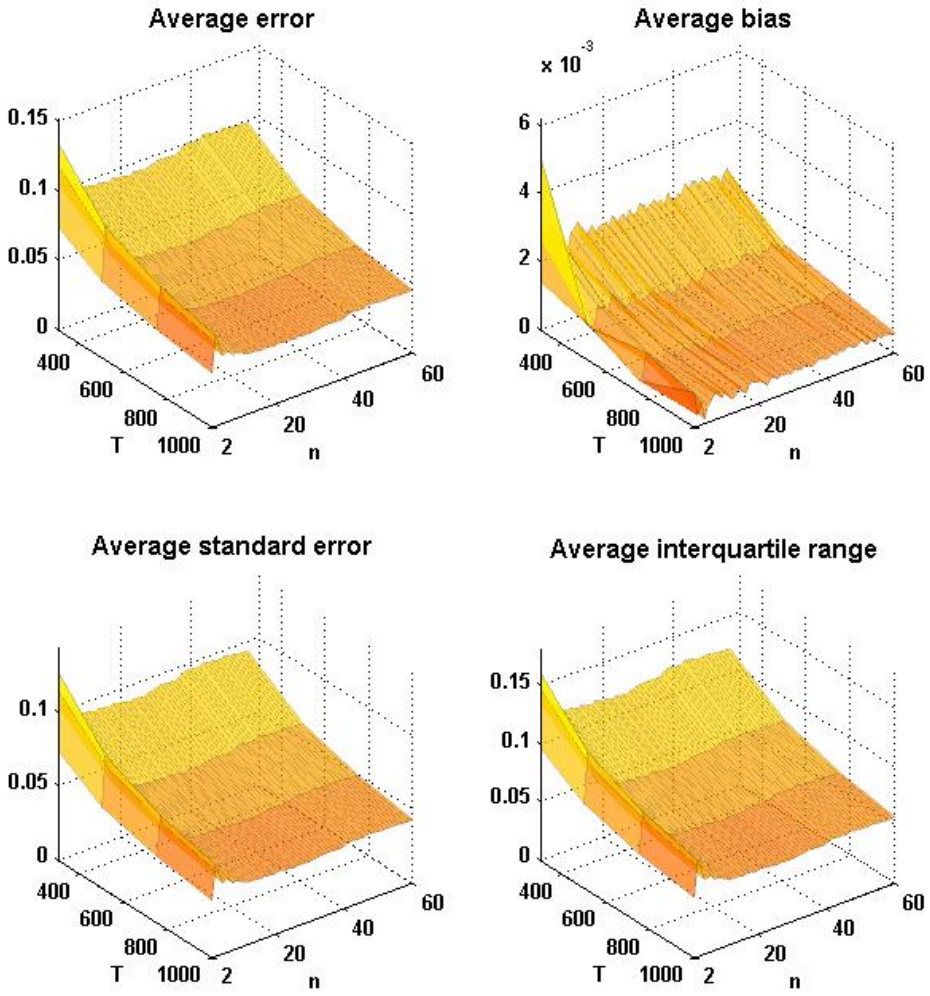

Figure 2.2: Average error, bias, standard error and interquartile range of the loadings estimates for increasing number of assets ($n = 2, \ldots, 60$) and increasing number of observations ($T = [250, 500, 750, 1000]$).

2.3.1.2 Simulation study: the 'all NIG' model

The NIG model allows us to perform direct maximum likelihood estimation in Step 1 and 3, since its probability density function is available in closed form (1.18). We introduce a drift to the standard model described in Section 1.3 to obviate the drawback that in the standard model mean and skewness are concordant, as Equations (1.23)-(1.26) show (since $k > 0$, c_1 and c_3 both have the sign of θ). As we would like to model properly return distributions with positive mean and negative skewness as well (and vice versa), we introduce a drift parameter μ. If X is the standard NIG process introduced in Section 1.3, hereafter we will call NIG the process

$$\tilde{X}_t = \mu t + X_t. \tag{2.16}$$

The characteristic function of (2.16) differs from (1.17) by a multiplicative factor $e^{i\mu t}$, the probability density function is (1.18) computed in $(x-\mu t)$ and only the first cumulant (1.23) changes, becoming $c_1 = (\mu + \theta)t$ (see Schoutens (2003)). We use the first four theoretical cumulants in order to obtain the method of moment estimates, to which we initialize the maximum likelihood optimization search, performed by means of the MATLAB® function `fminunc`. We recall from Section 2.3 that to fit the systematic component parameters we rely on the simulated series $z_1, \dots, z_t$, $t = 1 \dots, T$, while to fit the j^{th} unobservable idiosyncratic component parameters we perform a univariate estimation on the series $x_t^{(j)} - \hat{a}_j z_t$, $t = 1 \dots, T$.

The parameters used for the simulation were chosen applying the three-step routine to a real dataset of daily returns of 30 stocks in the S&P500 index, using the index returns as proxy for Z. Once fixed the parameters, we simulate the NIG processes Z and Y_j, $j = 1, \dots, n$, and then we obtain the multivariate process X as it was built in Proposition 1.

We assess the estimation procedure in 16 cases, varying the length of the simulated series ($T = [250, 500, 750, 1000]$) and the number of assets ($n = [5, 10, 15, 30]$). In each case we repeat the simulation and estimation 10000 times and, for each parameter, we compute the distribution of the loss (1.47), root mean square error (1.48), bias (1.45) and inefficiency (1.46). Since the number of parameters is high (if $n = 5$ the total number of

parameters is $4(Z)+5(a)+5\times 4(Y)=29$, if $n=30$ they are $4(Z)+30(a)+30\times 4(Y)=154$) we cannot display detailed results for each parameter; for illustrative purpose, we show only the assessment results for the estimation of Z and $Y^{(1)}$, having already examined the loadings estimation. Complete results are available upon request.

Step 1. Systematic component.

We present below the results of the estimation assessment relative to the parameters of the systematic factor Z, i.e. $\mu_Z, \theta_Z, \sigma_Z$ and k_Z.

Table 2.2 displays root mean square error (1.48), bias (1.45) and inefficiency (1.46) of the maximum likelihood estimators as the length of the simulated series varies in $T = [250, 500, 750, 1000]$. Obviously, for this step the number of assets n is irrelevant. To visualize the results we plot in Figure 2.3 the distributions of the estimators (on the left) and of the quadratic losses (1.47) (on the right) for each parameter when $T = 250$. On the left hand plots we highlight with a red dot the true value of each parameter and with a green dot the mean of the estimator. On the right hand plots we highlight the mean square error with a blue bar and the bias squared with an orange bar, reminding that the difference among them represents the squared inefficiency. The red dots and the orange bars are almost not visible in most of the plots revealing a low level of bias, and the loss distributions are sharply peaked around zero, meaning that the maximum likelihood estimators for the NIG model are suitable for the first step of our procedure. Both from Table 2.2 and from Figure 2.3 a relatively high level of inefficiency can be perceived; however, the goal of our simulation study is not to assess the maximum likelihood estimators for the NIG model, but rather to verify the effectiveness of the three step procedure we worked out. As a positive signal in this direction, we expect the errors obtained in the assessment of the third step to be comparable with those presented in this section[4].

[4] Mind that errors are expressed in absolute terms.

Z	*T=250*	*T=500*	*T=750*	*T=1000*
μ **= 0.0014**				
RMSE	9.85E-04	6.72E-04	5.42E-04	4.65E-04
Bias	4.33E-05	1.21E-05	1.84E-05	6.73E-06
Inefficiency	9.84E-04	6.71E-04	5.41E-04	4.65E-04
θ **= -0.0014**				
RMSE	1.47E-03	1.02E-03	8.20E-04	7.12E-04
Bias	3.09E-05	2.42E-05	1.87E-05	4.90E-06
Inefficiency	1.47E-03	1.02E-03	8.20E-04	7.12E-04
σ **= 0.0168**				
RMSE	1.76E-03	1.23E-03	1.01E-03	8.77E-04
Bias	1.77E-04	8.60E-05	6.41E-05	4.70E-05
Inefficiency	1.75E-03	1.22E-03	1.01E-03	8.75E-04
k **= 3.32**				
RMSE	1.30E+00	8.97E-01	7.26E-01	6.32E-01
Bias	1.91E-02	8.17E-03	5.85E-04	5.79E-03
Inefficiency	1.30E+00	8.97E-01	7.26E-01	6.32E-01

Table 2.2: Estimation assessment results for the parameters driving the systematic component Z (ML estimation).

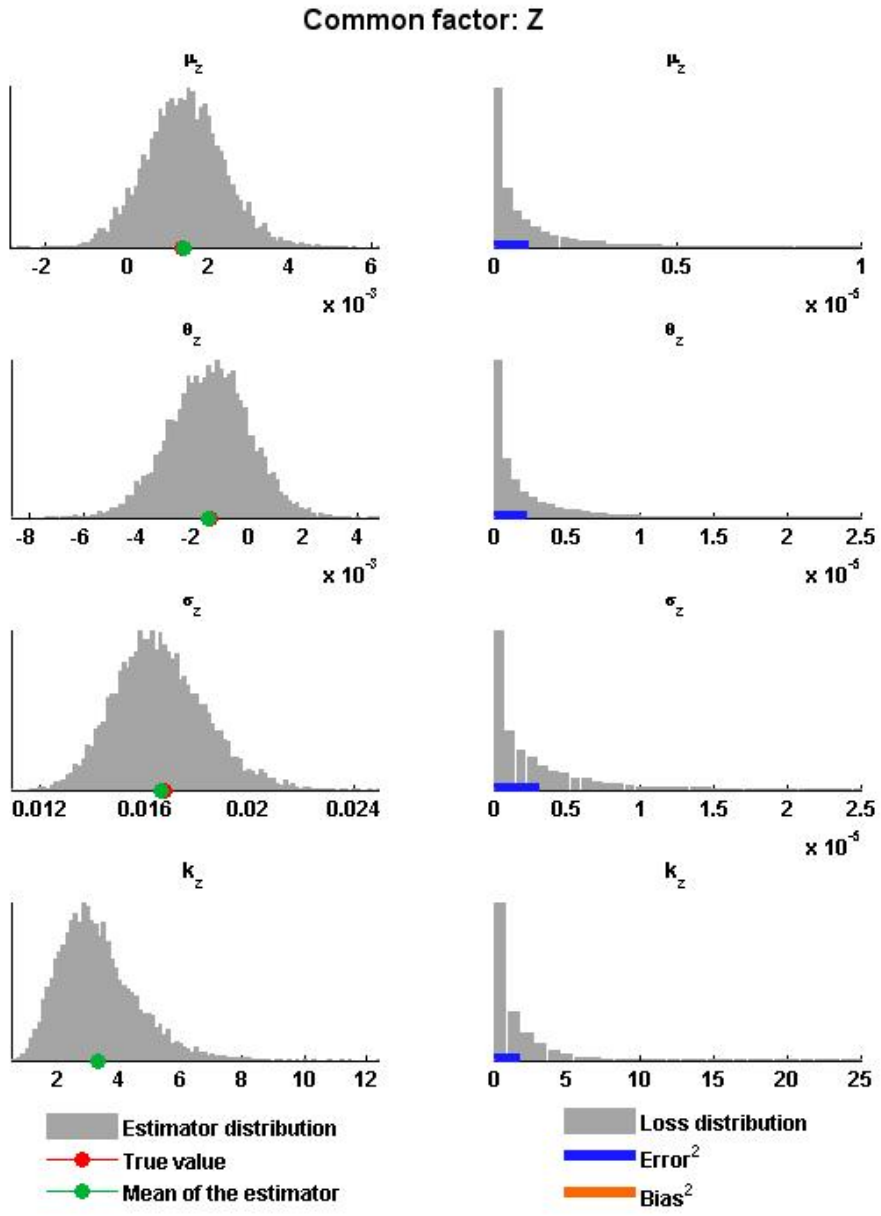

Figure 2.3: Distributions of the ML estimators and of the square losses for the parameters of the common factor Z. Number of simulations: 10000; length of the simulated time series: 250.

The spectral generalized method of moments can be also applied in the first step, especially when the probability density function is not available in closed form.
Here we adopt the discrete approach described in Section 1.4.4, implemented by means of a MATLAB® package developed by Cliff (2003). Table 2.3 presents results relative to sGMM estimators for the common factor's parameters. Figure 2.4 illustrates the case T=750. We can notice that increasing the length of the time series, the estimates for μ, θ and σ become quite accurate, while particularly hard is the estimation of k, the parameter controlling kurtosis.
Comparing Tables 2.3 and 2.2 we can see that ML estimators are preferable, displaying lower bias and inefficiency than sGMM estimators in correspondence of all the parameters.

Z	*T=250*	*T=500*	*T=750*	*T=1000*
μ **= 0.0014**				
RMSE	1.14E-01	7.27E-03	1.22E-03	9.53E-04
Bias	6.58E-04	3.54E-04	1.73E-04	1.89E-04
Inefficiency	1.14E-01	7.26E-03	1.21E-03	9.34E-04
θ **= -0.0014**				
RMSE	1.02E-01	7.33E-03	1.49E-03	1.24E-03
Bias	7.19E-04	7.50E-04	5.50E-04	5.20E-04
Inefficiency	1.02E-01	7.29E-03	1.39E-03	1.13E-03
σ **= 0.0168**				
RMSE	2.28E-03	1.78E-03	1.53E-03	1.35E-03
Bias	1.74E-03	1.40E-03	1.22E-03	1.09E-03
Inefficiency	1.48E-03	1.10E-03	9.26E-04	8.04E-04
k **= 3.32**				
RMSE	1.94E+00	1.76E+00	1.62E+00	1.51E+00
Bias	1.63E+00	1.39E+00	1.28E+00	1.21E+00
Inefficiency	1.06E+00	1.08E+00	9.81E-01	8.97E-01

Table 2.3: Estimation assessment results for the parameters driving the systematic component Z (spectral GMM estimation).

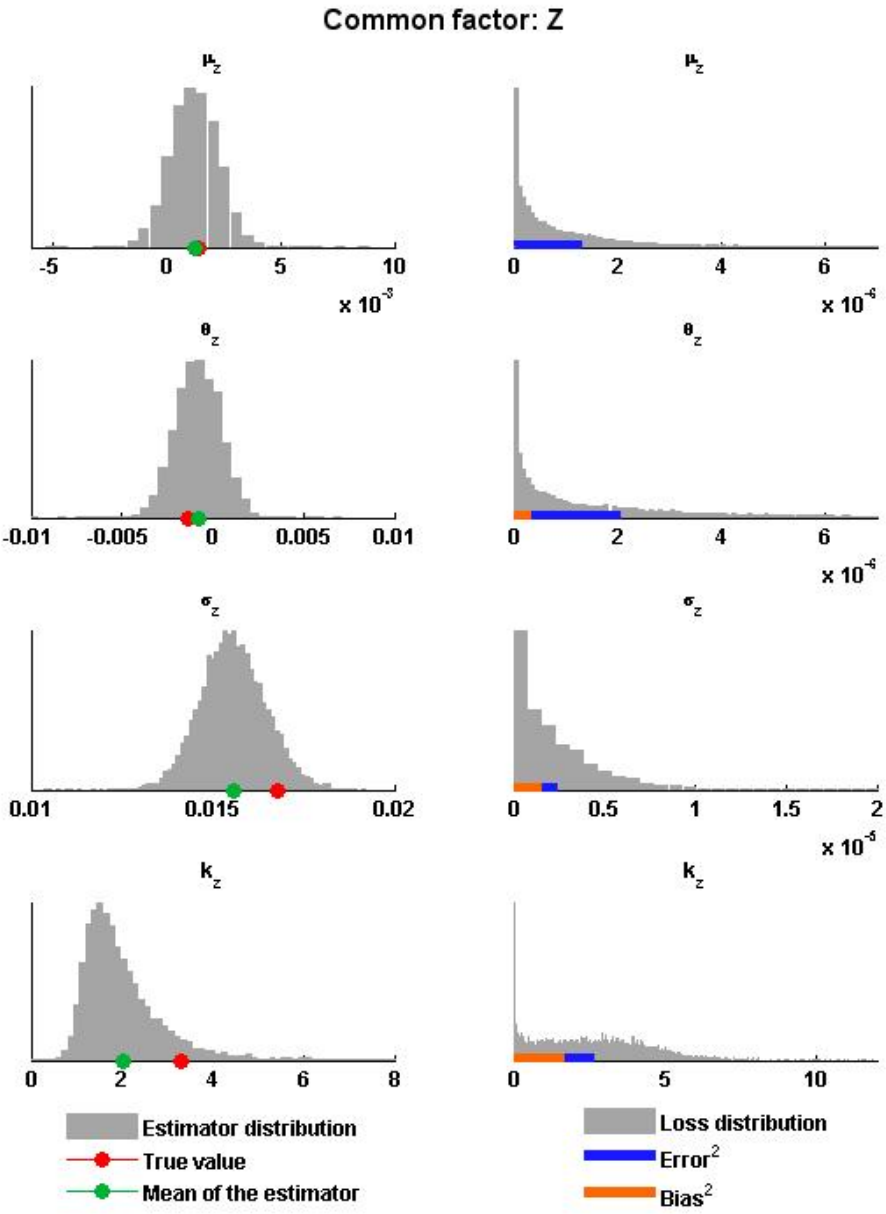

Figure 2.4: Distributions of the spectral GMM estimators and of the square losses for the parameters of the common factor Z. Number of simulations: 10000; length of the simulated time series: 750.

Step 3. First idiosyncratic component.

We present below the results of the estimation assessment for the parameters of the idiosyncratic component relative to the first stock in our simulated database, i.e. $\mu_{Y^{(1)}}, \theta_{Y^{(1)}}, \sigma_{Y^{(1)}}$ and $k_{Y^{(1)}}$.

Table 2.4 displays root mean square error (1.48), bias (1.45) and inefficiency (1.46) of the maximum likelihood estimators when the total number of assets is fixed ($n = 30$) and the length of the simulated series varies in $T = [250, 500, 750, 1000]$. In Table 2.5 we instead show the assessment results for a fixed $T = 500$, varying the number of assets. Although the estimation of each $Y^{(j)}$, $j = 1 \dots, n$, is performed univariately, the number of assets plays a key role in the estimation of the loadings, which enter and affect the estimation of the $Y^{(j)}$ parameters; however Table 2.5 reveals almost uniform estimation errors for $n = [5, 10, 15, 30]$.

To visualize the results we plot in Figure 2.5 the distributions of the estimators (on the left) and of the quadratic loss (1.47) (on the right) for each parameter when $T = 250$ and $n = 30$. As in Figure 2.3, on the left hand plots we highlight with a red dot the true value of each parameter and with a green dot the mean of the estimator. On the right hand plots we highlight the mean square error with a blue bar and the bias squared with an orange bar. The red dots and the orange bars are almost not visible in most of the plots revealing a low level of bias, and the loss distributions are squeezed above zero, meaning the third step of our estimation procedure works well. As we noticed in relation to the estimation of the systematic factor, for all of the parameters almost the whole root mean square error is due to inefficiency. Nevertheless we observe estimation errors and inefficiency levels in line with those obtained in Step 1[5], therefore splitting the estimation procedure in three steps, besides being simple, seems also to be effective.

[5]Mind that errors are expressed in absolute terms.
Analogous observations hold true for components other than the first.

$Y^{(1)}$	*T=250*	*T=500*	*T=750*	*T=1000*
μ **= 9.92E-04**				
RMSE	2.17E-03	1.13E-03	9.00E-04	7.69E-04
Bias	1.09E-05	2.80E-06	2.58E-05	4.71E-06
Inefficiency	2.17E-03	1.13E-03	9.00E-04	7.69E-04
θ **= 2.15E-04**				
RMSE	2.45E-03	1.37E-03	1.09E-03	9.40E-04
Bias	8.74E-06	9.50E-06	3.46E-05	9.08E-06
Inefficiency	2.45E-03	1.37E-03	1.09E-03	9.40E-04
σ **= 0.0173**				
RMSE	1.39E-03	9.71E-04	7.97E-04	6.74E-04
Bias	2.08E-04	1.03E-04	7.91E-05	6.38E-05
Inefficiency	1.37E-03	9.66E-04	7.93E-04	6.71E-04
k **= 1.483**				
RMSE	6.19E-01	4.31E-01	3.48E-01	3.02E-01
Bias	2.04E-02	7.46E-03	1.50E-02	7.11E-03
Inefficiency	6.19E-01	4.31E-01	3.47E-01	3.02E-01

Table 2.4: Estimation assessment results for the parameters of the idiosyncratic component $Y^{(1)}$ when $n = 30$, varying the length of the series T.

$Y^{(1)}$	*n=5*	*n=10*	*n=15*	*n=30*
μ **= 9.92E-04**				
RMSE	1.13E-03	1.12E-03	1.14E-03	1.13E-03
Bias	5.71E-06	1.19E-07	1.18E-05	2.80E-06
Inefficiency	1.13E-03	1.12E-03	1.14E-03	1.13E-03
θ **= 2.15E-04**				
RMSE	1.39E-03	1.37E-03	1.40E-03	1.37E-03
Bias	1.17E-05	8.77E-06	1.81E-05	9.50E-06
Inefficiency	1.39E-03	1.37E-03	1.40E-03	1.37E-03
σ **= 0.0173**				
RMSE	9.65E-04	9.60E-04	9.61E-04	9.71E-04
Bias	1.05E-04	1.05E-04	9.35E-05	1.03E-04
Inefficiency	9.60E-04	9.54E-04	9.56E-04	9.66E-04
k **= 1.483**				
RMSE	4.29E-01	4.28E-01	4.28E-01	4.31E-01
Bias	1.41E-02	1.53E-02	1.11E-02	7.46E-03
Inefficiency	4.29E-01	4.27E-01	4.28E-01	4.31E-01

Table 2.5: Estimation assessment results for the parameters of the idiosyncratic component $Y^{(1)}$ when $T = 500$, varying the total number of assets n.

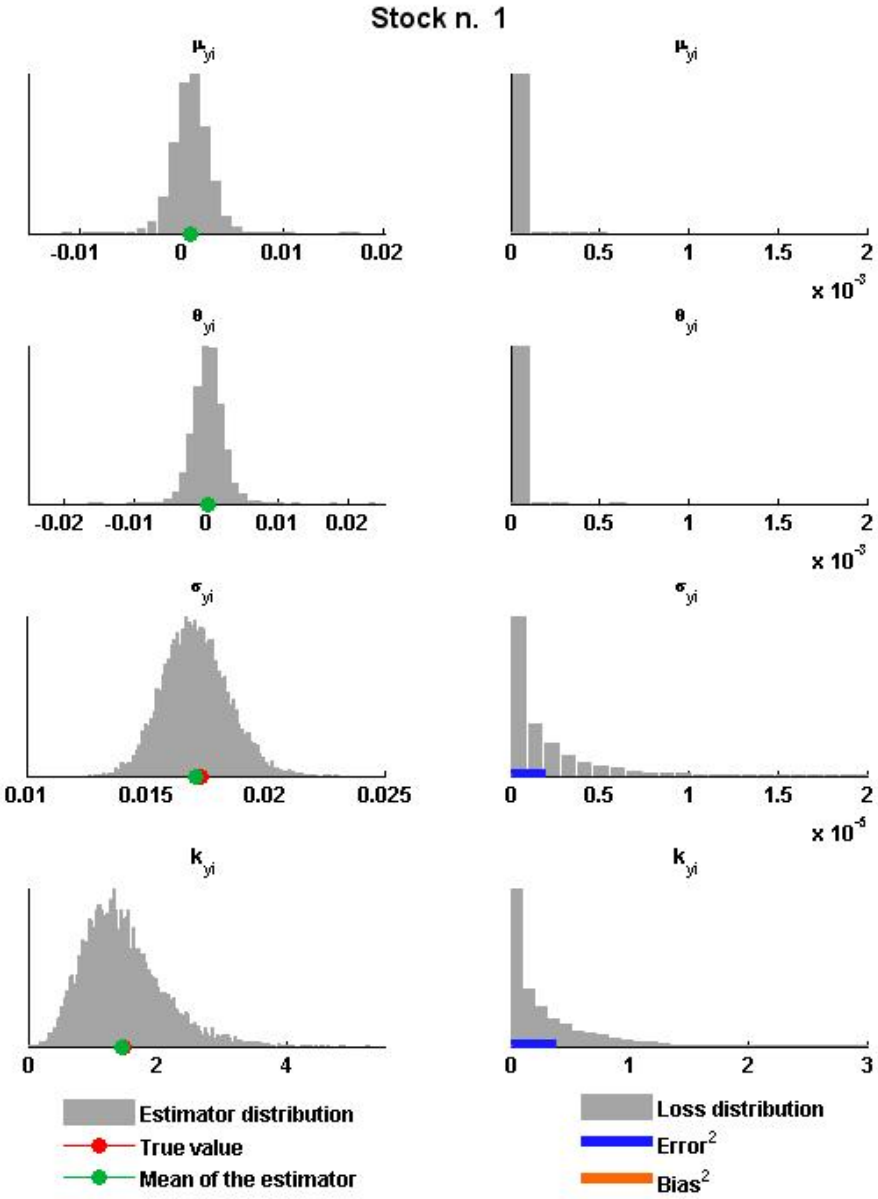

Figure 2.5: Distributions of the estimators and of the square losses for the parameters of the first stock idiosyncratic factor $Y^{(1)}$. Number of simulations: 10000; length of the simulated time series: 250; number of assets: 30.

2.3.1.3 Simulation study: the 'all MJD' model

In this section we present part of the results of a simulation study similar to the above, where all the involved processes are supposed to be Merton's JD.

As the estimation of the loadings is model independent[6], here we focus only on Step 1 and Step 3. To fit the Merton's model we implement the EM algorithm in the formulation proposed by Duncan et al. (2009), since ML estimation for jump-diffusion processes is not straightforward and requires a careful numerical optimization (see for example Honoré (1998)). The EM algorithm of Duncan et al. (2009) is particularly efficient as they provide simple closed form solutions for the M-step. For a detailed description of the algorithm we refer to Appendix B.

Step 1. Systematic component.

We present below the results of the estimation assessment relative to the parameters of the systematic factor Z, i.e. μ_Z, σ_Z^2, ν_Z, τ_Z^2 and λ_Z.

As we did for the NIG model in Section 2.3.1.2, we show in Table 2.6 the root mean square error (1.48), bias (1.45) and inefficiency (1.46) of each estimator when the estimation is performed on time series of increasing length. In Figure 2.6 we report the distribution of each estimator and of the respective loss (1.47). Even in the MJD case we can see that the ML estimators obtained by EM are almost unbiased and we can take the errors and inefficiency levels as terms for comparison to evaluate Step 3[7].

[6] We chose to use the sample variance of Z_1 in expression 2.13, to compute the theoretical covariance matrix.
[7] Mind that errors are expressed in absolute terms.

Z	*T=250*	*T=500*	*T=750*	*T=1000*
μ **= 0.0012**				
RMSE	8.24E-04	5.83E-04	4.66E-04	4.05E-04
Bias	2.80E-05	1.83E-05	2.49E-05	2.66E-05
Inefficiency	8.23E-04	5.83E-04	4.65E-04	4.04E-04
σ^2**= 5.75E-05**				
RMSE	2.23E-05	1.59E-05	1.58E-05	1.38E-05
Bias	3.22E-06	2.70E-06	3.17E-06	2.74E-06
Inefficiency	2.21E-05	1.57E-05	1.54E-05	1.35E-05
ν **= -0.0025**				
RMSE	3.14E-03	1.90E-03	1.51E-03	1.28E-03
Bias	1.35E-04	1.72E-04	6.60E-05	6.62E-05
Inefficiency	3.13E-03	1.89E-03	1.51E-03	1.28E-03
τ^2 **= 0.0004**				
RMSE	1.90E-04	1.52E-04	1.35E-04	1.23E-04
Bias	4.44E-05	3.36E-05	3.34E-05	2.90E-05
Inefficiency	1.85E-04	1.48E-04	1.31E-04	1.19E-04
λ**=0.47**				
RMSE	1.50E-01	1.07E-01	8.72E-02	7.88E-02
Bias	1.48E-02	1.94E-02	2.33E-02	2.01E-02
Inefficiency	1.49E-01	1.05E-01	8.40E-02	7.62E-02

Table 2.6: Estimation assessment results for the parameters driving the systematic component Z.

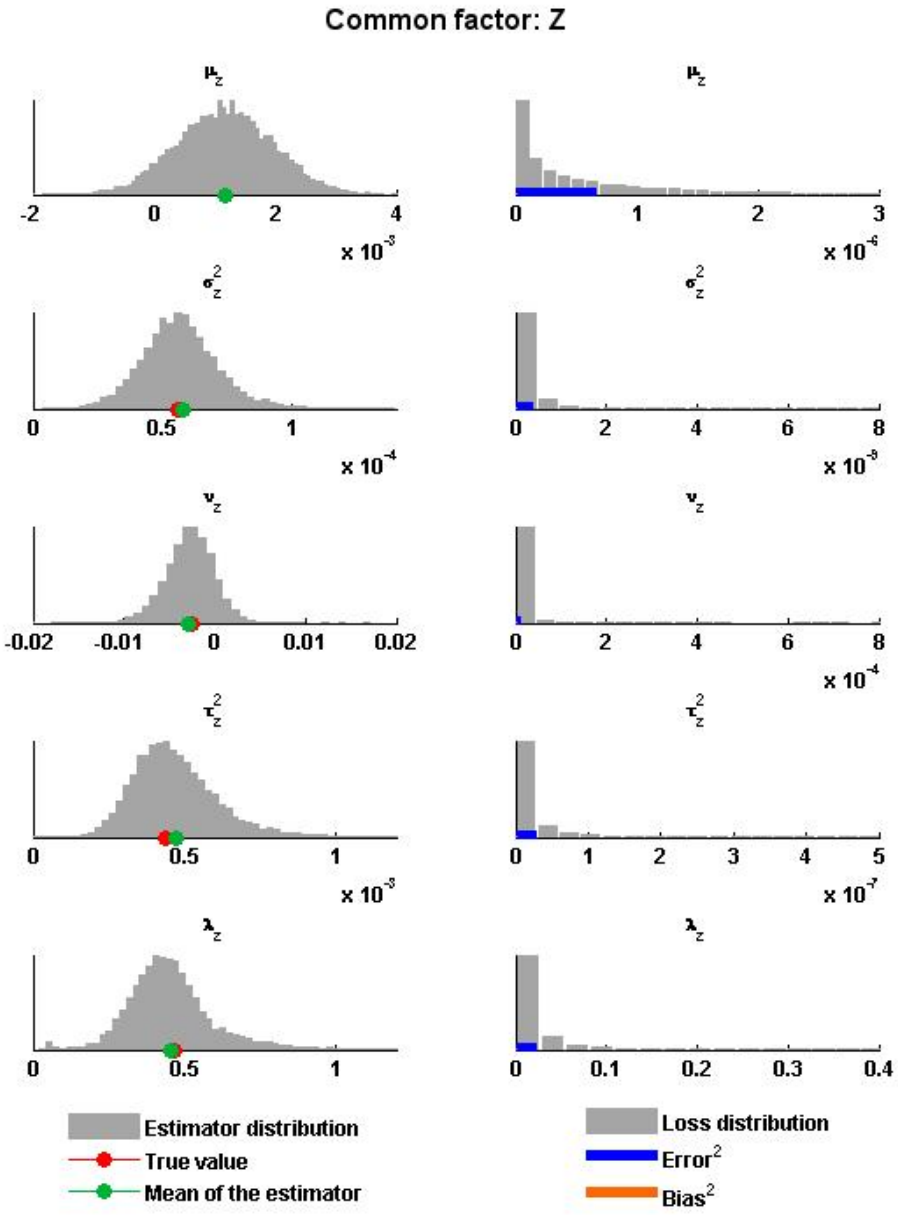

Figure 2.6: Distributions of the estimators and of the square losses for the parameters of the common factor Z. Number of simulations: 10000; length of the simulated time series: 250.

Step 3. First idiosyncratic component.

We present below the results of the estimation assessment for the parameters of the idiosyncratic component relative to the first stock in our simulated database, i.e. $\mu_{Y^{(1)}}, \sigma^2_{Y^{(1)}}, \nu_{Y^{(1)}}, \tau^2_{Y^{(1)}}$ and $\lambda_{Y^{(1)}}$.[8]

Table 2.7 displays root mean square error (1.48), bias (1.45) and inefficiency (1.46) of the maximum likelihood estimators when the total number of assets is fixed ($n = 30$) and the length of the simulated series varies in $T = [250, 500, 750, 1000]$. In Table 2.8 the length of the time series is fixed ($T = 500$) and the number of assets varies in $n = [5, 10, 15, 30]$. As observed in relation to the 'all NIG' case, the number of assets, which enters in the loadings estimation, has only little impact on the estimation errors of the idiosyncratic terms. Figure 2.7, analogously to the previous ones, shows estimators and losses distributions referring to the case $T = 250$ and $n = 30$.

Both from graphs and tables we can notice that estimation errors and inefficiency levels are again in line with those obtained in Step 1[9], standing in support of our three-step routine for the 'all-MJD' model as well.

[8] Complete results are available upon request.

[9] Mind that errors are expressed in absolute terms.
Analogous observations hold true for components other than the first.

$Y^{(1)}$	*T=250*	*T=500*	*T=750*	*T=1000*
μ **= 0.00133**				
RMSE	1.10E-03	7.61E-04	6.12E-04	5.33E-04
Bias	9.29E-06	3.76E-06	7.53E-06	1.45E-07
Inefficiency	1.10E-03	7.61E-04	6.12E-04	5.33E-04
σ^2**= 0.00012**				
RMSE	3.02E-05	2.29E-05	2.00E-05	1.93E-05
Bias	1.22E-07	5.15E-07	3.35E-07	7.67E-07
Inefficiency	3.02E-05	2.29E-05	2.00E-05	1.93E-05
ν **= -0.0004**				
RMSE	7.78E-03	3.26E-03	2.43E-03	2.03E-03
Bias	1.62E-04	3.30E-05	3.83E-06	5.59E-06
Inefficiency	7.78E-03	3.26E-03	2.43E-03	2.03E-03
τ^2 **=0.00059**				
RMSE	3.30E-04	2.36E-04	2.07E-04	2.02E-04
Bias	4.81E-05	3.03E-05	2.55E-05	2.90E-05
Inefficiency	3.27E-04	2.34E-04	2.06E-04	2.00E-04
λ **=0.29214**				
RMSE	1.61E-01	1.21E-01	1.03E-01	9.06E-02
Bias	1.62E-02	1.56E-02	1.17E-02	5.44E-03
Inefficiency	1.60E-01	1.20E-01	1.02E-01	9.04E-02

Table 2.7: Estimation assessment results for the parameters of the idiosyncratic component Y_1 when $n = 30$, varying the length of the series T.

$Y^{(1)}$	*n=5*	*n=10*	*n=15*	*n=30*
μ **= 0.00133**				
RMSE	7.55E-04	7.63E-04	7.57E-04	7.61E-04
Bias	1.59E-06	1.28E-05	2.36E-06	3.76E-06
Inefficiency	7.55E-04	7.63E-04	7.57E-04	7.61E-04
σ^2**= 0.00012**				
RMSE	2.33E-05	2.32E-05	2.30E-05	2.29E-05
Bias	3.61E-08	1.80E-07	6.70E-07	5.15E-07
Inefficiency	2.33E-05	2.32E-05	2.30E-05	2.29E-05
ν **= -0.0004**				
RMSE	3.25E-03	3.18E-03	3.09E-03	3.26E-03
Bias	4.24E-05	4.42E-05	1.10E-05	3.30E-05
Inefficiency	3.25E-03	3.18E-03	3.09E-03	3.26E-03
τ^2 **=0.00059**				
RMSE	2.42E-04	2.38E-04	2.35E-04	2.36E-04
Bias	3.34E-05	3.44E-05	2.99E-05	3.03E-05
Inefficiency	2.40E-04	2.36E-04	2.33E-04	2.34E-04
λ **=0.29214**				
RMSE	1.21E-01	1.20E-01	1.21E-01	1.21E-01
Bias	1.54E-02	1.39E-02	1.62E-02	1.56E-02
Inefficiency	1.20E-01	1.19E-01	1.19E-01	1.20E-01

Table 2.8: Estimation assessment results for the parameters of the idiosyncratic component Y_1 when $T = 500$, varying the total number of assets n.

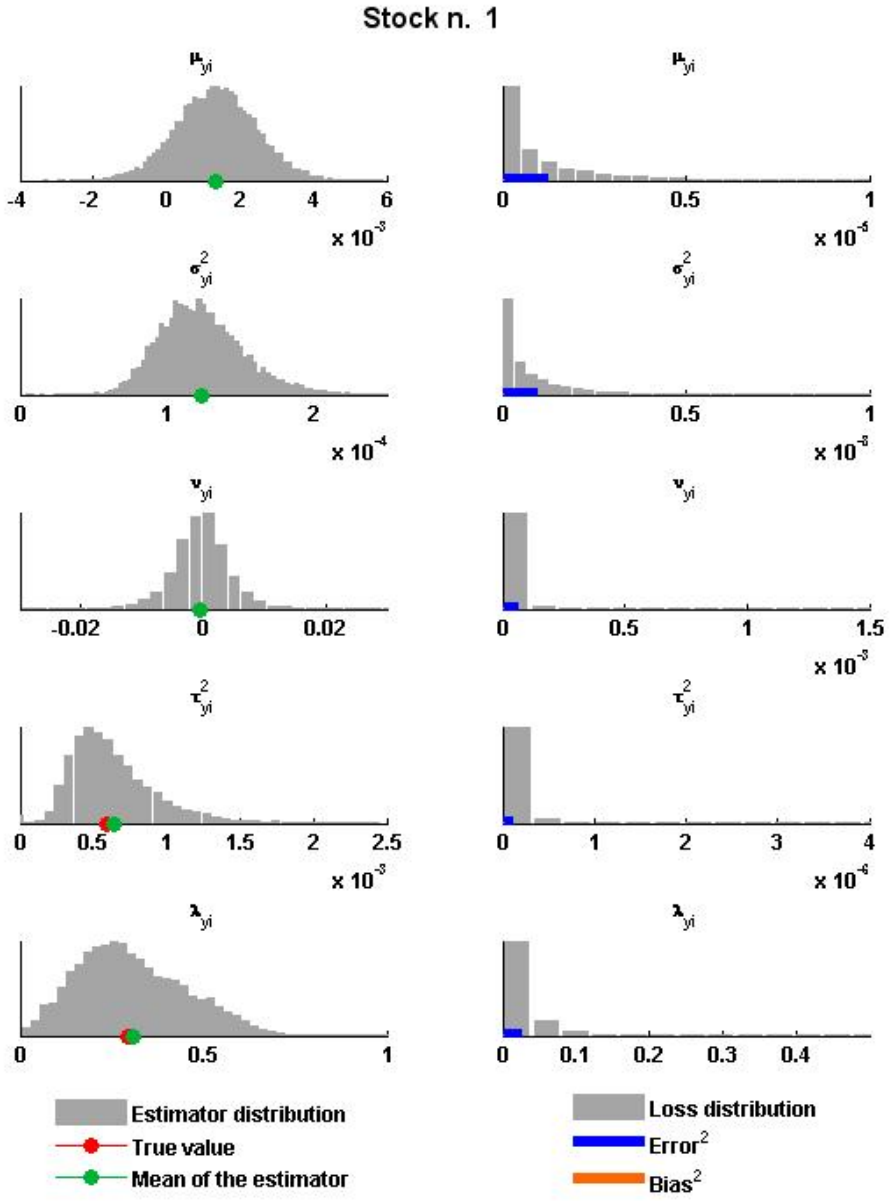

Figure 2.7: Distributions of the estimators and of the square losses for the parameters of the first stock idiosyncratic component $Y^{(1)}$. Number of simulations: 10000; length of the simulated time series: 250; number of assets: 30.

2.3.2 One-step approach: maximum likelihood

In this section we show how to perform the maximum likelihood estimation of all the parameters involved in our multivariate Lévy model (2.1) on a single step, avoiding the assumption of the common factor Z being observable. This approach is feasible especially when the probability density function of the Lévy models chosen for the systematic and idiosyncratic factors are available in closed form, as in the 'all-NIG' and 'all-MJD' specifications. If this is not the case, though still viable, the one-step approach may become computationally challenging.

The joint probability density function of $(x_{1,t}, \dots, x_{n,t})$ reads

$$\begin{aligned} f_X(x_{1,t}, \dots, x_{n,t}) &= \int_{-\infty}^{\infty} f_{X|Z}((x_{1,t}, \dots, x_{n,t})|z) f_Z(z) dz \\ &= \int_{-\infty}^{\infty} f_{Y_1}(x_{1,t} - a_1 z) \cdot \ldots \cdot f_{Y_n}(x_{n,t} - a_n z) f_Z(z) dz, \end{aligned} \tag{2.17}$$

where we used the independence of the Y_j, $j = 1 \dots n$, among themselves and of Z. The likelihood function of the sample $\mathbf{x} = (x_{1,t}, \dots, x_{n,t})_{t=1\dots T}$ then follows as

$$L(\mathbf{x}, \boldsymbol{\theta}) = \prod_{t=1}^{T} \left[\int_{-\infty}^{\infty} f_{Y_1}(x_{1,t} - a_1 z; \boldsymbol{\theta}_{Y_1}) \cdot \ldots \cdot f_{Y_n}(x_{n,t} - a_n z; \boldsymbol{\theta}_{Y_n}) f_Z(z; \boldsymbol{\theta}_Z) dz \right], \tag{2.18}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_{Y_1}, \dots, \boldsymbol{\theta}_{Y_n}, \boldsymbol{\theta}_Z, \mathbf{a}]$.

To compute the likelihood (2.18) we fix a grid of values for Z and we numerically evaluate the integral by means of the MATLAB® function `trapz`; to perform the maximization we use the MATLAB® function `fminsearch`, exploiting the three-step procedure described in Section 2.3.1 to initialize the vector of parameters.

2.3.2.1 Simulation study: the 'all-NIG' model

The maximum likelihood estimation under the 'all-NIG' model consists in maximizing the likelihood function (2.18), where all the probability density function involved are NIG (1.18)[10], with respect to the $5 \times n + 4$ parameters of the model.

In this section we present part of the results of a simulation study aimed at assessing

[10] with drift (see Section 2.3.1.2).

the effectiveness of the one-step ML approach in the estimation of the 'all-NIG' model parameters. Due to the computational cost of the procedure, here we evaluate the estimation for a small number of assets ($n = 5$, i.e. 24 parameters to be estimated) repeating the simulation 1000 times; we then perform 100 simulations to evaluate the estimation for $n = 15$ assets (i.e. 79 parameters), leaving to future studies a deeper assessment of the overall-ML estimation method for the all-NIG model.
Results relative to the common factor Z are reported in Table 2.9; those relative to the first idiosyncratic component $Y^{(1)}$, including the loading a_1, are displayed in Table 2.10. Complete results are available upon request.
The estimators and loss distributions for $n = 5$ are plotted in Figures 2.8, 2.9 and 2.10.

Keeping in mind the different number of simulations performed[11], the results reported in Table 2.9 can be compared to those displayed in the second column of Table 2.2, corresponding to estimates based on $T = 500$ observations, while the results in Table 2.10 can be compared with those in the first and third columns of Table 2.5; the errors relative to the first loading estimates can be compared to those reported in Table 2.1, where $T = 500$, $n = 5, 15$. We observe that the errors obtained with the three-step procedure, using ML estimation, are in line with those obtained with the one-step ML approach, which in principle should be the most effective method, exploiting all at once the whole information contained in the data.

[11] A higher number of simulations leads in general to higher errors, due to higher inefficiency, since the variability of the estimates tends to increase.

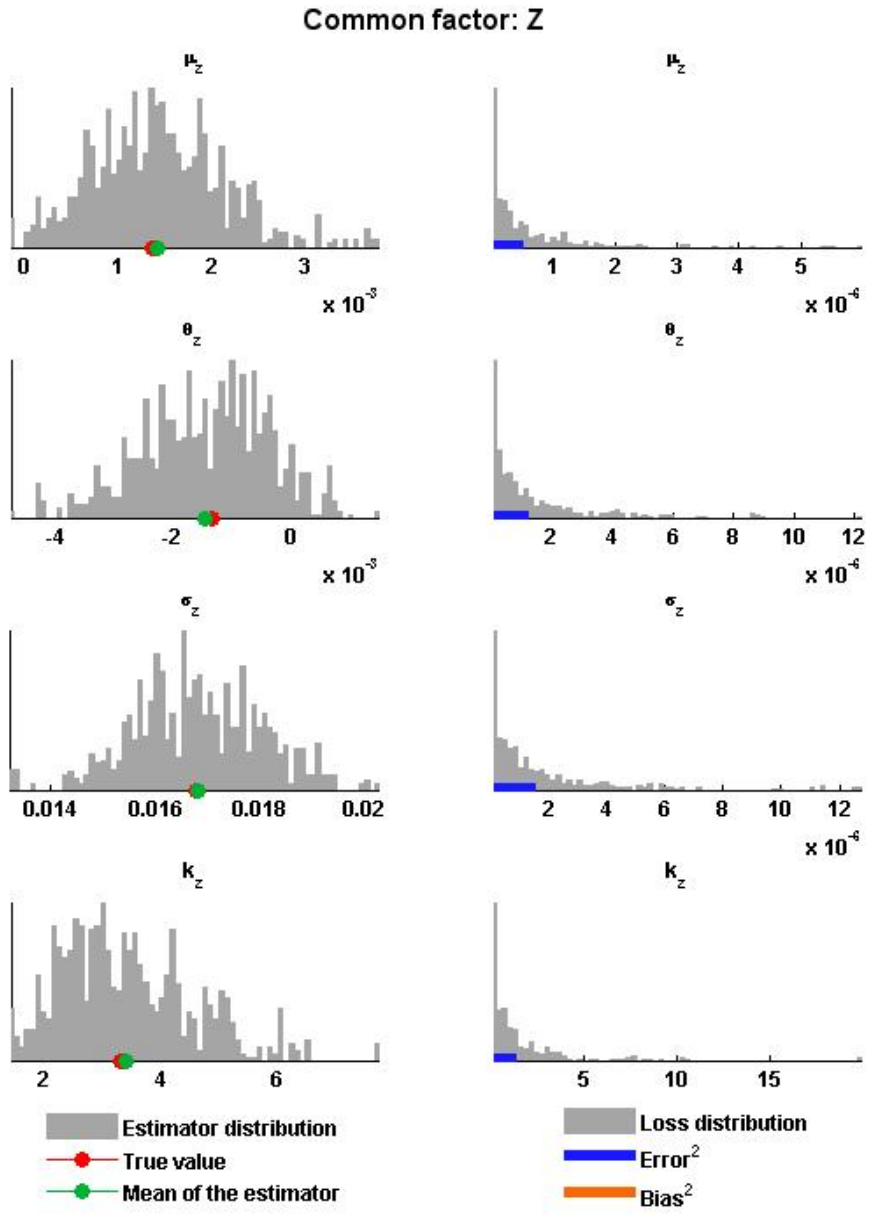

Figure 2.8: Distributions of the estimators and of the square losses for the parameters of the common factor Z. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

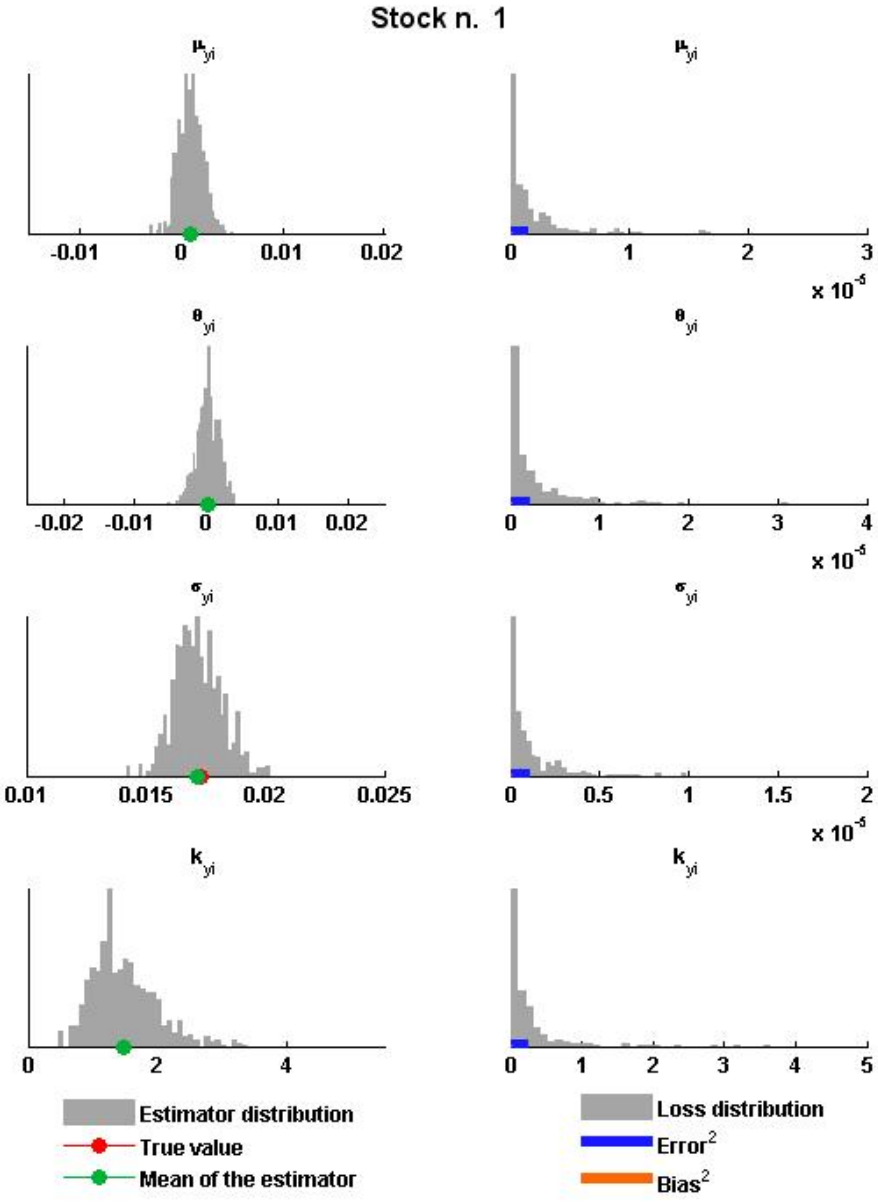

Figure 2.9: Distributions of the estimators and of the square losses for the parameters of the first idiosyncratic component $Y^{(1)}$. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

Z	*n=5* (1000 sim.)	*n=15* (100 sim.)
μ **= 0.0014**		
RMSE	7.07E-04	5.67E-04
Bias	5.82E-05	4.18E-05
Inefficiency	7.04E-04	5.65E-04
θ **= -0.0014**		
RMSE	1.11E-03	8.64E-04
Bias	9.64E-05	5.63E-05
Inefficiency	1.10E-03	8.62E-04
σ **= 0.0168**		
RMSE	1.24E-03	1.25E-03
Bias	3.60E-05	7.30E-05
Inefficiency	1.24E-03	1.25E-03
k **= 3.32**		
RMSE	1.15E+00	8.35E-01
Bias	8.90E-02	1.29E-01
Inefficiency	1.14E+00	8.25E-01

Table 2.9: Estimation assessment results for the parameters of the common factor Z for $T = 500$.

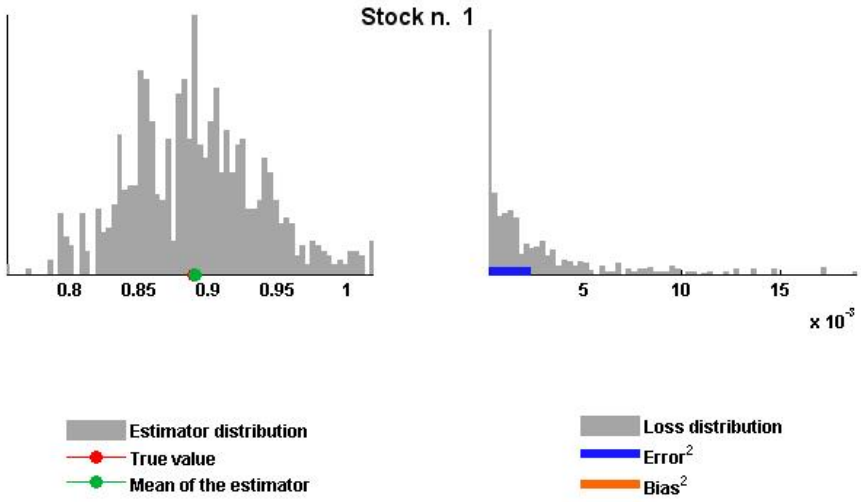

Figure 2.10: Distributions of the estimator and of the square loss for first loading a_1. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

$Y^{(1)}$	*n=5* (1000 sim.)	*n=15* (100 sim.)
μ **= 9.92E-04**		
RMSE	1.19E-03	1.11E-03
Bias	2.51E-05	1.58E-04
Inefficiency	1.19E-03	1.10E-03
θ **= 2.15E-04**		
RMSE	1.47E-03	1.52E-03
Bias	1.61E-05	1.43E-04
Inefficiency	1.47E-03	1.52E-03
σ **= 0.0173**		
RMSE	9.95E-04	1.12E-03
Bias	1.04E-04	1.97E-04
Inefficiency	9.90E-04	1.10E-03
k **= 1.483**		
RMSE	4.87E-01	4.59E-01
Bias	1.22E-02	4.56E-02
Inefficiency	4.87E-01	4.57E-01
a_1**= 0.8898**		
RMSE	4.72E-02	3.85E-02
Bias	2.36E-04	1.64E-03
Inefficiency	4.72E-02	3.85E-02

Table 2.10: Estimation assessment results for the parameters of the first idiosyncratic component $Y^{(1)}$ and the first loading a_1 for $T = 500$.

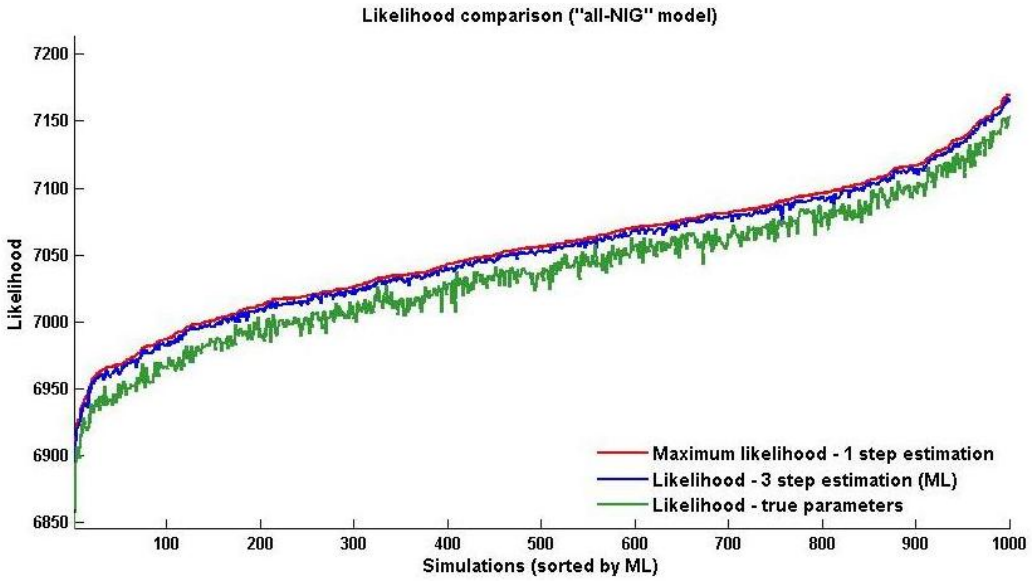

Figure 2.11: Likelihood comparison ('all-NIG' model)

As a further comparison between the three-step and one-step procedures, we simulate 1000 samples, each made of 500 observations from an 'all-NIG' model with 5 components, and we estimate the parameters with both methods. For each simulated sample we then compare the maximum likelihood achieved using the one-step approach with the likelihood based on the parameters estimated by the three step approach (where univariate estimations are performed by ML). Figure 2.11 displays the results; we sort the simulations by increasing values of the maximum likelihood for more clarity and we also plot the likelihood obtained using the true parameters, i.e. those adopted in the simulation. We can notice how, in every simulation, the estimates obtained by means of the three-step procedure lead to a likelihood which is very close to the maximum one. Therefore, if ML is a viable method for the univariate estimations, the three-step procedure is an effective, simple and fast alternative to the overall ML estimation.

2.3.2.2 Simulation study: the 'all-MJD' model

The maximum likelihood estimation under the 'all-MJD' model consists in maximizing the likelihood function (2.18), where all the probability density function involved are MJD (1.29), with respect to the $6 \times n + 5$ parameters of the model.

In this section we present the results of a simulation study aimed at assessing the effectiveness of the one-step ML approach in the estimation of the 'all-MJD' model parameters. Due to the computational cost of the procedure, here we evaluate the estimation for a small number of assets ($n = 5$, i.e. 35 parameters to be estimated) repeating the simulation 1000 times; we then perform 100 simulations to evaluate the estimation for $n = 15$ assets (i.e. 95 parameters), leaving to future studies a deeper assessment of the overall-ML estimation method for the all-MJD model.
Results relative to the common factor Z are reported in Table 2.11; those relative to the first idiosyncratic component $Y^{(1)}$, including the loading a_1, are displayed in Table 2.12. Complete results are available upon request.
The estimators and loss distributions for $n = 5$ are plotted in Figures 2.12, 2.13 and 2.14.

Keeping in mind the different number of simulations performed[12], we can compare the errors relative to the one-step ML approach with the corresponding errors obtained by means of the three-step procedure, where we used the EM algorithm to obtain ML estimators. In particular, the results reported in Table 2.11 can be compared to those displayed in the second column of Table 2.6, while the results in Table 2.12 can be compared with those in the first and third columns of Table 2.8; the errors relative to the first loading estimates can be compared to those reported in Table 2.1, where $T = 500$, $n = 5, 15$. We observe that for the 'all-MJD' model the errors of the three-step procedure are just a bit higher than those obtained with the one-step ML approach.

[12] A higher number of simulations leads in general to higher errors, due to higher inefficiency, since the variability of the estimates tends to increase.

Z	*n=5* (1000 sim.)	*n=15* (100 sim.)
μ **= 0.0012**		
RMSE	6.24E-04	6.23E-04
Bias	9.65E-05	1.06E-04
Inefficiency	6.16E-04	6.13E-04
σ^2**= 5.75E-05**		
RMSE	1.60E-05	9.63E-06
Bias	3.72E-06	3.07E-06
Inefficiency	1.56E-05	9.13E-06
ν **= -0.0025**		
RMSE	1.69E-03	1.85E-03
Bias	8.17E-05	2.73E-04
Inefficiency	1.69E-03	1.83E-03
τ^2 **= 0.0004**		
RMSE	1.11E-04	1.43E-04
Bias	2.55E-05	5.90E-05
Inefficiency	1.08E-04	1.31E-04
λ**=0.47**		
RMSE	2.33E-01	1.49E-01
Bias	1.63E-01	8.84E-02
Inefficiency	1.66E-01	1.20E-01

Table 2.11: Estimation assessment results for the parameters of the common factor Z for $T = 500$.

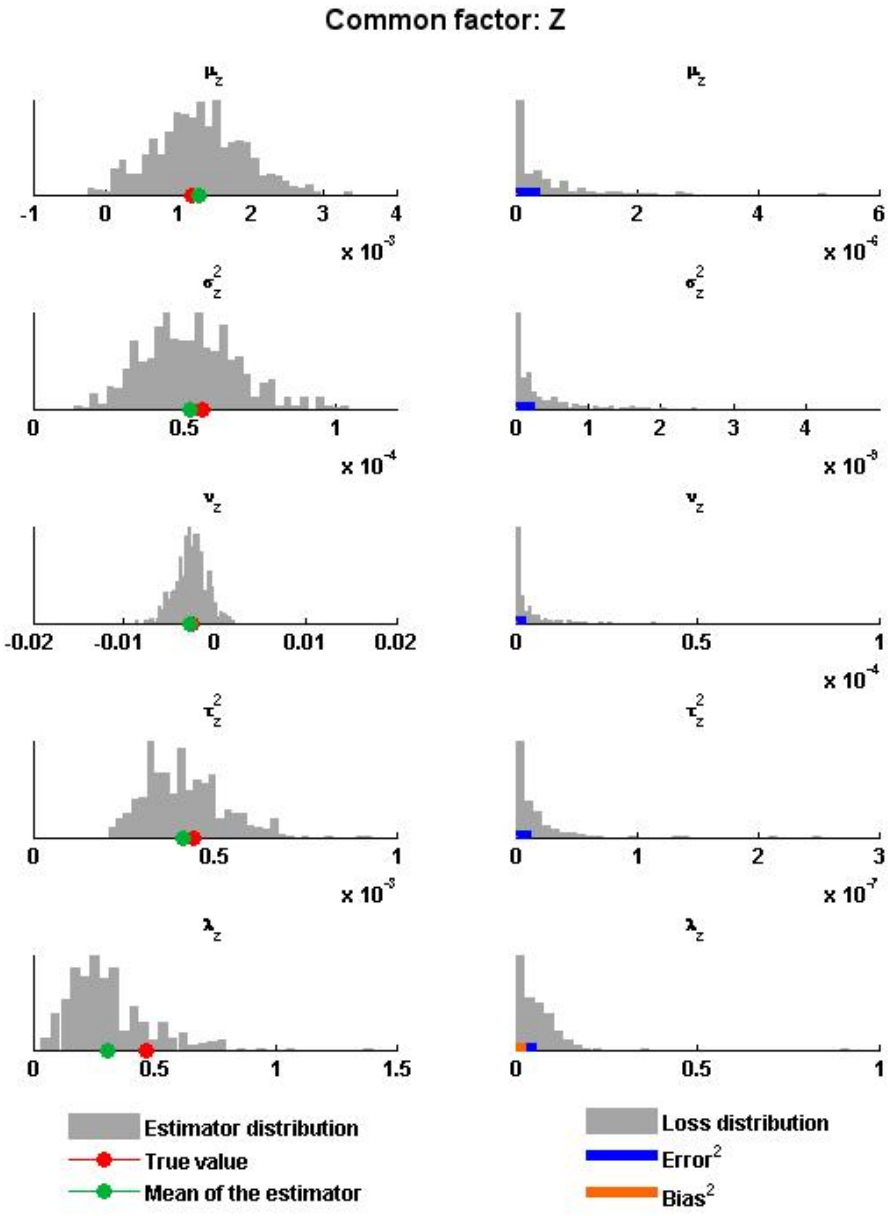

Figure 2.12: Distributions of the estimators and of the square losses for the parameters of the common factor Z. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

$Y^{(1)}$	*n=5* (1000 sim.)	*n=15* (100 sim.)
μ **= 0.00133**		
RMSE	7.85E-04	7.68E-04
Bias	4.54E-05	6.12E-05
Inefficiency	7.84E-04	7.65E-04
σ^2**= 0.00012**		
RMSE	2.41E-05	2.21E-05
Bias	1.02E-06	2.12E-06
Inefficiency	2.41E-05	2.20E-05
ν **= -0.0004**		
RMSE	3.03E-03	3.19E-03
Bias	9.54E-05	7.16E-05
Inefficiency	3.03E-03	3.19E-03
τ^2 **=0.00059**		
RMSE	2.41E-04	2.33E-04
Bias	3.63E-05	6.06E-05
Inefficiency	2.39E-04	2.25E-04
λ **=0.29214**		
RMSE	1.36E-01	1.15E-01
Bias	2.15E-02	1.20E-02
Inefficiency	1.35E-01	1.14E-01
a_1**= 0.8921**		
RMSE	5.44E-02	3.98E-02
Bias	1.08E-02	4.52E-03
Inefficiency	5.33E-02	3.95E-02

Table 2.12: Estimation assessment results for the parameters of the first idiosyncratic component $Y^{(1)}$ and the first loading a_1 for $T = 500$.

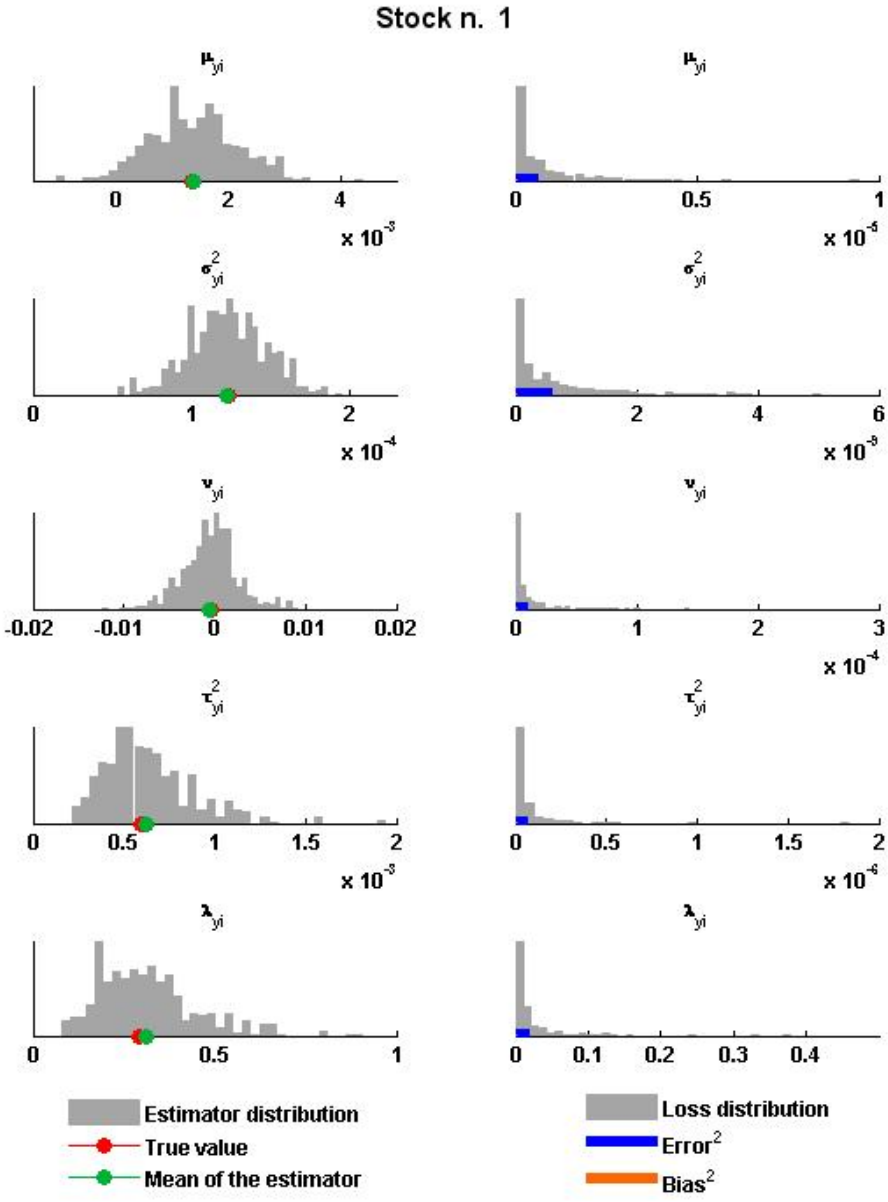

Figure 2.13: Distributions of the estimators and of the square losses for the parameters of the first idiosyncratic component $Y^{(1)}$. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

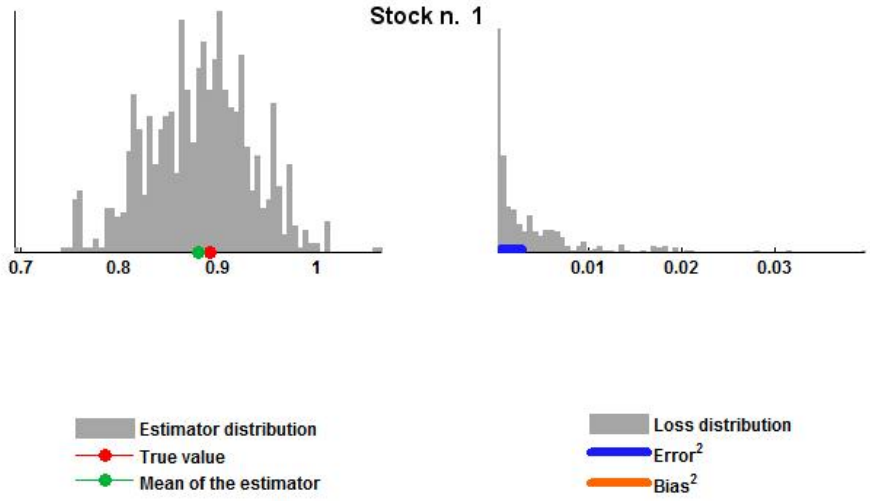

Figure 2.14: Distributions of the estimator and of the square loss for first loading a_1. Number of simulations: 1000; length of the simulated time series: 500; number of assets: 5.

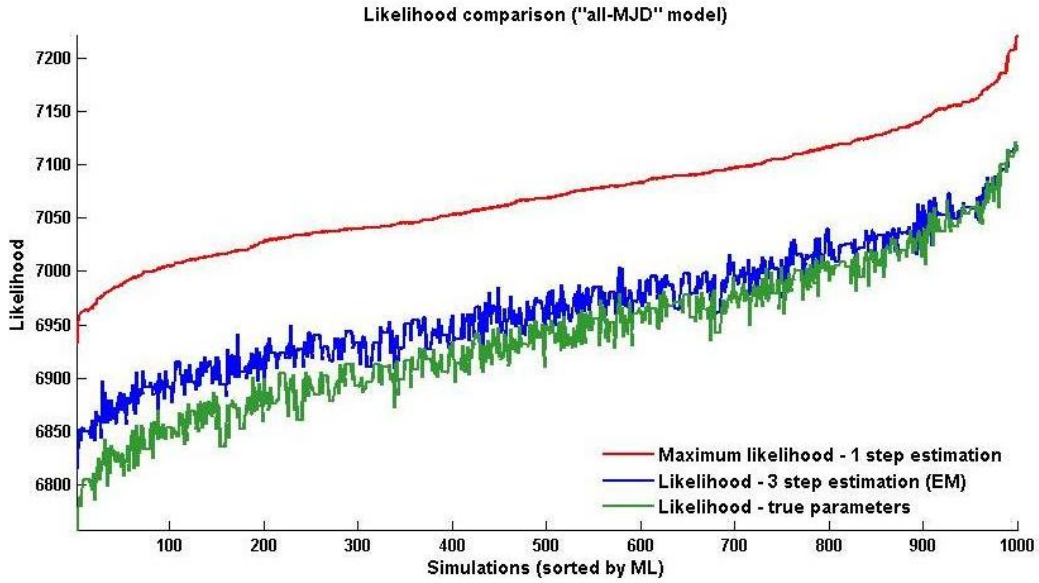

Figure 2.15: Likelihood comparison ('all-MJD' model)

As we did for the 'all-NIG' model, we then simulate 1000 samples, each made of 500 observations from an 'all-MJD' model with 5 components, and we compare the likelihoods achieved using the parameters estimated by the three-step procedure with the maximum likelihoods. Figure 2.15, displays the results; we sort the simulations by increasing values of the maximum likelihood for more clarity and we also plot the likelihood obtained using the true parameters, i.e. those adopted in the simulation. We can notice that, with respect to the 'all-NIG' model, where the univariate estimations in the three-step approach were performed by ML, here the likelihoods resulting from the three-step routine, where the univariate estimations are performed by EM algorithm, are less close to the maximum ones, but still high, being bigger than the likelihoods obtained with the true parameters. Jointly with the observations made above with regards to the 'all-NIG' model, this suggests that the loss of efficiency in the three-step approach for the 'all-MJD' model is mostly due to the lower efficiency of the univariate estimations rather than to the splitting process.

2.3.3 Estimation on real data

As a further experiment we estimate the multivariate Lévy model (2.1) (in both the 'all-NIG' and 'all-MJD' specifications) on a real dataset of stock returns (see Section 3.5) to investigate if the estimated model well describes the actual asset return distribution. The estimation is performed via the three-step procedure presented in Section 2.3.1.
For purely illustrative purpose, in Figure 2.16 we plot the histogram of the log-returns of 'Apple Inc.', the first of the twenty stocks in our database, with superimposed the estimated marginal probability density function resulting from our models.
Moreover, in Table 2.13 we compare the first four moments with their theoretical counterparts, obtained through relations (2.3) and (1.10).

Finally, we compare sample and theoretical (2.13) covariance matrices, to check if we manage to catch the correlation structure. Since reporting and comparing big matrices may be cumbersome, in Figure 2.17 we show the two color-coded matrices, where the estimated one is under the 'all-NIG' specification of the multivariate model[13]. Each entry of the two matrices is colored according to its value; the conversion color-value is provided in the lateral color bars (which are the same to make the comparison immediate). We can notice that the estimated matrix colors resemble those of the sample covariance, meaning that the multivariate Lévy model accurately reproduces the covariance among the assets. It is worth stressing that, even if in our dataset all of the correlations happen to be positive, the simulation study presented in Section 2.3.1.1, and in particular the results in Figure 2.2, involves the presence of negative loadings.

	mean	**variance**	**skewness**	**kurtosis**
sample	0.00084	0.00055	-0.497	9.53
all-NIG model	0.00084	0.00054	-0.262	6.54
all-MJD model	0.00084	0.00052	-0.152	4.88

Table 2.13: First four moments of Apple Inc. returns distribution: sample and multivariate Lévy estimates.

[13] Under the 'all-MJD' model only the diagonal entries change, due to the different estimation of the idiosyncratic components variance.

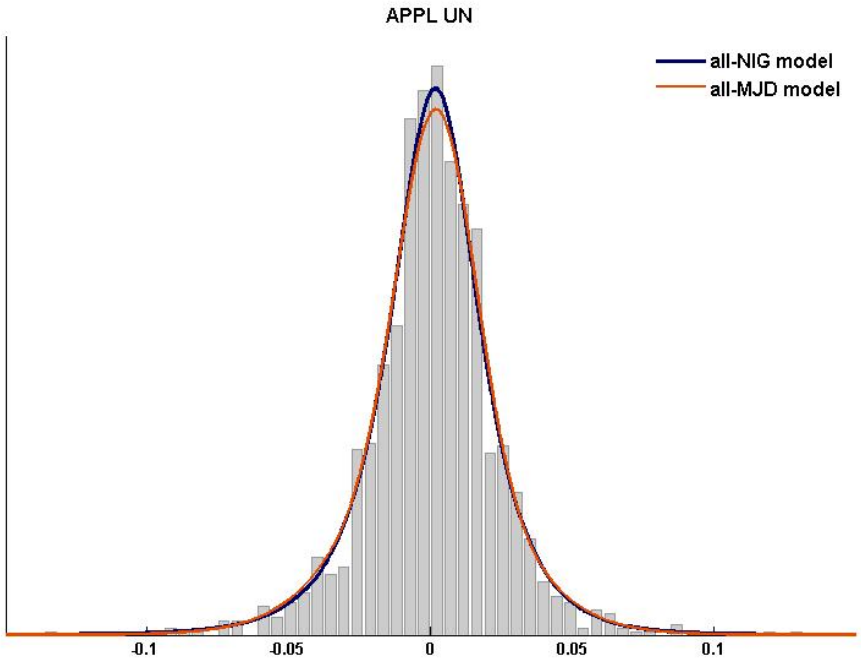

Figure 2.16: Apple Inc. returns distribution. The estimated marginal probability density function under two specifications of the multivariate Lévy model are superimposed to the histogram.

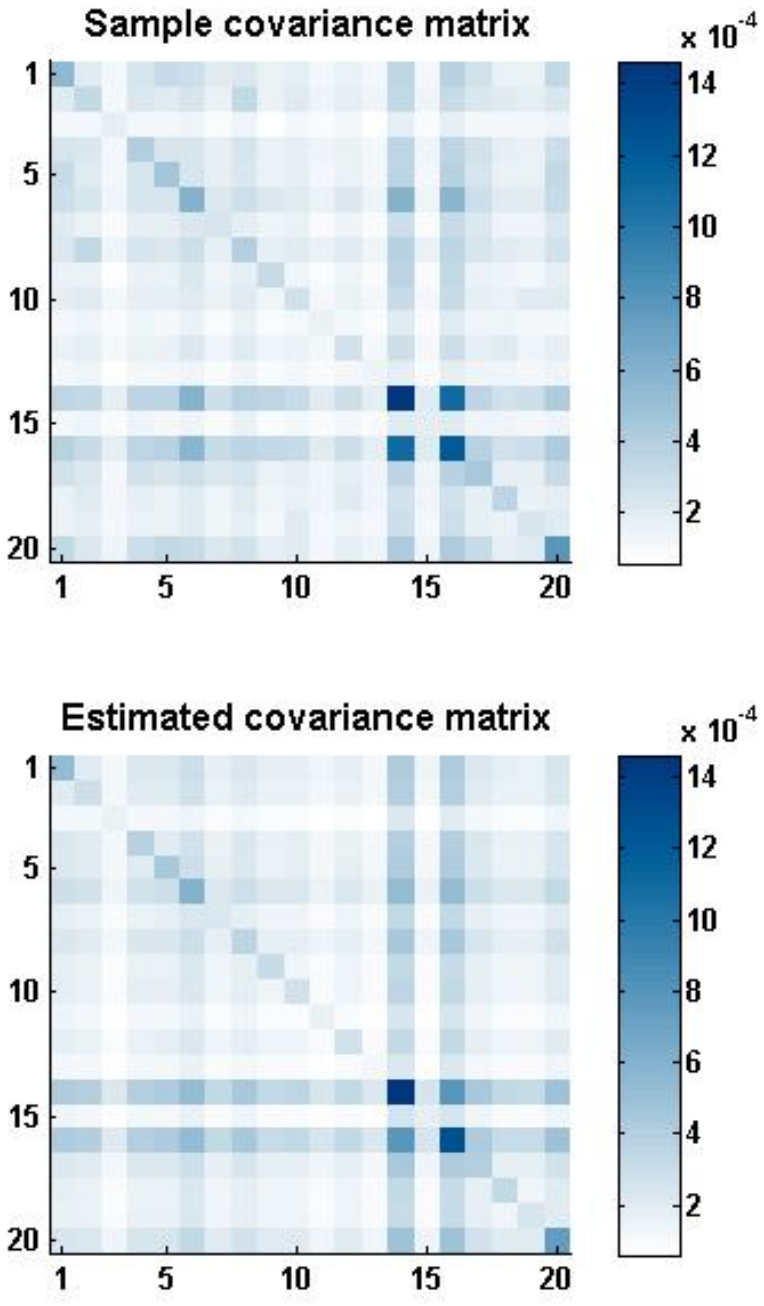

Figure 2.17: Sample and estimated covariance matrices ('all NIG' model). The conversion color-values is provided in the lateral color bars.

Chapter 3

Application to portfolio allocation

In this chapter we apply the model presented in Chapter 2, in the particular 'all NIG' and 'all MJD' specifications we adopted in the simulation studies (Sections 2.3.1.2 and 2.3.1.3), for portfolio selection purposes. We focus on the stock market, considering stocks belonging to the S&P500 index. In Section 3.1 we define our assumptions on the investor's problem. In Section 3.2 we show how to perform portfolio selection according to the multivariate Lévy model (2.1). In the subsequent part we describe in details data, methodology and results of the application we carry out, illustrating the benchmark strategies and the measures of performance we take into account. In particular, we include among our benchmarks a strategy that we call 'non-parametric portfolio optimization' (see Section 3.3.4), which to our knowledge has never been employed, although being straightforward and exhibiting quite good out of sample performances.

3.1 The investor's problem

In our empirical analysis we consider a standard expected utility maximization framework, i.e. the investor goal is to maximize the expected utility of terminal wealth subject to a set of investment constraints.

Given n risky assets, we indicate by $\mathbf{w}_t$ the vector of weights at time t, meaning that $w_{i,t}$ is the proportion of wealth invested on the i-th asset at time t; we consider only positive weights, not allowing for short selling.

We indicate by $r_{i,t}$ the log-return of asset i at time t and by $r_{p,t}$ the portfolio return at time

t.[1]

We will work on daily or weekly time steps, for which the distributions of linear and compounded returns are similar[2], therefore we can approximate the portfolio return $r_{p,t}$ as

$$r_{p,t} \approx \sum_{i=1}^{n} w_{i,t} r_{i,t} = \mathbf{w_t}' \mathbf{r_t}. \tag{3.1}$$

Moreover we focus on a short investment horizon, daily or weekly, interpreting the point of view of a professional investor who frequently rebalances her positions.

The investor's problem reads

$$\begin{cases} w^* = & \underset{w}{\text{argmax}}\ E[U(W_T)] \\ \text{s.t.} & \sum_{i=1}^{n} w_i = 1 \\ & w_i \geq 0\ \forall i = 1 \dots n, \end{cases} \tag{3.2}$$

where we assume that the wealth W is entirely determined by the portfolio outcome r_p. We will generally refer to the above constraints as IC (investment constraints), using the compact notation $\mathbf{w} \in IC$.

In the application we deal with two different objective functions, the exponential (or CARA, acronym for constant absolute risk aversion) utility function

$$U(W) = -\mathbf{exp}(-\lambda W) \tag{3.3}$$

and the power (or CRRA, acronym for constant relative risk aversion) utility function

$$U(W) = \begin{cases} \frac{W^{(1-\lambda)}}{1-\lambda} & \text{if } \lambda \neq 1 \\ \log(W) & \text{if } \lambda = 1, \end{cases} \tag{3.4}$$

where in both cases $\lambda > 0$ represents the investor's risk aversion coefficient.
Since the exponential utility function is increasing and concave on all of the real line, when dealing with it we identify wealth with portfolio outcome, maximizing $E[U(R)]$.
For the CRRA utility function, which is increasing and concave only for positive level of wealth we consider instead $W_1 = W_0(1 + r_{p,1})$, taking the initial wealth W_0 as numéraire,

[1] When there is no ambiguity we may omit time subscripts to simplify notation.
[2] If returns are very volatile or the time step is longer, it becomes essential to work with linear returns, for which relation (3.1) holds exactly. For more details, see Meucci (2010).

i.e. setting $W_0 = 1$, and assuming that for daily or weekly portfolio returns $r_p = -1$ is a reasonable lower bound.

When the estimation step is daily and the horizon is weekly, we perform the optimization considering the portfolio returns distribution projected to the horizon. To do this we rely on approximation (3.1) and we project the characteristic function of portfolio returns distribution (3.6) (see Meucci (2005), Meucci (2011)).

3.2 Asset allocation with the multivariate Lévy model

In this section we illustrate how we perform portfolio selection based on the multivariate Lévy model presented in Chapter 2.

Relying on approximation (3.1) we can write the random variable representing portfolio return as

$$R_p \approx \sum_{j=1}^{n} w_j X^{(j)} = \sum_{j=1}^{n} w_j Y^{(j)} + Z \sum_{j=1}^{n} w_j a_j. \tag{3.5}$$

The characteristic function of the portfolio returns distribution can then be easily obtained, exploiting the independence of the $Y^{(j)}$, $j = 1, \ldots, n$, among themselves and of Z, as

$$\begin{aligned} E\left[\exp\left(i\gamma R_p\right)\right] &= E\left[\exp\left(i\gamma\left(\sum_{j=1}^{n} w_j Y^{(j)} + Z\sum_{j=1}^{n} w_j a_j\right)\right)\right] \\ &= E\left[\exp\left(i\gamma\left(\sum_{j=1}^{n} w_j Y^{(j)}\right)\right)\right] E\left[\exp\left(i\gamma Z\sum_{j=1}^{n} w_j a_j\right)\right] \\ &= \left(\prod_{j=1}^{n} E\left[\exp\left(i\gamma w_j Y^{(j)}\right)\right]\right) E\left[\exp\left(i\gamma Z\sum_{j=1}^{n} w_j a_j\right)\right] \\ &= \left(\prod_{j=1}^{n} \phi_{Y^{(j)}}\left(\gamma w_j\right)\right) \phi_Z\left(\gamma\sum_{j=1}^{n} w_j a_j\right) \quad \forall\gamma \in \mathbb{R}. \end{aligned} \tag{3.6}$$

In particular, in the 'all NIG' specification of the model, ϕ_Z and $\phi_{Y^{(j)}}$, $j = 1, \ldots, n$, are NIG characteristic functions (1.17), while in the 'all MJD' specification they are all MJD characteristic functions (1.35).

For a generic utility function $U(W)$, to obtain the best allocation we maximize with respect to the vector of weights the expected utility of the associated final wealth (see

(3.2)), which explicitly reads

$$E(U(W)) = \int_D U(W) f_W(W) dW, \tag{3.7}$$

where the wealth W depends on the portfolio return r_p as specified in Section 3.1 (as $W = r_p$ or $W = 1 + r_p$); hence we can write

$$E(U(W)) = \int_{\tilde{D}} U(W) f_{R_p}(r) dr, \tag{3.8}$$

where the integration sets in (3.7) and (3.8) are the support of the random variables W and R_p respectively.

To recover the probability density function f_{R_p} we invert the characteristic function (3.6) using the discrete Fourier transform, computed with a fast Fourier transform algorithm by the MATLAB® function `fft`.

To evaluate (3.8) we adopt the trapezoidal numerical integration given by the function `trapz`.

In the particular case of an exponential utility function (3.3), the computations are simplified and exact instead, since from (3.6) we can derive an explicit expression for the expected utility. In fact

$$E(U(W)) = E[-\exp(-\lambda R_p)] = -E[\exp(-\lambda R_p)], \tag{3.9}$$

which can be computed from (3.6) replacing $i\gamma$ by $-\lambda$, i.e. the expected utility is obtained from the expression of the characteristic function of the portfolio distribution imposing $\gamma = i\lambda$.

3.3 Benchmarks

3.3.1 Equally weighted portfolio [EQ]

We take as first benchmark the equally-weighted portfolio, in which a fraction $1/n$ of wealth is allocated to each of the n assets available for investment. This naive rule is easy to implement not relying either on estimation of the asset returns distribution or on optimization. Despite the sophisticated theoretical models developed in the last 60 years and the advances in methods for estimating the parameters of these models, investors con-

tinue to use such simple allocation rules for allocating their wealth across assets (see Benartzi and Thaler (2001) and Huberman and Jiang (2006)). Moreover, some studies have shown that equally-weighted portfolios, which are not exposed to estimation errors, outperformed portfolios constructed using optimization techniques (see for example DeMiguel et al. (2009) and Duchin and Levy (2009)).

Independently of market data and investor's preferences, the 'EQ' strategy reads:

$$w_i^{EQ} = \frac{1}{n}, \quad \forall\, i = 1, \dots, n. \tag{3.10}$$

3.3.2 Mean-variance portfolio [MV]

As a second benchmark we use the well known mean-variance portfolio, based on the modern portfolio theory developed by Markowitz (1952). The MV portfolio is obtained by a two-step approach: in the first step the so called mean-variance efficient frontier, i.e. the set of portfolios offering the highest expected return for a given level of risk (measured in terms of variance), is computed; the second step finds, among the portfolios belonging to the efficient frontier, the one giving the maximum level of satisfaction. Expected returns and variances are estimated through their sample counterparts; given the approximation (3.1), we compute the mean and the variance of portfolio returns distribution as

$$\mu_p \approx \mathbf{w}'\mu \tag{3.11}$$

$$\sigma_p^2 \approx \mathbf{w}'\Sigma\mathbf{w}, \tag{3.12}$$

where μ and Σ are respectively the vector of the means and the covariance matrix of the assets returns.

Therefore we can write the first step as

$$\mathbf{w}(v) = \underset{\substack{\mathbf{w}\in IC \\ \sigma_p^2 = v}}{\operatorname{argmax}}\ \mu_p. \tag{3.13}$$

We evaluate the investor's satisfaction related to a given portfolio $\mathbf{w}$ by means of the certainty equivalent ($CE(\mathbf{w})$).

Hence the second step reads:

$$\mathbf{w}^{MV} = \underset{v\geq 0}{\operatorname{argmax}}\ CE(\mathbf{w}(v)). \tag{3.14}$$

In particular, we focus on Arrow-Pratt approximation of the certainty equivalent:

$$CE(\mathbf{w}) \approx E(W) + \frac{1}{2}\frac{u''(E(W))}{u'(E(W))}\text{Var}(W), \tag{3.15}$$

where W is the final wealth as we specified in Section 3.1.

Espression (3.15) becomes

$$CE(\mathbf{w}) \approx \mu_p - \frac{\lambda}{2}\sigma_p^2, \tag{3.16}$$

for the CARA utility function (3.3)[3], and

$$CE(\mathbf{w}) \approx (1+\mu_p) - \frac{\lambda}{2}\frac{\sigma_p^2}{(1+\mu_p)}, \tag{3.17}$$

for the CRRA utility function (3.4).

3.3.3 Four-moments based allocation (Single Factor approach) [SF]

As a third benchmark we consider the allocation resulting from a fourth-order Taylor expansion of the utility function, using the single-factor estimators for higher-order moments proposed by Martellini and Ziemann (2010). We summarize the main ideas and we report the formulas needed for the implementation; refer to the original paper for a deeper discussion.

Using a fourth-order Taylor expansion of the utility function, the expected utility can be approximated as

$$E[U(W)] \approx U(E(W)) + \frac{U^{(2)}(E(W))}{2}\mu^{(2)} + \frac{U^{(3)}(E(W))}{6}\mu^{(3)} + \frac{U^{(4)}(E(W))}{24}\mu^{(4)} \tag{3.18}$$

being $\mu^{(n)}$ the n-th central moment:

$$\mu^{(n)} = E[(W - E(W))^n]. \tag{3.19}$$

We will here identify final wealth with portfolio return; thanks to the translation invariance property of central moments, if $W = 1 + r_p$, the only moment affected is the first, becoming $E(W) = 1 + E(R_p)$. Therefore the approximated expected utility (3.18) depends on the derivatives of the utility function and on the first four moments of the portfolio

[3] In this case, under the assumption of Gaussian returns, the direct optimization of expression (3.16) leads to an exact one-step optimal allocation, as proved in Appendix C.

returns distribution, which in turn are functions of portfolio weights and of the first four moments and comoments of assets returns distributions. More precisely, portfolio moments are given by

$$\mu^{(1)} = \mu_p \tag{3.20}$$

$$\mu^{(2)} = \sigma_p^2 = \mathbf{w}' M_2 \mathbf{w}; \tag{3.21}$$

$$\mu^{(3)} = s_p = \mathbf{w}' M_3 (\mathbf{w} \otimes \mathbf{w}); \tag{3.22}$$

$$\mu^{(4)} = k_p = \mathbf{w}' M_4 (\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}); \tag{3.23}$$

where $\otimes$ denotes the Kronecker product and M_2, M_3, M_4 are the so called higher order moment tensors (Jondeau and Rockinger (2003)) representing covariance, coskewness and cokurtosis respectively[4]. To estimate the higher moment tensors we choose the single-factor approach, which goes back to the single-factor linear model of Sharpe (1963) for asset returns

$$r_{i,t} = c + \beta_i F_t + \epsilon_{i,t}, \tag{3.24}$$

where ϵ_i is the residual term relative to asset i and the factor F is taken to be a broad-based index (the S&P500 in our case). The residuals are assumed as homoscedastic and cross-sectionally uncorrelated:

$$\epsilon \sim (0, \Psi), \tag{3.25}$$

where Ψ is a diagonal matrix containing the residual variances ('idiosyncratic' risks). Then the covariance matrix M_2 can be written as

$$M_2 = \beta \beta' \mu_0^{(2)} + \Psi, \tag{3.26}$$

where β is the column vector containing the regression coefficients and $\mu_0^{(2)}$ is the variance of the single-factor. Martellini and Ziemann (2010) derive analogue decompositions for the coskewness and cokurtosis:

$$M_3 = (\beta \beta' \otimes \beta') \mu_0^{(3)} + \Phi \tag{3.27}$$

$$M_4 = (\beta \beta' \otimes \beta' \otimes \beta') \mu_0^{(4)} + \Upsilon, \tag{3.28}$$

[4]We follow the notation in Martellini and Ziemann (2010) for the tensors M_3 and M_4, stacking the subcomoment matrices column-wise.

where $\mu_0^{(3)}$ and $\mu_0^{(4)}$ are respectively the skewness and kurtosis of the single factor. The values in Ψ, Φ and Υ are explicitly assessed under the assumption that all the cross-sectional residuals ϵ_i and ϵ_j $(i \neq j)$ are independent, and exploiting the independence of the factor return process from the residual return process which comes from the least-squares regression technique. The structure of the $n \times n$ matrix Ψ is

$$\psi_{ii} = E(\epsilon_i^2), \quad \text{with the sample estimate } \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_{i,t}^2, \tag{3.29}$$

$$\psi_{ij} = 0 \; \forall i \neq j. \tag{3.30}$$

The structure of the $n \times n^2$ matrix Φ is

$$\phi_{iii} = E(\epsilon_i^3), \quad \text{with the sample estimate } \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_{i,t}^3, \tag{3.31}$$

$$\psi_{iij} = 0 \tag{3.32}$$

$$\psi_{ijk} = 0 \; \forall i \neq j \neq k. \tag{3.33}$$

The $n \times n^3$ matrix Υ entries are

$$\upsilon_{iiii} = E(\epsilon_i^4), \quad \text{with the sample estimate } \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_{i,t}^4, \tag{3.34}$$

$$\upsilon_{iiij} = 3\beta_i\beta_j\mu_0^{(2)}\psi_{ii} \tag{3.35}$$

$$\upsilon_{iijj} = \beta_i^2\mu_0^{(2)}\psi_{jj} + \beta_j^2\mu_0^{(2)}\psi_{ii} + \psi_{ii}\psi_{jj} \tag{3.36}$$

$$\upsilon_{iijk} = \beta_j\beta_k\mu_0^{(2)}\psi_{ii} \tag{3.37}$$

$$\upsilon_{ijkl} = 0 \; \forall i \neq j \neq k \neq l. \tag{3.38}$$

Summarizing, to apply the 'SF' strategy we perform the linear regressions (3.24) and we estimate the moments and comoments of the assets returns as in (3.30), (3.33) and (3.38). For a given allocation $\mathbf{w}$ we can then compute the first four moments of portfolio returns distribution as in (3.23) and obtain the approximated expected utility value through (3.18). The optimal allocation is the one resulting from the constrained maximization of (3.18). More precisely, writing explicitly the derivatives, and according to our wealth definitions

(see Section 3.1), expression (3.18) becomes

$$E[U(W)] \approx -e^{-\lambda\mu_p}\left(1+\frac{1}{2}\lambda^2\sigma_p^2-\frac{1}{6}\lambda^3 s_p+\frac{1}{24}k_p\right) \tag{3.39}$$

for the CARA utility function (3.3), and

$$U(E(W)) = \begin{cases} \frac{\bar{W}^{(1-\lambda)}}{(1-\lambda)} - \frac{\lambda\bar{W}^{-(\lambda+1)}\sigma_p^2}{2} + \frac{\lambda(\lambda+1)\bar{W}^{-(\lambda+2)}s_p}{6} - \frac{\lambda(\lambda+1)(\lambda+2)\bar{W}^{-(\lambda+3)}k_p}{24} & \text{if } \lambda \neq 1 \\ \log(\bar{W}) - \frac{1}{2}\frac{\sigma_p^2}{\bar{W}^2} + \frac{1}{3}\frac{s_p}{\bar{W}^3} - \frac{1}{4}\frac{k_p}{\bar{W}^4} & \text{if } \lambda = 1 \end{cases} \tag{3.40}$$

for the CRRA utility function (3.4), where $\bar{W} = 1 + \mu_p$.

3.3.4 Non-parametric optimization [NP]

As a further benchmark we introduce a strategy based on the Gaussian kernel estimation of the portfolio returns distribution. Given a vector of weights $\mathbf{w}$ and the time series of assets prices, collected in a $T \times n$ matrix $\mathbf{S}$, we can compute the portfolio log-returns as

$$r_{p,t} = \log\frac{V_t}{V_{t-1}}, \tag{3.41}$$

where V_t is the value of the portfolio at time t, obtained as $V_t = \mathbf{S}_t\mathbf{w}$.

Based on the sample $[r_{p,1}, \ldots, r_{p,T}]$ of portfolio returns obtained through (3.41), we can perform a kernel estimation of the returns density as

$$\hat{f}_{R_p}(r) = \frac{1}{Th}\sum_{t=1}^{T}\mathrm{K}\left(\frac{r - r_{p,t}}{h}\right), \tag{3.42}$$

where h is the kernel bandwidth and we take the kernel density K to be standard normal. Then, for a generic utility function we can compute the expected utility numerically evaluating the integral (3.8), where f_{R_p} is replaced by $\hat{f}_{R_p}$. This is indeed the way we implement the strategy for the CRRA utility function (3.4), for $W = 1 + r_p$, performing the kernel estimation by means of the MATLAB® function `ksdensity` and evaluating the integral through the function `trapz`.

In the particular case of exponential utility function (3.3) we don't need the numerical approximation of the integral (3.8) instead, since we can express the expected utility in a closed form. Consider the moment generating function of a random variable R_p distributed according to the density $\hat{f}_{R_p}$ in equation (3.42); as we show below, it can be writ-

ten as a (rescaled) sum of T normal moment generating functions with means $r_{p,t}$, for $t = 1 \dots T$, and variance h.

$$\begin{aligned} E(e^{\alpha R_p}) &= \int_D e^{\alpha r} \hat{f}_{R_p}(r) dr = \int_D e^{\alpha r} \frac{1}{Th} \sum_{t=1}^{T} \mathbf{K}\left(\frac{r - r_{p,t}}{h}\right) dr \\ &= \frac{1}{Th} \sum_{t=1}^{T} \int_D e^{\alpha r} \mathbf{K}\left(\frac{r - r_{p,t}}{h}\right) dr = \frac{1}{Th} \sum_{t=1}^{T} E(e^{\alpha \tilde{R}}) \\ &= \frac{1}{Th} \sum_{t=1}^{T} e^{\alpha r_{p,t} + \frac{1}{2} h \alpha^2} = \frac{1}{Th} e^{\frac{1}{2} h \alpha^2} \sum_{t=1}^{T} e^{\alpha r_{p,t}}, \end{aligned} \tag{3.43}$$

where in the fourth passage we denoted with $\tilde{R}$ a random variable with distribution N($r_{p,t}, h$), and in the fifth one we made explicit the expression of its moment generating function. The expression of the expected utility, i.e. our objective function, then follows directly from (3.43) as:

$$E(-e^{-\lambda R_p}) = -E(e^{(-\lambda) R_p}) = -\frac{1}{Th} e^{\frac{1}{2} h \lambda^2} \sum_{t=1}^{T} e^{-\lambda r_{p,t}}. \tag{3.44}$$

3.4 Performance measures

We calculate several performance measures to evaluate the optimized portfolios.

Portfolio moments

First, we report mean, variance, skewness and kurtosis of the out-of-sample portfolio returns for each optimization strategy.

Monetary Utility Gain/Loss Measure (MUG)

A measure of performance to compare two strategies (let us call them 'strategy A' and 'strategy B') is the Monetary Utility Gain/Loss Measure (MUG), proposed by Ang and Bekaert (2002). The goal is to measure the economic gain/loss that results from holding portfolio A instead of portfolio B. We implicitly define the quantity W_A as follows:

$$E(U(W_T^{(strategyB)}) | W_0 = 1) = E(U(W_T^{(strategyA)}) | W_0 = W_A). \tag{3.45}$$

When dealing with an exponential utility function (3.3), the equation above reads

$$\sum_{t=1}^{T} -e^{-\lambda(1+r_t^{(strategyB)})} = \sum_{t=1}^{T} -e^{-\lambda W_A(1+r_t^{(strategyA)})}, \tag{3.46}$$

while when dealing with a CRRA utility function (3.4) with $\lambda \neq 1$, Equation (3.45) reads

$$\sum_{t=1}^{T} (1+r_t^{(strategyB)})^{1-\lambda} = \sum_{t=1}^{T} W_A(1+r_t^{(strategyA)})^{1-\lambda}, \tag{3.47}$$

where $r_t^{(strategyA/B)}$ are the annualized geometric out-of-sample portfolio returns obtained using strategy A/B. If $W_A > 1$ we prefer strategy B, since we have to invest more than one euro in strategy A to reach the same utility we obtain from investing one euro in strategy B; on the other hand, if $W_A < 1$ we prefer strategy A[5].

The MUG of strategy B with respect to strategy A is given by $MUG = 100(W_A - 1)$, and it represents the percentage annual excess amount that should be invested in portfolio A to reach the same terminal utility achieved by investing one euro in portfolio B. Therefore, if $MUG > 0$ we prefer strategy B, if $MUG < 0$ we prefer strategy A.

An equivalent interpretation of MUG says that it represents the percentage increase in the certain equivalent from moving from strategy A to strategy B.

Sharpe ratio

Further, we compute the out-of-sample Sharpe ratio as the mean of out-of-sample returns divided by their standard deviation.

$$SR = \frac{\mu_{r_p}}{\sigma r_p} \tag{3.48}$$

We use the two-sample statistic for comparing Sharpe ratios as proposed by Opdyke (2007) to test if the difference in Sharpe ratios of two portfolios is significant. This test can be applied under the very general conditions of stationary and ergodic returns; it permits auto-correlated and non-normal distributed returns and allows for a likely high correlation between the portfolio returns of different strategies.

Moreover we compute the multivariate statistic introduced by Leung and Wong (2006) to test the hypothesis that the portfolios obtained with the strategies we implemented have

[5] Given this immediate interpretation of the quantity W_A we report it, instead of MUG, in our result tables.

no significantly different Sharpe ratios, under the assumption of i.i.d. returns. Both the test statistics are reported in Appendix D.

Portfolio turnover

We compute the portfolio turnover, which quantifies the amount of trading required to implement a certain strategy, in line with Wolff et al. (2012). The portfolio turnover is the average sum of the absolute values of the trades across the n available assets:

$$TO = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{n}\left(|w_{j,t+1} - w_{j,t^+}|\right), \tag{3.49}$$

where $w_{j,t}$ is the portfolio weight in asset j at time t and w_{j,t^+} is the weight of asset j before rebalancing at $t+1$, i.e.

$$w_{j,t^+} = \frac{w_{j,t}S_{j,t+1}}{\sum_{j=1}^{n} w_{j,t}S_{j,t+1}}. \tag{3.50}$$

For example, in an equally weighted portfolio strategy $w_{j,t} = w_{j,t+1} = 1/n$, but w_{j,t^+} may be different due to changes in asset prices between t and $t+1$.

The turnover can be interpreted as the average percentage of wealth traded in each period.

Maximum drawdown

As a further risk measure, we compute the maximum drawdown, as proposed by Grossman and Zhou (1993), which reflects the maximum accumulated loss that an investor may suffer in the worst case during the whole investment period.

We compute the percentage maximum drawdown (MDD) as:

$$MDD = \max_{\tau\in(0,T)}\left[\max_{t\in(0,\tau)}\left(\frac{P_t - P_\tau}{P_t}\right)\right] \tag{3.51}$$

where P_t is the portfolio price at time t, when the portfolio is bought, and P_τ is the portfolio price at time τ, when the portfolio is sold.

3.5 Data

The dataset employed in our empirical analysis is composed by 20 assets chosen among the most capitalized stocks belonging to the S&P500 index. We collected the dividend-

adjusted prices from 10 September 2007 and 20 May 2013. In Table 3.1 we report mean, standard deviation, skewness and kurtosis of the daily log-returns of the index, which we use as a proxy for the common factor Z in our model (2.1), and of the 20 assets during the reference period. In Figures 3.1-3.1 we plot the time series of the stocks daily log-returns and the respective histograms, superimposing the estimated marginal probability density functions under the 'all-NIG', 'all-MJD'[6] and Gaussian model.

	mean	**st. dev**	**skewness**	**kurtosis**
S&P500	9.62E-05	0.0159	-0.2663	10.4532
AAPL UW	8.35E-04	0.0235	-0.4975	9.5293
XOM UN	1.55E-04	0.0182	0.1268	17.1272
WMT UN	5.13E-04	0.0133	0.1432	10.7217
MSFT UW	2.39E-04	0.0198	0.3312	11.1282
GOOG UW	3.97E-04	0.0213	0.3380	11.9304
GE UN	-2.00E-04	0.0244	0.0063	10.2653
IBM UN	4.79E-04	0.0156	0.0178	7.8491
CVX UN	3.85E-04	0.0198	0.1474	16.3137
BRK/B UN	2.51E-04	0.0176	0.7678	14.7178
T UN	1.94E-04	0.0162	0.6211	13.7416
PG UN	2.47E-04	0.0126	-0.2324	10.4033
PFE UN	3.13E-04	0.0165	-0.0824	8.2781
JNJ UN	3.73E-04	0.0112	0.6151	17.4169
WFC UN	1.90E-04	0.0381	0.7104	15.4729
KO UN	4.24E-04	0.0136	0.6199	14.1810
JPM UN	2.11E-04	0.0348	0.2856	11.8390
ORCL UW	4.08E-04	0.0210	-0.1182	7.5823
MRK UN	1.01E-04	0.0190	-0.5751	12.8927
VZ UN	4.47E-04	0.0158	0.3343	10.9179
AMZN UW	8.14E-04	0.0281	0.5348	10.3585

Table 3.1: Data sample moments.

[6] To estimate the parameters of the 'all-NIG' and 'all-MJD' models we adopt the three-step estimation procedure introduced in Section 2.3.1.

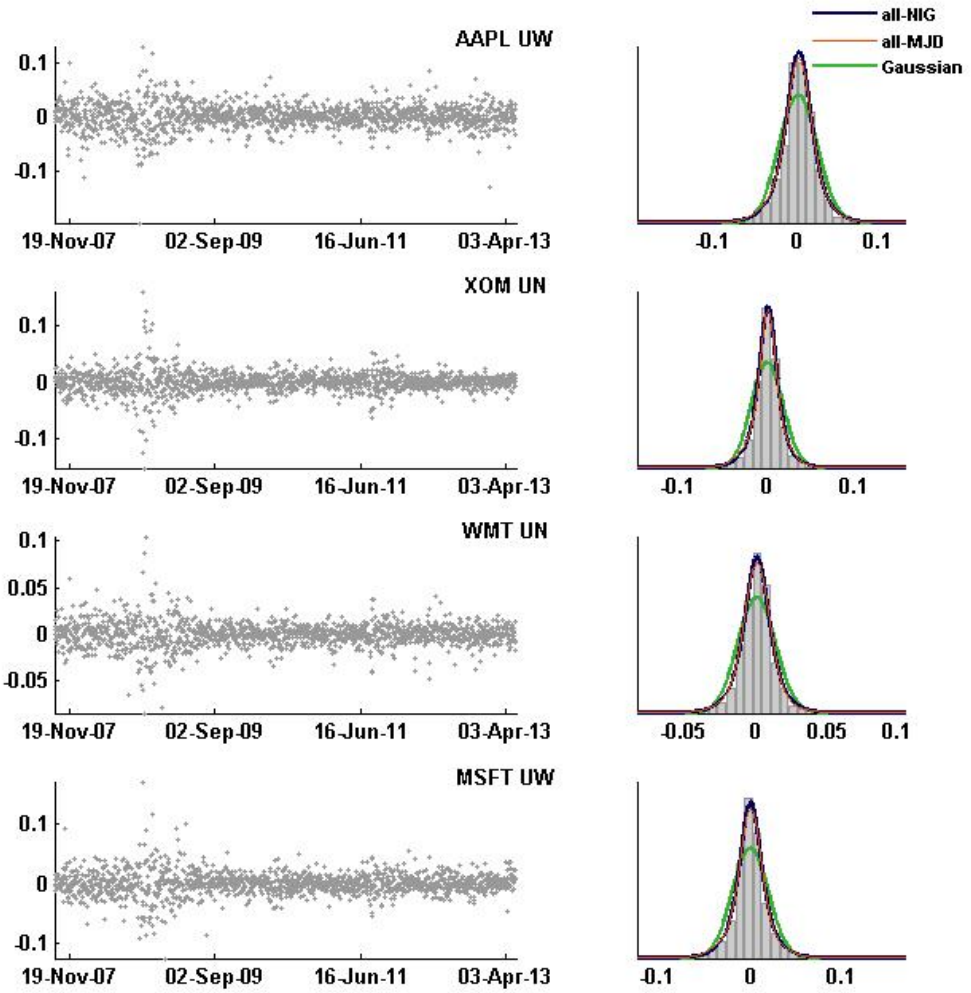

Figure 3.1: Time series and histograms: Apple, Exxon Mobil Corporation, Wal-Mart Stores, Microsoft Corporation.

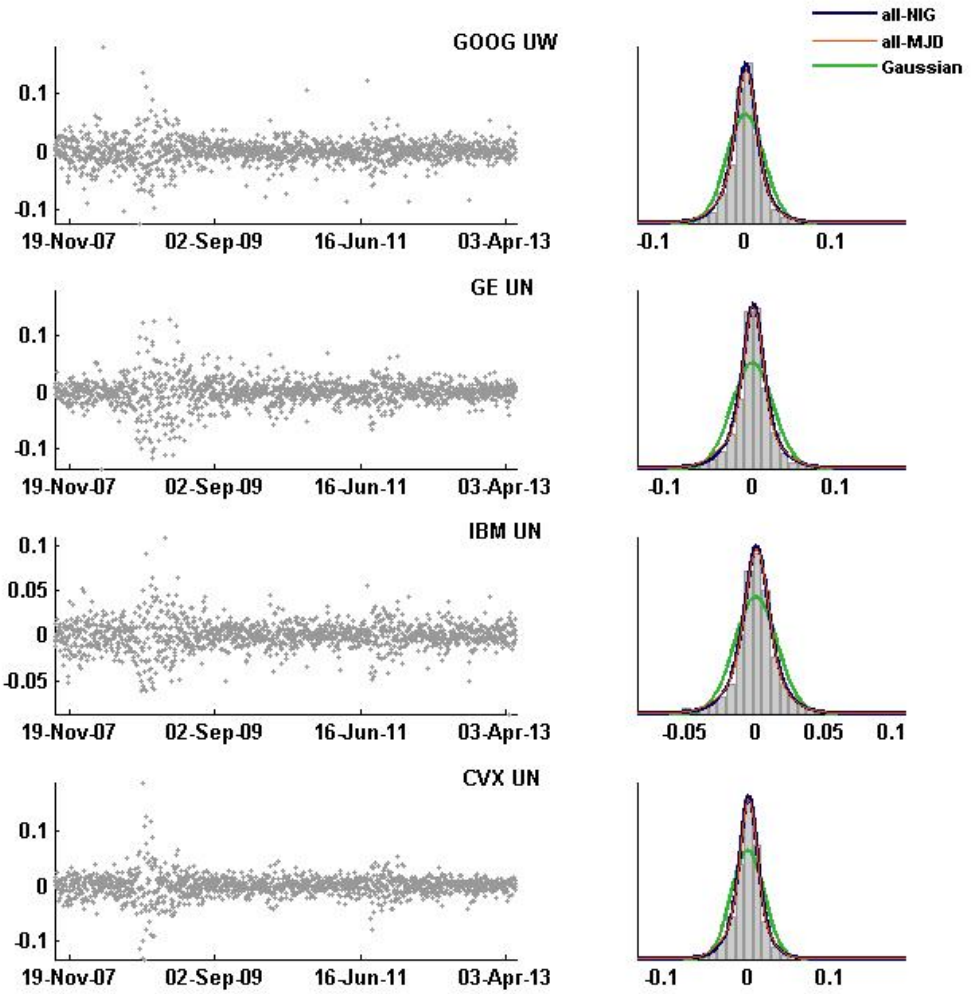

Figure 3.2: Time series and histograms: Google, General Electric, IBM, Chevron Corporation.

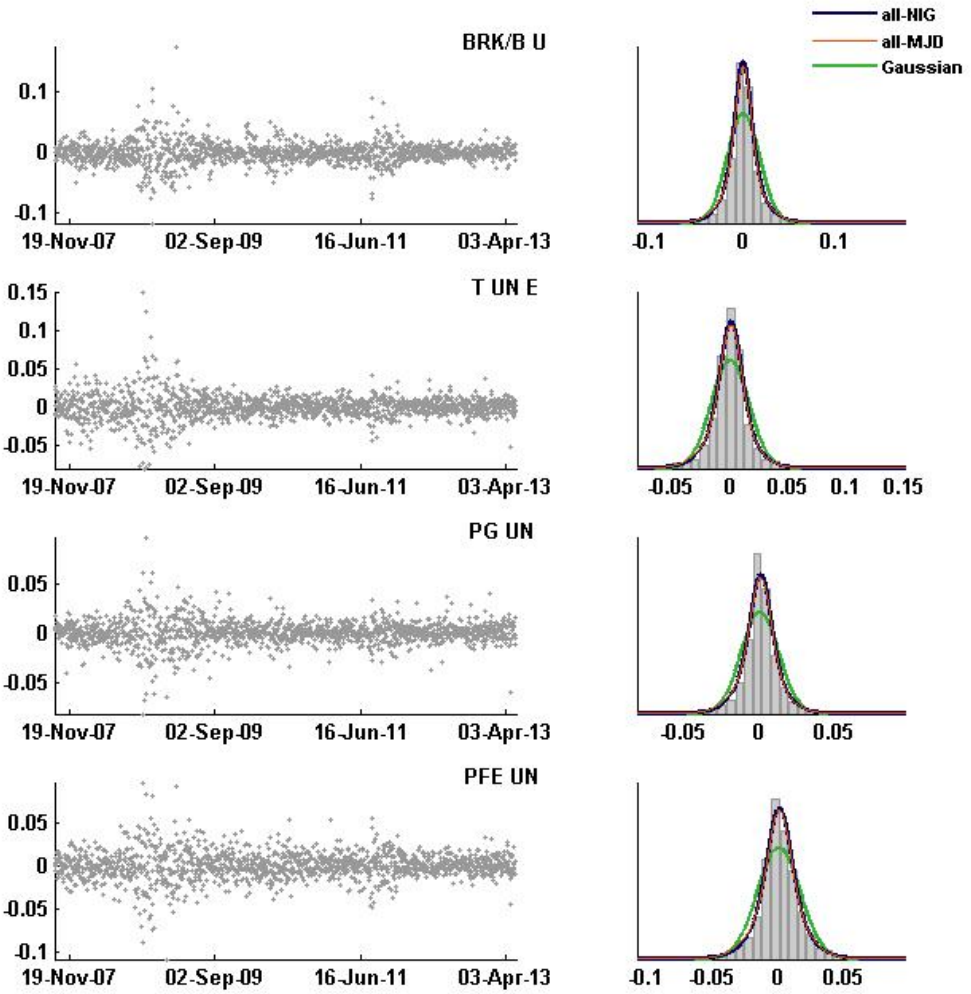

Figure 3.3: Time series and histograms: Berkshire Hathaway, AT&T, Procter & Gamble, Pfizer.

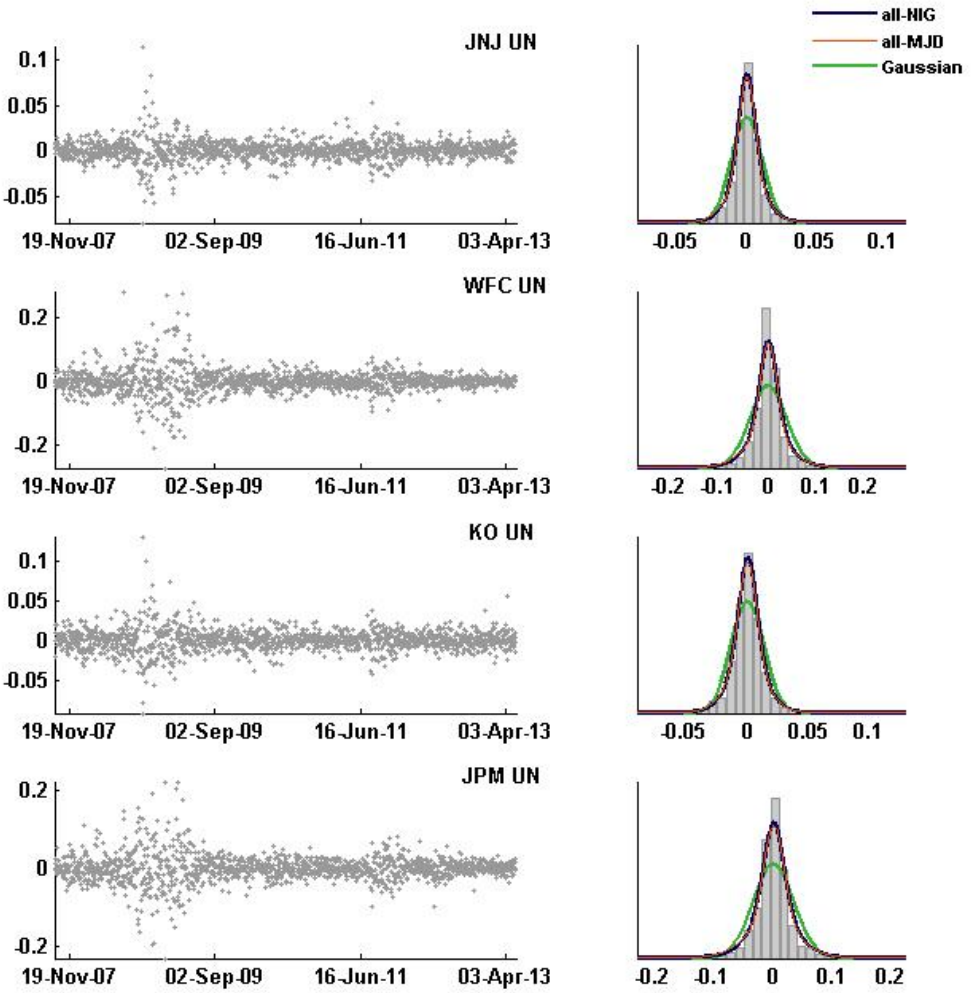

Figure 3.4: Time series and histograms: Johnson & Johnson, Wells Fargo & Co., Coca-Cola, JPMorgan Chase & Co..

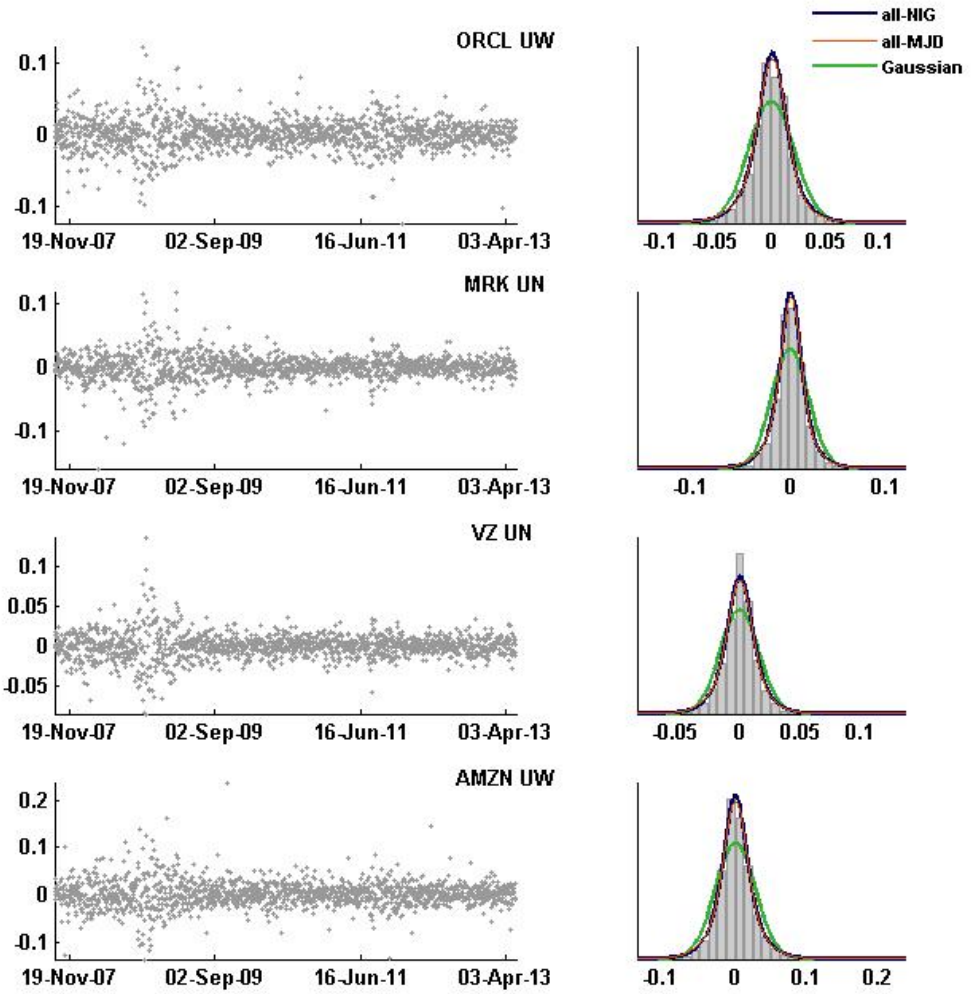

Figure 3.5: Time series and histograms: Oracle, Merck & Co., Verizon Communications, Amazon.com.

3.6 Methodology and results

In this section we present methodology and results relative to our empirical investigation. We test the out-of-sample performance of the allocations based on the multivariate Lévy model (2.1), in its 'all-NIG' and 'all-MJD' specifications, compared to those obtained applying the benchmark strategies introduced in Section 3.3.
The out-of-sample results are evaluated by means of the performance measures presented in Section 3.4.
Sections 3.6.1 and 3.6.2 illustrate the rolling window approaches adopted to test allocations over daily and weekly horizons respectively, and point out some expedients adopted in the portfolio construction, integrating the basic framework introduced in Section 3.1.
Results are presented distinguishing whether the maximization of expected utility is based on exponential (3.3) or power (3.4) function.
Finally, we summarize and comment results in Section 3.6.3.

3.6.1 Daily horizon

To test allocations over a daily horizon, we adopt the following rolling window strategy: estimation is performed on the most recent 500 daily returns; the weights resulting from the allocation decision are then kept fixed for the subsequent 5 days, and portfolio returns on these 5 days are used to evaluate what we call out-of-sample performance. Then a new allocation decision takes place. Given our dataset, made of 1434 observations, we get 186 estimation periods and 930 allocations to evaluate the strategies presented in Sections 3.2 and 3.3. Figure 3.6 illustrates the rolling window approach.

It has long be recognized that the estimation of expected returns is a challenging task, with the consequence that in mean-variance portfolio allocations, for example, global minimum variance portfolio often achieve higher out-of-sample Sharpe ratio. This is why many works focus on global minimum risk portfolio to pursue the optimal allocation (see Martellini and Ziemann (2010) and references therein). We decide to adopt the same expedient and neutralize the impact of the expected return parameters in the strategies we implement. This choice is supported by evaluation results; in fact, it leads to better results in term of Sharpe ratio, turnover and maximum drawdown for every strategy and situ-

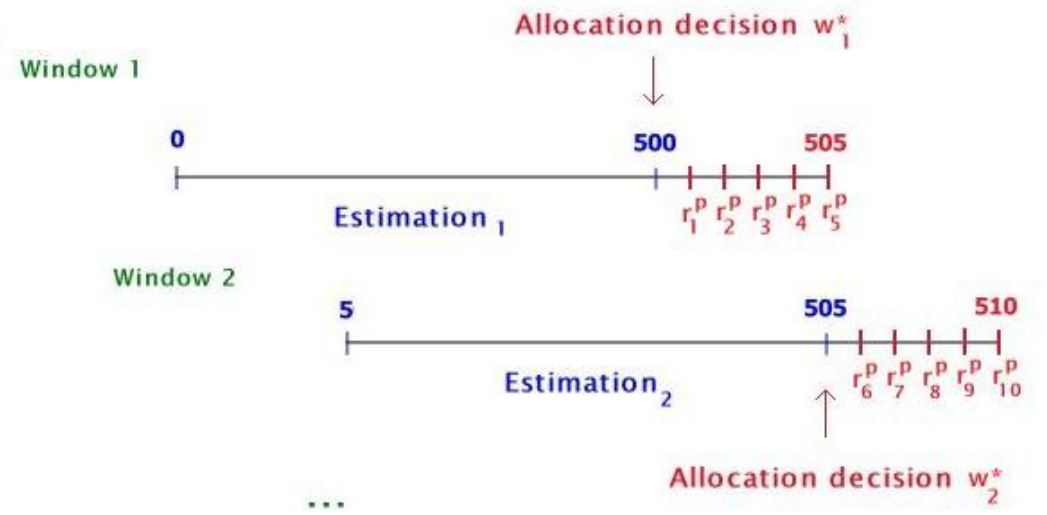

Figure 3.6: Rolling window strategy (daily horizon)

ation (number of assets, horizon, utility function) we investigated. Table 3.6.1, reported as an example, refers to an allocation with daily horizon, CARA utility function (3.3) with risk aversion coefficient $\lambda = 15$, and 20 available assets. Evaluation results when allocation takes into account the estimates of expected returns are highlighted in pink; the performance measures obtained when expected returns are neglected are highlighted in light-blue.

The impact of the expected return parameters in mean-variance (MV) and single factor (SF) strategies is neutralized simply setting the expected return of all assets to the same constant value, e.g. $\mu = 0$, as in Martellini and Ziemann (2010).

For the non-parametric (NP) approach instead, we perform kernel density estimation on demeaned portfolio returns.

In the 'all-NIG' model case we nullify the expected return effect setting the following conditions on drift parameters: $\mu_Z = -\theta_Z$ and $\mu_{Y^{(j)}} = -\theta_{Y^{(j)}}$, for all $j = 1, \dots, n$.

Similarly, in the 'all-MJD' model we set $\mu_Z = -\lambda_Z \nu_Z$ and $\mu_{Y^{(j)}} = -\lambda_{Y^{(j)}} \nu_{Y^{(j)}}$, for all $j = 1, \dots, n$.

Sections 3.6.1.1 and 3.6.1.2 present allocation and evaluation results for CARA (3.3)

	EQ	MV	SF	NP	all-NIG	all-MJD
SR	0.06895	0.07991	0.08558	0.08469	0.07879	0.07950
	0.06895	0.08245	0.08831	0.08958	0.08301	0.08419
TO	0.00752	0.00491	0.00599	0.00606	0.00474	0.00471
	0.00752	0.00487	0.00455	0.00435	0.00407	0.00409
MDD	0.12953	0.63950	0.52852	0.53648	0.69741	0.71206
	0.12953	0.09737	0.22324	0.08781	0.09471	0.09835

Table 3.2: Comparison of Sharpe ratio, turnover and maximum drawdown when allocation takes into account the estimates of expected returns (pink) or neglects it (light-blue), for different strategies. Daily horizon, CARA utility function, $\lambda = 15$, number of assets $n = 20$.

and CRRA (3.4) utility functions respectively. With respect to the exponential utility function we consider for the risk aversion parameter the values: $\lambda = 10$ and $\lambda = 15$; for the power function we investigate the cases $\lambda = 5$ and $\lambda = 10$.

On results tables we highlight in bold font the best value achieved in correspondence of each performance measure.

Each table is followed by an image showing the evolution of the weights, allocation by allocation, obtained applying the different strategies.

3.6.1.1 CARA utility function

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	6.72E-04	6.25E-04	6.34E-04	6.08E-04	6.22E-04	**6.30E-04**
Variance	9.84E-05	6.61E-05	**6.48E-05**	6.72E-05	6.59E-05	6.55E-05
Skewness	**-0.3727**	-0.4484	-0.4279	-0.4648	-0.4532	-0.4471
Kurtosis	6.4684	6.3326	6.2860	**6.0480**	6.1013	6.2354
W(all-NIG)	4.3712	1.3348	1.4003	1.0353	**1.0000**	1.1290
W(all-MJD)	4.3712	1.1823	1.2402	0.9169	**0.8857**	1.0000
SR	0.0678	0.0769	**0.0787**	0.0741	0.0766	0.0778
TO	0.0067	0.0061	0.0057	**0.0050**	0.0052	0.0054
MDD	**0.1323**	0.2850	0.4807	0.4045	0.3660	0.3450

Table 3.3: Daily horizon. CARA utility ($\lambda = 10$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.01.

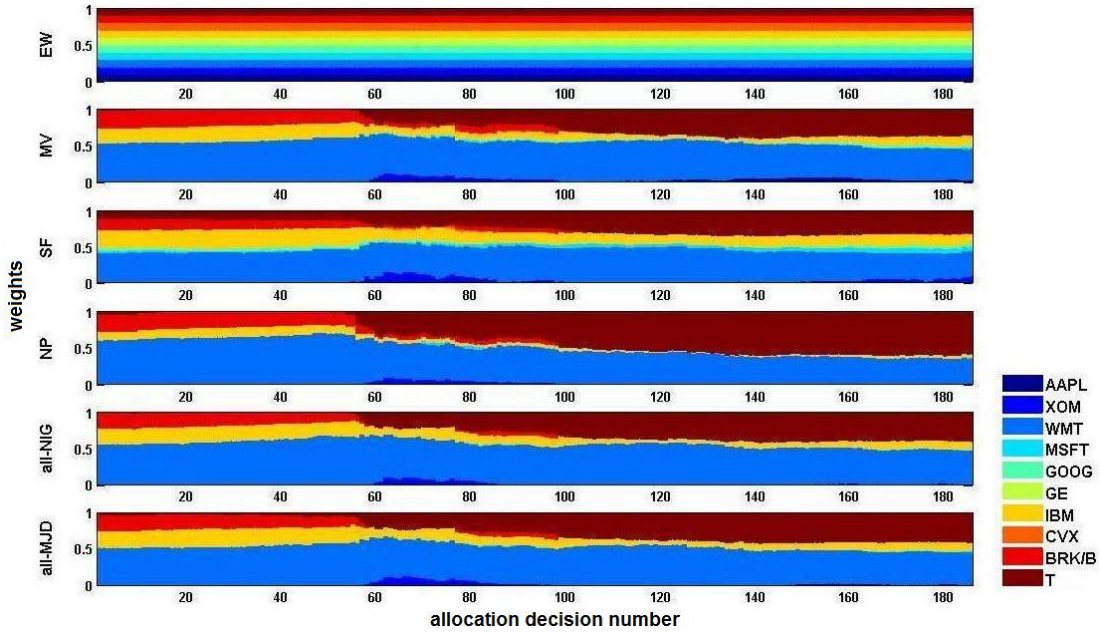

Figure 3.7: Weights: daily horizon; CARA utility ($\lambda = 10$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.72E-04**	6.25E-04	6.48E-04	6.08E-04	6.19E-04	6.29E-04
Variance	9.84E-05	6.61E-05	6.88E-05	6.73E-05	6.60E-05	**6.55E-05**
Skewness	**-0.3727**	-0.4470	-0.4347	-0.4657	-0.4536	-0.4463
Kurtosis	6.4684	6.3207	6.5762	**6.0453**	6.0788	6.2335
W(all-NIG)	2.6272	1.3163	2.5953	1.0703	**1.0000**	1.1495
W(all-MJD)	2.6272	1.1450	2.4560	0.9311	**0.8699**	1.0000
SR	0.0678	0.0769	**0.0782**	0.0741	0.0762	0.0777
TO	0.0067	0.0060	0.0062	**0.0049**	0.0051	0.0053
MDD	**0.1323**	0.2935	0.4941	0.4070	0.3718	0.3472

Table 3.4: Daily horizon. CARA utility ($\lambda = 15$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.01.

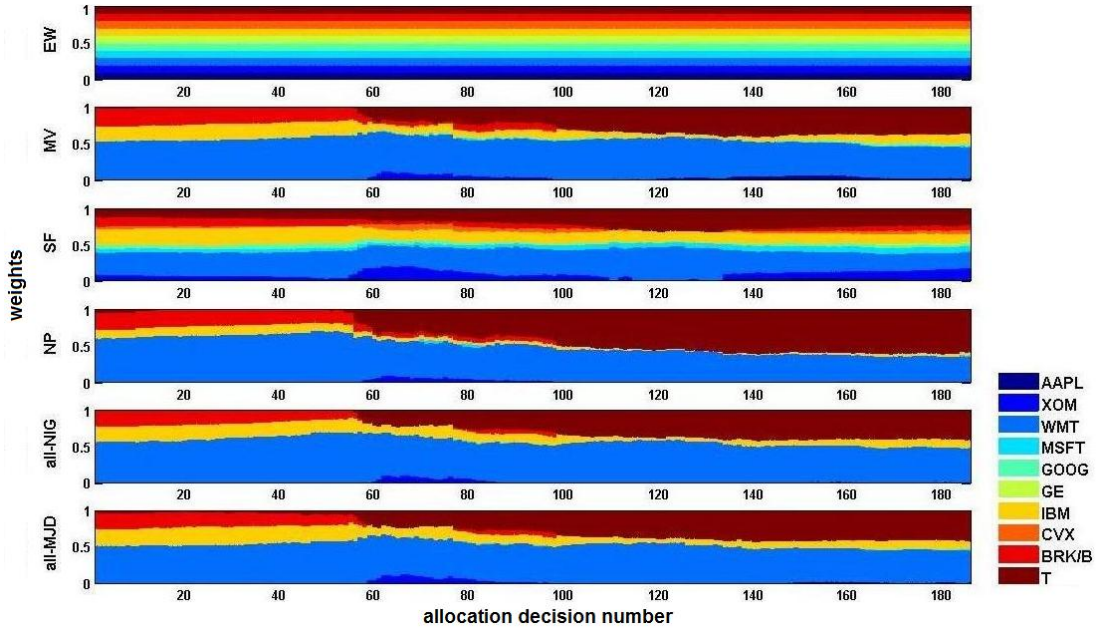

Figure 3.8: Weights: daily horizon; CARA utility ($\lambda = 15$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.76E-04**	5.97E-04	6.28E-04	6.49E-04	5.95E-04	6.01E-04
Variance	9.62E-05	5.25E-05	5.12E-05	5.29E-05	5.12E-05	**5.10E-05**
Skewness	-0.3603	-0.3382	**-0.3179**	-0.3417	-0.3451	-0.3341
Kurtosis	6.6800	6.1803	5.8610	5.9702	**5.8214**	5.8224
W(all-NIG)	2.4021	1.0290	1.0056	1.0372	1.0000	**0.9948**
W(all-MJD)	2.4217	1.0344	1.0108	1.0427	1.0052	**1.0000**
SR	0.0690	0.0824	0.0877	**0.0893**	0.0832	0.0842
TO	0.0075	0.0050	0.0044	0.0044	**0.0041**	0.0041
MDD	0.1295	**0.0868**	0.1957	0.1013	0.0915	0.0964

Table 3.5: Daily horizon. CARA utility ($\lambda = 10$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.01.

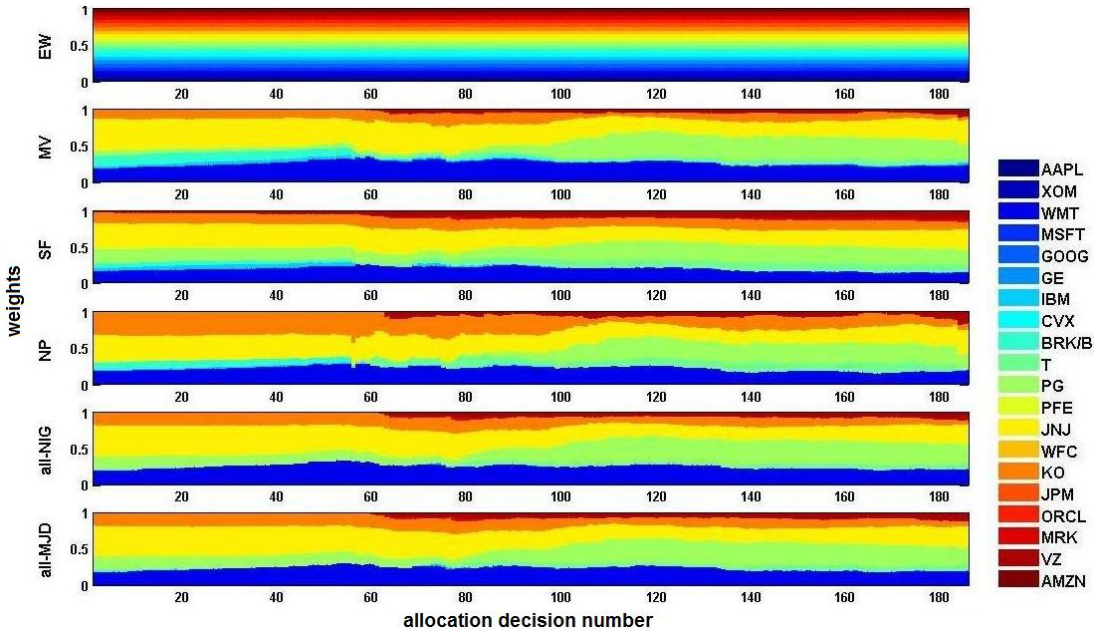

Figure 3.9: Weights: daily horizon; CARA utility ($\lambda = 10$); $n = 20$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.76E-04**	5.97E-04	6.36E-04	6.52E-04	5.94E-04	6.01E-04
Variance	9.62E-05	5.24E-05	5.19E-05	5.30E-05	5.12E-05	**5.10E-05**
Skewness	-0.3603	-0.3337	**-0.3110**	-0.3389	-0.3471	-0.3336
Kurtosis	6.6800	6.1482	5.8459	5.9534	**5.8127**	5.8130
W(all-NIG)	3.0054	1.0280	1.0263	1.0466	1.0000	**0.9928**
W(all-MJD)	3.0391	1.0354	1.0337	1.0542	1.0072	**1.0000**
SR	0.0690	0.0825	0.0883	**0.0896**	0.0830	0.0842
TO	0.0075	0.0049	0.0045	0.0044	**0.0041**	0.0041
MDD	0.1295	0.0974	0.2232	**0.0878**	0.0947	0.0983

Table 3.6: Daily horizon. CARA utility ($\lambda = 15$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.01.

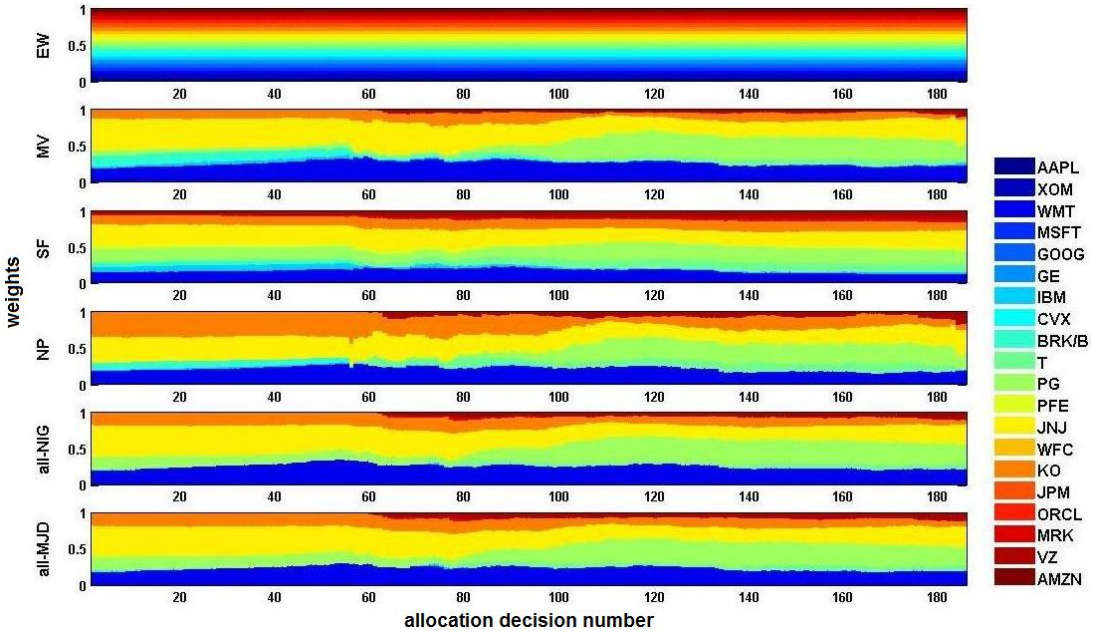

Figure 3.10: Weights: daily horizon; CARA utility ($\lambda = 15$); $n = 20$.

3.6.1.2 CRRA utility function

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	6.72E-04	6.25E-04	6.34E-04	**6.94E-04**	6.23E-04	6.29E-04
Variance	9.84E-05	6.61E-05	**6.44E-05**	6.74E-05	6.58E-05	6.55E-05
Skewness	-0.3727	-0.4448	-0.4306	**-0.3350**	-0.4516	-0.4454
Kurtosis	**6.4684**	6.2770	6.1461	5.6022	6.1143	6.2338
W(all-NIG)	22.1032	1.2550	1.0862	1.0119	**1.0000**	1.1147
W(all-MJD)	21.7794	1.1258	0.9744	0.9077	**0.8971**	1.0000
SR	0.0678	0.0768	0.0790	**0.0846**	0.0768	0.0777
TO	0.0067	0.0059	0.0053	0.0055	**0.0051**	0.0053
MDD	**0.1323**	0.3245	0.3945	0.5671	0.3658	0.3460

Table 3.7: Daily horizon. CRRA utility ($\lambda = 5$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.1.

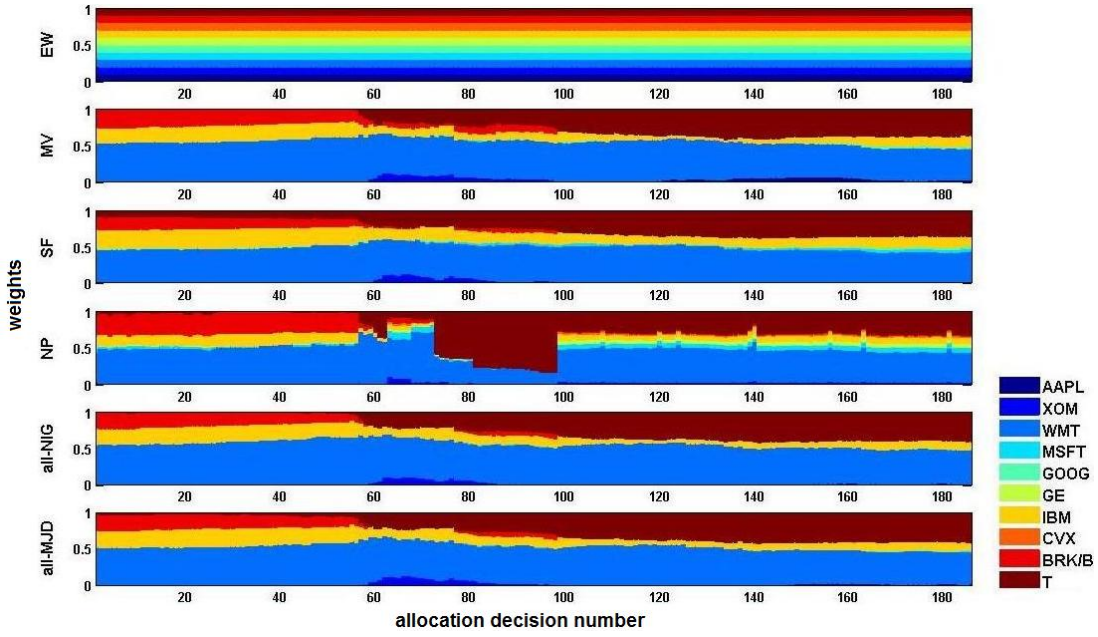

Figure 3.11: Weights: daily horizon; CRRA utility ($\lambda = 5$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.72E-04**	6.25E-04	6.34E-04	6.08E-04	6.20E-04	6.28E-04
Variance	9.84E-05	6.61E-05	**6.48E-05**	6.72E-05	6.60E-05	6.55E-05
Skewness	**-0.3727**	-0.4484	-0.4279	-0.4648	-0.4540	-0.4487
Kurtosis	**6.4684**	6.3326	6.2860	6.0480	6.0931	6.2375
W(all-NIG)	4.3712	1.3456	1.4116	1.0436	**1.0000**	1.1418
W(all-MJD)	4.3712	1.1785	1.2363	0.9140	**0.8758**	1.0000
SR	0.0678	0.0769	**0.0787**	0.0741	0.0764	0.0776
TO	0.0067	0.0061	0.0057	**0.0050**	0.0051	0.0053
MDD	**0.1323**	0.2850	0.4807	0.4045	0.3623	0.3517

Table 3.8: Daily horizon. CRRA utility ($\lambda = 10$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.01.

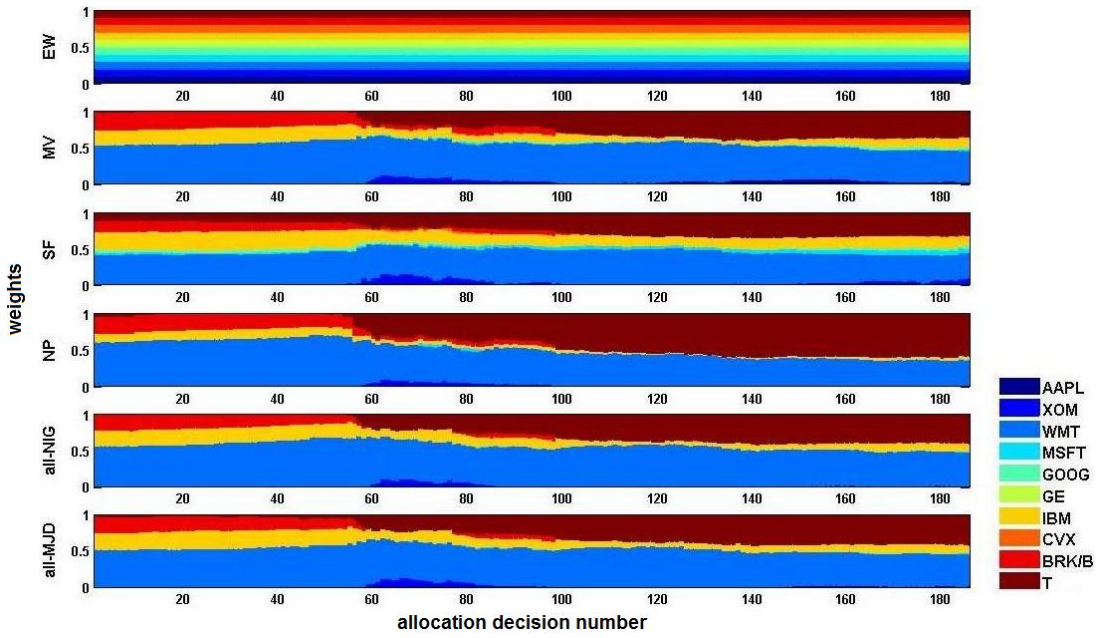

Figure 3.12: Weights: daily horizon; CRRA utility ($\lambda = 10$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.76E-04**	5.97E-04	6.18E-04	6.21E-04	5.96E-04	6.00E-04
Variance	9.62E-05	5.21E-05	**5.08E-05**	5.66E-05	5.11E-05	5.19E-05
Skewness	-0.3603	**-0.3223**	-0.3236	-0.3496	-0.3425	-0.3498
Kurtosis	6.6800	6.0588	5.8546	6.4443	**5.8090**	6.0149
W(all-NIG)	312.2649	1.3194	**0.9542**	4.4790	1.0000	1.3752
W(all-MJD)	227.0617	0.9594	**0.6938**	3.2569	0.7271	1.0000
SR	0.0690	0.0826	**0.0866**	0.0825	0.0833	0.0833
TO	0.0075	0.0045	0.0042	0.0053	**0.0041**	0.0047
MDD	0.1295	0.1440	0.1591	0.2093	0.0927	**0.0809**

Table 3.9: Daily horizon. CRRA utility ($\lambda = 5$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.01.

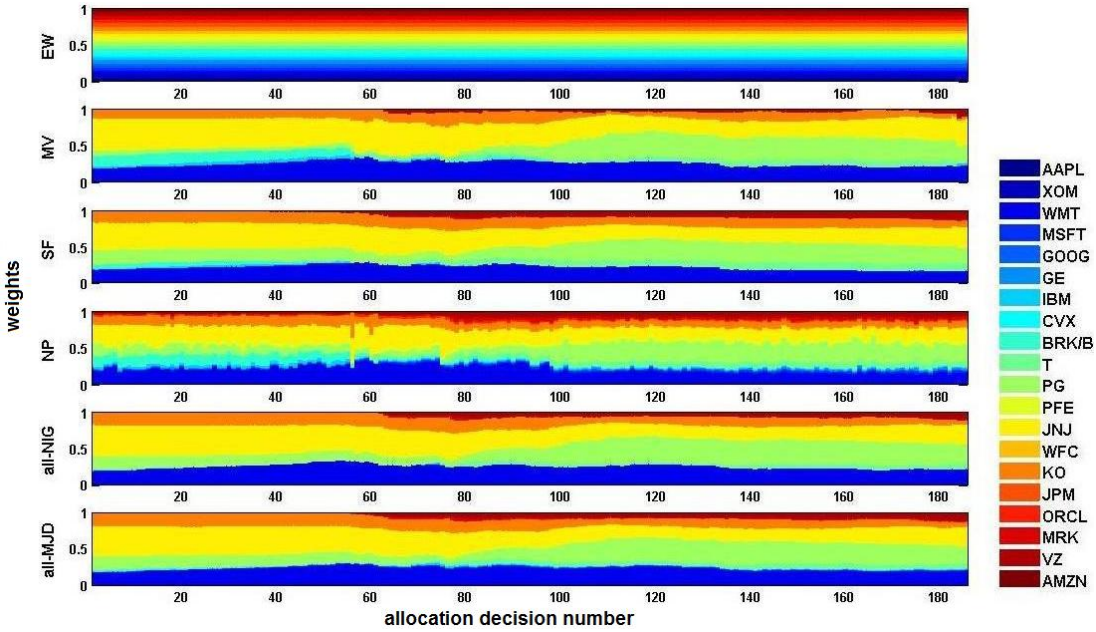

Figure 3.13: Weights: daily horizon; CRRA utility ($\lambda = 5$); $n = 20$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**6.76E-04**	5.97E-04	6.31E-04	6.63E-04	5.94E-04	6.01E-04
Variance	9.62E-05	5.21E-05	5.14E-05	6.00E-05	**5.12E-05**	5.15E-05
Skewness	-0.3603	-0.3223	**-0.3157**	-0.3332	-0.3447	-0.3440
Kurtosis	6.6800	6.0588	**5.8411**	5.8533	5.8129	5.9264
W(all-NIG)	313.2932	1.3092	**0.9925**	1.9199	1.0000	1.1704
W(all-MJD)	267.6810	1.1186	**0.8480**	1.6403	0.8544	1.0000
SR	0.0690	0.0826	**0.0880**	0.0856	0.0831	0.0838
TO	0.0075	0.0045	0.0044	0.0044	**0.0041**	0.0044
MDD	0.1295	0.1440	0.2006	0.2373	0.0946	**0.0846**

Table 3.10: Daily horizon. CRRA utility ($\lambda = 10$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.1.

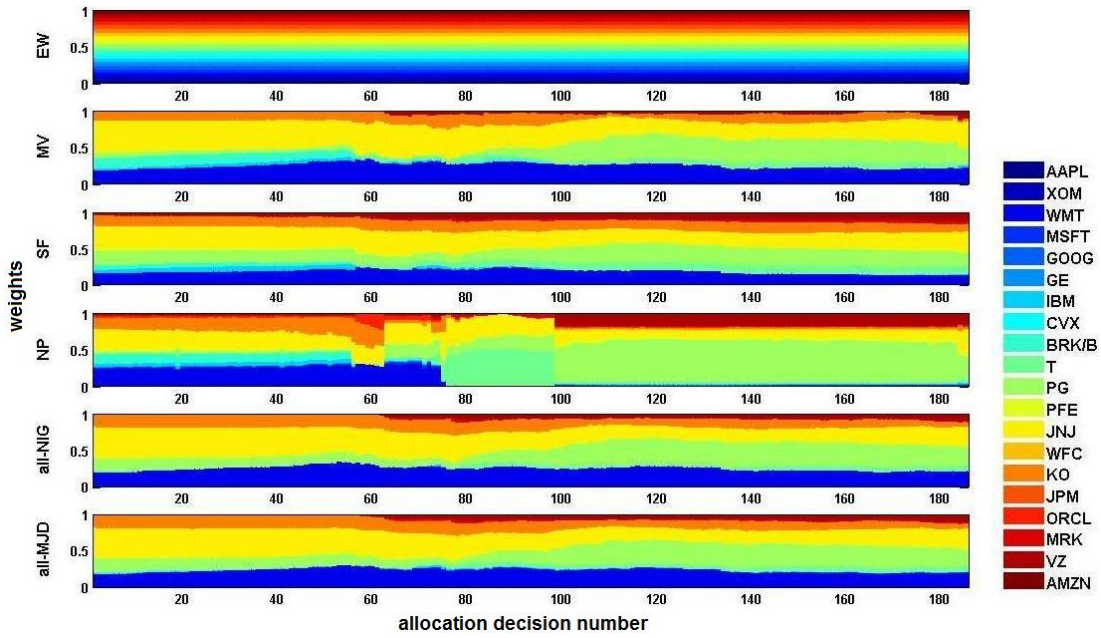

Figure 3.14: Weights: daily horizon; CRRA utility ($\lambda = 10$); $n = 20$.

3.6.2 Weekly horizon

To test allocations over a weekly horizon, we adopt the following rolling window strategy: the multivariate Lévy models estimation is performed on 500 daily returns and the resulting distribution is then projected to a weekly horizon; the implementation of mean-variance, single factor and non-parametric approaches is based on 100 weekly observations instead. Out-of-sample performance is evaluated on the week following the estimation. Given our dataset, made of 1434 daily returns, corresponding to 286 weekly observations, we get 186 allocations to evaluate the different strategies. Figure 3.15 illustrates the rolling window approach.

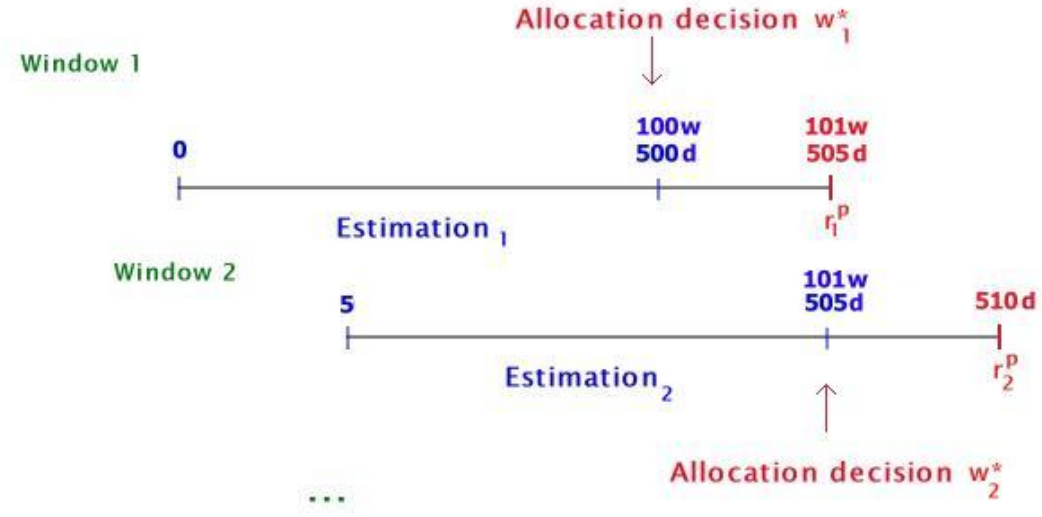

Figure 3.15: Rolling window strategy (weekly horizon)

We choose to neutralize the effect of expected return parameters as we did for allocations with daily horizon, for the same reasons discussed above.

Sections 3.6.2.1 and 3.6.2.2 present allocation and evaluation results for CARA (3.3), with $\lambda = 10, 15$, and CRRA (3.4), with $\lambda = 5, 10$, utility functions respectively.
On results tables we highlight in bold font the best value achieved in correspondence of each performance measure.
Each table is followed by an image showing the evolution of the weights obtained with the different strategies.

3.6.2.1 CARA utility function

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**5.09E-04**	3.71E-04	2.34E-04	3.34E-05	2.55E-04	2.77E-04
Variance	1.18E-04	7.67E-05	7.72E-05	7.63E-05	**7.21E-05**	7.28E-05
Skewness	**-0.5611**	-0.8016	-0.8075	-0.8654	-0.7945	-0.7882
Kurtosis	**7.8057**	5.9877	6.4377	6.1428	5.7601	5.8484
W(all-NIG)	1.1426	**0.9788**	1.0053	1.0927	1.0000	0.9991
W(all-MJD)	1.1438	**0.9797**	1.0062	1.0937	1.0009	1.0000
SR	**0.0469**	0.0424	0.0266	0.0038	0.0301	0.0324
TO	0.0165	0.0136	0.0133	**0.0099**	0.0117	0.0123
MDD	**0.1474**	0.4778	0.4845	0.3716	0.3224	0.3047

Table 3.11: Number of assets: $n = 10$. Risk aversion parameter: $\lambda = 10$. Kernel bandwidth for NP strategy: h=0.1.

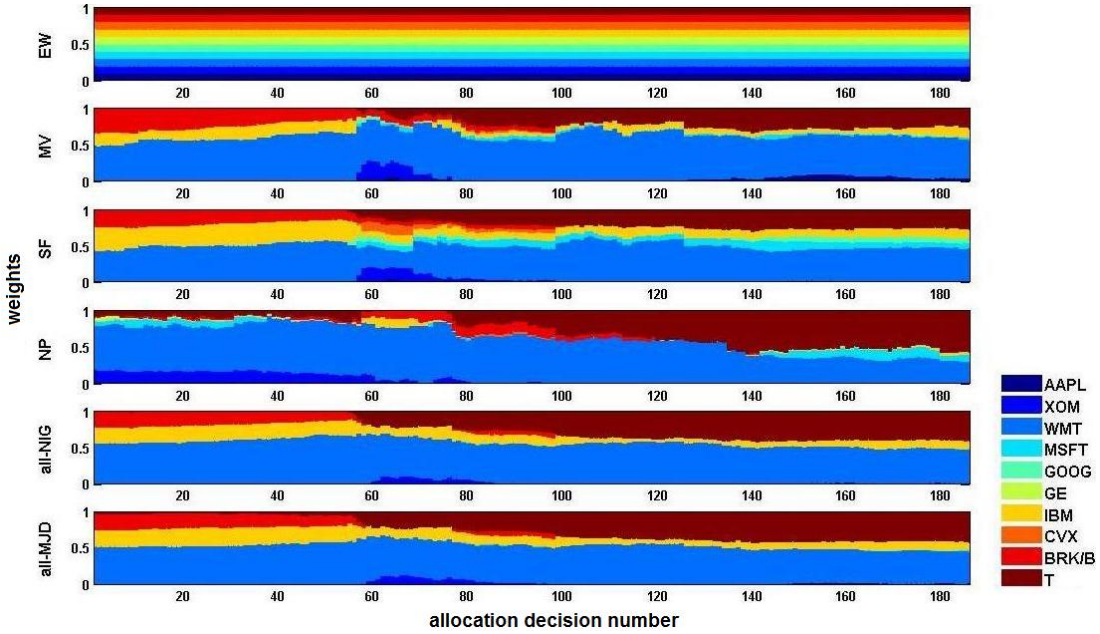

Figure 3.16: Weights: weekly horizon; CARA utility ($\lambda = 10$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**5.09E-04**	3.72E-04	1.98E-04	3.40E-05	2.54E-04	2.78E-04
Variance	1.18E-04	7.67E-05	7.93E-05	7.62E-05	**7.20E-05**	7.28E-05
Skewness	**-0.5611**	-0.7990	-0.8770	-0.8790	-0.8015	-0.7866
Kurtosis	7.8057	5.9541	7.1237	6.1444	**5.7529**	5.8402
W(all-NIG)	1.2327	1.0147	1.0372	1.1159	**1.0000**	1.0067
W(all-MJD)	1.2228	1.0075	1.0296	1.1079	**0.9934**	1.0000
SR	**0.0469**	0.0424	0.0223	0.0039	0.0299	0.0326
TO	0.0165	0.0134	0.0145	**0.0099**	0.0116	0.0123
MDD	**0.1474**	0.4917	0.4032	0.3875	0.3241	0.3066

Table 3.12: Weekly horizon. CARA utility ($\lambda = 15$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.1.

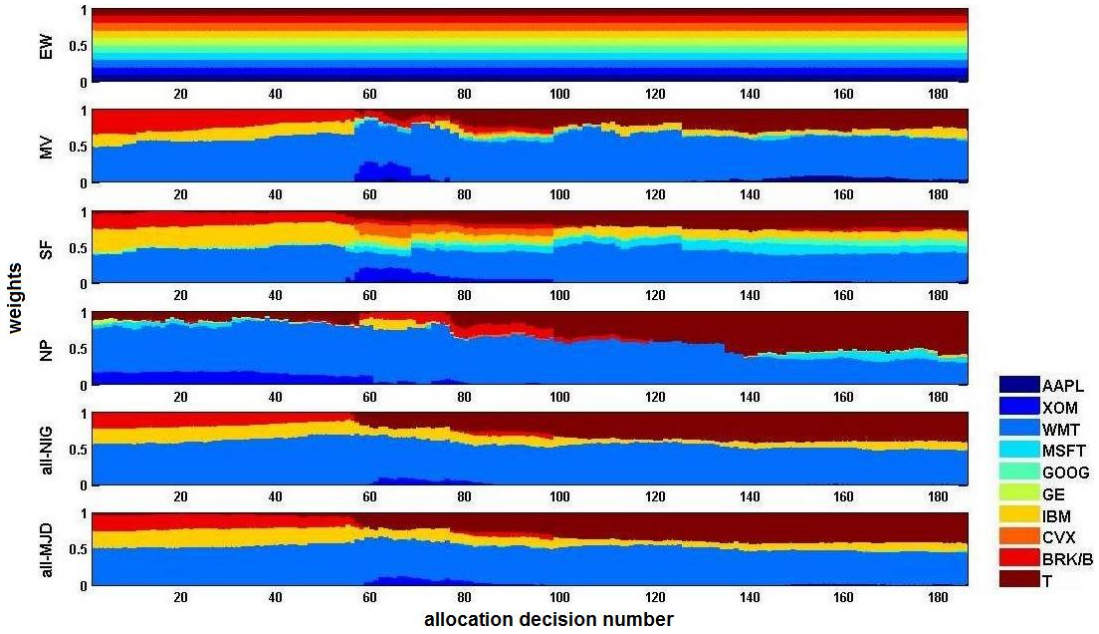

Figure 3.17: Weights: weekly horizon; CARA utility ($\lambda = 15$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	7.30E-04	7.12E-04	7.25E-04	5.02E-04	7.96E-04	**8.18E-04**
Variance	1.20E-04	6.17E-05	6.05E-05	6.20E-05	5.77E-05	**5.76E-05**
Skewness	-0.5072	-0.3386	-0.3334	-0.6071	-0.3297	**-0.3186**
Kurtosis	7.6798	**4.7614**	5.2593	5.5785	4.9364	4.9765
W(all-NIG)	2.0503	1.1300	1.0606	1.1925	1.0000	**0.9906**
W(all-MJD)	2.0795	1.1406	1.0703	1.2050	1.0097	**1.0000**
SR	0.0666	0.0907	0.0932	0.0637	0.1048	**0.1077**
TO	0.0182	0.0113	0.0106	0.0101	**0.0097**	0.0097
MDD	0.1412	0.2372	0.2587	0.2105	**0.1401**	0.1448

Table 3.13: Number of assets: $n = 20$. Risk aversion parameter: $\lambda = 10$. Kernel bandwidth for NP strategy: h=0.2.

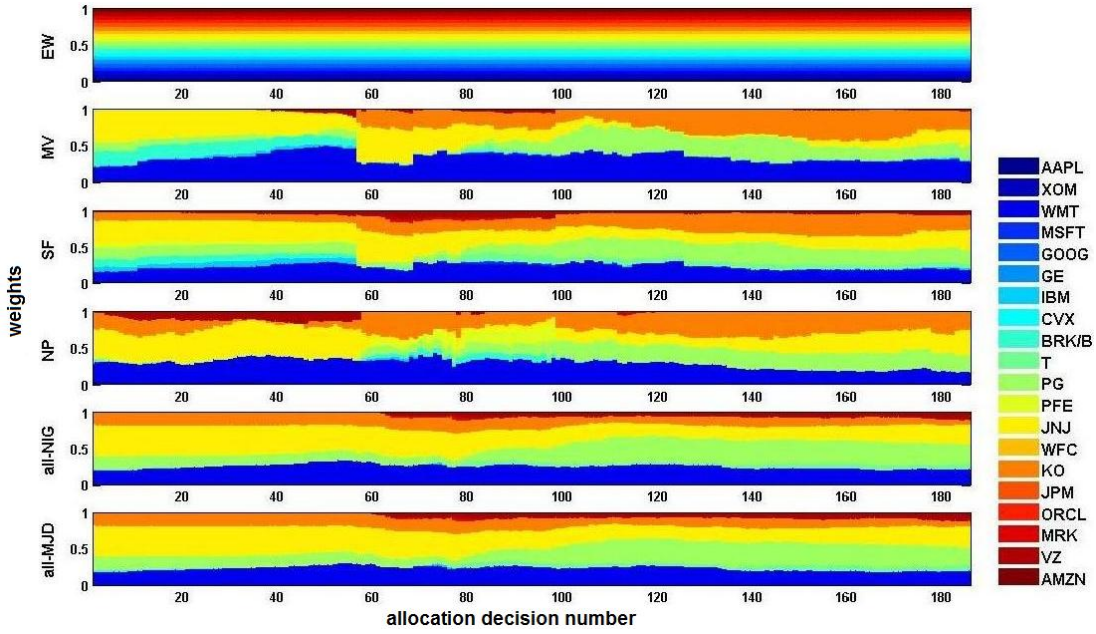

Figure 3.18: Weights: weekly horizon; CARA utility ($\lambda = 10$); $n = 20$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	7.30E-04	7.15E-04	7.08E-04	4.88E-04	7.92E-04	**8.17E-04**
Variance	1.20E-04	6.15E-05	6.27E-05	6.19E-05	5.77E-05	**5.76E-05**
Skewness	-0.5072	-0.3293	-0.3448	-0.6009	-0.3323	**-0.3186**
Kurtosis	7.6798	**4.6983**	5.4039	5.6126	4.9248	4.9727
W(all-NIG)	2.5193	1.1073	1.0630	1.2279	1.0000	**0.9798**
W(all-MJD)	2.6043	1.1296	1.0837	1.2560	1.0211	**1.0000**
SR	0.0666	0.0911	0.0894	0.0620	0.1043	**0.1076**
TO	0.0182	0.0110	0.0112	0.0101	**0.0097**	0.0097
MDD	0.1412	0.2449	0.2862	0.1854	**0.1391**	0.1448

Table 3.14: Number of assets: $n = 20$. Risk aversion parameter: $\lambda = 15$. Kernel bandwidth for NP strategy: h=0.2.

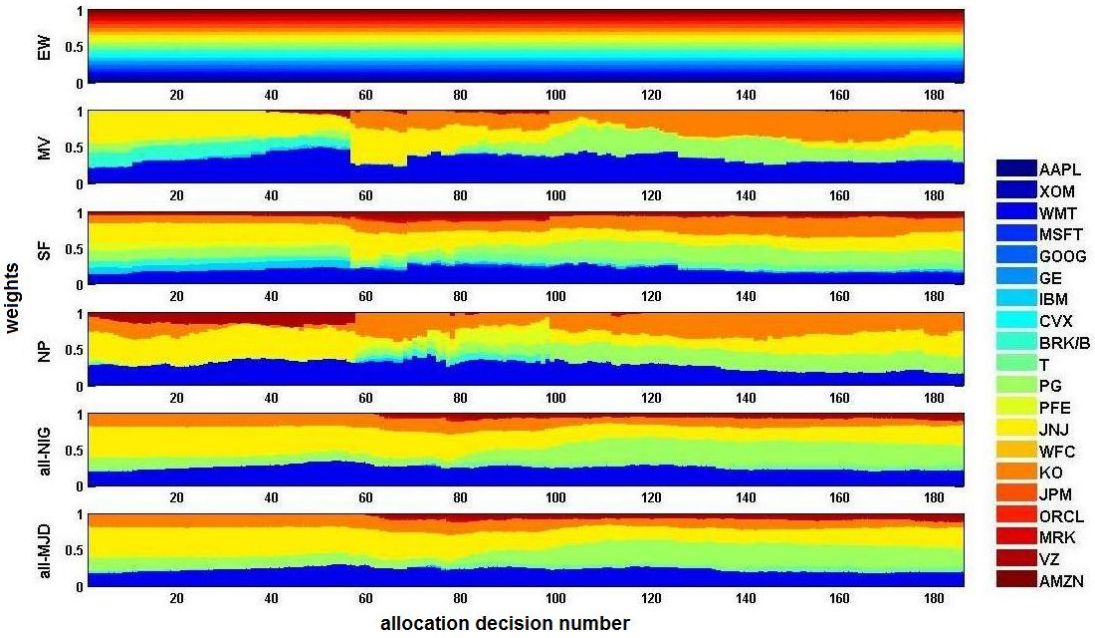

Figure 3.19: Weights: weekly horizon; CARA utility ($\lambda = 15$); $n = 20$.

3.6.2.2 CRRA utility function

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**5.09E-04**	3.79E-04	2.90E-04	8.91E-05	2.58E-04	2.77E-04
Variance	1.18E-04	7.67E-05	7.54E-05	7.43E-05	**7.21E-05**	7.28E-05
Skewness	**-0.5611**	-0.7788	-0.7426	-0.7932	-0.7917	-0.7874
Kurtosis	7.8057	**5.8090**	5.8360	5.5826	5.7674	5.8408
W(all-NIG)	22.5092	1.1758	1.2334	**0.8728**	1.0000	1.0941
W(all-MJD)	22.4125	1.0747	1.1273	**0.7977**	0.9140	1.0000
SR	**0.0469**	0.0433	0.0334	0.0103	0.0304	0.0325
TO	0.0165	0.0130	0.0125	**0.0098**	0.0118	0.0123
MDD	**0.1474**	0.5230	0.5246	0.4060	0.3201	0.3041

Table 3.15: Weekly horizon. CRRA utility ($\lambda = 5$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.1.

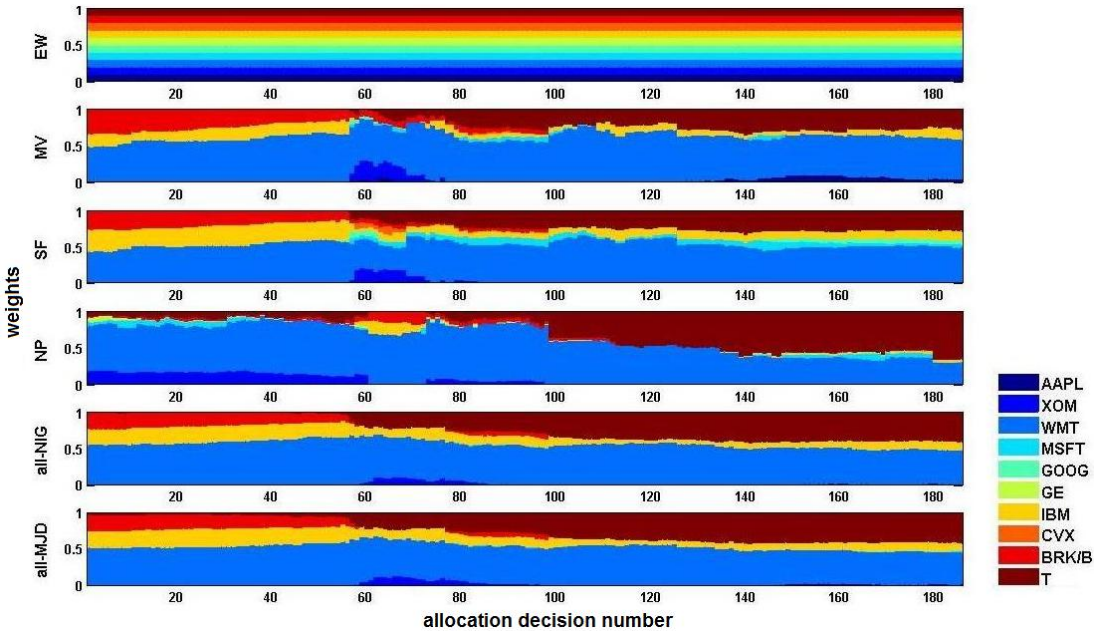

Figure 3.20: Weights: weekly horizon; CRRA utility ($\lambda = 5$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	**5.09E-04**	3.79E-04	2.15E-04	4.71E-05	2.55E-04	2.77E-04
Variance	1.18E-04	7.67E-05	7.80E-05	7.49E-05	**7.21E-05**	7.28E-05
Skewness	**-0.5611**	-0.7788	-0.8307	-0.8189	-0.7955	-0.7859
Kurtosis	7.8057	5.8090	6.6804	**5.7394**	5.7543	5.8366
W(all-NIG)	4.3712	1.1881	3.2863	1.1143	**1.0000**	1.1020
W(all-MJD)	4.3712	1.0781	2.9821	1.0112	**0.9075**	1.0000
SR	**0.0469**	0.0433	0.0244	0.0054	0.0300	0.0324
TO	0.0165	0.0130	0.0137	**0.0098**	0.0117	0.0123
MDD	**0.1474**	0.5230	0.4603	0.4038	0.3215	0.3044

Table 3.16: Weekly horizon. CRRA utility ($\lambda = 10$). Number of assets: $n = 10$. Kernel bandwidth for NP strategy: h=0.07.

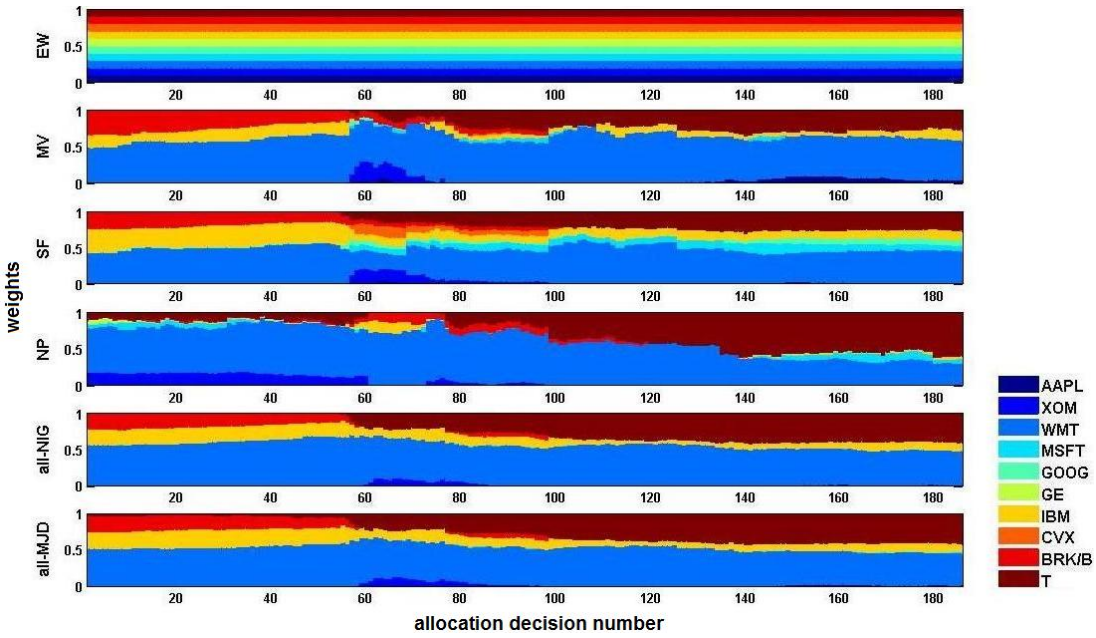

Figure 3.21: Weights: weekly horizon; CRRA utility ($\lambda = 10$); $n = 10$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	7.30E-04	7.30E-04	7.64E-04	5.15E-04	7.99E-04	**8.18E-04**
Variance	1.20E-04	6.12E-05	5.95E-05	6.58E-05	5.76E-05	**5.76E-05**
Skewness	-0.5072	**-0.3081**	-0.3468	-0.5255	-0.3272	-0.3178
Kurtosis	7.6798	**4.5648**	5.1381	5.5372	4.9391	4.9734
W(all-NIG)	22.7374	**0.9894**	1.6764	4.9333	1.0000	1.0461
W(all-MJD)	22.7374	**0.9493**	1.6025	4.7159	0.9559	1.0000
SR	0.0666	0.0933	0.0991	0.0636	0.1052	**0.1078**
TO	0.0182	0.0102	0.0103	0.0096	0.0097	**0.0097**
MDD	**0.1412**	0.2612	0.2587	0.2104	0.1416	0.1457

Table 3.17: Weekly horizon. CRRA utility ($\lambda = 5$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.1.

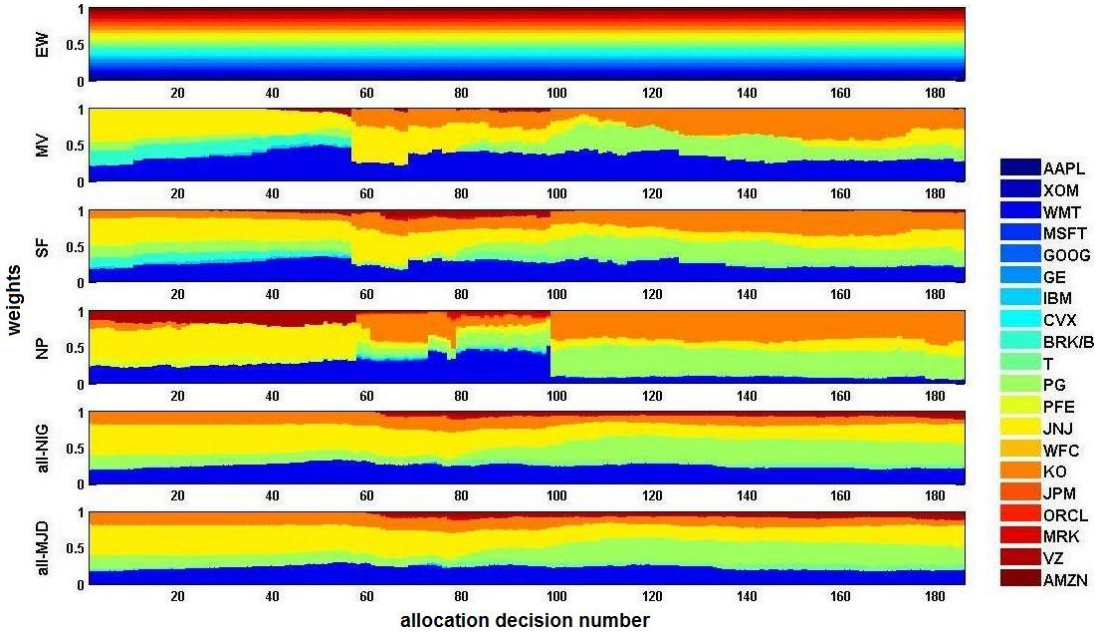

Figure 3.22: Weights: weekly horizon; CRRA utility ($\lambda = 5$); $n = 20$.

	EQ	MV	SF	NP	all-NIG	all-MJD
Mean	7.30E-04	7.30E-04	7.16E-04	4.90E-04	7.95E-04	**8.17E-04**
Variance	1.20E-04	6.12E-05	6.10E-05	6.84E-05	5.77E-05	**5.77E-05**
Skewness	-0.5072	**-0.3081**	-0.3297	-0.5886	-0.3315	-0.3195
Kurtosis	7.6798	**4.5648**	5.2870	6.1413	4.9402	4.9745
W(all-NIG)	693.7070	**0.9814**	2.0008	11.5704	1.0000	1.0395
W(all-MJD)	667.3671	**0.9472**	1.9248	11.1311	0.9620	1.0000
SR	0.0666	0.0933	0.0917	0.0593	0.1047	**0.1075**
TO	0.0182	0.0102	0.0107	**0.0095**	0.0097	0.0097
MDD	0.1412	0.2612	0.2576	0.2026	0.1403	**0.1311**

Table 3.18: Weekly horizon. CRRA utility ($\lambda = 10$). Number of assets: $n = 20$. Kernel bandwidth for NP strategy: h=0.1.

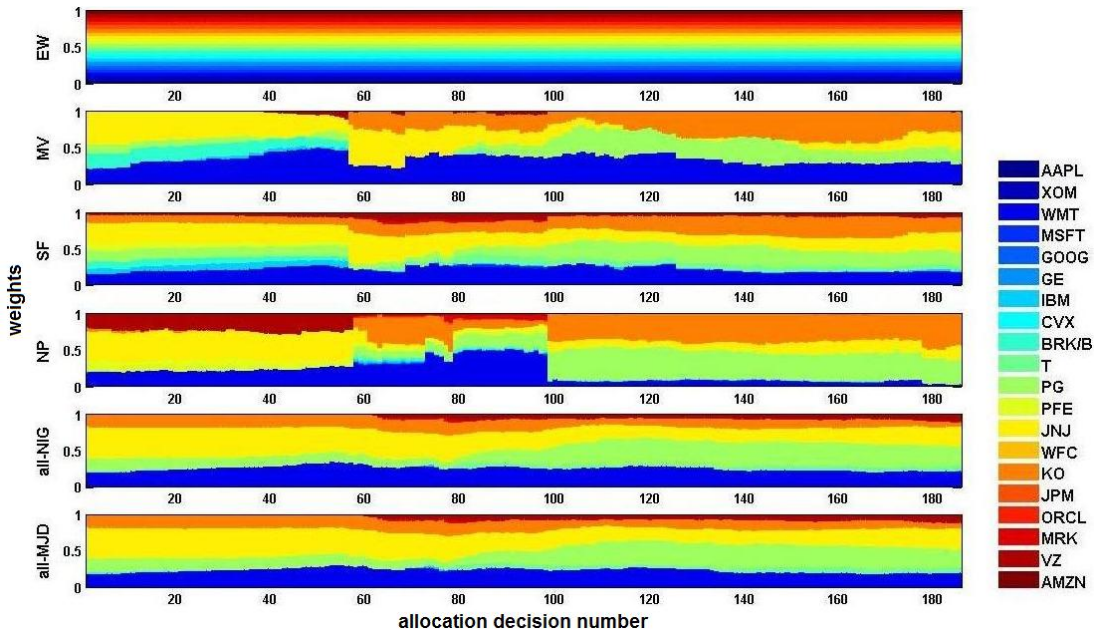

Figure 3.23: Weights: weekly horizon; CRRA utility ($\lambda = 10$); $n = 20$.

3.6.3 Summarizing allocation results

To summarize and better understand the results of our empirical analysis, we report in Tables 3.19 and 3.20 the strategies giving respectively the best and the worst values according to the different performance measures, in correspondence of each situation we tested. In the losers' table we display two strategies, since we want to verify if there are strategies, other than the naive equally-weighted portfolio, which often produce bad results.

First of all we observe that the 'all-NIG' and 'all-MJD' specifications of the multivariate Lévy model often appear in the winners' table, especially when focusing on MUG and turnover. On the other hand, they are almost missing from the losers' one, the only exception being the presence of the 'all-NIG' model as second worst according to the SR criterion[7] when considering a weekly horizon, 10 assets and power utility function with risk aversion coefficient $\lambda = 5$.

The naive EQ strategy always gives the worst results according to MUG and turnover criteria, and bad results are obtained using Sharpe ratio as well, except for weekly horizon allocations with 10 available assets.

The strong point of equally weighted portfolios is that they have a small maximum drawdown, but the allocations based on multivariate Lévy models tend to beat them when the number of available assets increases from 10 to 20.

While on the mean-variance approach we can easily tell that it appears in the losers' table much more frequently than in the winners' one, the single-factor and non-parametric strategies exhibit ambivalent outcomes.

The SF approach displays good Sharpe ratios when investments focus on a daily horizon, but low ones when investment is set over a weekly horizon; it is always among the worst strategies when evaluation is based on maximum drawdown.

Portfolios obtained by means of the non-parametric approach are characterized by low turnover, but often display low Sharpe ratios and poor outcomes with respect to the MUG

[7] As we mentioned in Section 3.4, for each set of results in Sections 3.6.1 and 3.6.2 we tested the hypothesis that the Sharpe ratios obtained applying different strategies be equal. We performed both pairwise and multivariate tests, described in Appendix D, with a significance level $\alpha = 5\%$. Whereas the pairwise Sharpe ratio test shows that the performances obtained with different strategies are in most cases indistinguishable, the multivariate test results reject the equality of the Sharpe ratios. Therefore we can at least rely on the fact that the worst and the best strategies according to the SR criterion produces significantly different results.

THE WINNERS

MUG	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	**all-NIG**	**all-MJD**	MV	**all-MJD**
λ=15	**all-NIG**	**all-MJD**	**all-NIG**	**all-MJD**
CRRA				
λ=5	**all-NIG**	SF	NP	MV
λ=10	**all-NIG**	SF	**all-NIG**	MV

SR	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	SF	NP	EQ	**all-MJD**
λ=15	SF	NP	EQ	**all-MJD**
CRRA				
λ=5	NP	SF	EQ	**all-MJD**
λ=10	SF	SF	EQ	**all-MJD**

TO	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	NP	**all-NIG**	NP	**all-NIG**
λ=15	NP	**all-NIG**	NP	**all-NIG**
CRRA				
λ=5	**all-NIG**	**all-NIG**	NP	**all-MJD**
λ=10	NP	**all-NIG**	NP	NP

MDD	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	EQ	MV	EQ	**all-NIG**
λ=15	EQ	NP	EQ	**all-NIG**
CRRA				
λ=5	EQ	**all-MJD**	EQ	EQ
λ=10	EQ	**all-MJD**	EQ	**all-MJD**

Table 3.19: Summary of results: best strategies

THE LOSERS

MUG	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	EQ, SF	EQ,NP	EQ,NP	EQ,NP
λ=15	EQ, SF	EQ,NP	EQ,NP	EQ,NP
CRRA				
λ=5	EQ,MV	EQ,NP	EQ, SF	EQ,NP
λ=10	EQ,SF	EQ,NP	EQ, SF	EQ,NP

SR	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	EQ,NP	EQ,MV	NP,SF	NP,EQ
λ=15	EQ,NP	EQ,MV	NP,SF	NP,EQ
CRRA				
λ=5	EQ,MV	EQ,NP	NP,**all-NIG**	NP,EQ
λ=10	SR,NP	EQ,MV	NP,SF	NP,EQ

TO	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	EQ,MV	EQ,MV	EQ,MV	EQ,MV
λ=15	EQ,SF	EQ,MV	EQ,SF	EQ,SF
CRRA				
λ=5	EQ,MV	EQ,NP	EQ,MV	EQ,SF
λ=10	EQ,MV	EQ,MV	EQ,SF	EQ,SF

MDD	**Daily horizon**		**Weekly horizon**	
	n=10	**n=20**	**n=10**	**n=20**
CARA				
λ=10	SF,NP	SF,EQ	SF,MV	SF,MV
λ=15	SF,NP	SF,EQ	MV,SF	SF,MV
CRRA				
λ=5	NP,SF	NP,SF	SF,MV	MV,SF
λ=10	SF,NP	NP,SF	MV,SF	MV,SF

Table 3.20: Summary of results: worst strategies

criterion. However, the NP strategy could be further refined and improved by a more careful choice of the kernel bandwidth h, which in the present work has been set simply according to graphical considerations arising from the weights plots. In fact, with too high or too low values for the bandwidth parameters the weights look very unstable; nevertheless, among the range of good values we may have missed the optimal ones, which is a point requiring deeper investigation.

Chapter 4

Downside risk measures for Lévy portfolios

In this chapter we deal with downside risk measures for portfolios made of multivariate Lévy assets.

In particular, in Section 4.1 we present formulas to compute value at risk (VaR) and tail conditional expectation (TCE), and we develop a new approach to compute the corresponding marginal measures, exploiting the expression for the conditional characteristic function introduced by Bartlett (1938).

In Section 4.2 we discuss intra-horizon risk measures, i.e. intra-horizon value at risk (VaR-I) and intra-horizon tail conditional expectation (TCE-I), which take into account the magnitude of losses that can incur throughout the investment term, not only at the end of the horizon. To compute these measure we adopt the Fourier Space Time-stepping (FST) algorithm, introduced by Jackson et al. (2008) for option pricing purposes.

Section 4.3 handles portfolio optimization when the objective is to minimize CVaR and asset returns follow the multivariate Lévy model (2.1), following the approach of Rockafellar and Uryasev (2000).

Finally, in Section 4.4, we present an application. For the 'all-NIG', 'all-MJD' and Gaussian model we build the optimal portfolio based on CVaR objective and we compute all the risk measures mentioned above.

4.1 Value at risk, expected shortfall and marginal measures

In this section we describe how to efficiently compute value at risk, expected shortfall, and their respective marginal measures when portfolio returns are assumed to follow exponential Lévy models.

4.1.1 Value at risk (VaR)

A primary tool for financial risk assessment is the value at risk (VaR), defined as the amount lost on a position (or portfolio) with a given small probability at the end of a fixed horizon. More precisely, the $(1-\alpha)$-VaR over an horizon $[0, T]$ is implicitly defined as

$$P(S_T - S_0 \leq -\text{VaR}) = \alpha, \tag{4.1}$$

where S_t is the value of the position (or portfolio) at time t.
We consider a VaR defined on the return distribution, as

$$P(X_T \leq -\text{VaR}) = \alpha, \tag{4.2}$$

where $X_t = \log(S_t) - \log(S_0)$, and the relation among the two VaRs defined in (4.1) and (4.2) is immediate.

To compute the α-quantile of the return distribution when assuming an exponential Lévy model, we adopt the approach of Le Courtois and Walter (2009), followed also by Ramponi (2012). In particular, they suggest how to compute the cumulative distribution function from the characteristic function, without first recovering the probability density function.
The following formula is available to bypass the use of the density

$$F(x) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \frac{\phi(u)}{iu} du; \tag{4.3}$$

however, the use of (4.3) can be problematic, because it is not stable around $u = 0$. Le Courtois and Walter (2009) developed a new formula to compute directly the cumulative distribution associated with any Lévy process as a Fourier transform of a well behaved function

of the characteristic function. For any positive real number a, they proved the following relation

$$F(x) = \frac{e^{ax}}{2\pi} \int_{-\infty}^{+\infty} e^{iux} \frac{\phi(-u + ia)}{a + iu} du. \tag{4.4}$$

Therefore, a Fast Fourier Transform (FFT) has to be performed to obtain F_X from ϕ_X and then a root search allows to compute the VaR, i.e. the absolute value of the α-quantile.

4.1.2 Expected shortfall (ES/CVaR/TCE)

The expected shortfall (ES) at tail probability α is defined as the average of the VaRs which are larger than the $(1-\alpha)$-VaR, i.e. it is focused on the losses in the tail which are larger than the corresponding VaR level. Formally

$$\text{ES}_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_p(X) dp. \tag{4.5}$$

Unlike VaR, that lacks subadditivity and convexity, the ES satisfies all the axioms of coherent risk measures (see Artzner et al. (1997), Acerbi and Tasche (2002)). The ES is convex for all possible portfolios, which means that it always accounts for the diversification effect.

Moreover, it can be easily proved that, for continuous random variables, the ES coincides with the tail conditional expectation (TCE), which is defined as

$$\text{TCE}_\alpha(X) = E(-X | X \leq -\text{VaR}_\alpha(X)). \tag{4.6}$$

To compute the ES under exponential Lévy models, Le Courtois and Walter (2009) developed a formula similar to (4.4), which bypasses the use of probability density function. In particular, they proved that

$$E(X | X \leq x) = \frac{E(X 1_{\{X \leq x\}})}{P(X \leq x)} = \frac{G(x)}{F(x)}, \tag{4.7}$$

where F is given in (4.4) and G can be written as

$$G(x) = \frac{e^{ax}}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \frac{\phi'(u + ia)}{u + ia} du, \tag{4.8}$$

for any $a > 0$.

This formula can always be applied, however it may be unhandy since the first derivative

of the characteristic function need to be computed. For a portfolio of assets following our multivariate Lévy model (2.1), whose characteristic function is given in (3.6), this is indeed the case. Therefore we first compute the VaR as described in Section 4.1.1, then we recover the probability density function of X, and finally we numerically evaluate the integral

$$\text{TCE}_\alpha = -\frac{1}{\alpha}\int_{-\infty}^{-\text{VaR}} y f_X(y)dy, \tag{4.9}$$

where $1-\alpha$ is the VaR level (e.g. for the 99%-VaR, $\alpha = 0.01$).

If the ES in monetary terms, instead of return terms, is needed, we have to compute

$$\text{ES}_\alpha = E(S_T - S_0 \,|S_T - S_0 \leq -\text{VaR}_\alpha), \tag{4.10}$$

where VaR_α is expressed in monetary terms, as in (4.1).

Based on the approach of Le Courtois and Walter (2009), we achieve the following formula

$$\text{ES}_\alpha = S_0\left[H\left(\log\left(\frac{-VaR_\alpha + S_0}{S_0}\right)\right) - 1\right], \tag{4.11}$$

where

$$H(x) = E\left[e^{X_T}\,|X_T \leq x\right] = \frac{e^{ax}}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\frac{\phi(u+i(a-1))}{a-iu}du, \tag{4.12}$$

a is a strictly positive real number and ϕ is the characteristic function of $X_T = \log(S_T/S_0)$. The detailed proof is given in Appendix E.

To implement the formula we only need to perform a FFT on ϕ_{X_T}.

4.1.3 Marginal value at risk (M-VaR)

When the VaR of a portfolio is defined on a return basis, as in (4.2), the marginal VaR (M-VaR henceforth) relative to the i-th asset is the change in VaR resulting from a marginal change in the relative position in instrument i. Hence, the M-VaR relative to component i is

$$\text{M-VaR}_i = \frac{\partial \text{VaR}_p}{\partial w_i}. \tag{4.13}$$

Relying only on approximation (3.1) and on the assumption that the first moment of all the relevant return distributions are finite, Tasche (1999) proves that M-VaR can be written

as

$$\text{M-VaR}_i = -E[X_i|R_p = -\text{VaR}]. \tag{4.14}$$

The intuition behind Equation (4.14) is immediate: when there is strong positive (negative) dependence between the asset return X_i and the portfolio return R_p, then large negative portfolio returns will be associated with large negative (positive) asset returns. Increasing the weight w_i of the asset will then lower (increase) the portfolio return, leading to an increase (decrease) in portfolio's VaR.

From Equation 4.14 we know that

$$\text{M-VaR} = -\int_{-\infty}^{\infty} x f_{X_i|R_p=-\text{VaR}}(x)dx, \tag{4.15}$$

where VaR can be computed as discussed in Section 4.1.1.

To compute the M-VaR in our exponential Lévy model we need the conditional distribution $f_{X_i|R_p}$, which we obtain exploiting the formula derived by Bartlett (1938), on the characteristic function of a conditional statistic.

We then recover the conditional probability density function $f_{X_i|R_p}$ via FFT transformation from the conditional characteristic function (see Bartlett (1938))

$$\phi_{X_i|R_p=-\text{VaR}}(u) = E[e^{iuX_i}|R_p = -\text{VaR}] = \frac{\int e^{iv\text{VaR}} M(u,v)dv}{\int e^{iv\text{VaR}} M(0,v)dv}, \tag{4.16}$$

where $M(u,v)$ denotes the bivariate characteristic function defined as

$$M(u,v) = E[e^{i(uX_i+vR_p)}]. \tag{4.17}$$

For a portfolio made of assets following the multivariate Lévy model (2.1) we can compute $M(u,v)$ from the expressions of the characteristic functions of the idiosyncratic factors, $Y^{(j)}$, $j = 1,\dots,n$, and the common factor Z, as

$$\begin{aligned} M(u,v) &= E[e^{iu(Y^{(i)}+a_iZ)+iv(\sum_{j=1}^n w_jY_j+Z\sum_{j=1}^n w_ja_j)}] \\ &= \left\{\prod_{j\neq i} E[e^{ivw_jY_j}]\right\} E[e^{i(u+vw_i)Y^{(i)}}]E[e^{iZ(ua_i+v\sum_{j=1}^n w_ja_j)}] \\ &= \left\{\prod_{j\neq i} \phi_{Y_j}(vw_j)\right\} \phi_{Y^{(i)}}(u+vw_i)\phi_Z(ua_i+v\sum_{j=1}^n w_ja_j). \end{aligned} \tag{4.18}$$

We then approximate the integrals in Equations (4.15) and (4.16) by trapezoidal integration, using the MATLAB® function `trapz`.

4.1.4 Marginal expected shortfall (M-ES)

A marginal measure similar to M-VaR can be defined with respect to the expected shortfall: the M-ES relative to the i-th asset is the change in ES resulting from a marginal change in the relative position in instrument i. Formally,

$$\text{M-ES}_i = \frac{\partial \text{ES}_p}{\partial w_i}. \tag{4.19}$$

Tasche (1999) proves a relation similar to (4.14), under the same assumptions:

$$\text{M-ES}_i = -E[X_i | R_p \leq -\text{VaR}]. \tag{4.20}$$

In the previous section we saw a way to compute the conditional probability density function $f_{X_i|R_p}$. Here we need the conditional distribution of $X_i|R_p \leq -\text{VaR}$, which we recover from the joint probability density function $f(X_i, R_p)$ as

$$f_{X_i|R_p \leq y}(x) = \frac{\int_{-\infty}^{y} f_{X_i,R_p}(x,z)dz}{F_{R_p}(y)}, \tag{4.21}$$

where

$$f_{X_i,R_p}(x,y) = f_{X_i|R_p=y}(x) f_{R_p}(y). \tag{4.22}$$

Therefore

$$\text{M-ES}_i = -E[X_i|R_p \leq -\text{VaR}] = -\frac{1}{\alpha}\int_{-\infty}^{\infty} x \int_{-\infty}^{-\text{VaR}} f_{X_i|R_p}(x|R_p = y) f_{R_p}(y) dy dx. \tag{4.23}$$

To compute the M-ES we approximate numerically the integrals in (4.23) by means of the MATLAB®function `trapz`.

4.2 Intra-horizon risk

In this section we show how to calculate the intra-horizon risk of a portfolio of assets following the multivariate Lévy model (2.1). In particular, we compute the intra-horizon value at risk (VaR-I hereafter), the intra-horizon tail conditional expectation (TCE-I) and the intra-horizon probability of breaching (PB-I).

Traditional risk measures, as VaR or TCE, focus on possible losses at the end of a predetermined time horizon; nevertheless, investors care about exposure to loss throughout the horizon, since they often have thresholds that cannot be breached for the investment to survive. Taking into account the magnitude of losses that can incur before the end of the horizon is of paramount importance, for example, for monitored asset managers, leveraged investors, borrowers required to maintain particular level of reserves as a condition of a loan agreement or securities lenders required to deposit collateral.

The emphasis on intra-horizon risk was first placed by Stulz (1996); Kritzman and Rich (2002) and Boudoukh et al. (2004) deal with intra-horizon risk assuming Gaussian distributed returns and considering a multi-year investment horizon, while Bakshi and Panayotov (2010) focus on the 10 days horizon relevant for regulatory purposes and consider three Lévy jump models for asset returns: the Merton's jump diffusion model (Merton (1976)), the finite moment log-stable model (FMLS) by Carr and Wu (2003) and the two sided pure-jump CGMY model of Carr et al. (2003).

As we will show, intra-horizon risk measures are based on the distribution of the minimum return, which is in turn strictly connected to the first passage probability of the return process to a lower barrier. While under the Gaussian assumption this distribution is analytically known (see Kritzman and Rich (2002)), in general it must be recovered numerically. In particular, while Bakshi and Panayotov (2010) resort to an approach based on partial integro-differential equations (PIDEs), we adopt an equivalent but simpler method, exploiting the Fourier Space Time-stepping (FST) algorithm introduced by Jackson et al. (2008) for option pricing purposes.

4.2.1 Definition of intra-horizon risk measures

In this section we formalize the definitions of the intra-horizon risk measures we are going to compute: VaR-I, TCE-I and PB-I. Let X_t, for $t \in [0, T]$, be the real-valued random process describing possible paths of an asset or portfolio return; the initial value is $X_0 = 0$, and we consider the random variable $X_T^{\min} := \min_{0 \le t \le T} X_t$.

VaR-I is the absolute value of a quantile of the distribution of the random variable $X_T^{\min}$. For example, the 10 days $(1-\alpha)$-VaR-I is the absolute value of the loss level exceeded at any point in time during the 10-day horizon with probability α.

Formally:

$$P(X_T^{\min} \leq -\text{VaR-I}) = \alpha, \tag{4.24}$$

where T = 10 days.

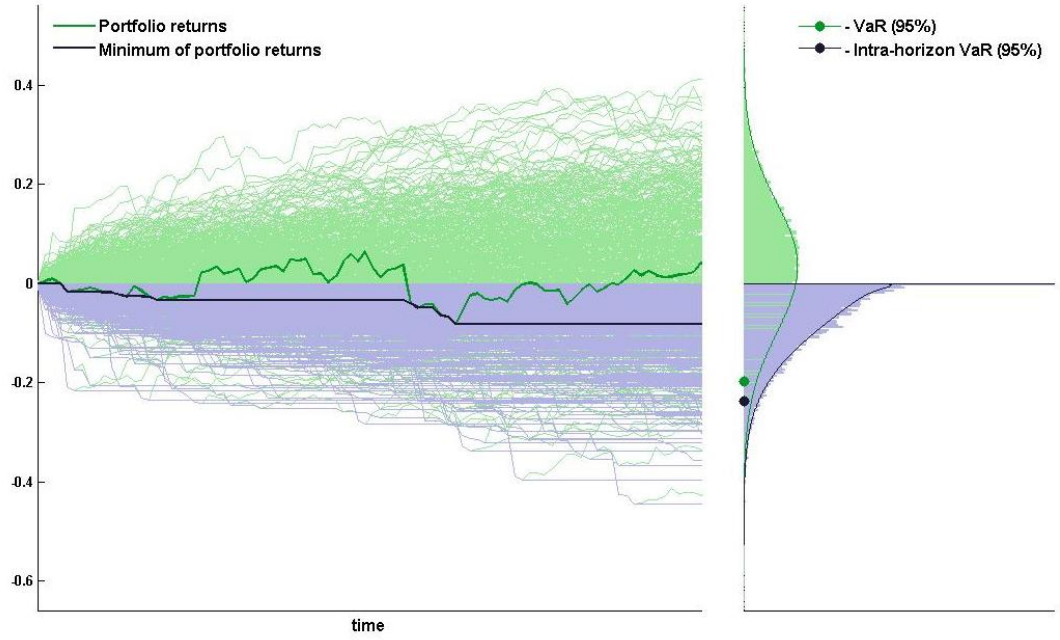

Figure 4.1: VaR and VaR-I: the idea.

Figure 4.1 illustrates the idea behind the definition of VaR-I; we plotted trajectories over a long horizon for the best understanding of the concept. We simulate 10000 trajectories for the returns of an equally-weighted portfolio where the assets follow a multivariate Lévy model with all NIG components. Some of these trajectories are plotted in light-green. The histogram of the return distribution at the end of the horizon is displayed in green and the estimated probability density function is superimposed. A dark-green dot identifies the 5%-quantile of this distribution, i.e. the opposite of the 95%-VaR. Then, for each trajectory X_t we build the corresponding trajectory of $X_t^{\min}$. For each light-green trajectory we report the corresponding path of the minimum in light-blue; we highlight one example of this construction with darker colors. The trajectories of minimum returns give rise to the distribution of $X_T^{\min}$, i.e. the distribution of the minimum at the horizon. A blue dot identifies the 5%-quantile of this distribution, i.e. the opposite of the 95%-VaR-I.

The idea is that, even when the return at the end of the horizon is positive, during the investment life the path of returns can reach high negative values, which investors may care about. In such cases the left tail of the minimum distribution better represents risk than the tail of return distribution itself.

Notice how the definition of VaR-I, given in expression (4.24), is closely related to the first-passage time $\mathcal{T}_y$ of the return process X_t to a lower level y:

$$\mathcal{T}_y := \inf\{t > 0 : X_t \leq y\} \quad \text{where } y < X_0; \tag{4.25}$$

if X_t never reaches the level y within the horizon, i.e. if $X_t > y \; \forall t \in [0, T]$, we have that $\mathcal{T}_y > T$. The two events $\{\mathcal{T}_y > T\}$ and $\left\{X_T^{\min} > y\right\}$ are therefore equivalent, and we can write

$$P(\mathcal{T}_y > T) = P(X_T^{\min} > y). \tag{4.26}$$

In particular, we indicate with PB-I$_{y,T}$ the probability of breaching a certain threshold y during the time horizon

$$\text{PB-I}_{y,T} = 1 - P(\mathcal{T}_y > T). \tag{4.27}$$

Then the α-quantile defining VaR-I can also be written as the level y such that

$$\text{PB-I}_{y,T} = \alpha. \tag{4.28}$$

Analogously to its traditional counterpart, the intra-horizon tail conditional expectation (TCE-I) is defined as the average of the larger-than-VaR-I losses that might occur within the horizon

$$\text{TCE-I} = E[-X_T^{\min} | X_T^{\min} \leq -\text{VaR-I}]. \tag{4.29}$$

4.2.2 Estimating the first-passage probability

To estimate the first passage probability or equivalently the distribution of the minimum of X, for a portfolio of assets following the multivariate Lévy model (2.1), we adopt the Fourier Space Time-stepping (FST) algorithm introduced by Jackson et al. (2008) for option pricing purposes. Our problem is indeed equivalent to finding the value of a down-and-out binary option, that is an option paying 1 if the underlying doesn't hit a certain

lower barrier within a given time period and 0 otherwise. However our computations are performed under the physical probability measure.

More precisely, let S_t be an exponential Lévy process describing the price process, then we can write $S_t = S_0 e^{X_t}$, where X_t is a Lévy process with Lévy exponent φ, representing the log-return over the period $[0, t]$. We want to compute the probability (4.26):

$$P(X_T^{\min} > y) = P(S_0 e^{X_T^{\min}} > S_0 e^y) = P(S_T^{\min} > \widetilde{y}) = E[1_{\{S_T^{\min} > \widetilde{y}\}}], \tag{4.30}$$

where we indicated the threshold for $S_T^{\min}$ by $\widetilde{y} = S_0 e^y$.

Without loss of generality we set $S_0 = 1$.

The quantity (4.30) can be seen as the value at time $t = 0$ of the down-and-out barrier version of a contract with payoff

$$\Psi(S_T) = 1_{\{S_T > \widetilde{y}\}}, \tag{4.31}$$

i.e.

$$\Psi_{DO}(S_T) = \begin{cases} \Psi(S_T) & \text{if } S_T^{\min} > \widetilde{y} \\ 0 & \text{otherwise} \end{cases} = 1_{\{S_T > \widetilde{y}\} \cap \{S_T^{\min} > \widetilde{y}\}} = 1_{\{S_T^{\min} > \widetilde{y}\}}, \tag{4.32}$$

where '*DO*' stands for down-and-out and the evaluation is made under the physical probability measure.

The FST algorithm allows us to compute the value of an option by means of a backward procedure based on the fact that the value function is a martingale[1] and that the process X_t has stationary and independent increments. In particular, we have

$$\begin{aligned} v(t_1, X(t_1)) &= E_{t_1}[v(t_2, X(t_2))] \\ &= \int_{-\infty}^{\infty} v(t_2, X_{t_1} + x) f_{X_{t_2} - X_{t_1}}(x) dx \\ &= \int_{-\infty}^{\infty} v(t_2, X_{t_1} + x) f_{X_{(t_2 - t_1)}}(x) dx, \end{aligned} \tag{4.33}$$

where $f_{X_t}(x)$ is the probability density function of the process X_t. Transforming (4.33) into the Fourier space leads to

$$\mathcal{F}[v](t_1, \omega) = \mathcal{F}[v](t_2, \omega) e^{\varphi(\omega)(t_2 - t_1)}, \tag{4.34}$$

[1] In the option pricing world the value function is a martingale by the risk neutral valuation formula; for us the value function is simply the conditional expectation of (4.31) and thus it is a martingale by construction.

where $\mathcal{F}$ is the Fourier transform (1.1) and the following fact are used: $\mathcal{F}[f_{X_t}](\omega) = e^{\varphi(\omega)}$, which follows by the definition of characteristic function (1.2) and by the Lévy-Khintchine representation (1.12); convolution in real space corresponds to multiplication in Fourier space (see Section 1.2).

Therefore

$$v(t_1, x) = \mathcal{F}^{-1}\left\{\mathcal{F}[v](t_2, \omega)e^{\varphi(\omega)(t_2-t_1)}\right\}(x). \tag{4.35}$$

To implement the algorithm, the time and space domain as well as the time and frequency domain are partitioned into a finite mesh of points and the discrete Fourier transform is computed through the FFT algorithm; starting from $t_m = T$, where the value function coincides with the final payoff ($v^T = v(T, x) = \Psi(S_0 e^x)$), at each backward time stepping the value v^{m-1} is recovered from v^m as

$$v^{m-1} = FFT^{-1}[FFT[v^m]e^{\varphi \Delta t}], \tag{4.36}$$

and the step is repeated until v^0 is reached.

To apply the algorithm to barrier options it is enough to continuously enforce the barrier boundary condition at each time step.

Summarizing, we start form $v^T = 1_{\{x>y\}}$ and we apply the FST algorithm for barrier options, whose step reads

$$v^{m-1} = H_{DO}(FFT^{-1}[FFT[v^m]e^{\varphi \Delta t}), \tag{4.37}$$

where $H_{DO}(v) = v1_{\{x>y\}}$.

For further details on the FST algorithm, refer to Jackson et al. (2008).

At this stage, through the FST algorithm, we can recover the value function $v(t, X(t))$; therefore, given a barrier y we are able to compute the probability of breaching the barrier within the interval $[0, T]$ (4.27) as

$$\text{PB-I}_{y,T} = 1 - P(X_T^{\min} > y | X(0) = 0) = 1 - v(0, 0). \tag{4.38}$$

Moreover, we can find the VaR-I (4.24) exploiting the following observation:

$$v(0, x) \equiv P(X_T^{\min} > y | X(0) = x) = P(X_T^{\min} > (y - x) | X(0) = 0), \tag{4.39}$$

which follows from the translation invariance property of Lévy processes[2]. In particular, to compute the $(1-\alpha)$-VaR-I we follow the procedure below:

- choose an arbitrary threshold y (say $y = 0$ for simplicity);
- compute the function $v(0, x)$ by means of the FST algorithm;
- spot the value x such that $v(0, x) = 1 - \alpha$;
- compute the VaR-I as VaR-I$= -(y - x)$.

Finally, we can numerically compute the TCE-I (4.29) after recovering the distribution of the minimum return from the value function, exploiting again the relation (4.39). In fact

$$F_{X_T^{\min}}(y) = 1 - P(X_T^{\min} > y) = 1 - P(X_T^{\min} > 0 | X(0) = -y) = 1 - v(0, -y). \tag{4.40}$$

Therefore we can find the probability density function of the minimum return approximating by finite differences the derivative of $F_{X_T^{\min}}(y)$, and then numerically evaluate the integral

$$\text{TCE-I} = -\frac{1}{\alpha} \int_{-\infty}^{-\text{VaR-I}} y f_{X_T^{\min}}(y) dy, \tag{4.41}$$

where $1 - \alpha$ is the VaR-I level (e.g. for the 99%-VaR-I, $\alpha = 0.01$).

4.3 Portfolio optimization with CVaR

This section handles portfolio optimization when the objective is to minimize CVaR, i.e. to reduce the risk of high losses, and asset returns follow the multivariate Lévy model (2.1). We follow the approach suggested by Rockafellar and Uryasev (2000) and applied to Lévy processes by Yu et al. (2009), who considered the case of a variance gamma copula model. Their method allows to optimize CVaR and calculate VaR at the same time.

The choice to optimize CVaR, instead of VaR, is due to the better properties of the former, in particular its convexity. VaR can be ill-behaved as a function of portfolio positions and can exhibit multiple local extrema (see Pflug (2000), Uryasev (2000)). However, since by construction the VaR at level α of a portfolio cannot exceed the respective CVaR, minimizing CVaR is closely related to minimizing VaR. Moreover, as showed by Rockafellar

[2] See Kolokoltsov (2011), page 118 ff.

and Uryasev (2000) and Uryasev (2000), CVaR can be efficiently optimized using linear programming techniques.

4.3.1 Framework

Let $w = (w_1, w_2, \ldots, w_n)'$ be the vector of portfolio weights and $X = (X_1, X_2, \ldots, X_n)'$ be sample asset returns modeled by the multivariate Lévy model (2.1). The loss associated to the decision w is defined as negative portfolio return: $f(w, X) = -w'X$. We consider an optimization problem where the objective function is CVaR, defined in (4.5) or equivalently, since we deal with continuous distribution, in (4.6)[3], which in the present notation reads

$$\text{CVaR}_\beta = E(f(w, X) | f(w, X) \geq \text{VaR}_\beta). \tag{4.42}$$

The optimization problem becomes

$$\begin{cases} \min_w & \text{CVaR}_\beta \\ \text{s.t.} & \sum_{i=1}^n w_i = 1 \\ & w_i \geq 0 \text{ for } i = 1, \ldots, n \\ & E[w'X] \geq r_0 \\ & X_i = Y_i + a_i Z, \text{ with } Y_i, Z \text{ indep. Lévy processes, for } i = 1, \ldots, n, \end{cases} \tag{4.43}$$

where the third constraint reflects the requirement that only portfolios that can be expected to return at least a given amount r_0 will be admitted.

4.3.2 CVaR optimization: general theory

Here we briefly review the theory developed by Rockafellar and Uryasev (2000), which allows to reduce the CVaR optimization problem (4.43) to a convex programming problem. The loss $f(w, X)$ is a random variable with a distribution induced by that of X. The assumption that X admits a probability density function $p(X)$ is made; however an analytical expression for $p(X)$ is not needed, it is enough to have an algorithm to generate samples from it.

[3] Here we assume that VaR and CVaR are defined on a return basis.

The probability of $f(w, X)$ not exceeding a threshold α is given by

$$\Psi(w, \alpha) = \mathbf{Pr}(f(w, X) \leq \alpha) = E[1_{f(w,X)\leq\alpha}] = \int_{f(w,X)\leq\alpha} p(X)dX. \tag{4.44}$$

As a function of α, for a fixed decision vector w, Ψ is the cumulative distribution function for the loss associated with w. Rockafellar and Uryasev (2000) assume that the function Ψ, which is non-decreasing with respect to α and continuous from the right, is everywhere continuous as a function of α.

VaR (4.2) and CVaR (4.42) at level β can then be written respectively as

$$\mathrm{VaR}_\beta(w) = \min\{\alpha \in \mathbb{R} : \Psi(w, \alpha) = \beta\} \tag{4.45}$$

$$\mathrm{CVaR}_\beta(w) = \frac{1}{1-\beta} \int_{f(w,X)\geq \mathrm{VaR}_\beta(w)} f(w, X)p(X)dX, \tag{4.46}$$

being $(1 - \beta)$ the probability of the loss exceeding VaR_β.

Rockafellar and Uryasev (2000) introduce an auxiliary function F_β, defined as

$$F_\beta(w, \alpha) = \alpha + \frac{1}{1-\beta} \int_{X\in\mathbb{R}^n} [f(w, X) - \alpha]^+ p(X)dX \tag{4.47}$$

for which they prove the following theorems.

Theorem 1. As a function of α, $F_\beta(w, \alpha)$ is convex and continuously differentiable. The CVaR_β of the loss associated with any w can be determined from the formula

$$\mathrm{CVaR}_\beta(w) = \min_{\alpha\in\mathbb{R}} F_\beta(w, \alpha). \tag{4.48}$$

In this formula the set consisting of the values of α for which the minimum is attained, namely

$$A_\beta(w) = \operatorname*{argmin}_{\alpha\in\mathbb{R}} F_\beta(w, \alpha) \tag{4.49}$$

is a nonempty, closed, bounded interval, and the VaR_β of the loss is given by

$$\mathrm{VaR}_\beta(w) = \text{left endpoint of } A_\beta(w). \tag{4.50}$$

In particular, one always has

$$\mathrm{VaR}_\beta(w) \in \operatorname*{argmin}_{\alpha\in\mathbb{R}} F_\beta(w, \alpha) \text{ and } \mathrm{CVaR}_\beta(w) = F_\beta(w, \mathrm{VaR}_\beta(w)). \tag{4.51}$$

Theorem 2. Minimizing the CVaR_β of the loss associated with w over all the admissible decisions w is equivalent to minimizing $F_\beta(w, \alpha)$ over all (w, α), in the sense that

$$\min_w \text{CVaR}_\beta(w) = \min_{(w,\alpha)} F_\beta(w, \alpha), \tag{4.52}$$

where moreover a pair (w^*, α^*) achieves the second minimum if and only if w achieves the first minimum and $\alpha \in A_\beta(w^*)$ (4.49). In particular, therefore, in circumstances where the interval $A_\beta(w^*)$ reduces to a single point (as is typical), the minimization of $F_\beta(w, \alpha)$ over (w, α) produces a pair (w^*, α^*), not necessarily unique, such that w^* minimizes CVaR_β and α^* gives the corresponding VaR_β.

Furthermore, $F_\beta(w, \alpha)$ is convex with respect to (w, α), and $\text{CVaR}_\beta(w)$ is convex with respect to w, when $f(w, X)$ is convex with respect to w, in which case, if the constraints are such that the set of feasible decisions is convex, the joint minimization is an instance of convex programming.

Since the function F_β is convex and continuously differentiable, as stated in Theorem 1, it can be easily minimized numerically. Moreover, expression (4.48) gives a way to compute CVaR without first calculating VaR.

The integral appearing in the definition of F_β can be approximated by sampling the distribution of X according to its probability density function $p(X)$. If the sampling generates a collection of vectors $X^{(1)}, X^{(2)}, \ldots, X^{(q)}$ (where $X^{(k)} = (X_1^{(k)}, \ldots, X_n^{(k)})$, for $k = 1, \ldots, q$), then the approximation reads

$$\tilde{F}_\beta(w, \alpha) = \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} [f(w, X^{(k)}) - \alpha]^+, \tag{4.53}$$

where $[t]^+ = t \cdot 1_{\{t>0\}}$. $\tilde{F}_\beta(w, \alpha)$ is a convex and piecewise linear function of α and, if the set of feasible decisions is convex, it is convex with respect to w as well. Although $\tilde{F}_\beta$ it is not differentiable with respect to α, it can readily be minimized, either by line search techniques or by representation in terms of an elementary linear programming problem, as we see in the application below.

4.3.3 CVaR optimization: multivariate Lévy model

In this section we show how to apply the general theory described in Section 4.3.2 to solve the portfolio optimization problem (4.43).

First, we observe that the assumptions required for Theorems 1 and 2 to hold are all fulfilled:

- The distribution of X admits a probability density function $p(X)$ which, although not always available in analytical form, can be recovered from the characteristic function (2.2). However, as long as we can simulate the Lévy processes Z and Y_i, for $i = 1, \ldots, n$, which compose the multivariate model (2.1), we can easily generate random samples from the distribution of X; hence we can implement the approach without first recovering $p(X)$.
- Under our multivariate Lévy model, X has a continuous distribution, therefore the cumulative distribution function for the loss associated with a given w (4.44) is itself continuous.
- The loss $f(w, X)$ is convex with respect to w (in fact, it is linear).
- The set of feasible portfolio is convex (in fact, polyhedral due to the linearity of all the constraints).

Therefore, to solve the allocation problem (4.43), we consider the function

$$F_\beta(w, \alpha) = \alpha + \frac{1}{1-\beta} \int_{X \in \mathbb{R}^n} [-w'X - \alpha]^+ p(X) dX, \tag{4.54}$$

whose minimization with respect to (w, α) reduces to a convex programming problem and is equivalent to CVaR_β minimization, as granted by Theorems 1 and 2.

More precisely we focus on the approximation given in equation (4.53), based on random return samples generated according to our model for X.

Introducing the auxiliary variables u_k, $k = 1, \dots, q$, problem (4.43) can be represented in term of the following linear programming problem

$$\begin{cases} \min_{\alpha, w, u_k} & \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k \\ \text{s.t.} & u_k \geq 0, \quad k = 1, \dots, q \\ & w' X^{(k)} + \alpha + u_k \geq 0, \quad k = 1, \dots, q \\ & \sum_{i=1}^{n} w_i = 1 \\ & w_i \geq 0 \text{ for } i = 1, \dots, n \\ & w' E[X] \geq r_0, \end{cases} \tag{4.55}$$

where $E[X]$ is the expected return according to model (2.1)[4] and $X^{(k)}$, $k = 1, \dots, q$, are random vectors sampled according to the distribution of X.

Summarizing, we need to go through the following steps:

- Step 1. Choice of the model: define all the Lévy processes involved in the multivariate model (2.1).
- Step 2. Estimation of model parameters (refer to Section 2.3).
- Step 3. Computation of $E[X] = E[Y] + \mathbf{a}' E[Z]$, given the parameters estimated in Step 2.
- Step 4. Scenario generation: given the parameters estimated in Step 2, simulate a large number q of samples for X. More precisely, simulate $Z^{(k)}$ and $Y_i^{(k)}, i = 1 \dots, n$, for $k = 1, \dots, q$, from the respective distributions; build each $X^{(k)}$ as $X_i^{(k)} = Y_i^{(k)} + \hat{a}_i Z^{(k)}$.
- Step 5. Solve the linear programming problem (4.55). This can be easily and quickly achieved by means of the MATLAB®function `linprog`.

4.4 Application

As an example we apply the procedure described in Section 4.3 to build a portfolio composed by the 20 stocks considered in Section 3.5, where the objective is to minimize 95%-CVaR over a weekly horizon.

[4] We still rely on approximation 3.1.

We consider the multivariate Lévy model (2.1) both in its 'all-NIG' and 'all-MJD' specifications, comparing the results with an 'all-Gaussian' (hence Gaussian) version of the model.

The estimation of parameters is performed by means of the three-step procedure presented in Section 2.3.1, on weekly data.

The expectation $E[X]$ is obtained as $E[X] = E[Y] + \mathbf{a}'E[Z]$ where each $E[Y_i]$ for $i = 1, \dots, n$ and $E[Z]$ is computed according to the first equation of formula (1.26), in the NIG case, or (1.34), in the MJD case.

Scenario generation is achieved as described in Step 4, simulating each $Z^{(k)}$ and each $Y_i^{(k)}, i = 1 \dots, n$, for $k = 1, \dots, 10000$ from NIG or MJD models, with the estimated parameters.

We set $r_0 = 0$ in the definition of problem (4.55), meaning that we exclude allocations giving a negative expected return, and we solve it using the MATLAB®function `linprog`.

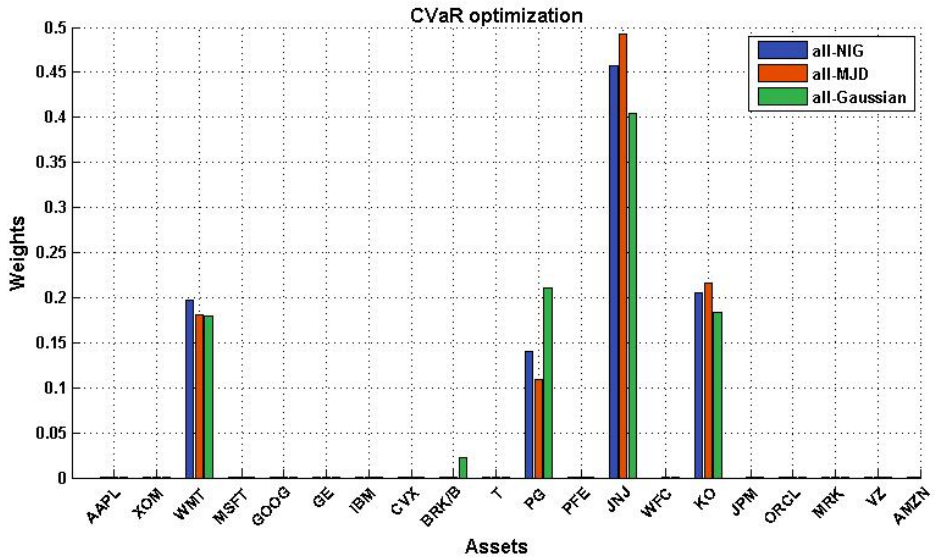

Figure 4.2: Optimal allocations under 'all-NIG', 'all-MJD' and 'all-Gaussian' specifications of our multivariate Lévy model, when objective is to minimize CVaR over a weekly horizon.

In Figure 4.2 we show the resulting allocations. We can notice that, even if the allocations are quite different, the 'all-NIG' and 'all-MJD' models agree on which stocks

should be overweighted in order to reach the goal of minimizing risk in terms of CVaR: Johnson&Johnson, Coca-Cola, Procter&Gamble and Walmart. The Gaussian model suggests a significant investment in Berkshire Hathaway as well, and a lower investment in Procter&Gamble.

In Figure 4.3 we plot the frontier of optimal 95%-CVaR as a function of the minimum required expected return r_0, obtained using the three models: 'all-NIG', 'all-MJD' and 'all-Gaussian'. The top plot, with $n = 10$, refers to the first ten assets in our dataset, while the bottom plot is built using the whole dataset, so that $n = 20$.
Knowing that the multivariate Lévy models in 'all-NIG' and 'all-MJD' specifications better describe our dataset than the Gaussian model, which appears clearly from Figures (3.1)-(3.5), we notice how relying on the Gaussian model leads to a significant underestimation of risk. On the other hand, the 'all-NIG' model appears to be the most conservative one.

Moreover, from the solution of the linear programming problem we can obtain 95%-VaR and 95%-CVaR corresponding to the optimal allocations w^*, as illustrated in Theorems 1 and 2: 95%-VaR is the optimum α and 95%-CVaR is the value of the objective function computed in correspondence of (w^*, α^*). In Table 4.1 we compare these values with those computed according to the definitions, as exposed in Sections 4.1.1 and 4.1.2 for the 'all-NIG' and 'all-MJD' models, analytically for the Gaussian model.
We notice that VaR and CVaR computed during the optimization are very close to those computed by definition.
Table 4.1 reports also the intra-horizon measures, VaR-I and TCE-I, computed as discussed in Section 4.2. While the VaR-I of the optimal portfolios computed under the three models are similar, the TCE-I computed under the 'all-NIG' model is considerably higher then the others. Of course intra horizon measures are greater than traditional measures, taking into account the possible losses that may incur within the (weekly) investment horizon.
Finally, we compute the marginal risk measures, M-VaR and M-ES, as described in

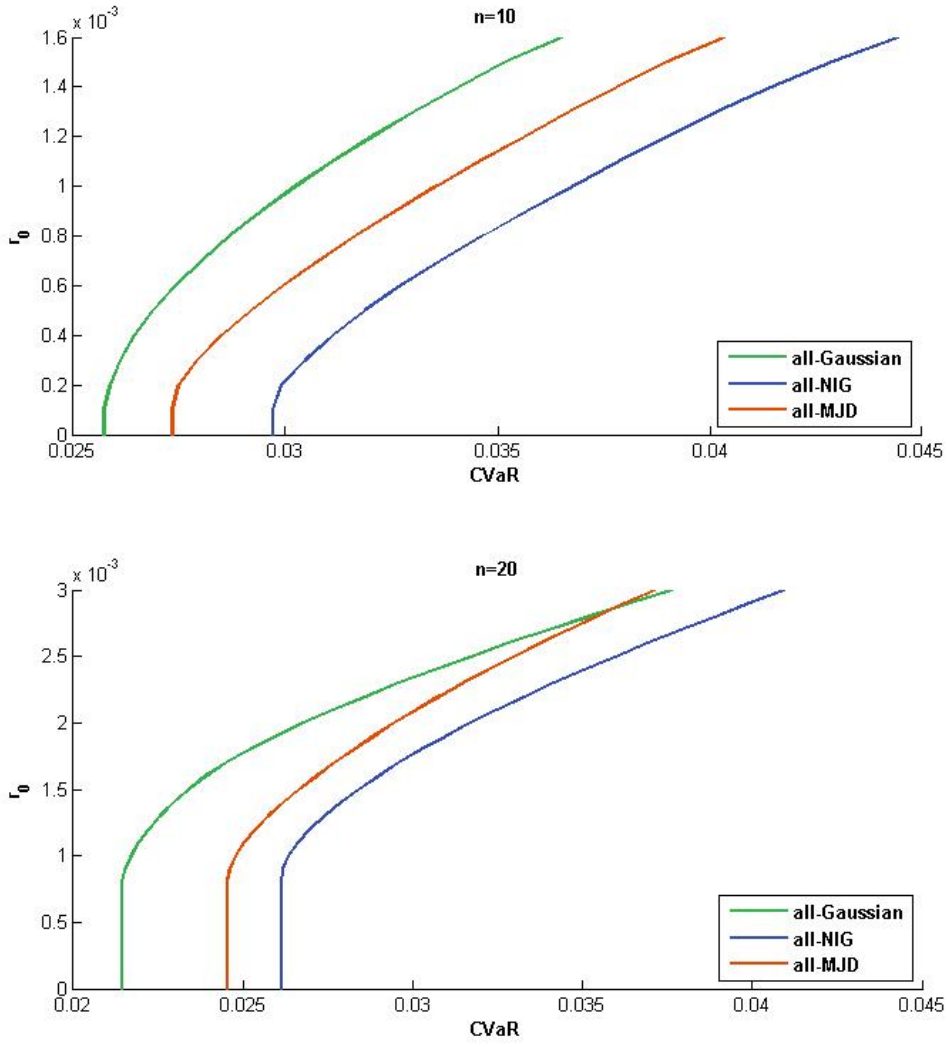

Figure 4.3: Minimum expected return - CVaR frontier: Gaussian, 'all-NIG' and 'all-MJD' models.

Sections 4.1.3 and 4.1.4 for the 'all-NIG' and 'all-MJD' models, analytically for the Gaussian model. The results are displayed in Table 4.2. First of all we observe that all the marginal measures are positive, meaning that alterations of the optimal weights would lead to an increase in VaR and, as expected, CVaR. Secondly, we notice that the stocks overweighted in the optimal portfolios are those displaying lower marginal risk.

	all-NIG	**all-MJD**	**Gaussian**
VaR (def)	0.0150	0.0158	0.0170
VaR (opt)	0.0151	0.0154	0.0171
CVaR (def)	0.0250	0.0243	0.0215
CVaR (opt)	0.0253	0.0245	0.0214
VaR-I	0.0390	0.0385	0.0392
CVaR-I	0.0541	0.0497	0.0484

Table 4.1: Traditional and intra-horizon risk measures relative to the portfolio optimizing CVaR.

	all-NIG		all-MJD		Gaussian	
	M-VaR	**M-ES**	**M-VaR**	**M-ES**	**M-VaR**	**M-ES**
APPL	0.0235	0.0453	0.0258	0.0447	0.0246	0.0313
XOM	0.0185	0.0360	0.0203	0.0355	0.0240	0.0305
WMT	0.0156	0.0244	0.0161	0.0237	0.0174	0.0218
MSFT	0.0192	0.0373	0.0211	0.0368	0.0241	0.0307
GOOG	0.0202	0.0378	0.0220	0.0374	0.0235	0.0295
GE	0.0238	0.0445	0.0259	0.0439	0.0239	0.0300
IBM	0.0152	0.0291	0.0167	0.0287	0.0179	0.0227
CVX	0.0210	0.0405	0.0230	0.0400	0.0272	0.0345
BRK/B	0.0141	0.0264	0.0154	0.0261	0.0177	0.0222
T	0.0188	0.0363	0.0206	0.0358	0.0246	0.0312
PG	0.0144	0.0247	0.0148	0.0239	0.0191	0.0241
PFE	0.0155	0.0295	0.0170	0.0291	0.0208	0.0263
JNJ	0.0150	0.0246	0.0160	0.0243	0.0182	0.0230
WFC	0.0379	0.0709	0.0414	0.0701	0.0371	0.0466
KO	0.0148	0.0250	0.0155	0.0244	0.0189	0.0241
JPM	0.0394	0.0734	0.0430	0.0726	0.0382	0.0479
ORCL	0.0208	0.0391	0.0227	0.0386	0.0257	0.0323
MRK	0.0163	0.0313	0.0178	0.0309	0.0235	0.0297
VZ	0.0141	0.0278	0.0156	0.0274	0.0181	0.0231
AMZN	0.0182	0.0382	0.0203	0.0376	0.0180	0.0238

Table 4.2: Marginal risk-measures relative to the portfolios optimizing CVaR respectively under the 'all-NIG', 'all-MJD' and Gaussian models.

Conclusion and future work

In conclusion, believing in the crucial importance of modeling as realistically as possible the joint distribution of asset returns in risk and portfolio management applications, we suggested and developed some tools to extend and simplify the usage of multivariate Lévy models in this area.

We chose as reference model the one introduced by Ballotta and Bonfiglioli (2014), due to its generality, flexibility and tractability.

We proposed a fast three-step estimation procedure whose complexity does not increase with the number of assets n, since it basically reduces to $(n+1)$ univariate estimations and a least squares estimation on the covariance matrix. Simulation studies revealed that this approach is as effective as a more computationally intensive overall maximum likelihood estimation of the whole set of parameters, as long as proper univariate estimation methods are used.

We then illustrated how the multivariate Lévy model can be employed for asset allocation purposes in a standard utility maximization framework, obtaining a closed form expression for the expected utility in the exponential case. For power utility function, and in general for any utility other than exponential, we suggest to resort to numerical integration. We then made an empirical test to check whether the allocations based on the multivariate Lévy model, in two particular specifications taken as representative examples, perform better than those obtained applying the most common allocation strategies and a new non-parametric approach. The results of our test say that the Lévy-based allocations often perform better and never perform worse than the others in terms of monetary utility, Sharpe ratio, turnover and maximum drawdown. Portfolio based on the non-parametric approach are characterized by low turnover, but their performance could be further improved with a more accurate choice of the bandwidth for the kernel estimation, point

which should be more carefully investigated.

Exploiting the results developed in Rockafellar and Uryasev (2000), we dealt also with a different approach to portfolio selection, describing how to solve, in our multivariate Lévy framework, optimization problems which have as objective the minimization of conditional value at risk while requiring a minimum expected return. Moreover we introduced new formulas and method to compute as efficiently as possible some downside risk measures. In particular we developed a simple formula for CVaR in monetary terms, following Le Courtois and Walter (2009), formulas for marginal-VaR and marginal-CVaR, using the conditional characteristic function by Bartlett (1938), and we illustrated how to apply the algorithm developed by Jackson et al. (2008), with the purpose of pricing barrier options, in order to compute intra-horizon risk measures.

Future study will compare the performance of allocations based on Ballotta and Bonfiglioli (2014) model with results obtained using other multivariate Lévy models, as those built by subordination.

Moreover, it would be interesting to estimate the model using non-parametric procedures, avoiding the need of a priori specifying the nature of the idiosyncratic and common components.

Appendix A

Proof of statements in Chapter 2

A.1 Proof of Proposition 1

The process $\mathbf{X}_t$ defined in (2.1) is a Lévy process in $\mathbb{R}^n$.

Each component of $\mathbf{X}_t$ is defined as

$$X_t^{(j)} = Y_t^{(j)} + a_j Z_t, \tag{A.1}$$

where $Y^{(j)}$ and Z are Lévy processes.

- $\mathbf{X}_t$ is adapted and càdlàg. It follows from Z and $Y^{(j)}$, for every $j = 1, \dots, n$, being Lévy processes, hence adapted and càdlàg.
- $X_0 = 0$ almost surely. $X_0^{(j)} = Y_0^{(j)} + a_j Z_0$, with Z_0 and $Y_0^{(j)}$, for every $j = 1, \dots, n$, almost surely null.
- Increments are independent from the past. Increments are defined as

  $$\mathbf{X}_t - \mathbf{X}_s = \left\{ X_t^{(j)} - X_s^{(j)} \right\}_{j=1,\dots,n} = \left\{ (Y_t^{(j)} - Y_s^{(j)}) + a_j (Z_t - Z_s) \right\}_{j=1,\dots,n}, \tag{A.2}$$

  with $0 \leq s < t < \infty$. Since the increments of Z and the increments of $Y^{(j)}$, for every $j = 1, \dots, n$, are independent of $\mathcal{F}_s$, the increments of $\mathbf{X}$ are independent of $\mathcal{F}_s$ as well.
- Increments are stationary. The distribution of the increments of $\mathbf{X}_t$ is determined by the distributions of the increments of Z and $\mathbf{Y}$. Since $(Z_t - Z_s)$ is distributed as Z_{t-s} and $(Y_t^{(j)} - Y_s^{(j)})$ is distributed as $Y_{t-s}^{(j)}$, for $j = 1, \dots, n$ and for any $0 \leq s < t < \infty$, then $(X_t^{(j)} - X_s^{(j)})$ is distributed as $X_{t-s}^{(j)}$, for $j = 1, \dots, n$ and for any $0 \leq s < t < \infty$.

- $\mathbf{X}$ is continuous in probability.

$$\lim_{s\to t} X_s^{(j)} = \lim_{s\to t}(Y_s^{(j)} + a_j Z_s) = Y_t^{(j)} + a_j Z_t = X_t^{(j)} \text{ for } \quad j = 1, \dots, n, \tag{A.3}$$

  where all the limits are taken in probability.

The characteristic function of $\mathbf{X}_t$ is (2.2).
Using the independence of the processes $Y^{(j)}$, $j = 1, \dots, n$, from Z and among themselves, we can write:

$$\begin{aligned}
\phi_{\mathbf{x}}(\mathbf{u}; t) &= E\left[e^{i\mathbf{u}\mathbf{X}_t}\right] = E\left[e^{i\sum_{j=1}^n u_j X_t^{(j)}}\right] = E\left[e^{i\left(\sum_{j=1}^n u_j Y_t^{(j)} + Z_t \sum_{j=1}^n u_j a_j\right)}\right] \\
&= E\left[e^{i\sum_{j=1}^n u_j Y_t^{(j)}}\right] E\left[e^{iZ_t \sum_{j=1}^n u_j a_j}\right] = E\left[\prod_{j=1}^n e^{iu_j Y_t^{(j)}}\right] E\left[e^{iZ_t \sum_{j=1}^n u_j a_j}\right] \\
&= \left(\prod_{j=1}^n E\left[e^{iu_j Y_t^{(j)}}\right]\right) E\left[e^{iZ_t \sum_{j=1}^n u_j a_j}\right] = \left(\prod_{j=1}^n \phi_{Y^{(j)}}(u_j; t)\right) \phi_Z\left(\sum_{j=1}^n a_j u_j; t\right).
\end{aligned} \tag{A.4}$$

A.2 Proof of Corollary 1

(i) It follows from direct differentiation of the cumulant generating function of the components of $\mathbf{X}_t$ and from the fact that Lévy processes have an infinitely divisible distribution.

In particular, the cumulant generating function of $X_t^{(j)}$, for $j = 1, \dots, n$, can be written as

$$\begin{aligned}
k_{X_t^{(j)}}(u) &= \log m_{X_t^{(j)}}(u) = \log E\left[e^{uX_t^{(j)}}\right] \\
&= \log\left(E\left[e^{uY_t^{(j)}}\right] E\left[e^{ua_j Z_t}\right]\right) \log\left(E\left[e^{uY_t^{(j)}}\right]\right) + \log\left(E\left[e^{ua_j Z_t}\right]\right) \\
&= \log m_{Y_t^{(j)}}(u) + \log m_{Z_t}(a_j u) = \log(m_{Y_1^{(j)}}(u))^t + \log(m_{Z_1}(a_j u))^t \\
&= t(\log m_{Y_1^{(j)}}(u) + \log m_{Z_1}(a_j u)),
\end{aligned} \tag{A.5}$$

where m_X indicates the moment generating function of the process X.

Differentiating (A.5) with respect to u and evaluating derivatives in correspondence of $u = 0$ easily gives expression (2.3) for the cumulants of the j-th component of $\mathbf{X}_t$, in terms of the cumulants of the j-th component of $\mathbf{Y}_t$, the cumulants of Z, and the j-th loading a_j.

(ii) It follows from the assumption that the process $\mathbf{Y}_t$ has independent components and it is independent of Z_t, and from the fact that Lévy processes have an infinitely divisible distribution. In fact, for $j \neq l$,

$$\begin{aligned} Cov(X_t^{(j)}, X_t^{(l)}) =& Cov(Y_t^{(j)} + a_j Z_t, Y_t^{(l)} + a_l Z_t) = Cov(Y_t^{(j)}, Y_t^{(l)}) + a_j a_l Cov(Z_t, Z_t) \\ &= a_j a_l Var(Z_t) = a_j a_l Var(Z_1) t. \end{aligned} \tag{A.6}$$

A.3 Proof of Corollary 2

The proof of Corollary 2 is given in full details in Ballotta and Bonfiglioli (2014).

Appendix B

EM algorithm for Merton's JD model

To fit the Merton's jump-diffusion model (1.27) we implemented the EM algorithm in the formulation proposed by Duncan et al. (2009). We report here the main ideas and the formulas needed to implement the procedure, while referring to the original work for further details.

The independent random vectors

$$\mathbf{C}_t = \begin{cases} (D_t, N_t) & (N_t = 0) \\ (D_t, N_t, J_{t,1}, \dots, J_{t,N_t}) & (N_t > 0) \end{cases} \qquad t = 1 \dots, T \tag{B.1}$$

completely determine the jump diffusion process X in (1.27). In the EM terminology, C_t are the *complete* data, while X_t are the *incomplete* data. The complete log-likelihood of $C_1, \dots, C_T$ is given by

$$\begin{aligned} \ln(L_c(\boldsymbol{\theta})) = &-\frac{1}{2} T \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (D_t - \mu)^2 - T \ln(\sqrt{2\pi}) \\ &- \frac{1}{2} \ln(\tau^2) \sum_{t=1}^{T} N_t - \frac{1}{2\tau^2} \sum_{t=1}^{T} \sum_{k=1}^{N_t} (J_{t,k} - \nu)^2 - \ln(\sqrt{2\pi}) \sum_{t=1}^{T} N_t \\ &- T\lambda + \ln\left(\lambda \sum_{t=1}^{T} N_t\right) - \sum_{t=1}^{T} \ln(N_t!), \end{aligned} \tag{B.2}$$

where $\boldsymbol{\theta} = (\mu, \sigma^2, \nu, \tau^2, \lambda)$ is the vector of parameters and we interpret $\sum_{k=1}^{N_t} (J_{t,k} - \nu)^2 = 0$ if $N_t = 0$. As we have seen in Section 1.4.3, starting from an initial guess $\boldsymbol{\theta}_0$, in the E-step we should compute the best (quadratic loss) predictor of $\ln(L_c(\boldsymbol{\theta}))$, i.e. the conditional expectation (1.39), where we condition on the available data $\mathbf{X}$. The M-step gives a new estimate $\boldsymbol{\theta}^{(1)}$, which maximizes the conditional expectation $Q(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$. Under certain

general conditions the sequence of estimates obtained in this way yields monotonically increasing values of the likelihood and converges to the ML estimator for the incomplete data $X_1, \ldots, X_T$.

The algorithm proposed by Duncan et al. (2009) is particularly efficient since they provide simple closed form solutions for the M-step. The values of μ, σ^2, ν, τ^2 and λ that maximize $Q(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$ are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0}(D_t | X_t) \tag{B.3}$$

$$\hat{\sigma^2} = \frac{1}{T} \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0} \left\{ (D_t - \hat{\mu})^2 X_t \right\} \tag{B.4}$$

$$\hat{\nu} = \frac{1}{T\hat{\lambda}} \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0}(N_t J_{t,k} | X_t) \tag{B.5}$$

$$\hat{\tau^2} = \frac{1}{T\hat{\lambda}} \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0} \left\{ N_t (J_{t,k} - \hat{\nu})^2 | X_t \right\}, \tag{B.6}$$

where

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} E_{\boldsymbol{\theta}_0}(N_t | X_t) \tag{B.7}$$

The formulas are made operational by evaluating the conditional expectations.

First, $\hat{\lambda}$ (B.7) and the two functions

$$a_t(\beta) = E_{\boldsymbol{\theta}_0} \left(\frac{1}{1 + N_t \beta^2} \middle| X_t \right) \tag{B.8}$$

$$b_t(\beta) = Var_{\boldsymbol{\theta}_0} \left(\frac{1}{1 + N_t \beta^2} \middle| X_t \right) \tag{B.9}$$

must be evaluated in $\beta_0 = \sigma_0 / \tau_0$, knowing the conditional probability

$$P_{\boldsymbol{\theta}_0}(N_t = k | X_t) = \frac{\varphi(X_t; \mu_0 + k\nu_0, \sigma_0^2 + k\tau_0^2)}{\sum_{k=0}^{\infty} \varphi(X_t; \mu_0 + k\nu_0, \sigma_0^2 + k\tau_0^2)} \quad k = 0, 1, \ldots \tag{B.10}$$

where φ is the normal probability density function. The estimates for μ (B.3) and σ^2 (B.4) can then be computed using

$$E_{\boldsymbol{\theta}_0}(D_t | X_t) = \mu_0 - \frac{\nu_0}{\beta_0^2} + a_t(\beta_0) \left(X_t - \mu_0 + \frac{\nu_0}{\beta_0^2} \right) \tag{B.11}$$

$$E_{\boldsymbol{\theta}_0} \left\{ (D_t - \hat{\mu})^2 | X_t \right\} = (E_{\boldsymbol{\theta}_0}(D_t | X_t) - \hat{\mu})^2 + \sigma_0^2 (1 - a_t(\beta_0)) + b_t(\beta_0) \left(X_t - \mu_0 + \frac{\nu_0}{\beta_0^2} \right)^2. \tag{B.12}$$

To find the estimates for ν (B.5) and τ^2 (B.6) the following quantities are needed:

$$c_t(\beta_0) = \beta_0^2 a_t(\beta_0)(1 - a_t(\beta_0)) - \beta_0^2 b_t(\beta_0) - \frac{(1 - a_t(\beta_0))^2}{E_{\boldsymbol{\theta}_0}(N_t|X_t)} \tag{B.13}$$

$$E_{\boldsymbol{\theta}_0}(N_t J_{t,k}|X_t) = (1 - a_t(\beta_0)) \left(X_t - \mu_0 + \frac{\nu_0}{\beta_0^2} \right) \tag{B.14}$$

$$\begin{aligned} E_{\boldsymbol{\theta}_0}\left\{ N_t (J_{t,k} - \hat{\nu})^2 | X_t \right\} &= \tau_0^2 (E_{\boldsymbol{\theta}_0}(N_t|X_t)) - 1 + a_t(\beta_0)) + c_t(\beta_0) \left(X_t - \mu_0 + \frac{\nu_0}{\beta_0^2} \right)^2 \\ &+ E_{\boldsymbol{\theta}_0}(N_t|X_t) \left(\hat{\nu} - \frac{E_{\boldsymbol{\theta}_0}(N_t J_{t,k}|X_t)}{E_{\boldsymbol{\theta}_0}(N_t|X_t)} \right)^2 . \end{aligned} \tag{B.15}$$

To initialize the algorithm, Duncan et al. (2009) suggest to fix the value of $\nu_0 = 0$, assuming symmetric returns, and of $\lambda_0 = 0.2$ or smaller, and then to recover μ_0, β_0 and σ_0^2 exploiting respectively the moment conditions

$$E(X_t) = \mu \tag{B.16}$$

$$\beta^2 = \sqrt{\frac{\hat{\gamma}}{3\lambda}} \left(1 - \lambda \sqrt{\frac{\hat{\gamma}}{3\lambda}} \right)^{-1} \tag{B.17}$$

$$\mathbf{Var}(X_t) = \sigma^2 (1 + \beta^2 \lambda), \tag{B.18}$$

where $\hat{\gamma}$ is the sample excess kurtosis of X_t.

Appendix C

CARA utility and normal returns

Here we prove that under the assumptions of exponential utility function (3.3) and normally distributed returns, the optimization of expression (3.16) leads to an exact one-step optimal solution for the allocation problem (3.2).

Given the assumptions above, the expected utility reads

$$EU(W) = \frac{1}{\sigma_p\sqrt{2\pi}} \int_{-\infty}^{\infty} -e^{-\lambda W} e^{-\frac{(W-\mu_p)^2}{2\sigma_p^2}} dW = \frac{1}{\sigma_p\sqrt{2\pi}} \int_{-\infty}^{\infty} -e^{-\left(\lambda W + \frac{(W-\mu_p)^2}{2\sigma_p^2}\right)} dW. \quad \text{(C.1)}$$

We can now rewrite the exponent grouping terms that depend on W and terms that do not depend on W. Note that

$$\lambda W + \frac{(W-\mu_p)^2}{2\sigma_p^2} = \frac{(W-\mu_p+\lambda\sigma_p^2)^2}{2\sigma_p^2} + \lambda\left(\mu_p - \frac{\lambda\sigma_p^2}{2}\right). \quad \text{(C.2)}$$

Substituting (C.2) in the expected utility expression (C.1) we have

$$EU(W) = -\frac{e^{-\lambda\left(\mu_p - \frac{\lambda\sigma_p^2}{2}\right)}}{\sigma_p\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(W-\mu_p+\lambda\sigma_p^2)^2}{2\sigma_p^2}} dW, \quad \text{(C.3)}$$

where

$$\frac{1}{\sigma_p\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(W-\mu_p+\lambda\sigma_p^2)^2}{2\sigma_p^2}} dW = 1, \quad \text{(C.4)}$$

being the integral of the probability density function of a normal random variable with mean $\mu' = \mu_p - \lambda\sigma_p^2$ and variance σ_p^2 over the entire support.

Therefore

$$EU(W) = -e^{-\lambda\left(\mu_p - \frac{\lambda\sigma_p^2}{2}\right)}, \quad \text{(C.5)}$$

and the investor objective becomes to maximize expression (3.16).

Appendix D

Testing the equality of Sharpe ratios

D.1 Pairwise Sharpe ratio test

Let SR_a and SR_b the Sharpe ratios for the returns $\{R_{at}\}_{t=1,\dots,T}$ and $\{R_{bt}\}_{t=1,\dots,T}$ of portfolios a and b. Opdyke (2007) obtain the asymptotic distribution of

$$\hat{SR}_{\text{diff}} = (\hat{SR}_a - \hat{SR}_b) - (SR_a - SR_b)$$

under the null hypothesis

$$H_0 : SR_a = SR_b. \tag{D.1}$$

In particular, under very general, 'real world' financial conditions, i.e. the assumptions of stationary and ergodic returns, the following holds true

$$\sqrt{T}(\hat{SR}_{\text{diff}}) \overset{a}{\sim} N\left(0, \text{Var}_{\text{diff}}\right) \tag{D.2}$$

where T is the number of observations and Var_{diff} is given by

$$\begin{aligned}\text{Var}_{\text{diff}} =& 1 + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] - SR_a\frac{\mu_{3a}}{\sigma_a^3} + 1 + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - SR_b\frac{\mu_{3b}}{\sigma_b^3} \\ & - 2\left[\rho_{a,b} + \frac{SR_aSR_b}{4}\left(\frac{\mu_{2a,2b}}{\sigma_a^2\sigma_b^2} - 1\right) - \frac{1}{2}SR_a\frac{\mu_{1b,2a}}{\sigma_b\sigma_a^2} - \frac{1}{2}SR_b\frac{\mu_{1a,2b}}{\sigma_a\sigma_b^2}\right].\end{aligned} \tag{D.3}$$

In equation (D.3),σ_a (σ_b), μ_{3a} (μ_{3b}), μ_{4a} (μ_{4b}) are respectively the standard deviation, skewness and kurtosis of the returns of portfolio a (b), $\rho_{a,b}$ indicates the correlation coefficient among the returns of portfolio a and b, and the joint central moments are defined as

$$\begin{aligned}\mu_{1a,2b} &= E[(R_a - \mu_a)(R_b - \mu_b)^2] \\ \mu_{2a,2b} &= E[(R_a - \mu_a)^2(R_b - \mu_b)^2].\end{aligned} \tag{D.4}$$

Minimum variance unbiased estimators of $\mu_{1a,2b}$, $\mu_{1b,2a}$ and $\mu_{2a,2b}$ are the respective h-statistics $h_{1a,2b}$, $h_{1b,2a}$ and $h_{2a,2b}$, where

$$\begin{aligned} h_{1,2} =& \frac{2s_{01}^2 s_{10} - Ts_{02}s_{10} - 2s_{01}s_{11} + T^2 s_{12}}{T(T-1)(T-2)} \\ h_{2,2} =& \frac{-3s_{01}^2 s_{10}^2 + Ts_{02}s_{10}^2 + 4Ts_{01}s_{10}s_{11} - 2(2T-3)s_{11}^2 - 2(T^2-2T+3)s_{10}s_{12}}{T(T-1)(T-2)(T-3)} + \\ & \frac{s_{01}^2 s_{20} - (2T-3)s_{02}s_{20} - 2(T^2-2T+3)s_{01}s_{21} + T(T^2-2T+3)s_{22}}{T(T-1)(T-2)(T-3)} \end{aligned} \tag{D.5}$$

and $s_{xy} = \sum_{t=1}^{T} R_{at}^x R_{bt}^y$.

Therefore, we reject the hypothesis $SR_a = SR_b$ with a $(1-\alpha)$ confidence level if

$$\frac{\sqrt{T}\left|\hat{SR}_a - \hat{SR}_b\right|}{\sqrt{\text{Var}_{\text{diff}}}} > z_{(1-\frac{\alpha}{2})} \tag{D.6}$$

The proof, based on the delta-method, is given in full detail in Opdyke (2007).

D.2 Multivariate Sharpe ratio test

Leung and Wong (2006) develop a multivariate Sharpe ratio statistic to test the hypothesis of the equality of multiple Sharpe ratios, working out the asymptotic distribution of the statistic and its properties, under the assumption of i.i.d. returns. Let $u = (SR_1, \ldots, SR_k)'$ be the vector of Sharpe ratios of k portfolios and $\hat{u}$ its sample counterpart.

The null hypothesis reads

$$H_0 : SR_1 = \ldots = SR_k. \tag{D.7}$$

We first need to define the $(k-1) \times k$ matrix

$$C = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & -1 \end{pmatrix} \tag{D.8}$$

so that the null hypothesis (D.7) can be equivalently written as

$$H_0 : Cu = 0. \tag{D.9}$$

The multivariate Sharpe ratio statistic reads

$$S = nk(C\hat{u})'(C\hat{\Omega}C')^{-1}(C\hat{u}). \tag{D.10}$$

By $\hat{\Omega}$ we indicate the estimate of

$$\Omega = \frac{1}{2}\begin{pmatrix} 2+SR_1^2 & 2\rho_{12}+SR_1SR_2\rho_{12}^2 & \dots & 2\rho_{1k}+SR_1SR_k\rho_{1k}^2 \\ 2\rho_{12}+SR_1SR_2\rho_{12}^2 & 2+SR_2^2 & \dots & 2\rho_{2k}+SR_2SR_k\rho_{2k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ 2\rho_{k1}+SR_kSR_1\rho_{k1}^2 & 2\rho_{k2}+SR_kSR_2\rho_{k2}^2 & \dots & 2+SR_k^2 \end{pmatrix} \tag{D.11}$$

such that the unknown Sharpe ratios and correlation coefficients are replaced by their sample estimates.

For the α level, we reject the null hypothesis (D.9) if

$$S > \frac{(T-1)(k-1)}{(T-k+1)}F_{k-1,T-k+1}(\alpha), \tag{D.12}$$

where $F_{k-1,T-k+1}(\alpha)$ is the upper $(100\alpha)^{th}$ percentile of an F-distribution with $k-1$ and $T-k+1$ degrees of freedom.

For more details refer to Leung and Wong (2006).

Appendix E

Formula for ES in monetary terms

In this appendix we prove formula (4.12), useful to compute the ES in monetary terms under exponential Levy models (4.10)-(4.11).

We have

$$H(x) = E\left[e^{X_T} \,|X_T \leq x\right] = \int_{-\infty}^{x} e^s f_{X_T}(s)ds. \tag{E.1}$$

Suppressing the subscript in f_{X_T} to simplify the notation and introducing a strictly positive number a, let us define

$$\zeta(x) = e^{-ax}H(x) = e^{-ax}\int_{-\infty}^{x} e^s f(s)ds. \tag{E.2}$$

The Fourier transform of (E.2) reads

$$\hat{\zeta}(u) = \int_{-\infty}^{\infty} e^{iux}\zeta(x)dx, \tag{E.3}$$

so that

$$\hat{\zeta}(u) = \int_{-\infty}^{\infty} e^{iux}\left(e^{-ax}\int_{-\infty}^{x} e^s f(s)ds\right)dx \tag{E.4}$$

and

$$\hat{\zeta}(u) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} e^{iux}e^{-ax}e^s f(s)dsdx. \tag{E.5}$$

Noting that

$$-\infty < s < x < \infty,$$

we can swap the integrals as follows

$$\hat{\zeta}(u) = \int_{-\infty}^{\infty}\int_{s}^{\infty} e^{iux}e^{-ax}e^s f(s)dxds. \tag{E.6}$$

Therefore

$$\begin{aligned}\hat{\zeta}(u) &= \int_{-\infty}^{\infty} e^s f(s) \left(\int_s^{\infty} e^{iux} e^{-ax} dx \right) ds = \int_{-\infty}^{\infty} e^s f(s) \left[\frac{e^{-(a-iu)x}}{-(a-iu)} \right]_s^{\infty} ds \\ &= \int_{-\infty}^{\infty} e^s f(s) \left(-\frac{e^{-(a-iu)s}}{-(a-iu)} \right) ds = \frac{1}{a-iu} \int_{-\infty}^{\infty} e^{is(i(a-1)+u)} f(s) ds \\ &= \frac{1}{a-iu} \phi_{X_T}(i(a-1)+u),\end{aligned} \tag{E.7}$$

where we used the fact that $|e^{-(a-iu)x}| = e^{-ax}$ tends to zero when x becomes infinite.

The function (E.2) being the inverse Fourier transform of $\hat{\zeta}(u)$, we get

$$\zeta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{\phi_{X_T}(i(a-1)+u)}{a-iu} du. \tag{E.8}$$

Hence, from (E.2), we obtain

$$H(x) = \frac{e^{ax}}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{\phi_{X_T}(i(a-1)+u)}{a-iu} du. \tag{E.9}$$

Bibliography

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

T. Ane and H. Geman. Order flow, transaction clock, and normality of asset returns. *Journal of Finance*, 55:2259–2284, 2000.

A. Ang and G. Bekaert. International asset allocation with regime shifts. *Review of Financial studies*, 15(4):1137–1187, 2002.

P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Thinking coherently. *Risk Magazine*, 10 (11):68–71, 1997.

L. Bachelier. Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Superieure*, 3(17):21–86, 1900.

G. Bakshi and G. Panayotov. First-passage probability, jump models, and intra-horizon risk. *Journal of Financial Economics*, 95:20–40, 2010. doi: 10.1016/j.jfineco.2009.01.003.

L. Ballotta and E. Bonfiglioli. Multivariate asset models using Lévy processes and applications. *The European Journal of Finance*, Forthcoming, 2014.

O. E Barndorff-Nielsen. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, 24(1):1–13, 1997.

O. E. Barndorff-Nielsen and A. M. Lindner. Some aspects of Lévy copulas. 2004.

O.E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2:41–68, 1998.

M. S. Bartlett. The characteristic function of a conditional statistic. *Journal of the London Mathematical Society*, 1(1):62–67, 1938.

M. Baxter. Lévy simple structural models. *International Journal of Theoretical and Applied Finance*, 10, 2007.

S. Benartzi and R. Thaler. Naive diversification strategies in defined contribution saving plans. *American Economic Review*, pages 79–98, 2001.

C. Bertini, S. Lozza Ortobelli, and A. Staino. Discrete time portfolio selection with Lévy processes. In *Intelligent Data Engineering and Automated Learning-IDEAL 2007*, pages 1032–1041. Springer, 2007.

J. Boudoukh, M. Richardson, R. Stanton, and R. Whitelaw. Maxvar: Long horizon value at risk in a mark-to-market environment. *Journal of Investment Management*, 2(3):1–6, 2004.

P. Carr and L. Wu. The finite moment log stable process and option pricing. *Journal of Finance*, pages 753–777, 2003.

P. Carr, H. Geman, D. H. Madan, and M. Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13:345–382, 2003.

M. Carrasco and J. P. Florens. Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16(06):797–834, December 2000.

G. Chacko and L.M. Viceira. Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, 116(1–2):259 – 292, 2003.

M. T. Cliff. GMM and MINZ program libraries for Matlab. 2003.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. 2001.

R. Cont and P. Tankov. *Financial modelling with jump processes*, volume 133. 2004.

E. A. Cornish and R. A. Fisher. Moments and cumulants in the specification of distributions. *Extrait de la Revue de l'Institute International de Statistique*, 4:1–14, 1937.

D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164, 1955.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.

A. P. Dempster, M. N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–22, 1977.

R. Duchin and H. Levy. Markowitz versus the talmudic portfolio diversification strategies. *Journal of Portfolio Management*, 35:71–74, 2009.

J. Duncan, J. Randal, and P. Thomson. Fitting jump diffusion processes using the EM algorithm. In *Contributed talk at the Australasian Meeting of the Econometric Society, Canberra, Australia*, 2009.

A. Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society*, 43(1):20–27, 1981.

R.A. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725, 1925.

S. Grossman and Z. Zhou. Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3:241–276, 1993.

A. Hitaj and L. Mercuri. Portfolio allocation using multivariate variance gamma. *Financial Markets and Portfolio Management*, 27(1):65–99, 2013.

P. Honoré. Pitfalls in estimating jump-diffusion models. *Available at SSRN 61998*, 1998.

G. Huberman and W. Jiang. Offering versus choice in 401(k) plans: Equity exposure and number of funds. *Journal of Finance*, 61(2):763–801, 04 2006.

G. Iaquinta, F. Lamantia, I. Massabò, and S. Ortobelli. Moment based approaches to value the risk of contingent claim portfolios. *Annals of Operations Research*, 165(1):97–121, 2009.

K. Jackson, S. Jaimungal, and V. Surkov. Fourier space time-stepping for option pricing with Lévy models. *Journal of Computational Finance*, 12(2):1–29, 2008.

G. J. Jiang and J. L. Knight. Estimation of continuous time processes via the empirical characteristic function. *Journal of Business and Economic Statistics*, 20:198–212, 2002.

E. Jondeau and M. Rockinger. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10): 1699–1737, 2003.

J. Kallsen and P. Tankov. Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572, 2006.

V. N. Kolokoltsov. *Markov processes, semigroups, and generators*, volume 38. Walter de Gruyter, 2011.

S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48:1086–1101, 2002.

M. Kritzman and D. Rich. The mismeasurement of risk. *Financial Analysts Journal*, pages 91–99, 2002.

O. Le Courtois and C. Walter. A study on value-at-risk and Lévy processes. 2009. URL http://ssrn.com/abstract=1598360.

P. Leoni and W. Schoutens. Multivariate smiling. 7(05), 2007.

P. L. Leung and W. K. Wong. On testing the equality of the multiple sharpe ratios, with application on the evaluation of ishares. *Available at SSRN 907270*, 2006.

E. Luciano and W. Schoutens. A multivariate jump-driven financial asset model. *Quantitative finance*, 6(5):385–402, 2006.

E. Luciano and P. Semeraro. Multivariate time changes for Lévy asset models: Characterization and calibration. *J. Computational Applied Mathematics*, pages 1937–1953, 2010a.

E. Luciano and P. Semeraro. A generalized normal mean-variance mixture for return processes in finance. *International Journal of Theoretical and Applied Finance*, 13(03): 415–440, 2010b.

D. Madan and E. Seneta. The variance gamma (vg) model for share market returns. *Journal of business*, pages 511–524, 1990.

D. Madan and J. Yen. Asset allocation for cara utility with multivariate Lévy returns. 2007.

D. Madan, P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2:79–105, 1998.

B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

L. Martellini and V. Ziemann. Improved estimates of higher-order comoments and implications for portfolio selection. *Review of Financial Studies*, 23:1467–1502, 2010. doi: 10.1093/rfs/hhp099.

R. C. Merton. Option pricing when underlying stocks are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

A. Meucci. *Risk and Asset Allocation*. Springer, 2005. URL http://symmys.com.

A. Meucci. Linear vs. compounded returns - common pitfalls in portfolio management. *GARP Risk Professional*, April:52–54, 2010. URL http://symmys.com/node/141.

A. Meucci. The Prayer: Ten-step checklist for advanced risk and portfolio management. *GARP Risk Professional*, April/June:54–60/55–59, 2011. URL http://symmys.com/node/63.

A. Meucci. *Advanced Risk and Portfolio Management*. Forthcoming, 2014.

I. Monroe. Processes that can be embedded in Brownian motion. *The Annals of Probability*, 6(1):42–56, 1978.

T. Moosbrucker. Pricing CDOs with correlated variance gamma distributions. Technical report, 2006.

S. Ng, T. Krishnan, and G.J. McLachlan. The EM algorithm. In J. E. Gentle, W. K. Härdle, and Y. Mori, editors, *Handbook of Computational Statistics*, Springer Handbooks of Computational Statistics, pages 139–172. Springer Berlin Heidelberg, 2012.

J. D. Opdyke. Comparing Sharpe ratios: So where are the p-values? *Journal of Asset Management*, 8(5):308–336, 2007.

M. Osborne. Brownian motion in the stock market. *Operations research*, 7(2):145–173, 1959.

A. J. Patton. Copula-based models for financial time series. *Handbook of Financial Time Series*, pages 767–785, 2009.

K. Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.

G. C. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, pages 272–281. Springer, 2000.

K. Prause. *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*. PhD thesis, PhD thesis, University of Freiburg, 1999.

S. J. Press. A compound events model for security prices. *Journal of business*, pages 317–335, 1967.

A. Ramponi. Computing quantiles in regime-switching jump-diffusions with application to optimal risk management: a Fourier transform approach. *arXiv preprint arXiv:1207.6759*, 2012.

R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

P. A. Samuelson. Rational theory of warrant pricing. *Industrial management review*, 6: 13–31, 1965.

K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999. ISBN 9780521553025.

W. Schoutens. *Lévy Processes in Finance*. Wiley., 2003.

P. Semeraro. A multivariate variance gamma model for financial applications. *Int. J. Theor. Appl. Finance*, 11(1):1–18, 2008.

W. F. Sharpe. A simplified model for portfolio analysis. *Management science*, 9(2):277–293, 1963.

K. J. Singleton. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102(1):111–141, 2001.

A. Staino, S. Ortobelli, and I. Massabò. A comparison among portfolio selection strategies with subordinated Lévy processes. *IJCSNS*, 7(7):224, 2007.

R. M. Stulz. Rethinking risk management. *Journal of applied corporate finance*, 9(3): 8–25, 1996.

P. Tankov. Simulation and option pricing in Lévy copula models. *Mathematical Modelling of Financial Derivatives, IMA Volumes in Mathematics and Applications, Springer*, 2006.

D. Tasche. Risk contributions and performance measurement. *Working Paper, Technische Universitaet Muenchen.*, 1999.

S. Uryasev. Conditional value-at-risk: Optimization algorithms and applications. In *Computational Intelligence for Financial Engineering, 2000.(CIFEr) Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on*, pages 49–57. IEEE, 2000.

M. Wallmeier and M. Diethelm. Multivariate downside risk: Normal versus variance gamma. *Journal of Futures Markets*, 32(5):431–458, 2012.

J. Wang. The multivariate variance gamma process and its applications in multi-asset option pricing. 2009.

L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference (Springer Texts in Statistics)*. Springer, 2003.

D. Wolff, W. Bessler, and H. Opfer. Multi-asset portfolio optimization and out-of-sample performance: An evaluation of Black-Litterman, mean variance and naïve diversification approaches. 2012.

J. Yu. Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123, 2004.

J. Yu, X. Yang, and S. Li. Portfolio optimization with CVaR under vg process. *Research in International Business and Finance*, 23(1):107–116, 2009.