Abstract

In this thesis, we try to provide a broad econometric analysis of a class of risk measures, distortion risk measures (DRM). With carefully selected functional form, the Value-at-Risk (VaR) and Tail-VaR (TVaR) are special cases of DRMs. Besides, the DRM also admits interpretation in the sense of non-expected utility type of preferences.

We first provide a unified statistical framework for the nonparametric estimators of the DRMs in a univariate case. The asymptotic properties of both the DRMs and their sensitivities with respect to the parameters representing risk aversion and/or pessimism are derived. Moreover, the relationships between the VaR and TVaR are also investigated in detail, which, we hope, can shed new lights on the way passing one risk measure to another. Then, the analysis of DRMs are extended to a multi-dimensional framework, where the DRM is computed for a portfolio consisting of many primitive assets. Analogous to the mean-variance frontier analysis, we study the efficient portfolio frontier when both objective and constraint are replaced by the DRMs. We call this the DRM-DRM framework. Under a nonparametric setting, we propose three asymptotic test statistics for evaluating the efficiency of a given portfolio. Finally, we discuss the criteria used for evaluating models used to forecast the VaRs. More precisely, we propose a criterion which takes into account the loss levels beyond the VaRs.
I would like to thank Christian Gourieroux, my supervisor, for his generous
guidance, many suggestions and constant support during this research. Without
him this work would never have come into existence. This thesis is constituted
by our joint works and I am deeply grateful for him allowing me use them in
my thesis.

I am also thankful to Jin-Chuan Duan and Angelo Melino for their detailed
comments and corrections on the earlier versions of this work. I also benefit
from personal talks with René Garcia and Linda Khalaf and comments from
the seminar participants in CREST, Carleton University and University of Mon-
treal.
# Table of Contents

Table of Contents \hspace{1cm} iv

List of Tables \hspace{1cm} vii

List of Figures \hspace{1cm} viii

Introduction \hspace{1cm} 1

0.1 Contribution of the thesis \hspace{1cm} 3

0.2 Outline \hspace{1cm} 5

1 Preference, Stochastic Dominance and Risk Measures \hspace{1cm} 6

1.1 Expected utility \hspace{1cm} 7

1.2 Dual theory \hspace{1cm} 8

1.3 Stochastic dominance \hspace{1cm} 10

1.3.1 Stochastic dominance and expected utility \hspace{1cm} 10

1.3.2 Dual stochastic dominance \hspace{1cm} 11

1.3.3 Second-order stochastic dominance \hspace{1cm} 12

1.4 Coherent risk measures \hspace{1cm} 13

1.5 Basel II and Solvency II \hspace{1cm} 15

2 Sensitivity Analysis of Distortion Risk Measures \hspace{1cm} 18

2.1 Introduction \hspace{1cm} 18

2.2 The distortion risk measures and their sensitivities \hspace{1cm} 19

2.2.1 Choquet expected utility and distortion risk measures \hspace{1cm} 19

2.2.2 Relationship between VaR and Tail-VaR \hspace{1cm} 25

2.2.3 Sensitivity of a distortion risk measure with respect to a distortion parameter \hspace{1cm} 28

2.3 Nonparametric estimation of functional distortion risk measures and their sensitivities \hspace{1cm} 31

2.3.1 Asymptotic distribution of the nonparametric quantile estimator \hspace{1cm} 31

2.3.2 Estimation of distortion risk measure \hspace{1cm} 32

2.3.3 Estimation of the sensitivity \hspace{1cm} 35

2.3.4 Examples \hspace{1cm} 36

2.3.5 Implied pessimism parameter \hspace{1cm} 38

2.4 Tail-VaR versus VaR \hspace{1cm} 39

2.4.1 Nonparametric estimator of $g$ \hspace{1cm} 39
List of Tables

1 Relationship between VaR and Tail-VaR . . . . . . . . . . . . . . . . . . . . . . 124
2 Sensitivity of the VaR and Tail-VaR . . . . . . . . . . . . . . . . . . . . . . . . 124
3 Asymptotic variance and relative accuracy of the estimated $VaR(p)$ . . . . . . 125
4 Asymptotic variance and relative accuracy of the estimated $TVaR(p)$ . . . . . . 126
5 Summary statistics of currency portfolio returns . . . . . . . . . . . . . . . . . 127
6 Summary statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
7 Autocorrelations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
8 Test results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 129
List of Figures

1. $\frac{TVaR(p)}{VaR(p)}$ as function of $p$. ......................... 130
2. $\frac{TVaR(p)}{VaR(p)}$ as function of $a$ when $X \sim Pareto(a, b)$. ......................... 131
3. $g(p)$ when $X \sim N(0, 1)$. ......................... 131
4. $\alpha$ as function of $a$ when $X \sim Pareto(a, b)$. ......................... 131
5. Plot of the weights $w(u, p)$ as a function of $u$ in the sensitivity of $PH(p)$. ......................... 132
6. Plot of the weights $w(u, p)$ as a function of $u$ in the sensitivity of $EX(p)$. ......................... 132
7. Variance and relative accuracy for $\tilde{VaR}_T(p)$. ......................... 134
8. Variance and relative accuracy for $TVaR_T(p)$. ......................... 136
9. Variance and relative accuracy for $\tilde{PH}_T(p)$. ......................... 138
10. Value-at-Risk of the currency portfolio. ......................... 139
11. Tail-VaR of the currency portfolio. ......................... 140
12. $\frac{TVaR(p)}{VaR(p)}$ of the equally weighted currency portfolio. ......................... 141
13. Estimated link $g(p)$ for the equally weighted currency portfolio. ......................... 142
14. Estimated link $g(p)$ for the currency portfolios and varying allocations. ......................... 143
15. Term structure of slope parameter $\alpha$ ......................... 144
16. Estimators of delta-TVaR and delta-PH ......................... 145
17. Finite sample variances of estimators of delta-TVaR and delta-PH ......................... 145
18. Finite sample distribution of estimators of $(a^1)^*$. ......................... 146
19. Finite sample distribution of estimators of standard Sharpe ratio ......................... 147
20. Joint Simulated Path of $(r_t, \sigma^2_t)$ ......................... 148
21. Conditional and Unconditional Quantiles ......................... 148
22. Joint Evolution of $\Psi(t, \hat{z}_t)$, $\Psi(t, z)$ and $c(t, z_t)$ ......................... 149
23. Disaggregate reserve level, function of $\rho$. ......................... 149
24. Value of weights as function of $(\rho, y_1)$ and modified cdf when $\rho = \pm 0.9$. ......................... 150
25. Disaggregate reserve level as function of $\rho$ with aggregate reserve constraint ......................... 150
26. Relative allocation as function of $\rho$ and $\sigma^2$. ......................... 151
Introduction

New rules for fixing the reserves needed to balance a risky investment have recently been introduced both in Banking industry (Basel II) and Insurance sector (Solvency II). The level of required capital depends on the type of investment, but also on the selected risk measure. It is recognized that a standard mean-variance approach is generally inappropriate for risk control and other risk measures have been considered. For instance, since 1996, the Basle Committee has proposed to use the Value-at-Risk (VaR), which is a quantile of the profit and loss (P&L) distribution. Since then, a great deal of effort has been devoted to the study of the applications of VaR in related literatures. These include determining capital reserves, portfolio management, hedging and so on [see Gourieroux and Jasiak (2005) for an overview]. For instance, the capital reserve at a given risk level can be determined by taking the sum of VaR and initial wealth. VaR has a few nice properties as a measure of risk. For example, i) VaR summarizes the risks as a number representing an extreme event with a certain confidence level and thus is easy to understand; ii) by directly measuring the tail of the distribution of the profit and loss of a portfolio, VaR is compatible with non-Gaussian distributional properties such as the fat tail commonly encountered in asset returns. However, focusing only on VaR for measuring risk can be misleading since it only takes into account one point of the distribution. This can lead to counterintuitive behavior. For instance, Boyle et al. (2005) show that under a VaR constraint, a trader has the incentive to hold a riskier portfolio which can be subject to huge losses with small probability. Basak and Shapiro (2001) study portfolio selection under a VaR constraint and find that an agent tends to invest more in the risky asset than she would in the absence of this constraint. Furthermore, the VaR is not convex in general, so that an investor may be
better off in some cases if she invests in the individual asset separately. Indeed, the VaR is convex only under additional restrictions on the conditional distribution of asset returns, for instance, when the returns are i.i.d. Gaussian or follow a Gaussian random walk with stochastic volatility [see Gourieroux et al. (2000)].

Better measures of risk are desired for robust risk management. These measures have to take into account not only the probability of a bad event, but also its magnitude. Artzner et al. (1999) follow a systematic approach and define “coherent risk measures”. A risk measure is coherent if it satisfies axioms such as monotonicity, invariance with respect to drift, homogeneity and subadditivity. Clearly, the VaR is not coherent and violates the subadditivity for most distributions.\(^1\) The appropriateness of these axioms is still a matter for debate; nevertheless, they build a standard for introducing new risk measures. The authors propose in particular to replace the standard VaR by the Tail-VaR, which takes into account not only the probability of loss, but also the magnitude of the loss, when a loss occurs. The application of Tail-VaR has gained increasing interest both in the academic literature and in industry. For instance, it is used to derive the efficient portfolio frontier [see Rockafellar and Uryasev (1999), (2000), Bassett et al. (2004) for example] as in the mean-variance framework [Markowitz (1952)], to calculate required capital [Manistre and Hancock (2005)], or to perform a sensitivity analysis of portfolio risk [Tasche (2002), Laurent (2003), Fermanian and Scaillet (2005)]. Tail-VaR has been recommended as the standard measure for calculating the capital requirement. For example, in 2002, the Life capital subcommittee of American Academy of Actuaries has suggested the use of Tail-VaR to set the risk-based capital requirement.

Both VaR and Tail-VaR are closely related to distortion risk measures considered in the insurance literature [Wang (1996), Wang and Young (1998)]. The distortion risk measure (DRM) is a special case of the so-called Choquet expected utility, that is, an expected utility calculated under a modified probability measure [Bassett et al. (2004)]. The distortion risk measure distorts the probability measure while specifying the utility as an identity function.

\(^1\)Recently, independent works by Ibragimov (2005) and Garcia et al. (2006) show that VaR may satisfy the subadditivity requirement if the tails of the marginal distributions are reasonably thin and equally asymmetric.
This special class has various alternative names, such as spectral risk measure [Acerbi (2002)], or pessimistic risk measure [Bassett et al. (2004)].

0.1 Contribution of the thesis

In this thesis, we provide a broad econometric analysis of the distortion risk measures and their applications in finance, which covers their estimations and asymptotic theory, application in portfolio choices and criteria in model selections.

We start with a unified statistical framework for analyzing the asymptotic properties of the nonparametric estimators of DRMs. This unified framework is based on the properties of empirical process and the analogy principle. Thus, all limiting processes can be written as stochastic integrals with respect to Brownian bridges. Closed form expressions for the asymptotic variances of the estimators of the VaR, TVaR and Proportional Hazard DRM are derived. Comprehensive risk analysis requires the joint consideration of not only different risk measures, but also risk measures at several levels. Indeed, many distortion risk measures can be characterized by parameters representing risk aversion and/or pessimism. For example, the loss probability associated with the VaR or Tail-VaR represents the risk level selected by the regulator or the investor. Thus, knowledge about the sensitivity of risk measures with respect to a slight modification of the risk level is useful for selecting an appropriate risk management strategy. This can be done by studying the partial derivative of the risk measures with respect to the risk level. We show that this sensitivity has a similar expression as the DRM itself. Thus, the statistical framework can be applied directly to the estimators of the sensitivities. We further study the relationship between two popular risk measures, VaR and TVaR. The VaR and TVaR are two most popular risk measures that have been proposed in recent literature of Finance and Insurance. A simple link between them allows us to easily pass one measure to the other, which simplifies the calculation of different risk measures and extends the interpretations. In this thesis, we study two types of relationship, the amplifying factor, which specifies their ratio as a function of the risk level, and the links of their associated risk levels. These are the
main elements that are usually considered by the regulator for adjusting the required capital. By considering a few examples, we show that different distributions imply different shape of the amplifying factor between the VaR and TVaR, and it is independent of risk level if and only, if it is a Pareto distribution. On the other hand, a common linear relationship of their risk levels often holds under various distributions. We characterize the distributions that imply this linear relationship and propose a simple test statistics of the linear relationship using the empirically estimated distribution function.

The nonparametric analysis of the DRMs is also extended to the multi-variate case. We consider the risk measures for a portfolio composed of many primitive assets. Thus, the DRMs are indexed by the allocation assigned to each primitive asset. We alternatively define the sensitivity of the DRMs as the first-order derivative of the DRM w.r.t. the allocation, which corresponds to the marginal risk contribution by a particular asset or a delta-DRM, say. We propose two nonparametric estimators of the delta-DRM: the first one is based on the Nadaraya-Watson kernel estimator and the other is a sample moment analogue. We derive the asymptotic properties of both estimators and explain how the efficiency can be improved by introducing a new estimator as a convex combination of these two. We further apply these estimators to the efficient portfolio analysis. Instead of the standard return-risk framework, we extend it to a DRM-DRM setup, as the expected return is also a special case of the DRM. The efficient portfolio allocations are estimated by solving the standard first-order conditions, which include delta-DRMs as crucial components. Thus, the properties of the estimated efficient portfolio allocations depend on the estimators of delta-DRMs used. In other words, the asymptotic properties of the efficient allocation estimators are fully determined by the properties of the delta-DRM estimators. We propose three asymptotic test statistics based on the asymptotic properties of the delta-DRMs. These test statistics are analogous to the standard Wald, LR and LM tests.

We finally study the criteria used for model validations. We emphasize on the VaR forecast and the models can be nonparametric or parametric, static or dynamic. This is motivated by the drawbacks of the current practice in selecting models, namely, counting the number of
times a particular sample exceeding the estimated VaRs. In a simple example, we show that two completely different models can yield similar number of exceedance. We propose a criterion which takes into account the values beyond VaR and is closely related to the objective function of a standard quantile regression. Therefore, two models can be compared by measuring the distances of their values of this criterion to the optimal one. Or simply, we can just select the model with the smallest criterion value. Rigorous statistical inference can be constructed based on a multidimensional inequality test, in which different measures of performance difference can be considered jointly. A statistics of this kind has been well studied in the literature and follows a weighted combination of $\chi^2$ distributions. In addition, an extension to dependent risk is considered in a two-dimensional case. In particular, the VaR estimation of one business line can be implemented by taking into account the risk information from the other business line and vice versa, with or without constraint on the aggregated level of reserve.

### 0.2 Outline

The rest of the thesis is structured as follows. In Chapter 1, I give a brief review of the preference theory, stochastic dominance and coherent risk measures. The regulation frameworks established in Banking industry (Basel I, II) and Insurance sector (Solvency II) are also discussed. Chapter 2 considers the nonparametric estimation of the DRMs in a univariate case, where the asymptotic distributions are derived. The relationships between the VaR and TVaR are also analyzed in detail. Chapter 3 focuses on the portfolio analysis using DRMs. The delta-DRM and the DRM-DRM framework are defined. Model validation criteria are discussed in Chapter 4. Lengthy proofs are provided in the end of each chapter accordingly. Figures and tables are listed at the end.
Chapter 1

Preference, Stochastic Dominance and Risk Measures

Since the distortion risk measure is closely related to preference type of decision makers, let us briefly review the main concepts of preference theories and related issues such as stochastic dominance. In Section 1.1, we recall the axioms associated with the expected utility theory. For convenience, we follow the notations of Yaari (1987). As paralleled to the expected utility theory, the dual theory [Yaari (1987)] also enjoys a similar set of axioms. The only modification takes place on the independence axiom. In Section 1.2, we explain the dual theory of choice introduced by Yaari (1987). Section 1.3 discusses how both preference theories are linked to partial ordering of risks. More specifically, the stochastic dominance is associated with expected utility theory and a dual stochastic dominance is associated with a dual theory of choice [Wang and Young (1998)]. The coherent risk measure is explained in Section 1.4 and the new standards imposed by Basel II and Solvency II are discussed in Section 1.5.
1.1 Expected utility

Let \( X \) be the set of all random variables on some given probability space, with non-negative values. For each \( x \in X \), define the cumulative distribution function (cdf) of \( x \) by \( F_x(t) = Pr(x \leq t) \). Here, we interpret the values of the random variables in \( X \) as payments, denominated in some given monetary unit. Thus, each \( x \in X \) is interpretable as a gamble or a lottery which a decision maker might consider holding. A preference relation \( \succeq \) is assumed to be defined on \( X \). Let the symbols \( \succ \) and \( \sim \) stand for strict preference and indifference, respectively. The axioms below define the expected utility theory.

**Axiom 1** (Neutrality:). For any \( x \) and \( y \) in \( X \), with respective cdfs \( F_x \) and \( F_y \). If \( F_x = F_y \), then \( x \sim y \).

This axiom implies an unambiguous preference manner among cdfs. In other words, for two cdfs \( F_1 \) and \( F_2 \), \( F_1 \) is said to be preferred to \( F_2 \) if and only, if there exists two elements, \( x \) and \( y \) such that \( F_x = F_1, F_y = F_2, \) and \( x \succeq y \).

**Axiom 2** (Complete weak order:). \( \succeq \) is reflexive, transitive, and connected.

**Axiom 3** (Continuity:). Let \( F, F', G, G' \) belong to the set of cdf functions; assume that \( F \succ F' \). Then, there exists an \( \varepsilon > 0 \) such that \( \| F - G \| < \varepsilon \) and \( \| F' - G' \| < \varepsilon \) imply \( G \succ G' \), where \( \| F \| = \int |F(t)| dt \) is the \( L_1 \)-norm.

**Axiom 4** (Monotonicity:). If \( F_x(t) \leq F_y(t) \) for all \( t \), then \( F_x \succeq F_y \).

**Axiom 5** (Independence:). If \( F, F' \) and \( G \) belong to the set of cdf functions and \( \lambda \) is a real number satisfying \( 0 \leq \lambda \leq 1 \), then \( F \succeq F' \) implies \( \lambda F + (1 - \lambda)G \succeq \lambda F' + (1 - \lambda)G \).

Then, we obtain the result that preferences are representable by expected utility comparisons.

---

1 We restrict \( X \) to be non-negative values in order to keep the definitions of this chapter consistent with those in Yaari (1987). This restriction is relaxed to include the entire real line in later chapters.
Theorem 1. A preference relation \( \succeq \) satisfies Axioms 1 – 5 if and only if there exists a continuous and nondecreasing real function \( u \), defined on the real line, such that, for all \( x \) and \( y \) belonging to \( X \),

\[
x \succeq y \iff Eu(x) \geq Eu(y).
\]

1.2 Dual theory

The dual theory of choice under risk was introduced by Yaari (1987) for two reasons: i) separating agent’s attitude towards wealth from that towards risk; ii) rationalizing the agent’s behavior treated as “paradoxical” under expected utility theory, e.g. Allais (1953) and Kahneman and Tversky (1979). As showed by Yaari, the Axioms 1 – 4 also apply to the dual theory of choice. However, the dual theory weakens the independence axiom in the sense that it is postulated for convex combinations, formed along the payment axis. For stating the independence axiom under the dual theory, it is convenient to consider quantile functions, i.e. the inverses of cdfs.

For \( p \in [0, 1] \), let \( Q(p) = F^{-1}(p) \) denote the quantile function associated with cdf \( F \). Then, we can define a harmonic convex combination of two cdfs, \( F_1 \) and \( F_2 \), as:

\[
\lambda F_1 \boxplus (1 - \lambda) F_2 = (\lambda F_1^{-1} + (1 - \lambda) F_2^{-1})^{-1}.
\]

A formal independence axiom is given below.

**Axiom 6** (Dual Independence;). If \( F, F' \) and \( G \) are cdf functions and \( \lambda \) is a real number satisfying \( 0 \leq \lambda \leq 1 \), then \( F \succeq F' \) implies \( \lambda F \boxplus (1 - \lambda) G \succeq \lambda F' \boxplus (1 - \lambda) G \).

A preference relationship dual to the expected utility theory is obtained and formally stated below.

**Theorem 2.** A preference relation \( \succeq \) satisfies Axioms 1 – 4 and 6 if and only if there exists a continuous and nondecreasing real function \( \nu \), defined on the unit interval, such that, for all \( x \) and \( y \) belonging to \( X \),

\[
x \succeq y \iff \int_{\mathbb{R}^+} \nu(1 - F_x(t))dt \geq \int_{\mathbb{R}^+} \nu(1 - F_y(t))dt.
\]
Let $x$ belong to $X$, with cdf $F_x$, and let $V(x)$ be defined by

$$V(x) = \int_{\mathbb{R}^+} \nu(1 - F_x(t))dt,$$

(1.2.1)

where $\nu$ is a continuous and nondecreasing function. The assumption of the dual theory is that agents will choose among random variables so as to maximize a function $V$. This is in analogy (and in contrast) with the assumption of expected utility theory, where an agent chooses among random variables so as to maximize the function $U$, given by

$$U(x) = Eu(x) = \int_{\mathbb{R}} u(t)dF_x(t).$$

(1.2.2)

The utility $V$ of the dual theory has two noteworthy properties: First, $V$ assigns to each random variable its certainty equivalent. In other words, if $x$ belongs to $X$, then $V(x)$ is equal to that sum of money which, when received with certainty, is considered by the agent equally as good as $x$. The second important property of $V$ is linearity in payments: When a random variable is subjected to some fixed positive affine transformation, the corresponding value of $V$ undergoes the same transformation. The following propositions provide a precise statement of these properties.

**Proposition 1** (Yaari (1987)). Under Axioms 1 – 4 and 6, the relationship

$$x \sim [V(x); 1]$$

(1.2.3)

holds for every $x \in X$, where $[t; p]$ denotes a lottery paying $t$ with probability $p$.

**Proposition 2** (Yaari (1987)). Let $x$ belong to $X$, let $a$ and $b$ be two real numbers, with $a > 0$, and $0 \leq ax(s) + b \leq 1$. Then, $V(ax + b) = aV(x) + b$.

That is, in dual theory, the preference is linear in wealth while having risk attitude preserved. The representation of preference is simply passed from the expected utility theory to the dual theory by replacing the independence axiom of the former with a weaker version. It is proved in
Yaari (1987), Proposition 3, that it is equivalent to imposing an ordinary convex combination of the wealth themselves rather than of the distributions.

Yaari (1987) showed that a paradoxical behavior under expected utility theory, such as those described in Allais (1953) and Kahneman and Tversky (1979), can be rationalized by the dual theory. This is due to the role exchange between the wealth and probability. Certainly, dual theory also produce paradoxical choices of agent which may be consistent with decisions under expected utility theory.

Moreover, both the expected utility and the dual theory are just special cases of a more generalized specification, called Choquet expected utility as initiated by Schmeidler (1989). In particular, a Choquet expected utility is an expected utility under modified probability measure. These relationships are important for us to link the distortion risk measures to the theories of preference and will be discussed in more detail in Chapter 2.

1.3 Stochastic dominance

In this section, we first review the concepts of stochastic dominance and its relationship with expected utility theory. Then an alternative partial ordering of risks is explained. It is defined analogously to the stochastic dominance and can be considered dual to stochastic dominance orderings [see e.g. Wang and Young (1998)]. Finally, we consider the special case of stochastic dominance at order 2.

1.3.1 Stochastic dominance and expected utility

The nth order stochastic dominance relation, denoted by $\succeq_n$, is defined by means of repeated integration of cdfs.

**Definition 1.** For $x$ and $y$ belong to $X$, with cdfs $F_x$ and $F_y$, we say $x \succeq_n y$ iff

$$
\int_0^t \int_0^{t_{n-1}} \cdots \int_0^{t_1} F_x(s) ds dt_1 \cdots dt_{n-1} \leq \int_0^t \int_0^{t_{n-1}} \cdots \int_0^{t_1} F_y(s) ds dt_1 \cdots dt_{n-1}, \text{ for all } t.
$$
If \( n = 1 \) for instance, \( y \) is first-order stochastically dominated by \( x \), iff \( F_y \geq F_x \). Therefore, \( n = 1 \) corresponds to a case where the payment of lottery \( x \) is everywhere better than lottery \( y \). Similarly, \( n = 2 \) corresponds to a case where \( y \) is more likely to have bad outcomes than \( x \), and so on.

The different stochastic dominance orderings can be expressed by means of utility functions.

**Proposition 3.** Letting \( U_n \) be the class of utility functions \( u \) such that \( u \) is at least \( n - 1 \) times differentiable, \( (-1)^{k+1}u^{(k)} \geq 0, k = 1, 2, \ldots, n - 1, \) and \( (-1)^{n-1}u^{(n-1)} \) nondecreasing, \( x \preceq_n y \) iff \( E[u(x)] \geq E[u(y)] \) for all \( u \in U_n \).

1. For \( n = 1 \), \( U_1 \) is the class of increasing utility functions; the agents prefer more wealth to less, that is the so called non-satiation property.
2. For \( n = 2 \), \( U_2 \) is the class of increasing, concave utility functions; it corresponds to the non-satiated agents who are also risk-averse.
3. For \( n = 3 \), \( U_3 \) is the class of utility functions of non-satiated, risk-averse agents with decreasing absolute risk aversion.

### 1.3.2 Dual stochastic dominance

A similar partial ordering of risks can be constructed by exchanging the roles of probability and wealth. As we will see in this section, this definition of risk ordering is linked with Yaari’s dual theory of choices (called dual stochastic dominance in Wang and Young (1998)) and denoted by \( \succ^*_n \). Let \( x \) and \( y \) belong to \( X \) with survivor functions \( S_x = 1 - F_x \) and \( S_y = 1 - F_y \), respectively. The dual stochastic dominance is defined as follows.

**Definition 2.** \( x \succ^*_n y \) if

\[
\int_0^p \int_0^{p_{n-1}} \cdots \int_0^{p_1} S_x^{-1}(q)dqdp_1 \cdots dp_{n-1} \geq \int_0^p \int_0^{p_{n-1}} \cdots \int_0^{p_1} S_y^{-1}(q)dqdp_1 \cdots dp_{n-1}, \text{ for all } p \in [0, 1].
\]
The partial ordering of risks defined above applies repeated integrations on the wealth with respect to probability. It admits an economic interpretation under the dual theory by classifying the function $\nu$.

**Proposition 4** (Wang and Young (1998), Definition 4.2). Letting $V_n$ be the class of function $\nu$ which are at least $n - 1$ times differentiable, and satisfy $\nu^{(k)} \geq 0, k = 1, 2, \ldots, n$. Then, $x \succeq_n^* y$ iff $V(x) \geq V(y)$ for all $\nu \in V_n$.

1. If $n = 1$, $V_1$ is the class of nondecreasing function, $\nu$.

2. If $n = 2$, $V_2$ is the class of nondecreasing, convex function, $\nu$. The agents assign relative larger weights to bad outcomes than to good outcomes.

3. If $n = 3$, $V_3$ is the class of functions, $\nu$, which assign increasingly higher weights as the outcomes get worse. Thus, an agent with $\nu$ in $V_3$ is even more pessimistic.

### 1.3.3 Second-order stochastic dominance

Most applications in Finance and Economics assume investors, who are either solely profit maximizers, or risk averse. Thus, risk orderings are usually based on stochastic dominance theory up to the second order. It can be shown that both stochastic dominance theories discussed above lead to equivalent ordering of risks up to the order two [see e.g. Yaari (1987); Wang and Young (1998)]. This provides the basis for applying dual theory to risk measurements.

The following proposition summarizes the relationship between the expected utility theory and the stochastic dominance of order one and two.

**Proposition 5.** i) $Eu(x) \geq Eu(y)$ for any increasing utility function iff $F_x^\prime(\tau) \leq F_y^\prime(\tau)$ for all $\tau$.

ii) $Eu(x) \geq Eu(y)$ for any increasing, concave utility function iff

$$\int_0^\tau (1 - F_x(t))dt \geq \int_0^\tau (1 - F_y(t))dt, \text{ for all } \tau. \quad (1.3.1)$$

---

2In Wang and Young (1998), the function $\nu$ is called distortion function since in the definition of dual preference, the probability measure is modified (or distorted) before taking expectation.
Thus, letting $\succeq$ be a preference relation on $X$, as before, we say that $\succeq$ is risk averse if $y = x + \text{noise}$ implies $x \succeq y$. Yaari (1987) shows that, under the dual theory, this is equivalent to say that $\succeq$ is risk averse if and only if the function $\nu$ representing $\succeq$ is convex.

The fact that risk aversion is characterized in the dual theory by the convexity of $\nu$ has a useful interpretation when $\nu$ is differentiable. Indeed, by integrating by parts, we obtain

$$V(x) = \int_{\mathbb{R}^+} t\nu'(1 - F_x(t))dF_x(t).$$

Since $\int \nu'(1 - F_x(t))dF_x(t) = 1$, $\{\nu'(1 - F_x(t))\}$ is a system of nonnegative weights summing to 1. Thus, $V(x)$ is the mean of $x$ under a modified probability, which assigns a weight $\nu'(1 - F_x(t))$ to payment level $t$. If $\nu$ is convex, then $\nu'$ is nondecreasing; i.e., those values of $t$ for which $F_x(t)$ is large receive relatively low weights and those values of $t$ for which $F_x(t)$ is small receive relatively high weights. Thus, agent with preference $V(x)$ behaves pessimistically, as though bad outcomes are more likely than they really are and good outcomes are less likely than they really are. By analogy, risk loving or risk neutrality can be defined by picking concave $\nu$ or identity function.

### 1.4 Coherent risk measures

A risk measure is a function, say $R(\cdot)$, that transforms the risks into a real number, which can correspond to the extra cost necessary for keeping the investment position under acceptable risk (a position with an acceptable future net worth). This acceptable risk is often decided by a supervisor such as:

1. a regulator who takes into account the unfavorable states when allowing a risky position that may draw on the resources of the government – for example as a guarantor of last resort;

2. an exchange’s clearing firm, which has to make good on the promises to all parties of transactions being securely completed;
3. an investment manager who knows that his firm has basically given to its traders an exit option in which the strike “price” consists in being fired in the event of big trading losses on one’s position. 

Requirements on the set of acceptable risks rationalize the selection of the class of risk measures. In line with this acceptable risk arguments, Artzner et al. (1999) developed the concept of coherent risk measure, which provides a standard for introducing new measure of risks. A coherent risk measure satisfies the following four axioms.

**Axiom C. 1** (Translation invariance). *For all x belonging to X and all real number c, we have R(x + c) = R(x) − c.*

Thus, if the future payoff of the portfolio increases (decreases) by a fixed amount c in every possible states, the risk measure should diminish (rise) by the same amount.

**Axiom C. 2** (Positive homogeneity). *For all x belonging to X and all nonnegative number c, we have R(cx) = cR(x).*

Axiom C.2 simply states that increase the position size of a portfolio will raise its risk proportionally. Thus, it reflects the possible situation where no netting or diversification occurs. In particular, a government or an exchange does not prevent many firms or investors from all taking the same position.

**Axiom C. 3** (Monotonicity). *For all x and y belonging to X, with x ≻_1 y, we have R(x) ≤ R(y).*

An investment with less risk, according to the investors opinion, should be confirmed by its measure of risk. Apparently, Axiom C.3 links risk measures to an investor’s risk attitude toward risk. And thus, it represents a subjective perspective of risk and associates to the an agent’s preference. Indeed, although the monotonicity axiom is initiated with an ordering of risk in the sense of first order stochastic dominance [see Artzner et al. (1999)], it can be generalized to stochastic dominance of any order.
Axiom C. 4 (Subadditivity). For all $x$ and $y$ belonging to $X$, we have $R(x+y) \leq R(x) + R(y)$.

Axiom C.4 means that a merger does not create extra risk. One example that is consistent with this logic is that, if an individual wishes to take the risk $x + y$, opening two accounts separately will not help him save margin requirement of an exchange.

Artzner et al. (1999) defines the coherent risk measure as below.

Definition 3 (Coherence). A risk measure is coherent iff it satisfies Axioms C.1 – C.4.

While the coherent risk measure provides a standard for constructing meaningful measure of risks, it is not sufficiently restrictive to specify a unique risk measure. Instead, it characterizes a large class of risk measures. For instance, a large group of distortion risk measures as defined in the next chapter satisfy the coherency requirements. However, the choice of precisely which measure to use should presumably be made on the basis of additional economic considerations and subjective preference specifications.

1.5 Basel II and Solvency II

One of the main applications of risk measures is the determination of the minimum levels of capital that banks or insurance companies must hold to hedge the extreme losses due to dramatic downturn of the economy or caused by force majeure such as earthquake, flood, and war. The adequacy of these required capitals is supervised by internationally established regulators, such as the BIS (the Bank for International Settlements) for banks and the CEA (the European Insurance Federation) for insurance sector. In this section, let us recall the development path of Basel committee and the newly issued Basel II. Since Solvency II is established on a similar spirit and purpose, we only mention it for comparison and contrast reasons. Indeed, the main body of this thesis was largely motivated by the challenge to the practice induced by both Basel II and Solvency II.

As banks were becoming increasingly international in their operations, it was decided in the late 1980s that an international regime was necessary to ensure that banks had adequate capital
to ensure their soundness and thereby protect the global financial system and their depositors. The BIS, based in Basel in Switzerland, was charged with establishing a framework for setting a minimum level of capital each bank should have to hold, which is determined with regard to the riskness of banks’ assets.

In the first Accord of Basel Committee (Basel I), this minimum level of capital was set to be a proportion of the value of banks’ risky exposures. Specifically, each asset on the balance sheet of a bank was given a weighting between 0% and 100%, where 0% represented the safest assets such as sovereign bonds and 100% the riskiest exposures such as corporate debt and unsecured personal loans. Thus, a risk weighted asset value is computed and acts as the base on which the minimum level of capital is determined.

Apparently, Basel I simply sets the required capital on a fixed level and is not sophisticated enough to prevent banks from taking too much risk. By the late 1990s, banks had become much more capable of finding ways to reduce required capitals without lowering their risk exposures (so called regulatory capital arbitrage). Thus, a new capital standard was demanded and work began on Basel II. Basel II came into effect in the European Union on 1 January, 2007 under the Capital Requirements Directive (CRD) and all lenders covered by the CRD will have to implement it from the beginning of 2008.

Two main modifications of Basel II are noteworthy: i) flexible model selection for required capital calculation; ii) regulator’s adjustment to the regulatory capital requirement.

i) According to Pillar 1 of Basel II, banks are encouraged to build their own models to decide the level of reserved capital necessary for hedging extreme risk. This allows a bank not only the freedom to choose the model specifications, but also the measures of risks. In this respect, the coherency requirements discussed in the last section can stand as criterion. Value-at-Risk and Expected Shortfall are two risk measures currently suggested by the regulator and their relationships are investigated in more detail in Chapter 2. Although the quantitative models used in Insurance industry may be different, the Pillar 1 of Solvency II shares the same spirit as

---

3A similar issue occurred in Insurance sector. The inadequacy of Solvency I in accurately directing capital to where the risks are motivates a more sophisticated framework for regulation purpose, Solvency II.
that of Basel II. In particular, firms’s internal models are favored by the regulator to calculate the two capital thresholds defined in Insurance business: the Solvency Capital Requirement (SCR) and the Minimum Capital Requirement (MCR).

ii) Basel II (and so does Solvency II) grants regulators the discretion to adjust the regulatory capital requirement against that calculated by banks. Thus, a capital level higher than that determined by banks are generally expected. The selection of the adjustment factor is a complex process which requires a deep understanding of properties of various risk measures, a comprehensive knowledge of estimation problems and an appropriate set of validation methods and criteria. The rest of this thesis is dedicated to a broad range analysis of these issues and the family of distortion risk measures is focused here.
Chapter 2

Sensitivity Analysis of Distortion Risk Measures

2.1 Introduction

This chapter provides a unified statistical framework for a nonparametric analysis of the functional distortion risk measures and of their sensitivities with respect to parameters representing risk aversion and/or pessimism. The properties of the estimated functional risk measures are based on the analogy principle and the asymptotic properties of the empirical processes (recalled in A.1). Then, we study in detail the relationship between the VaR and Tail-VaR. Indeed, a simple relationship can simplify the computation of the risk measures, extend the range of their interpretation, and facilitate empirical sensitivity analysis. We first show that the Tail-VaR can be approximated by multiplying VaR at the same risk level by an amplifying factor. Alternatively, the VaR and the Tail-VaR can be related through their risk levels. Moreover, we identify the condition such that this relationship is linear. One of the main contributions of this paper is the analysis of the asymptotic properties of an estimator of the function that relates their risk levels; a test statistic is also provided for the null hypothesis of linearity of this function. The analysis is performed in an i.i.d. framework not only for expository purpose, but also since it corresponds to the approach suggested by the regulator...
(e.g. BIS). First, the regulator proposes to define the risk measure by historical simulation. More precisely, the measure is replaced by its sample counterpart computed on a rolling window basis. This practice assimilates the marginal and conditional distributions, and thus, assumes implicitly i.i.d. returns.\footnote{See Appendix A.2 for a discussion of the non i.i.d. case.} Second, the i.i.d. assumption is also required to check for the accuracy of the risk measure and its sensitivity to downturn conditions for instance. Indeed, this is usually done by Monte-Carlo, that is, by i.i.d. drawings from the historical distribution.

The rest of this chapter is organized as follows. In Section 2.2, we describe the distortion risk measures and discuss the relationship between the VaR and Tail-VaR. We further analyze the sensitivity of the distortion risk measures with respect to the distortion parameter. Indeed, the sensitivity has an expression similar to the expression of the distortion risk measure. In Section 2.3, we derive the functional asymptotic properties of the estimators of the functional distortion risk measures and their sensitivities. Besides the general case, we consider three examples of distortion risk measures (VaR, Tail-VaR, and Proportional Hazard) and provide a closed form expression for their asymptotic variances and covariances. In Section 2.4, we focus on the function defining the change of risk level to pass from the VaR to the Tail-VaR and propose a test to check if the function is linear. We illustrate our analytical results by considering currency portfolios in Section 2.5. Concluding remarks are given in Section 4.4 and proofs are gathered in Appendices.

### 2.2 The distortion risk measures and their sensitivities

#### 2.2.1 Choquet expected utility and distortion risk measures

Before introducing and interpreting the distortion risk measures, it is necessary to fix a convention of profit and loss appropriate for the application to market finance, credit risk and insurance. Let us denote by $Y$ a portfolio value, corresponding to a zero initial investment. There is a profit if $Y$ is positive, a loss, otherwise. Let us now consider the standard way for computing the amount of reserve to hedge this risky investment. For a given loss probability
$u$, the $VaR(u)$, is defined by:

$$P[Y < -VaR(u)] = u \Leftrightarrow P[-Y \leq VaR(u)] = 1 - u.$$ 

The $VaR$ is the negative of the $u$-quantile of the profit and loss variable $Y$, as well as the $(1-u)$-quantile of the loss and profit variable $X = -Y$. For applications to insurance or regulation in credit risk, the focus is on the loss and profit variable $X = -Y$. This variable is positive in a lot of applications, such as the study of the loss component in an insurance contract, or the Loss-Given-Default (LGD) in credit risk. In the sequel, the variable of interest is the loss (and profit) variable, $X = -Y$.

I. Definitions

Expected utility theory was the first coherent approach introduced to compare risk variables. The risks are compared by means of a scalar expected utility:

$$EU(Y) = EU^*(X) = \int U^*(x) \, dF(x),$$

where $U$ is an increasing concave utility function and $U^*(x) = U(-x)$ is its decreasing concave counterpart associated with the loss (and profit) variable. For a continuous one-dimensional risk variable, the expected utility can be written as:

$$EU^*(X) = \int_0^1 U^*[Q(v)] \, dv = \int_0^1 U^*[Q(1 - u)] \, du,$$  \hspace{1cm} (2.2.1)

where the second equality is obtained by the change of variable $u = 1 - F(x)$, and $Q = F^{-1}$ denotes the quantile function. Different authors [Yaari (1987), Schmeidler (1989)] argue that “the independence axiom underlying the von-Neumann-Morgenstern axiomatization may be too powerful to be acceptable” and they propose another independence axiom valid for comonotonic
variables. The set of scalar risk measures is enlarged to the so-called Choquet expected utilities:

\[ \Pi(U^*, H; Q) = \int_0^1 U^*[Q(1-u)]dH(u). \]  \hspace{1cm} (2.2.2)

The risk measure involves a utility function \(U^*\) as in the standard expected utility framework and a distorted cumulative distribution function \( H \) (also called capacity in Choquet’s terminology). Function \(U^*\) represents the standard risk aversion (when \(U^*\) is concave); the distortion measure defines a change of probability, and represents the more or less pessimistic view on admissible risk levels. The extent of pessimism is determined by the level of concavity of the distortion function \( H \) [see e.g. Bassett et al. (2004)].

The limiting case, \(U^*(x) = x\), where only the distortion measure matters, has gained increasing attention recently due to its close relationship with many well recognized risk measures [Wang (1995), (1996), (2000), (2001), Acerbi and Simonetti (2002), Bassett et al. (2004)].

**Definition 4 (Wang (1996)).** A distortion risk measure (DRM) is defined as

\[ \Pi(H; Q) = \int_0^1 Q(1-u)dH(u), \]  \hspace{1cm} (2.2.3)

where \( H \) is a cdf on \([0, 1]\).

When \(Q\) is the quantile function of a loss (and profit) variable, a DRM is simply a weighted sum of VaR at level \(u\). This interpretation explains why DRMs have been proposed to measure the risk and compute risk premiums in the insurance literature in a series of papers by Wang and others [Wang (1995), (1996), (2000), Wang and Young (1998)]. Moreover, when the distortion cdf \( H \) is concave, the DRM is a coherent risk measure in the sense of Artzner et al. (1999) [see e.g. Wirch and Hardy (1999)], and a good candidate to define a level of required capital to balance a risky investment.
Finally, a DRM admits different equivalent expressions. Indeed, we get:

\[
\Pi(H; Q) = \int_0^{1-F(0)} Q(1-u)dH(u) + \int_{1-F(0)}^1 Q(1-u)d[H(u) - 1] \quad \text{(by splitting the interval)}
\]

\[
= -\int_0^{1-F(0)} H(u)dQ(1-u) - \int_{1-F(0)}^1 [H(u) - 1]dQ(1-u) \quad \text{(by integrating by parts)}
\]

\[
= \int_{F(0)}^1 H(1-u)dQ(u) + \int_0^{F(0)} [H(1-u) - 1]dQ(u) \quad \text{(by the change of variable } u \to 1-u).
\]

These expressions are greatly simplified, when the loss (and profit) variable \( X \) is nonnegative. Indeed, we get \( F(0) = 0 \), and:

\[
\Pi(H; Q) = \int_0^1 Q(1-u)dH(u) = \int_0^1 H(1-u)dQ(u). \quad \text{(2.2.4)}
\]

So when \( X \) is nonnegative, there is a symmetry between functions \( H \) and \( Q \).

II. Families of distortion risk measures

Many risk measures applied in finance and insurance literature, such as the VaR, or the Tail-VaR, are DRMs with carefully selected distortion functions. In practice, several risk measures have to be jointly considered in order to make risk management and risk control robust. This is done by introducing parameterized families of DRMs, or equivalently of distortion functions. Let us consider a family of distortion functions, \( H(\cdot ; p) \), where parameter \( p \) belongs to some interval. We get a family of DRMs:

\[
\Pi(p ; Q) = \int_0^1 Q(1-u)dH(u ; p), \quad p \in [a, b],
\]

indexed by \( p \). Thus, we are replacing the analysis of the distribution of the risk variable by the analysis of the functional parameter:

\[
\Pi(\cdot ; Q) : p \to \Pi(p ; Q),
\]
which is more appropriate for risk control. This functional risk measure can be in a one-to-one relationship with the underlying quantile function \( Q \), or can strictly summarize the corresponding information, if we focus on a special risk feature.

i) **VaR**

When \( H(u; p) = 1_{(u \geq p)} \) for \( p \in [0, 1] \), the distortion cdf corresponds to a point mass at \( p \). We have:

\[
\Pi(p; Q) = Q(1 - p), \tag{2.2.5}
\]

which is the VaR at risk level \( p \). Thus, the VaR is a special DRM associated with an indicator distortion function, which is not concave.

ii) **Tail-VaR**

When \( H(u; p) = (u/p) \wedge 1 \) for \( p \in [0, 1] \), the distortion function is the cdf of the uniform distribution on \([0, p]\). We get:

\[
\Pi(p; Q) = \int_0^p \frac{Q(1 - u)}{p} du = \int_0^\infty \frac{x}{Q(1 - p)} F(x) = E[X | X \geq \text{VaR}(p)]. \tag{2.2.6}
\]

Thus, \( \Pi(p; Q) \) is the Tail-VaR at level \( p \) (denoted by \( \text{TVaR}(p) \)) as defined in Artzner et al. (1999). Since the function \( u \rightarrow (u/p) \wedge 1 \) is concave, the Tail-VaR is a coherent risk measure. The Tail-VaR is an equally weighted average of all VaR at levels smaller than \( p \). Finally, note that the Tail-VaR is in a simple one-to-one relationship with the Lorenz Curve [Gastwirth (1971)], by \( \mathcal{L}(p) = p\Pi(p; Q)/E[X] \).

iii) **Proportional Hazard distortion risk measure**

If \( H(u; p) = u^p \) for \( p \in [0, \infty] \), the distortion function is the power-law transformation and can
be interpreted as a cdf on $[0, 1]$. The associated DRM is:

$$\Pi(p; Q) = \int_0^1 Q(1 - u) pu^{p-1} d u$$

$$= \int_0^{F(0)} (1 - u)^p d Q(u) + \int_0^{F(0)} [((1 - u)^p - 1] d Q(u)$$

$$= \int_{-\infty}^0 [(1 - F(x))^p - 1] d x + \int_0^\infty (1 - F(x))^p d x. \quad (2.2.7)$$

The interpretation of the distortion above is the following: The initial survivor function $S(x) = 1 - F(x)$ is replaced by the transformed survivor function $S_p^*(x) = S(x)^p$. Therefore, we have:

$$\Pi(p; Q) = \int_{-\infty}^0 [S_p^*(x) - 1] d x + \int_0^\infty S_p^*(x) d x = E_p^*[X],$$

where $E_p^*$ denotes the expectation with respect to the distribution with survivor function $S_p^*$.

The relationship between the initial and transformed survivor functions can also be written as: $\log S_p^*(x) = p \log S(x)$, and implies $\frac{d \log S_p^*(x)}{d x} = p \left( \frac{d \log S(x)}{d x} \right)$. Thus, the hazard functions associated with both distributions are proportional, which explains the name of the risk measure. The proportional hazard distortion risk measures are coherent risk measures, if parameter $p < 1$, that is, if the extreme losses are overweighted.

iv) **Exponential distortion risk measure**

If $H(u; p) = (1 - e^{-pu})/(1 - e^{-p})$, the distortion function is the cdf of the exponential distribution on $[0, 1]$. The associated DRM is:

$$\Pi(p; Q) = \int_0^1 Q(1 - u) \frac{pe^{-pu}}{1 - e^{-p}} d u$$

$$= \int_0^{F(0)} \left[ \frac{1 - e^{-p(1-u)}}{1 - e^{-p}} - 1 \right] d Q(u) + \int_0^{F(0)} \frac{1 - e^{-p(1-u)}}{1 - e^{-p}} d Q(u)$$

$$= \int_{-\infty}^0 \left[ \frac{1 - e^{-p(1-F(x))}}{1 - e^{-p}} - 1 \right] d x + \int_0^\infty \frac{1 - e^{-p(1-F(x))}}{1 - e^{-p}} d x.$$

The exponential distortion risk measure satisfies the coherency conditions when $p > 0$. 
2.2.2 Relationship between VaR and Tail-VaR

A main drawback of VaR is that it ignores the magnitude of loss. This problem may be partially solved by replacing the VaR by a more appropriate risk measure, such as the Tail-VaR. In this section, we analyze the link between VaR and Tail-VaR for different underlying distributions.

The second row in Table 1 provides the ratio between VaR and Tail-VaR for uniform, exponential, Pareto and Gaussian distributions, respectively. These ratios are independent of any scale parameter, are nondecreasing functions in the risk level $p$, and are larger than 1. Thus, the Tail-VaR is an amplified VaR with an amplifying factor which is a positive nondecreasing function of $p$, $TVaR(p) = [1 + L(p)]VaR(p)$. The value and pattern of this factor depends on the distribution (see Figures 1 for $p \in (0, 0.2)$).\footnote{All ratios approach 1 for $p \to 0$. Since we are generally interested in risk levels less than 10%, our range of $p$ is wide enough to cover all meaningful situations.} The exponential distribution features the widest range for the factor ($L(p)$ is between 0 and 140%), while the uniform distribution has the narrowest variation ($L(p)$ is between 0 and 12%). The Pareto distribution yields the simplest modification, in which the Tail-VaR is obtained by simply multiplying the VaR by a constant factor ($a/(a - 1)$), depending on the shape parameter. This constant factor is a decreasing function of $a \in (1, \infty)$ (See Figure 2). In fact, we have the following result:

**Proposition 6.** For a positive loss continuous random variable $X$, the ratio between Tail-VaR and VaR is constant in $p$, if and only, if the underlying distribution is a single parameter Pareto with a parameter $a$ and the lower bound of $X$ equal to $b$, $\text{Pareto}(a, b)$.
Proof. Let us rewrite the ratio as:

\[
\frac{E[X|X > \eta]}{\eta} = \frac{\frac{1}{1-F(\eta)} \int_{\eta}^{\infty} xdF(x)}{\int_{\eta}^{\infty} xdS(x)} = \frac{1}{S(\eta)} \int_{\eta}^{\infty} xdS(x) = \eta + \frac{1}{S(\eta)} \int_{\eta}^{\infty} S(x)dx \quad \text{(by integrating by part)}
\]

\[
= 1 + \frac{1}{\eta S(\eta)} \int_{\eta}^{\infty} S(x)dx.
\]

(2.2.8)

The ratio between Tail-VaR and VaR is constant, if and only, if

\[
\frac{1}{S(\eta)} \int_{\eta}^{\infty} S(x)dx = c\eta,
\]

(2.2.9)

where \(c\) is a positive constant. By integrating both sides of equation (2.2.9), we see that:

\[
\frac{d}{d\eta} \log \left( \int_{\eta}^{\infty} S(x)dx \right) = -\frac{1}{c} \frac{d}{d\eta} \log(\eta).
\]

Thus, there exists a positive constant \(A\), such that:

\[
\int_{\eta}^{\infty} S(x)dx = A \eta^{-1/c}.
\]

(2.2.10)

Taking derivative of both sides of (2.2.10) with respect to \(\eta\), we get:

\[
S(\eta) = \frac{A}{c} \eta^{-(c+1)/c},
\]

(2.2.11)

which corresponds to a Pareto\((a, b)\) distribution with \(a = (c + 1)/c\) and \(b = (\frac{A}{c})^{c/(c+1)}\).

An alternative way to describe the relationship between the VaR and Tail-VaR is based on the link between their risk levels. Indeed, the Tail-VaR at risk level \(p\) can be viewed as a VaR at a more Constraining risk level \(p^*\). This defines an increasing function \(p^* = g(p)\) smaller than \(p\), which depends on the underlying distribution, and satisfies \(TVaR(p) = VaR(p^*)\) (see the
third row of Table 1). Except for the standard normal distribution, \( g(p) \) is proportional to \( p \). Its behavior for the standard normal distribution is plotted in Figure 3. In fact, for small value of \( p \), say less than 0.5, the function \( g(p) \) is hardly distinguishable from linearity even under Gaussian assumption. This is a desired property from a practical point view. Indeed, after calculating the VaR at several risk levels, the related Tail-VaRs are obtained automatically, which simplifies the computation procedures. In addition, an internal or external regulator can interpret an extreme quantile value either as an amount that a given portfolio’s losses will not be likely to exceed under normal market conditions or as the expected loss of the same portfolio under adverse market conditions.\(^3\)

Let us characterize the distributions such that the function \( g \) is linear with coefficient \( \alpha \). We get:

\[
TVaR(p) = VaR(\alpha p) \iff \int_0^p Q(1 - u) d u = p Q(1 - \alpha p).
\]  

(2.2.12)

In particular, by taking \( p = 1 \), we get an interpretation of the slope parameter \( \alpha \) as: \( Q(1 - \alpha) = VaR(\alpha) = E[X] \), and note that \( TVaR(p) = VaR[pVaR^{-1}(E[X])] \). Typically, the level \( p \) has to be divided by 2, if the mean is equal to the median, by a number strictly larger than 2 (resp. smaller than 2) if the mean is smaller (resp. larger) than the median, that is, if the distribution is “right skewed” (resp. “left skewed”). By differentiating both sides of (2.2.12), we get:

\[
Q(1 - \alpha p) - Q(1 - p) = \alpha p q(1 - \alpha p),
\]

where \( q(u) = \partial Q(u)/\partial u \) is the quantile density. From Table 1, we see that the uniform, exponential and Pareto distributions satisfy the condition with \( \alpha = 1/2 \) (since the mean is equal to the median), \( 1/e \) and \( ((a-1)/a)^a \), respectively. The dependence of the slope coefficient \( \alpha \) with respect to the shape parameter \( a \) is given in Figure 4. This coefficient varies between 0 and \( 1/e \). Since the tail of a Pareto distribution is thinner as \( a \) rises, the fatter the tail, the smaller is \( \alpha \).

\(^3\)The reported \( p^* \) are smaller than half the risk level \( p \). This is expected since all distributions considered here (except the uniform distribution) have tails skewed to the right.
2.2.3 Sensitivity of a distortion risk measure with respect to a distortion parameter

The sensitivity of the DRM is:

\[
\frac{\partial \Pi}{\partial p}(p; Q) = \frac{\partial}{\partial p} \left[ \int_0^1 Q(1-u)dH(u; p) \right] \\
= \int_0^1 Q(1-u)d \left[ \frac{\partial}{\partial p} H(u; p) \right] \\
= -\int_0^1 \frac{\partial}{\partial p} H(u; p) dQ(1-u) \\
= \int_0^1 \frac{\partial H}{\partial p}(1-u; p) dQ(u),
\]

(2.2.13)

since \( \frac{\partial}{\partial p} H(1; p) = \frac{\partial}{\partial p} H(0; p) = 0 \).

This expression is similar to the expression of a DRM except that the distortion function \( H^*(u; p) = \frac{\partial}{\partial p} H(u; p) \) is not a cdf, since \( H^*(1; p) = H^*(0; p) = 0 \). Moreover, the alternative expressions (2.2.4) of the sensitivity are still valid even if the loss (and profit) variable is not necessarily positive:

\[
\frac{\partial \Pi}{\partial p}(p; Q) = \Pi(H^*; Q) = \int_0^1 Q(1-u)d H^*(u; p) = \int_0^1 H^*(1-u; p)dQ(u).
\]

(2.2.14)

The examples below illustrate the computation and interpretation of the sensitivity.

i) Tail-VaR
We have:

\[
\frac{\partial}{\partial p} TVaR(p) = -\int_0^1 \frac{1 - u}{p^2} 1_{(1-u \leq p)} dQ(u) \\
= -\int_{1-p}^1 \frac{1 - u}{p^2} dQ(u) \\
= -\frac{1}{p^2} (1 - u) Q(u) \bigg|_{1-p}^1 + \frac{1}{p} \int_{1-p}^1 Q(u) d \frac{1 - u}{p} \\
= \frac{1}{p} VaR(p) - \frac{1}{p} \int_{Q(1-p)}^\infty x d \frac{F(x)}{p} \\
= \frac{1}{p} [VaR(p) - TVaR(p)]. \tag{2.2.15}
\]

The sensitivity of Tail-VaR with respect to the distortion parameter is the opposite of the difference between the conditional expected loss and the lower bound of the loss per unit of risk level. This derivative is negative and the Tail-VaR increases when the risk level diminishes. As seen in the next section, this value measures the accuracy of the nonparametric estimator of Tail-VaR. Indeed, a large (absolute) sensitivity of the Tail-VaR can induce substantial estimation errors at small \( p \).

The sensitivities of the VaR and Tail-VaR with respect to risk level \( p \) are provided in Table 2 for the uniform, exponential, Pareto and standard normal distributions, respectively.

**ii) Proportional Hazard distortion risk measure**

We have \( H^*(u; p) = \frac{\partial}{\partial p} (u^p) = u^p (\log u) \) and deduce that:

\[
\frac{\partial}{\partial p} PH(p) = \int_0^1 Q(1 - u) d [u^p (\log u)] \tag{2.2.16} \\
= \int_0^1 Q(1 - u) u^{p-1} (p \log u + 1) \ d u \\
= \int_0^1 Q(1 - u) w(u, p) \ d u, \quad \text{say.}
\]

When \( PH \) is interpreted as a risk premium, the sensitivity is the marginal response of this
premium to a slight adjustment of the pessimism level. This marginal response is a weighted expectation of VaR with the weighting function \( w(u, p) = u^{p-1} (p \log u + 1) \) depending on the pessimism parameter \( p \). Figure 5 provides two examples of weighting functions when \( p = 0.2, \) and 2, respectively.

As expected, the marginal response is negative. Indeed, by integrating (2.2.16) by part, we get:

\[
\frac{\partial}{\partial p} PH(p) = -\int_0^1 u^p (\log u) dQ(1-u) \quad \text{(integration by part)}
\]

\[
= \int_{-\infty}^{\infty} [1 - F(x)]^p \log[1 - F(x)] d x \quad \text{(change of variable)}
\]

\[
< 0.
\]

iii) **Exponential distortion risk measure**

With exponential distortion function, we get:

\[
H^*(u; p) = \frac{e^{-pu}u}{1 - e^{-p}} - \frac{e^{-p}(1 - e^{-pu})}{(1 - e^{-p})^2}.
\]

Thus, the sensitivity of the exponential distortion risk measure is given by:

\[
\frac{\partial}{\partial p} EX(p) = \int_0^1 Q(1-u) w(u, p) d u,
\]

which is a weighted expectation of VaR with weighting function,

\[
w(u, p) = \frac{e^{-pu}}{1 - e^{-p}} - \frac{e^{-p} - pu}{(1 - e^{-p})^2} - \frac{e^{-pu} pu}{1 - e^{-p}}.
\]

Two examples of the shape of this weighting function are plotted in Figure 6.
2.3 Nonparametric estimation of functional distortion risk measures and their sensitivities

Let us consider a set of i.i.d. one-dimensional observations \( x_1, ..., x_T \), with common cdf \( F_0 \) and quantile function \( Q_0 = F_0^{-1} \). The quantile function \( Q_0 \) can be estimated by the sample quantile \( \tilde{Q}_T \) defined by:

\[
\tilde{Q}_T(u) = \inf \{ x : \frac{1}{T} \sum_{t=1}^{T} 1_{(x_t \leq x)} \geq u \}, \text{ for } u \in [0, 1].
\] (2.3.1)

We first recall the asymptotic distribution of \( \tilde{Q}_T \). Then, we introduce the distortion risk measures and their expressions in terms of quantile function. By applying the analogy principle (see A.1), we deduce functional nonparametric estimators of distortion risk measures and of their sensitivities with respect to parameters.

2.3.1 Asymptotic distribution of the nonparametric quantile estimator

The analysis is based on the Bahadur representation of the quantile estimator, which provides the expressions of the estimated quantiles in terms of the associated cdf [Koenker (2005), Section 4.3]. We get,

\[
\sqrt{T} [\tilde{Q}_T(u) - Q_0(u)] = -\frac{1}{f_0(Q_0(u))} \sqrt{T} [\tilde{F}_T(Q_0(u)) - u] + o_p(1), \quad (2.3.2)
\]

where \( \tilde{F}_T \) is the sample cdf, \( Q_0 \) the true quantile function and \( f_0 \) the true density, and the proposition below.

**Proposition 7.** For an i.i.d. random sample from a distribution with quantile function \( Q_0 \) and pdf \( f_0 \), we have:

\[
\sqrt{T} [\tilde{Q}_T(\cdot) - Q_0(\cdot)] \Rightarrow -\frac{1}{f_0(Q_0(\cdot))} B(\cdot),
\]
where \( B(u) \) is a univariate Brownian bridge and \( \Rightarrow \) denotes weak convergence of stochastic processes (see the Functional Limit Theorem in A.1).

### 2.3.2 Estimation of distortion risk measure

By the analogy principle, a nonparametric estimator of the DRM is defined by:

\[
\tilde{\Pi}_T(p) = \Pi(p; \tilde{Q}_T), \quad p \text{ varying.}
\]

For a given sample \( x_1, \ldots, x_T \), the observations can be ranked in an ascending order such that \( x^*_1 \leq x^*_2 \cdots \leq x^*_T \), and the estimated DRM is simply:

\[
\tilde{\Pi}_T(p) = \sum_{i=1}^{T} x^*_i \left[ H \left( 1 - \frac{i - 1}{T} \right) - H \left( 1 - \frac{i}{T} \right) \right]. \tag{2.3.3}
\]

Thus, the nonparametric estimator of the DRM is a linear combination of the order statistics \( x^*_i \) and, for each value of the pessimism parameter, this is an example of \( L \)-statistics [see e.g. Jones and Zitikis (2003), (2005)]. For instance, the nonparametric estimator of VaR at risk level \( p \) can be written as:

\[
\tilde{\text{VaR}}_T(p) = \sum_{i=1}^{T} x^*_i \left[ 1 \left( \frac{i - 1}{T} \leq 1 - p \right) - 1 \left( \frac{i}{T} \leq 1 - p \right) \right] = \begin{cases} 
  x^*_{T(1-p)}, & \text{if } T(1-p) \text{ is integer}, \\
  x^*_{\lceil T(1-p) \rceil + 1}, & \text{otherwise}, 
\end{cases}
\]

where \( \lfloor a \rfloor \) denotes the integer part of \( a \). For the Tail-VaR, the estimator is directly related to the estimator introduced in the literature for the Lorenz Curve or the Gini Index [Gastwirth (1972), Barrett and Donald (2000), Zitikis (2003)].

The proposition below is a direct consequence of the expression of the DRM, \( \Pi(H; Q) = \int_0^1 Q(1 - u) dH(u; p) \), and of the results of Section 2.3.1. The asymptotic behavior is not only a pointwise convergence result [see e.g. Jones and Zitikis (2003)], but concerns the process of DRM indexed by pessimism parameter. This functional result is needed for further analysis of links between the VaR and Tail-VaR for instance.
Proposition 8. For an i.i.d. random sample from a distribution with quantile function $Q_0$ and pdf $f_0$, we have:

\[
\sqrt{T} [\tilde{\Pi}_T(p) - \Pi(p; Q)] \Rightarrow \int_0^1 \frac{B(1 - u)}{f_0(Q_0(1 - u))} dH(u; p),
\]

where $B(\cdot)$ is a standard Brownian bridge. The process is asymptotically Gaussian with pointwise variance equal to:

\[
V(\sqrt{T} [\tilde{\Pi}_T(p) - \Pi(p)]) = \int_0^1 \int_0^1 \frac{(1 - u_1) \wedge (1 - u_2) - (1 - u_1)(1 - u_2)}{f_0(Q_0(1 - u_1))f_0(Q_0(1 - u_2))} dH(u_1; p)dH(u_2; p)
\]

\[
= 2 \int_0^1 \frac{u_2A(u_2, p)}{f_0(Q_0(1 - u_2))} dH(u_2; p),
\]

where

\[
A(v, p) = \int_v^1 \frac{1 - u}{f_0(Q_0(1 - u))} dH(v; p).
\]

Replacing the quantile function and the density by their nonparametric estimators, we get the Corollary below about the estimation of the asymptotic variance of the estimated DRM.

Corollary 1. For an i.i.d. random sample, the asymptotic variance of estimated DRM can be consistently estimated by:

\[
\hat{V}(\sqrt{T} [\tilde{\Pi}_T(p) - \Pi(p)]) = 2 \int_0^1 \frac{u_2\hat{A}(u_2, p)}{\hat{f}(Q_T(1 - u_2))} dH(u_2; p),
\]

where

\[
\hat{A}(v, p) = \int_v^1 \frac{1 - u}{\hat{f}(Q_T(1 - u))} dH(v; p),
\]

and $\hat{f}$ is a nonparametric consistent estimator of the density function.

A common choice of the density estimator is a kernel estimator. The estimated asymptotic variance of the estimated DRM can be computed numerically. However, a kernel estimator of the density converges rather slowly, which may render the application of the asymptotic theory questionable in finite sample. Fortunately, except for the VaR, the density function can be
eliminated from the variance expression above.

**Corollary 2.** When $H$ is continuous and almost everywhere differentiable, we have

\[
V(\sqrt{T}[\hat{\Pi}_T(p) - \Pi(p)]) = \int_{\mathbb{R}^2} \frac{F(x_1) \wedge F(x_2) - F(x_1)F(x_2)}{f_0(x_1)f_0(x_2)} \frac{\partial H(1 - F(x_1); p)}{\partial u} \frac{\partial H(1 - F(x_2); p)}{\partial u} dF(x_1)dF(x_2)
\]

\[
= \int_{\mathbb{R}^2} (F(x_1) \wedge F(x_2) - F(x_1)F(x_2)) \frac{\partial H(1 - F(x_1); p)}{\partial u} \frac{\partial H(1 - F(x_2); p)}{\partial u} dx_1 dx_2
\]

\[
= 2 \int_\mathbb{R} (1 - F(x_2)) A(x_2, p) \frac{\partial H(1 - F(x_2); p)}{\partial u} dx_2,
\]

where

\[
A(y, p) = \int_{-\infty}^y F(x) \frac{\partial H(1 - F(x); p)}{\partial u} dx.
\]

Thus, this pointwise variance can be estimated by substituting the empirical distribution function $\hat{F}_T(x)$ in the expression [see Jones and Zitikis (2003), Theorem 3.2]. The asymptotic variance above is estimated by:

\[
\hat{V}(\sqrt{T} [\hat{\Pi}_T(p) - \Pi(p)]) = \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \left( \frac{i}{T} \wedge \frac{j}{T} - \frac{i}{T} \right) w\left(1 - \frac{i}{T}; p\right) w\left(1 - \frac{j}{T}; p\right) (x_{i+1}^* - x_i^*)(x_{j+1}^* - x_j^*),
\]

where

\[
w(u; p) = \frac{\partial}{\partial u} H(u; p).
\]

Similarly, it is easy to derive the estimated covariance between either the estimators of a DRM with different values of $p$, or the estimators of two DRMs. For instance, we get:

\[
COV\left(\sqrt{T}[\Pi_T p - \Pi(p)], \sqrt{T}[\Pi_T p' - \Pi(p')]\right) = Q_{F, F} \left(\frac{\partial H}{\partial u}(1-\cdot; p), \frac{\partial H}{\partial u}(1-\cdot; p')\right),
\]

and,

\[
\hat{COV}(\sqrt{T} [\hat{\Pi}_T(p) - \Pi(p)], \sqrt{T} [\hat{\Pi}_T(p') - \Pi(p')])
\]

\[
= \sum_{i=1}^{T-1} \sum_{j=1}^{T-1} \left( \frac{i}{T} \wedge \frac{j}{T} - \frac{i}{T} \right) w\left(1 - \frac{i}{T}; p\right) w\left(1 - \frac{j}{T}; p'\right) (x_{i+1}^* - x_i^*)(x_{j+1}^* - x_j^*).\]

4 with respect to Lebesgue measure on $[0, 1]$.

5 The quantity, $\int_{\min(F(x_1), F(x_2))}^{\min(F(x_1), F(x_2))} \Psi_2(F(x_1))\Psi_2(F(x_2))dx_1dx_2$, can be denoted as $Q_{F, F}(\Psi_1, \Psi_2)$. Thus, the pointwise variance is $Q_{F, F} \left(\frac{\partial H}{\partial u}(1-\cdot; p), \frac{\partial H}{\partial u}(1-\cdot; p')\right)$. 


Since the derivative $\frac{\partial H}{\partial u}(\cdot; p)$ is positive for any $p$, we deduce from the expression of $Q_{F,F}$, that two estimated DRM are always positively correlated.

### 2.3.3 Estimation of the sensitivity

From expression (2.2.14), the sensitivity of the distortion risk measure has a similar expression as the DRM, except that the distortion function is replaced by its first-order derivative with respect to parameter $p$. The limiting properties of their estimators are also similar. They are given in the Corollary below.

**Corollary 3.** If $H(u; p)$ is differentiable in $p$, for an i.i.d. random sample with quantile function $Q_0$ and pdf $f_0$, we have:

$$\sqrt{T}\left[\tilde{\Pi}_T(p; H^*) - \Pi(p; H^*)\right] \Rightarrow \int_0^1 \frac{B(1-u)}{f_0(Q_0(1-u))} dH^*(u; p).$$

$\tilde{\Pi}_T(p; H^*)$ is asymptotically Gaussian with pointwise variance given by:

$$V\left(\sqrt{T}\left[\tilde{\Pi}_T(p; H^*) - \Pi(p; H^*)\right]\right) = 2 \int_0^1 \frac{u_2 A^*(u_2; p)}{f_0(Q_0(1-u_2))} dH^*(u_2; p),$$

where

$$A^*(v, p) = \int_v^1 \frac{1-u}{f_0(Q_0(1-u))} dH^*(u; p).$$

The asymptotic variance of the sensitivity can be estimated in the same way as for the DRM with or without density estimation. Indeed, if the cross-derivative $\frac{\partial^2 H}{\partial u \partial p}$ exists, we have:

$$V\left(\sqrt{T}[\tilde{\Pi}_T(p; H^*) - \Pi(p; H^*)]\right) = Q_{F,F} \left(\frac{\partial^2 H}{\partial u \partial p}(1-\cdot; p), \frac{\partial^2 H}{\partial u \partial p}(1-\cdot; p)\right).$$

The function $H^*$ associated with the sensitivity of Tail-VaR is noncontinuous and the estimation of the density function cannot be avoided. On the contrary, it is not necessary to estimate the density function for examples such as Proportional Hazard and Exponential distortion risk measures.

The nonparametric estimator of the sensitivity of VaR is not well defined even though the quantile function is differentiable in $p$ analytically. However, the sensitivity analysis of VaR
may be approximated by the sensitivity analysis of a Tail-VaR. For instance, if the two risk measures are related by $TVaR(p) = VaR(\alpha p)$, the marginal change of $VaR(\alpha p)$ with respect to $\alpha p$ is approximated by the marginal change of the $TVaR(p)$ with respect to $p$ divided by $\alpha$, which has a well defined nonparametric estimator.

2.3.4 Examples

The closed form of the asymptotic variance can be derived for specific distortion functions. We consider below the examples of VaR, Tail-VaR, and $PH$ and illustrate the accuracy of estimation by studying the asymptotic variances of their nonparametric estimators. The detailed proofs are provided in Appendices C and D.

i) VaR

For $H(u; p) = 1_{(u \geq p)}$, $p \in [0, 1]$, the asymptotic variances of the estimator of VaR is:

$$V(\sqrt{T} [\widehat{VaR}_T(p) - VaR(p)]) = \frac{p(1-p)}{[f_0(Q_0(1-p))]^2} = \left[\frac{q_0(1-p)}{f_0(Q_0(1-p))}\right]^2 = \left[q_0(1-p)\right]^2p(1-p),$$

where $q_0 = 1/f_0(Q_0)$ is the quantile density function. The tail behavior is classified into different categories in practice. For example, the Gaussian distribution has a thin Gaussian tail; exponential, Laplace and logistic distributions have thick exponential tails; Pareto, Lévy and Cauchy distributions have thicker Pareto tails. It is important to understand how tail behavior influences the estimation accuracy of risk measures. The asymptotic variance of the nonparametrically estimated VaR and its relative accuracy are given in Table 3 for uniform, exponential, Pareto, Gaussian, Lévy and Cauchy distributions, respectively. Since distributions with large absolute values for the limits of the support are likely to yield noisy estimates of the extreme quantiles, relative accuracy may be more informative. In Figure 7, we plot the asymptotic variance and relative accuracy as functions of $p$. Distributions with unbounded support tend to induce large estimation errors at the tails. This is evidenced by the variance patterns for both tails of Gaussian and Cauchy distributions and for the right tail of exponential, Pareto
and Lévy distributions. Moreover, distributions with extremely heavy tails tend to cause huge estimation error, which is evidenced by the magnitude of variance associated with both Lévy and Cauchy distributions.

ii) Tail-VaR
If $H(u; p) = (u/p)^{\wedge}1$, for $p \in [0, 1]$, the asymptotic variance of the nonparametrically estimated Tail-VaR is given by:

$$V\left(\sqrt{T} \left[ \widetilde{TVaR}_T(p) - TVaR(p) \right] \right) = V\left(X | X \geq VaR(p) \right) + \frac{(1 - p) \left[ TVaR(p) - VaR(p) \right]^2}{p}. \quad (2.3.5)$$

Column 3 and 4 of Table 4 provide the asymptotic variance and relative accuracy of the nonparametric estimator for uniform, exponential, Pareto and Gaussian distributions, respectively. The variance for Lévy and Cauchy distributions cannot be derived, since the associated moments are not defined. In fact, for these distributions, the Tail-VaR may not even exist.\(^6\)

The second component in the decomposition of $V \left[ \sqrt{T} \widetilde{TVaR}_T(p) \right]$ is proportional to the square of the sensitivity of Tail-VaR. Thus, larger sensitivity tends to imply larger variance for the nonparametric estimator. This effect is seen on Figure 8, which plots the accuracies as function of the pessimism parameter. With unbounded right tails, the exponential, Pareto and Gaussian distributions imply large estimation error, both in terms of variance and relative accuracy, for small values of $p$.

iii) $PH$
For the power-law distortion function, $H(u; p) = u^p$, the asymptotic variances of the estimated

\(^6\)It can be verified that the asymptotic variances reduces to the unconditional variances when $p = 1$.\n
\[ V\left(\sqrt{T} [PH_T(p) - PH(p)]\right) \]
\[ = p^2 E_{p-1}^*(X) E_p^*(X) - p^2 \left[ E_p^*(X) \right]^2 + \]
\[ \begin{cases} 
\frac{p^2}{2p-1} E_{2p-1}(X^2), & \text{if } p \geq 0.5 \\
p^2 \int_{-\infty}^{\infty} F(X) (1 - F(X))^{2p-1} dX^2 - 2p^2 \left[ \frac{p-1}{2p-1} E_{2p-1}(X^2) - \frac{1}{2} E_{2p}^*(X^2) \right], & \text{if } 0 < p < 0.5. 
\end{cases} \]

Figure 9 displays the accuracy of the nonparametric estimator of Proportional Hazard distortion risk measure for various distributions. Both variance (panel (a)) and relative accuracy (panel (b)) are considered for \(0.5 < p < 1\). The more pessimistic, the less accurate is the estimator. The only exception occurs for the standard normal distribution, where the denominator of the relative accuracy goes to zero as \(p \to 1\). Intuitively, a smaller \(p\) induces on average a larger modification weight both for the distorted mean and for the distorted variance.

### 2.3.5 Implied pessimism parameter

The regulator receives the reserve levels reported by banks on a regular basis to check the capital adequacies. These data can be used to get information on the behavior of a bank concerning the risk, and in particular to estimate the level of the selected pessimism parameter. More precisely, if we observe a reserve level \(\Pi^o\), the implied pessimism parameter is defined by:

\[ \Pi(p; Q_0) = \Pi^o, \]

and is consistently estimated by:

\[ \hat{p}_T = \Pi^{-1}(\Pi^o; \tilde{Q}_T). \]

The asymptotic property of the estimated implied parameter \(\hat{p}_T\) is determined by the limiting behavior of the estimated distortion risk measure. More precisely, we have (see A.6 for a
derivation):

\[
\sqrt{T} (\hat{p}_T - p) = - \left( \frac{\partial \Pi}{\partial p} (p; Q) \right)^{-1} \sqrt{T} [\Pi(p; \tilde{Q}_T) - \Pi(p; Q)] + o_p(1). \tag{2.3.6}
\]

These estimated implied pessimism parameters can be computed for any reported reserve level, that is, for different dates and banks. Their comparison allows to follow how pessimism varies in time, or to get a segmentation of the banks in terms of pessimism.

\section{2.4 Tail-VaR versus VaR}

The aim of this section is to introduce a nonparametric estimator of the function \( g \), which links the VaR and Tail-VaR, and to derive its asymptotic properties. In a second step, we explain how to test for the linearity of function \( g \) and estimate the associated slope coefficient.

\subsection{2.4.1 Nonparametric estimator of \( g \)}

Function \( g \) is defined by: \( TVaR(p) = VaR[g(p)] \), or equivalently:

\[
TVaR(p) = Q(1 - g(p)) \Leftrightarrow g(p) = 1 - F[TVaR(p)].
\]

By the analogy principle, a nonparametric estimator of function \( g \) is:

\[
\tilde{g}_T(p) = 1 - \tilde{F}_T \left[ TVaR_T(p) \right]. \tag{2.4.1}
\]

Under the appropriate regularity conditions, the estimator \( \tilde{g}_T \) is consistent and such that:

\[
\sqrt{T} \left[ \tilde{g}_T(p) - g(p) \right] = - \sqrt{T} \left[ \tilde{F}_T \left[ TVaR(p) \right] - F_0 \left[ TVaR(p) \right] \right] \\
- f_0 \left[ TVaR(p) \right] \sqrt{T} \left[ TVaR_T(p) - TVaR(p) \right] + o_p(1).
\]

We deduce the proposition below.
Proposition 9. For an i.i.d. random sample, we have

\[
\sqrt{T} \left[ \bar{g}_T(p) - g(p) \right] \Rightarrow -B \left[ F_0[TVaR(p)] \right] + \frac{f_0[TVaR(p)]}{p} \int_{1-p}^{1} \frac{B(u)}{f_0[Q_0(u)]} d u,
\]

which is asymptotically Gaussian with zero mean and pointwise variance

\[
V \left[ \sqrt{T} (\bar{g}_T(p) - g(p)) \right] = \left[ 1 - g(p) \right] g(p) + f_0[TVaR(p)]^2 \left\{ \begin{array}{l} 
\frac{1}{p} \left( g(p) \left( [1 - g(p)] TVaR(g(p)) - VaR(p)(1 - p) \right) \\
- [p - g(p)] E \left[ X | VaR(p) \leq X \leq VaR(g(p)) \right] \end{array} \right.
\]

with \( V \left( \sqrt{T} \widehat{TVaR}_T(p) \right) \) given by (2.3.5).

2.4.2 Test of the linearity hypothesis

Let us now consider the null hypothesis of linearity of function \( g \) in a given risk window \((p_0, p_1)\). This hypothesis concerns the underlying distribution of returns and the portfolio allocation. As seen in the examples, it can be specified for some distributions of portfolio returns, but of course it cannot be satisfied by all possible distributions and portfolio allocations; otherwise, the VaR would be a coherent risk measure. The linearity hypothesis \( H_0 = \{ \exists \alpha_0 : g(p) = \alpha_0 p \ \text{for any} \ p \in (p_0, p_1) \} \) can be tested as follows.

Let us introduce a measure of the distance to the linearity hypothesis:

\[
\tilde{L}_T(\mu) = \min_{\alpha} \int_{p_0}^{p_1} \left( \bar{g}_T(p) - \alpha p \right)^2 \mu(p) \, dp,
\]

where \( \mu \) is a weighting function and accept the null hypothesis if the measure is sufficiently
small. More precisely, let us first consider the optimal value of the slope parameter $\alpha$:

$$
\tilde{\alpha}_T = \arg \min_\alpha \int_{p_0}^{p_1} (\tilde{g}_T(p) - \alpha p)^2 \mu(p) \, dp
$$

$$
= \frac{\int_{p_0}^{p_1} p \tilde{g}_T(p) \mu(p) \, dp}{\int_{p_0}^{p_1} p^2 \mu(p) \, dp}.
$$

(2.4.3)

We get the following result (see A.4):

**Proposition 10.** Under the null hypothesis, the estimator $\tilde{\alpha}_T$ is consistent, asymptotically Gaussian:

$$
\sqrt{T}(\tilde{\alpha}_T - \alpha_0) \overset{d}{\to} N(0, \eta^2),
$$

where the variance $\eta^2$ is given by

$$
\eta^2 = 2A^2 \left\{ \int_{p_0}^{p_1} A^* \left[ (1 - g(\tilde{p}))(1 - g(p)) \right] d\tilde{p} - 2 \int_{p_0}^{p_1} A^* \left[ f_0[TVaR(p)]^2 COV(\sqrt{T}VaR_T^T(g(p)), \sqrt{T}VaR_T^T(\tilde{p})) \right] d\tilde{p} \\
\int_{p_0}^{p_1} A^* \left[ f_0[TVaR(\tilde{p})]f_0[TVaR(p)] COV(\sqrt{T}VaR_T^T(\tilde{p}), \sqrt{T}VaR_T^T(p)) \right] d\tilde{p} \right\},
$$

(2.4.4)

where

$$
A = \frac{1}{\int_{p_0}^{p_1} p^2 \mu(p) \, dp}, \quad \text{and} \quad A^*[\xi] = \int_{p_0}^{\tilde{p}} \tilde{p} p \mu(\tilde{p}) \mu(p) \xi \, d\tilde{p}.
$$

Then, the optimal value of the criterion function is:

$$
\tilde{L}_T(\mu) = \int_{p_0}^{p_1} (\tilde{g}_T(p) - \tilde{\alpha}_T p)^2 \mu(p) \, dp = \int_{p_0}^{p_1} (\tilde{g}_T(p))^2 \mu(p) \, dp - \left[ \frac{\int_{p_0}^{p_1} p \tilde{g}_T(p) \mu(p) \, dp}{\int_{p_0}^{p_1} p^2 \mu(p) \, dp} \right]^2.
$$

By applying the Functional Limit Theorem (see A.1), we deduce the asymptotic behavior of the test criterion.
Proposition 11. Under the null hypothesis of linearity,

\[ T \tilde{L}_T(\mu) = T \int_{p_0}^{p_1} \left( \tilde{g}_T(p) - g(p) - (\tilde{\alpha}_T - \alpha) p \right)^2 \mu(p) dp \]

\[ \Rightarrow \int_{p_0}^{p_1} \left\{ B_g(p) - \frac{p \int_{p_0}^{p_1} \tilde{p} B_g(\tilde{p}) \mu(\tilde{p}) d\tilde{p}}{\int_{p_0}^{p_1} \tilde{p}^2 \mu(\tilde{p}) d\tilde{p}} \right\}^2 dp, \]

where

\[ B_g(p) = -B \left[ F_0[T VaR(p)] \right] + \frac{f_0[T VaR(p)]}{p} \int_{1-p}^1 \frac{B(u)}{f_0[Q_0(u)]} du. \]

This is the distribution of a series of weighted \( \chi^2_1 \) random variables [see e.g. Freitag et al. (2003), Remark 2.7].

2.5 Application to currency portfolio

In this section, we apply the results of the previous sections to currency portfolios. The currencies introduced in the portfolio are the Hongkong Dollar and Japanese Yen. The US Dollar (resp. the Singapore Dollar) is chosen as the basic numeraire of the investor. Indeed, the financial features of a currency portfolio can depend on the numeraire. The main part of the literature on currency portfolios [see e.g. Akgiray and Booth (1988); Breymann et al. (2003); Chen et al. (2004); Patton (2006)] consider portfolios written in US Dollar, and exhibit a number of stylized facts such as asymmetry, fat tail and stochastic jumps. It is important to see if these stylized facts are due either to the currencies introduced in the portfolio, or to the portfolio allocation, or if they come from the chosen numeraire. The data set consists of daily data from November 1993 to December 2005 which provides about 3200 observations. Denoting \( S_{u,t}^i \) (resp. \( S_{s,t}^i \)) the exchange rate (at date \( t \)) of currency \( i \) \((i = 1, 2\) representing Hongkong Dollar and Japanese Yen, respectively) in US Dollar (resp. Singapore Dollar), the daily returns are:

\[ \tilde{x}_{u,t}^i = \frac{S_{u,t}^i - S_{u,t-1}^i}{S_{u,t-1}^i}, \quad \tilde{x}_{s,t}^i = \frac{S_{s,t}^i - S_{s,t-1}^i}{S_{s,t-1}^i}. \]
The returns are computed at daily and monthly (20 trading days) horizons. To comply with the independence assumption on returns, we avoid overlapping in constructing monthly horizon returns and thus, get \( 160 = 3200/20 \) monthly observations. Summary statistics on equally weighted (negative) portfolio returns are provided in Table 6. Whereas the (negative) mean portfolio returns under both numeraires and both horizons are not statistically different from zero, their tails behave differently. At daily horizon, the portfolio exhibits a rather symmetric pattern when it is written in US Dollar and is more skewed to the left when it is written in Singapore Dollar. Although less significant, the reverse is observed at horizon 20 days, corresponding to a trading month. Finally, both daily returns display fatter tails than Gaussian with much fatter tails for the portfolio written in Singapore Dollar. This fat tail phenomenon is substantially reduced for monthly returns.

For both numeraires, the VaR and Tail-VaR of the portfolios are estimated nonparametrically for daily and monthly horizons (see Figures 10 and 11). In each figure, panel (a) (resp. panel (b)) corresponds to the portfolio written in US Dollar (resp. Singapore Dollar). The solid line represents the estimated risk measure and the dashed lines are the lower and upper bounds of its confidence band. The symmetry of the distribution is evidenced for the estimated VaR in all cases. Moreover, the estimation errors are larger at the (upper) extreme tail. This feature is consistent with the variance pattern implied by exponential, Pareto and Gaussian distributions. Although data under both horizons are symmetric about zero, the monthly (negative) portfolio returns are more likely to reach higher values. Indeed, with longer holding period, the price tends to be more volatile and increase the possibility for extreme changes. Due to the smaller number of observations, estimations performed on monthly data have wider confidence intervals. A similar pattern is illustrated in the Tail-VaR estimation. The confidence bands are large when \( p \) is small and shrink gradually as the Tail-VaR converges to the mean, that is when \( p \) tends to 1. The Tail-VaR at monthly horizon are at least about twice those under daily horizon. Moreover, the estimations under daily (negative) returns are more accurate than their monthly counterparts. For instance, the estimation of the Tail-VaR for the portfolio written in Singapore Dollar underestimate the loss by about 40 basis points when the risk level is extreme.
and the holding period is 20 days.

The amplifying factor $TVaR(p)/VaR(p)$ is plotted for both numeraires in Figure 12, with thicker line representing daily data. The amplifying factor can be used as an alternative tool to kurtosis for identifying the distributional behavior of the tail. For the daily (negative) returns, the amplifying factor of the portfolio written in US Dollar starts with a roughly concave shape and increases linearly in $p$ afterward, which is compatible with the pattern implied by exponential distribution. The behavior of the portfolio written in Singapore Dollar exhibits patterns closer to the standard normal distribution, that is, concave at the beginning and slightly convex afterwards (see Figures 5.1(b) and 5.1(c) for a visual comparison). Since the sample size is too small to accurately estimate the distribution characteristics, it is more difficult to identify the distribution patterns when the holding period is 20 days. However, the amplifying factor calculated for monthly data are smaller than for daily data. This is consistent with the thinner tail featured by the long horizon data.

Figure 13 displays the shape of function $g$ used to pass from the Tail-VaR to the VaR for the equally weighted currency portfolio. Similarly, the thicker solid line represents the result at daily horizon. For a one-day period, the function $g$ is close to a linear function under both numeraires. However, the function $g$ features steps when the horizon increases to 20 days. This is due to the smaller number of observations.

To analyze the sensitivity of the result above to portfolio allocation, we now consider different portfolio allocations. More precisely, the (negative) portfolio return is constructed as:

$$x_{j,t} = -(a^1 \tilde{x}^1_{j,t} + (1 - a^1) \tilde{x}^2_{j,t}), \quad \text{for } j = u, s.$$

The weight $a^1$ is chosen to get portfolios with only Japanese Yen ($a^1 = 0$), more Japanese Yen ($a^1 = 0.2$), more Hongkong Dollar ($a^1 = 0.8$) and only Hongkong Dollar ($a^1 = 1$). The patterns of function $g$ are provided in Figure 14. Figures of the first column correspond to the results written in US Dollar and those of the second column show the outputs written in Singapore-Dollar. Figures in rows 1,2,3,4 display the patterns associated with portfolios
when $a^1 = 0, 0.2, 0.8$ and $1$, respectively. The near-linearity feature is preserved for almost all portfolios at daily horizon. The only exception occurs in the portfolio including Hongkong Dollar only. When using US Dollar as basic numeraire, we identify a few jumps in function $g$. For comparison, their twenty-day counterparts are also plotted. Because of the small sample size, it is not surprising that we observe step functions in all cases.

Finally, the slope parameter $\alpha$ is estimated for the equally weighted (negative) portfolio return with $p_0 = 0.005$ and $p_1 = 0.2$. We consider various horizons, that are $k = 1, 5, 10, 15$ and 20 days. The estimated values are provided in Figure 15. Returns with longer horizons tend to have higher $\alpha$, even though the pattern may not be monotonic. This can be due to the thinner tail featured by longer horizon data.

### 2.6 Concluding remarks

This paper provides a unified framework for analyzing distortion risk measures, including as special risk measures the VaR and Tail-VaR. Indexing the distortion risk measure as a function of the distortion parameter $p$, we study the sensitivity of the risk measure with respect to a change of $p$. Since $p$ can be interpreted as parameter representing risk aversion and/or pessimism, the sensitivity measures the marginal effects on risk measures of slight adjustment of risk (or pessimism) level. Moreover, for special examples such as Tail-VaR, the sensitivity also serves as partial measure of the accuracy for its nonparametric estimator. Applying a Functional Limit Theorem, we derive the asymptotic properties of the nonparametric estimators of distortion risk measures and their sensitivities with respect to the pessimism parameter. Under standard regularity conditions, both distortion risk measures and their sensitivities are asymptotically Gaussian. Closed-form expressions for the asymptotic variances are derived for specific examples such as VaR, Tail-VaR and Proportional Hazard distortion risk measure.

Robust risk management requires control of various risk measures. Thus, the knowledge of relationship between different risk measures is important for selecting appropriate risk control strategies. In this paper, we emphasize the link between the VaR and Tail-VaR. On the one
hand, for a given risk level $p$, the Tail-VaR can be derived by multiplying the VaR with an amplifying factor. We show that this amplifying factor is independent of $p$, if and only, if the underlying distribution is Pareto. Defining the amplifying factor as function of $p$, we observe that different distributions usually imply different patterns of the amplifying factor. Thus, the shape of the amplifying factor can be a criterion for identifying the proper underlying distribution. On the other hand, the VaR and Tail-VaR are related through their risk levels by some transformation $g$. We introduce a nonparametric estimator of this transformation, derive its asymptotic properties and propose a specification test for the hypothesis of linear transformation $g$.

The results are illustrated by considering currency portfolios written in different numeraires. The linearity of the transformation of risk levels is observed for a large range of portfolio allocations.

Whereas the analysis considered in this paper is based on the assumption of i.i.d. observations, the extensions to the dynamic setting can be considered. First, we can still consider the historical DRMs suggested by the regulators, but derive their asymptotic behaviors when the portfolio returns are serially dependent. Second, we can introduce dynamic version of the DRM based on dynamic quantile functions. These extensions will likely be based on parametric specifications such as the dynamic additive quantile model (DAQ) proposed by Gourieroux and Jasiak (2006b).

Appendices

A.1 Analogy principle and empirical process

Let us briefly review the relevant results of analogy principle and empirical process. For a systematic analysis, including the appropriate regularity conditions, we refer to Pollard (1984), Shorack and Wellner (1986), Manski (1988), and van der Vaart and Wellner (1996).
Analogy principle

The analogy principle has been popularized in econometrics by Manski (1988). Let us consider i.i.d. observations $x_1, ..., x_T$ with common cumulative distribution function $F$, and a parameter of interest $\theta$. The analogy principle looks for an interpretation of parameter $\theta$, that is, a relationship explaining how $\theta$ is related to distribution $F$. This relationship takes the form: $h(F; \theta) = 0$, where $h$ is a known function. Then, this relation is used to get an estimator $\hat{\theta}_T$ of $\theta$ by replacing $F$ by the sample cdf $\tilde{F}_T$. Thus, $\hat{\theta}_T$ is defined as a solution of $h(\tilde{F}_T; \hat{\theta}_T) = 0$. If $h$ is invertible with respect to $\theta$ and “continuous” with respect to $(F, \theta)$, this approach provides a consistent estimator of $\theta$; if $h$ is first-order “differentiable” with respect to $(F, \theta)$, the estimator $\hat{\theta}_T$ is asymptotically Gaussian. However, the estimator $\hat{\theta}_T$ is not necessarily asymptotically efficient. Indeed, a given parameter can admit a lot of alternative interpretations. In a second step, it is important to look for an “optimal” interpretation of the parameter, that is, an interpretation leading to asymptotic (semi-parametric or parametric ) efficiency. This is the so-called empirical likelihood approach introduced in econometrics by Kitamura (1997), Kitamura and Stutzer (1997).

Empirical process

Let $x_1, ..., x_T$ be a random sample of i.i.d. one-dimensional observations. Their common cumulative distribution function (cdf) is denoted by $F(x) = \mathcal{P}[x_t \leq x]$. The observations $x_1, ..., x_T$ can be used to define the empirical process $\tilde{F}_T$ by:

$$\tilde{F}_T(x) = \frac{1}{T} \sum_{t=1}^{T} 1_{(x_t \leq x)},$$  \hspace{2cm} (2-A-1)

where $1(A)$ is the indicator function of event $A$.

The use of the analogy principle is based on the limiting behavior of the empirical process $\tilde{F}_T$. Loosely speaking, under standard regularity conditions, the empirical process is consistent and asymptotically Gaussian. The convergence in distribution of the empirical process to a Gaussian process is with respect to the notion of weak convergence on $\mathcal{D}[0,1]$, the Skorohod
space of right-continuous functions on $[0,1]$ with left limits (see e.g. van der Vaart and Wellner (1996)). This type of convergence is denoted by $\Rightarrow$.

**Functional Limit Theorem.** Let $x_1, \ldots, x_T$ be i.i.d. one-dimensional random observations, we have:

$$\sqrt{T} [\tilde{F}_T(x) - F(x)] \Rightarrow B(F(x)),$$

where $B(u) = W(u) - uW(1)$ is a Brownian bridge, and $W(u)$ a Brownian motion on $[0,1]$.

Defined as a linear combination of values of standard Brownian motion, the process $B$ is also Gaussian with zero-mean. Its covariance operator is:

$$COV(B(u_1), B(u_2)) = u_1 \wedge u_2 - u_1 u_2, \quad \text{for } u_1, u_2 \in [0,1],$$

where $u_1 \wedge u_2$ denotes the minimum of $u_1$ and $u_2$. These properties of the Brownian bridge are useful in deriving the limiting distribution of sample moments. More precisely, the asymptotic normality of the empirical cdf implies the asymptotic normality of any sample moments (under integrability conditions). By analogy principle, any theoretical moment of $\int_R g(x)dF(x)$ of a $p$-dimensional integrable function $g$ can be estimated by the following stochastic integral, which equals the associated sample moment:

$$\int_R g(x)d\tilde{F}_T(x) = \frac{1}{T} \sum_{t=1}^T g(x_t). \quad (2-A-2)$$

By applying the Functional Limit Theorem, the sample moments are such that:

$$\sqrt{T} \left[ \int_R g(x)d\tilde{F}_T(x) - \int_R g(x)dF(x) \right] \Rightarrow \int_R g(x)dB(F(x)). \quad (2-A-3)$$

The stochastic integral $\int_R g(x)dB(F(x))$ is Gaussian, zero-mean, with variance-covariance matrix $V[\int_R g(x)dB(F(x))] = Vg(x)$, which is the standard Central Limit Theorem.
A.2 Relaxation of i.i.d. assumption

A regulator is often interested in risk measures calculated from the marginal empirical distribution. However, it is shown in the literature that financial (negative) returns are often serially dependent at least for the second-order moment. The empirical process under dependent time series still converges in distribution to a Gaussian process, whenever the time series is stationary and satisfies appropriate ergodicity condition [see e.g. Arcones and Yu (1994)]. The stationary version of the functional limit theorem is given below.

**Functional Limit Theorem for stationary process.** For a stationary sequence $x_1, \ldots, x_T$ with marginal cdf $F$, we have:

$$\sqrt{T} \left[ \tilde{F}_T(x) - F(x) \right] \Rightarrow Z(F(x)),$$

where $Z(F(x))$ is a zero-mean Gaussian process with variance:

$$V[Z(F(x))] = F(x)(1 - F(x)) + 2 \sum_{j=2}^{k} E \left[ (1_{\{X_1 \leq x\}} - F(x)) \left( 1_{\{X_j \leq x\}} - F(x) \right) \right],$$

and covariance:

$$COV[Z(F(x)), Z(F(x'))] = F(x') - F(x) F(x') + 2 \sum_{j=2}^{k} E \left[ (1_{\{X_1 \leq x\}} - F(x)) \left( 1_{\{X_j \leq x'\}} - F(x') \right) \right],$$

where $k << T$ denotes the largest lag where $Cov(1_{X_i \leq x}, 1_{X_{i-k} \leq x}) \neq 0$.

This theorem can be used to extend the result of the paper to serially dependent data.

A.3 Preliminary lemmas

**Lemma 1.** Let us consider a random variable $X$ with continuous cdf $F$ and quantile function $Q = F^{-1}$, we have:
(i). \[
\int_{Q(1-p)}^\infty \frac{1}{p} [1 - F(x)] \, dx = TVaR(p) - VaR(p);
\]

(ii). \[
2 \int_{Q(1-p)}^\infty \frac{1}{p^2} x [1 - F(x)] \, dx = \frac{1}{p} \left[ E[X^2 | X \geq VaR(p)] - (VaR(p))^2 \right];
\]

(iii). \[
2 \frac{1}{p} \int_{Q(1-p)}^\infty Q(1-p) [1 - F(x)] \, dx = \frac{2VaR(p)}{p} \left[ TVaR(p) - VaR(p) \right].
\]

Proof. (i). By integrating by part, we get:

\[
\int_{Q(1-p)}^\infty \frac{1}{p} [1 - F(x)] \, dx = \frac{1}{p} x [1 - F(x)]\bigg|_{Q(1-p)}^\infty + \int_{Q(1-p)}^\infty x d \frac{F(x)}{p}
\]

\[
= E[X | X \geq Q(1-p)] - Q(1-p)
\]

\[
= TVaR(p) - VaR(p),
\]

by definition of VaR(p) and TVaR(p).

(ii). We have:

\[
2 \int_{Q(1-p)}^\infty \frac{1}{p^2} x [1 - F(x)] \, dx = \frac{1}{p^2} \int_{Q(1-p)}^\infty [1 - F(x)] \, dx^2
\]

\[
= \frac{1}{p} \left\{ \left[ x^2 \frac{1 - F(x)}{p} \right]_{Q(1-p)}^\infty - \int_{Q(1-p)}^\infty x^2 d \frac{[1 - F(x)]}{p} \right\}
\]

\[
= \frac{1}{p} \left[ E[X^2 | X \geq VaR(p)] - (VaR(p))^2 \right].
\]
(iii). By integrating by part, we have:

\[
2 \frac{1}{p^2} \int_{Q(1-p)}^{\infty} Q(1 - p)[1 - F(x)] \, dx \\
= 2 \frac{1}{p^2} Q(1 - p) x [1 - F(x)] \Bigg|_{Q(1-p)}^{\infty} - 2 \frac{Q(1 - p)}{p^2} \int_{Q(1-p)}^{\infty} x \, d [1 - F(x)] \\
= 2 \frac{VaR(p)}{p} \left[ TVaR(p) - VaR(p) \right].
\]

\[\square\]

### A.4 Variances of \( \sqrt{T}[TVaR_T(p) - TVaR(p)] \)

We have:

\[
V(\sqrt{T}[TVaR_T(p) - TVaR(p)]) \\
= \int_{Q(1-p)}^{\infty} \int_{Q(1-p)}^{\infty} \frac{F(x_1) \land F(x_2) - F(x_1)F(x_2)}{f(x_1)f(x_2)} \frac{1}{p^2} \, d F(x_1) \, d F(x_2) \\
= \int_{Q(1-p)}^{\infty} \int_{Q(1-p)}^{\infty} [F(x_1) \land F(x_2) - F(x_1)F(x_2)] \frac{1}{p^2} \, d x_1 \, d x_2 \\
= 2 \int_{Q(1-p)}^{x_2} \int_{Q(1-p)}^{x_2} F(x_1)(1 - F(x_2)) \frac{1}{p^2} \, d x_1 \, d x_2. \\
= 2 \int_{Q(1-p)}^{\infty} \int_{Q(1-p)}^{x_2} [1 - F(x_2)] \frac{1}{p^2} \, d x_1 \, d x_2 - 2 \int_{Q(1-p)}^{\infty} \int_{Q(1-p)}^{x_2} [1 - F(x_1)] [1 - F(x_2)] \frac{1}{p^2} \, d x_1 \, d x_2 \\
= 2 \frac{1}{p^2} \int_{Q(1-p)}^{\infty} \left[ x - Q(1 - p) \right] [1 - F(x)] \, dx - \left[ \frac{1}{p} \int_{Q(1-p)}^{\infty} [1 - F(x)] \, dx \right]^2. 
\] (2-A-4)

Thus, using the results of Lemma 1, we deduce:

\[
V(\sqrt{T}[TVaR_T(p) - TVaR(p)]) \\
= \frac{1}{p} \left[ E[X^2 | X \geq VaR(p)] - (VaR(p))^2 \right] - \frac{2VaR(p)}{p} \left[ TVaR(p) - VaR(p) \right] - \left[ TVaR(p) - VaR(p) \right]^2 \\
= \frac{V(X | X \geq VaR(p)) + (1 - p) \left[ TVaR(p) - VaR(p) \right]^2}{p}.
\]
It is straightforward to calculate the covariance between two Tail-VaR at different levels. For \( \tilde{p} > p \), we get:

\[
\text{COV}(\sqrt{T} \text{TVaR}_T(p), \sqrt{T} \text{TVaR}_T(\tilde{p})) = \int_{Q(1-p)}^{\infty} \int_{Q(1-\tilde{p})}^{\infty} [F(x_1) \land F(x_2) - F(x_1)F(x_2)] \frac{1}{pp} \, dx_1 \, dx_2
\]

\[
= 2 \int_{Q(1-p)}^{\infty} \int_{Q(1-\tilde{p})}^{x_2} F(x_1)(1 - F(x_2)) \frac{1}{pp} \, dx_1 \, dx_2.
\]

\[
= 2 \int_{Q(1-p)}^{\infty} \int_{Q(1-\tilde{p})}^{x_2} [1 - F(x_2)] \frac{1}{pp} \, dx_1 \, dx_2 - 2 \int_{Q(1-p)}^{\infty} \int_{Q(1-\tilde{p})}^{x_2} [1 - F(x_1)] [1 - F(x_2)] \frac{1}{pp} \, dx_1 \, dx_2
\]

\[
= 2 \frac{1}{pp} \int_{Q(1-p)}^{\infty} [x - Q(1-\tilde{p})] [1 - F(x)] \, dx - \left[ \frac{1}{p} \int_{Q(1-p)}^{\infty} [1 - F(x)] \, dx \right] \left[ \frac{1}{p} \int_{Q(1-\tilde{p})}^{\infty} [1 - F(x)] \, dx \right].
\]

By applying Lemma 1, we get:

\[
\text{COV}(\sqrt{T} \text{TVaR}_T(p), \sqrt{T} \text{TVaR}_T(\tilde{p})) = \frac{1}{\tilde{p}} \left[ E \left[ X^2 \, | \, X > \text{VaR}(p) \right] - \text{(VaR}(p))^2 \right] - \frac{2 \text{VaR}(\tilde{p})}{\tilde{p}} [(\text{TVaR}(p) - \text{VaR}(p)]
\]

\[
- \left[ \text{TVaR}(\tilde{p}) - \text{VaR}(\tilde{p}) \right] \left[ \text{TVaR}(p) - \text{VaR}(p) \right]
\]

\[
= \frac{1}{\tilde{p}} \left[ E \left[ X^2 \, | \, X > \text{VaR}(p) \right] - \text{(VaR}(p))^2 \right] - \frac{(2 - \tilde{p}) \text{VaR}(\tilde{p})}{\tilde{p}} [\text{TVaR}(p) - \text{VaR}(p)]
\]

\[
- \text{TVaR}(\tilde{p}) [(\text{TVaR}(p) - \text{VaR}(p)].
\]

(2-A-5)
Moreover, the covariance between $\sqrt{T}VaR_T(g(p))$ and $\sqrt{T}VaR_T(\bar{p})$ is given by:

\[
\text{COV}\left(\sqrt{T}VaR_T(g(p)), \sqrt{T}VaR_T(\bar{p})\right)
= \frac{1}{\bar{p}} \int_{Q(1-\bar{p})}^{\infty} \frac{(1-g(p)) \wedge F(x) - (1-g(p))F(x)}{f(Q(1-g(p))} \, dx
= \frac{1}{\bar{p}f(Q(1-g(p))} \int_{Q(1-\bar{p})}^{Q(1-g(p))} g(p)F(x) \, dx + \frac{1}{\bar{p}f(Q(1-g(p))} \int_{Q(1-g(p))}^{\infty} (1-g(p))(1-F(x)) \, dx
= \frac{g(p)}{\bar{p}f(Q(1-g(p))} \left\{ (1-g(p))VaR(g(p)) - (1-\bar{p})VaR(\bar{p}) - E\left[ X|VaR(\bar{p}) < X < VaR(g(p)) \right] \right\}
+ \frac{g(p)(1-g(p))}{\bar{p}f(Q(1-g(p))} \left[ TVaR(g(p)) - VaR(T) \right]
= \frac{g(p)}{\bar{p}f(Q(1-g(p))} \left\{ (1-g(p))TVaR(g(p)) - (1-\bar{p})VaR(\bar{p}) - E\left[ X|VaR(\bar{p}) < X < VaR(g(p)) \right] \right\}
\]

(2-A-6)

### A.5 Variance of $\sqrt{T}[PH_T(p) - PH(p)]$

For Proportional Hazard distortion, we define the expectation and probability operators as:

$$E_p^*[g(X)] = - \int_{R} g(x) (1-F(x))^p \, dx.$$  

The asymptotic variance of $\sqrt{T}[PH_T(p)]$ is given by:

\[
V(\sqrt{T}[PH_T(p) - PH(p)])
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ F(x_1) \wedge F(x_2) - F(x_1)F(x_2) \right] p^2(1-F(x_1))^{p-1}(1-F(x_2))^{p-1} \, dx_1 \, dx_2
= 2p^2 \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} F(x_1)(1-F(x_1))^{p-1}(1-F(x_2))^p \, dx_1 \, dx_2.
\]

We can express variance in terms of the distortion operators with varying $p$, to make the estimation straightforward. Let us integrate by parts with respect to $x_1$, we get:

\[
V(\sqrt{T}[PH_T(p) - PH(p)])
\]
\[ 2p^2 \int_{-\infty}^{\infty} x_1 (1 - F(x_1))^{p-1} \left[ x_2 (1 - F(x_2))^p \right] d x_2 
- 2p^2 \int_{-\infty}^{\infty} x_1 d \left[ \int_{-\infty}^{\infty} x_2 (1 - F(x_2))^{p-1} (1 - F(x_2))^p \right] d x_2 
= 2p^2 \int_{-\infty}^{\infty} x F(x) (1 - F(x))^{2p-1} d x - 2p^2 \int_{-\infty}^{\infty} x_1 (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1} 
= p^2 \int_{-\infty}^{\infty} F(x) (1 - F(x))^{2p-1} d x - 2p^2 \int_{-\infty}^{\infty} x_1 (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1} 
= -p^2 \int_{-\infty}^{\infty} x^2 d F(x) (1 - F(x))^{2p-1} - 2p^2 \int_{-\infty}^{\infty} x_1 (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1} 
= p^2 \left[ \int_{-\infty}^{\infty} x^2 (1 - F(x))^{2p} - \int_{-\infty}^{\infty} x^2 d (1 - F(x))^{2p-1} \right] 
- 2p^2 \int_{-\infty}^{\infty} x_1 (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1}, 
\]

which is equal to:

\[
\begin{align*}
&\begin{cases}
p^2 \left[ E_{2p-1}^* (X^2) - E_{2p}^* (X^2) \right] - 2p^2 \int_{-\infty}^{\infty} x_1 \int_{x_1}^{\infty} (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1} & p \geq 0.5, \\
p^2 \int_{-\infty}^{\infty} F(x) (1 - F(x))^{2p-1} d x^2 - 2p^2 \int_{-\infty}^{\infty} x_1 \int_{x_1}^{\infty} (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1} & p < 0.5.
\end{cases}
\end{align*}
\]

By integrating by part the second term with respect to \(x_2\), we have:

\[
2p^2 \int_{-\infty}^{\infty} x_1 \int_{x_1}^{\infty} (1 - F(x_2))^p d x_2 d F(x_1) (1 - F(x_1))^{p-1}
\]

\[
= -2p^2 \int_{-\infty}^{\infty} x^2 (1 - F(x))^p d F(x) (1 - F(x))^{p-1} - 2p^2 \int_{-\infty}^{\infty} \int_{x_1}^{\infty} x_1 x_2 d (1 - F(x_2))^p d F(x_1) (1 - F(x_1))^{p-1}
\]

\[
= -2p^2 \left[ \frac{1}{2} E_{2p}^* (X^2) - \frac{p - 1}{2p - 1} E_{2p-1}^* (X^2) \right] - p^2 \left[ 2 \int_{-\infty}^{\infty} \int_{x_1}^{\infty} x_1 x_2 d (1 - F(x_2))^p d (1 - F(x_1))^{p-1} 
- 2 \int_{-\infty}^{\infty} \int_{x_1}^{\infty} x_1 x_2 d (1 - F(x_2))^p d (1 - F(x_1))^{p-1} \right] 
= 2p^2 \left[ \frac{p - 1}{2p - 1} E_{2p-1}^* (X^2) - \frac{1}{2} E_{2p}^* (X^2) \right] - p^2 E_{p}^* (X) E_{p}^* (X) + p^2 \left[ E_{p}^* (X) \right]^2.
\]

Putting together, this complete the proof.
A.6 Asymptotic expansion of the implied pessimism parameter

Since:

\[ \Pi^\circ = \Pi(p; Q) = \Pi(\hat{p}_T; \tilde{Q}_T), \]

we get:

\[
0 = \Pi(p; Q) - \Pi(\hat{p}_T; \tilde{Q}_T) \\
= \Pi(p; Q) - \Pi(p; \tilde{Q}_T) + \Pi(p; \tilde{Q}_T) - \Pi(\hat{p}_T; \tilde{Q}_T) \\
= \Pi(p; Q) - \Pi(p; \tilde{Q}_T) + \Pi(p; Q) - \Pi(\hat{p}_T; Q) + o_p(1) \\
= \sqrt{T}[\Pi(p; Q) - \Pi(p; \tilde{Q}_T)] + \frac{\partial \Pi}{\partial p}(p; Q)\sqrt{T}(p - \hat{p}_T) + o_p(1),
\]

which can be rewritten as:

\[
\sqrt{T}(\hat{p}_T - p) = - \left(\frac{\partial \Pi}{\partial p}(p; Q)\right)^{-1} \sqrt{T}[\Pi(p; \tilde{Q}_T) - \Pi(p; Q)] + o_p(1).
\]

A.7 Proof of Proposition 9

The limiting process comes from the properties of the empirical process and the asymptotic Gaussian distribution follows immediately. Let us now derive the asymptotic variance.

\[
V\left(\sqrt{T} \left[\tilde{g}_T(p) - g(p)\right]\right) = (1 - g(p))g(p) + f[TVaR(p)]^2 V\left(\sqrt{T} [\tilde{TVaR}_T(p) - TVaR(p)]\right) \\
- 2 f[TVaR(p)] \frac{1}{p} \int_{1-p}^{1} \frac{(1 - g(p)) \wedge u - (1 - g(p)) u}{f[Q(u)]} du.
\]

Since \( g(p) \leq p \), the covariance term can be rewritten as:

\[
2 f[TVaR(p)] \frac{1}{p} \int_{1-p}^{1} \frac{(1 - g(p)) \wedge u - (1 - g(p)) u}{f[Q(u)]} du
\]

\[
= 2 f[TVaR(p)] \frac{1}{p} \left\{ \int_{Q(1-g(p))}^{Q(1-p)} F(x) g(p) dx + \int_{Q(1-g(p))}^{\infty} (1 - g(p))(1 - F(x)) dx \right\}
\]
\[
= 2 f [TVaR(p)] \frac{1}{p} \left\{ g(p) \left[ x F(x) \right] Q(1-g(p)) - \int_{Q(1-p)}^{Q(1-g(p))} x d F(x) \right\} \\
+ g(p) [1 - g(p)] [TVaR(g(p)) - VaR(g(p))] \right\}
\]

\[
= 2 f [TVaR(p)] \frac{1}{p} \left\{ g(p) \left[ VaR(g(p)) (1 - g(p)) - VaR(p)(1 - p) \\
- [p - g(p)] E \left[ X \left| VaR(p) \leq X \leq VaR(g(p)) \right. \right] \right\} + g(p) [1 - g(p)] [TVaR(g(p)) - VaR(g(p))] \right\}
\]

\[
= 2 f [TVaR(p)] \frac{1}{p} \left\{ -g(p) VaR(p)(1 - p) + g(p) [1 - g(p)] TVaR(g(p)) \\
- [p - g(p)] E \left[ X \left| VaR(p) \leq X \leq VaR(g(p)) \right. \right] \right\}
\]

\[
= 2 f [TVaR(p)] \frac{1}{p} \left\{ g(p) \left[ [1 - g(p)] TVaR(g(p)) - VaR(p)(1 - p) \\
- [p - g(p)] E \left[ X \left| VaR(p) \leq X \leq VaR(g(p)) \right. \right] \right\}.
\]

The result follows.
Chapter 3

Efficient Portfolio Analysis Using Distortion Risk Measures

3.1 Introduction

Portfolio choice theory considers criteria representing the trade-off between portfolio returns and risks. These criteria can be an expected utility (see e.g. von Neumann and Morgenstern, 1944), which expresses investors’ preferences through the valuation of future wealth level, or can be the variance adjusted expected returns in the mean-variance framework (Markowitz, 1952). While expected utility based framework is theoretically appealing, the mean-variance framework is more attractive to practitioners, due to its simplicity. However, on the one hand, investors’ preferences may be too complicated to be represented by a class of utility function and expected utility theory can generate paradoxical decisions. On the other hand, the variance is not an appropriate measure of risk in general. It values equally positive and negative deviations from mean, which is counter-intuitive as capital gains are often preferred by an investor. Moreover, variance itself is not sufficient to capture the uncertainty of return if the underlying distribution is far from Gaussian. Other features, such as skewness and extreme tails, have to be taken into account. For these reasons, other measures of risk have been introduced for regulatory purpose in Basel II (and Solvency II), and, as a consequence, will likely be used for risk management.
These measures can be quantiles of the portfolio returns distribution (the so-called Value-at-Risk or VaR), other quantile based measures such as the Tail-VaR (also called expected shortfall, or conditional tail expectation), or more generally a distortion risk measure (DRM) (see Wang, 1996).

As these risk measures are often used to control the extreme losses, it is important to know how the distribution of portfolio returns can be influenced by a constraint on such a measure (see e.g. Basak and Shapiro, 2001, Bassett et al., 2004) and if a given portfolio is efficient under such a framework. This chapter considers the problem of portfolio management in a DRM-DRM setup, which uses one DRM as the objective function and another DRM to define the constraint. As both expected return and some popular risk measures (e.g. VaR and expected shortfall) are examples of DRM, the DRM-DRM framework embeds the mean-VaR, mean-TailVaR and TailVaR-VaR setups as special cases. Moreover, since the DRM is introduced based on a preference theory dual to the expected utility theory (see Yaari, 1987), the DRM-DRM framework is analogous to an expected utility optimization problem subject to a constraint imposed by a regulator, a problem studied by Gourieroux and Monfort (2005).

In Section 3.2, we introduce two nonparametric estimators of the sensitivity of distortion risk measures with respect to the portfolio allocation, called delta-DRM, and derive their asymptotic properties. Their finite sample properties are compared by a simulation study. In Section 3.3, we consider the portfolio allocation problem, which minimizes a distortion risk measure under a budget constraint and a restriction on another distortion risk measure. The framework includes the minimization of the VaR (or Tail-VaR) under an expected return restriction as well as the minimization of Tail-VaR under a VaR restriction. Then, we explain how to estimate the efficient portfolio allocation nonparametrically. Tests of efficiency of a given portfolio allocation are developed in Section 3.4. We discuss alternative test procedures based on a comparison of either risky allocations, or delta-DRM, or realized extended Sharpe (RES) performances. An analysis of finite sample properties of the estimators of the efficient allocation and of the extended Sharpe ratio is presented in Section 3.5 for a TailVaR-VaR setup. By considering Gaussian simulated returns, we can also compare the nonparametric estimators introduced in
our paper with the standard mean-variance parametric estimators. Section 4.4 concludes.

3.2 Sensitivity of distortion risk measures

In this section, we provide an expression of the sensitivity of distortion risk measures (or delta-DRM), which, loosely speaking, is a weighted expectation of the VaR sensitivity [called delta-VaR in the literature (see Garman, 1996)] with respect to the portfolio allocation. Then, we introduce two nonparametric estimators of delta-DRM and derive their asymptotic properties. For expository purpose, we drop the time index from the expression of returns and allocations in the definition of delta-DRM. The time index is reintroduced when the estimation problems are considered.

3.2.1 Expression of the delta-DRM

The DRMs are generally applied to portfolio loss (and profit) instead of profit (and loss). More precisely, we consider the negative of the portfolio return $y(a) = -a^\prime r$, where $a$ denotes the vector of allocations and $r$ the $d \times 1$ primitive asset returns. For convenience, we denote $x = -r$ as the primitive asset losses, and thus, $y(a) = a^\prime x$. Throughout the paper, we use $(X, Y)$ for the random variables and the lower cases for their realized values. A DRM for this portfolio is defined as:

$$\Pi(H, a; F) = \int_0^1 Q(1 - u; a) d H(u), \quad (3.2.1)$$

where $F$ is the joint distribution of $X$, $Q(\cdot; a)$ is the quantile function of $Y(a) = a^\prime X$, that is the inverse of the cdf $G(\cdot; a)$ of the (negative) portfolio return. $H$ is the cumulative distribution function of a positive measure, called the distortion function (or Choquet capacity). As special cases, the VaR with risk level $p$ is obtained when $H(u; p) = 1_{(u \geq p)}$ corresponds to a point mass at $p$, and the Tail-VaR with risk level $p$ is obtained when $H(u; p) = (u/p) \land 1$ corresponds to the uniform distribution on $[0, p]$ [the expectation corresponds to the limiting case $H(u) = u$]. In both examples above, the distortion function is parametrized by an additional parameter $p$.
and is used to define a family $\Pi(p; a, F)$ of DRMs when $p$ varies.

The delta-DRM provides the marginal contribution of a particular asset to the risk measure of the portfolio, and can be written as:

$$
\frac{\partial \Pi}{\partial a}(H, a; F) = \frac{\partial}{\partial a} \int_0^1 Q(1 - u; a) d H(u) \\
= \int_0^1 \frac{\partial Q}{\partial a}(1 - u; a) d H(u).
$$

(3.2.2)

An explicit expression of VaR sensitivity with respect to portfolio allocation $a$ has been derived in Gourieroux et al. (2000):

$$
\frac{\partial Q}{\partial a}(1 - u; a) = E[X|a'X = Q(1 - u; a)].
$$

(3.2.3)

Thus, we get:

$$
\frac{\partial \Pi}{\partial a}(H, a; F) = \int_0^1 E[X|a'X = Q(1 - u; a)] d H(u).
$$

(3.2.4)

Let us now discuss some features of the delta-DRM with useful implications for portfolio analysis.

i) The delta-DRM is homogeneous of degree zero in $a$. Indeed, the DRM is homogeneous of degree one. This is due to the homogeneity of the quantile function, since we have $a' \frac{\partial Q}{\partial a}(\cdot; a) = E[a'X|a'X = Q(\cdot; a)] = Q(\cdot; a)$, which is the Euler characterization of a homogeneous function of degree one.

ii) The delta-DRM is a weighted expectation of delta-VaRs at all risk levels. For instance, the delta-VaR($p$) itself is a weighted expectation with weight equal to 1 at level $p$ and equal to 0 otherwise. If $H$ is differentiable, the expression (3.2.4) implies that the weighting function is the distortion density defined as $\partial H(u)/\partial u$. As shown in the next subsection, this facilitates the estimation of the delta-DRM once a delta-VaR estimator is available. Indeed, we can simply plug the delta-VaR estimator within the expectation expression (3.2.4).

iii) The delta-DRM can be alternatively interpreted as the expectation of the product between the distortion density and the primitive asset loss. More precisely, if function $H$ is continuous
and differentiable, we get the following expression (see A.2):

\[
\frac{\partial \Pi}{\partial a}(H, a; F) = E \left[ X \frac{\partial H \left(1 - G(a'X; a)\right)}{\partial u} \right].
\] (3.2.5)

The delta-DRM is equal to the expected co-movement between the primitive asset loss and the subjective perception, which is represented by the weights assigned to the losses. Indeed, the risk contribution of a particular asset is influenced jointly by the potential loss of the asset and how the risk measure is defined. The delta-DRM associated with the identity distortion function is simply the expected loss of the primitive asset.

### 3.2.2 Nonparametric estimators of the delta-DRM

We have seen in Section 3.2.1 that the delta-DRM can be written as a linear combination of delta-VaRs. In this section, we introduce nonparametric estimators of the delta-DRM for i.i.d. returns. A first nonparametric estimator of the delta-DRM is derived by substituting a kernel estimator of the delta-VaR into the expression (3.2.4). An alternative estimator of the delta-DRM is deduced from expression (3.2.5), under additional conditions on the distortion function.

**I. Kernel estimator of the delta-DRM**

Let us consider a sequence of observed (negative) i.i.d. asset returns \(x_1, \ldots, x_T\). The delta-VaR admits the interpretation (3.2.3) as a conditional expectation. A kernel estimator of the delta-VaR is defined as follows:

\[
\frac{\partial \tilde{Q}_T}{\partial a}(\cdot; a) = \frac{1}{T h_T} \sum_{t=1}^{T} x_t \frac{1}{h_T} \sum_{t=1}^{T} k \left( \frac{a'x_t - \tilde{Q}_T(\cdot; a)}{h_T} \right),
\] (3.2.6)

---

1 The result still holds, if the distortion function \(H\) is continuous and almost everywhere differentiable. For instance, the distortion function associated with Tail-VaR is not differentiable when \(u = p\), but the equality (3.2.5) is still valid with the right derivative of \(H\). However, the relation is not valid for the application to the delta-VaR, in which \(H\) is not continuous.

2 An i.i.d. assumption is also used in the quantile regression framework of Bassett et al. (2004) and corresponds to the historical simulation approach suggested by the regulator in Basel II.
where \( \hat{Q}_T(u; a) = \inf \left\{ y : \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(a^\prime x_t \leq y) \geq u \right\} \) is the empirical quantile, \( k \) is a symmetric kernel function such that \( \int k(u)du = 1 \) and \( \int u k(u)du = 0 \), and \( h_T \) denotes the bandwidth. The estimator (3.2.6) is a Nadaraya-Watson estimator of the regression function \( q \rightarrow \mathbb{E}[X|a^\prime X = q] \) after substitution of the theoretical quantile \( Q(\cdot; a) \) by its sample counterpart (see e.g. Yatchew, 2003, Chapter 3, for the definition and properties of the Nadaraya-Watson estimator).

The kernel estimator of the delta-DRM is defined as:

\[
\frac{\partial \tilde{\Pi}_T}{\partial a}(H, a) = \int_0^1 \frac{\partial \tilde{Q}_T}{\partial a}(1 - u; a) \, d H(u) = \int_0^1 \frac{1}{T h_T} \sum_{t=1}^{T} x_t k \left( \frac{a^\prime x_t - \hat{Q}_T(1-u; a)}{h_T} \right) d H(u). \tag{3.2.7}
\]

Since the empirical quantile function is a stepwise function, the integral (3.2.7) can be replaced by a summation. We get the equivalent expression:

\[
\frac{\partial \tilde{\Pi}_T}{\partial a}(H, a) = \sum_{i=1}^{T} \sum_{t=1}^{T} x_t k \left( \frac{a^\prime x_t - y^*_i(a)}{h_T} \right) \left[ H(1 - \frac{i-1}{T}) - H(1 - \frac{i}{T}) \right], \tag{3.2.8}
\]

where \( y^*_1(a) < y^*_2(a) < \cdots < y^*_T(a) \) are the order statistics of the (negative) observed portfolio returns associated with allocation \( a \). In order to preserve the homogeneity property, a consistent kernel estimator of the DRM can be defined as \( \tilde{\Pi}_T(H, a) = a^{\frac{\partial \tilde{\Pi}_T}{\partial a}}(H, a) \).

The expression (3.2.7) is valid for any distortion function \( H \), implying that the kernel estimator can be applied in the general framework, which includes delta-VaR as a special case. Although the rate of convergence of the kernel estimator of delta-VaR is \( \sqrt{T h_T} \), the kernel estimator of a delta-DRM will converge at rate \( \sqrt{T} \) due to its integral expression as shown in Section 3.2.3 (Similar results can be found in Ait-Sahalia (1993) and Gagliardini and Gourieroux (2006)).

II. Empirical estimator of delta-DRM

Alternatively, if the distortion function \( H \) is continuous and differentiable, the delta-DRM can be estimated by the sample analog of expression (3.2.5). The empirical estimator of the delta-DRM is defined as the sample average of the products between the observed (opposite) primitive
asset returns and distortion densities,\(^3\)

\[
\frac{\partial \hat{\Pi}_T}{\partial a}(H; a) = \frac{\partial \Pi}{\partial a}(H, a, \hat{F}_T) = \frac{1}{T} \sum_{t=1}^{T} x_t \left( \frac{\partial H \left( 1 - \hat{G}_T(a'x_t; a) \right)}{\partial u} \right),
\]  

(3.2.9)

where \(\hat{G}_T(y; a) = \frac{1}{T} \sum_{t=1}^{T} 1_{(a'x_t \leq y)}\) is the sample cdf of \(Y(a)\). This estimator is also equal to:

\[
\frac{\partial \hat{\Pi}_T}{\partial a}(H; a) = \frac{1}{T} \sum_{t=1}^{T} x^*_t \left( \frac{\partial H \left( 1 - t/T \right)}{\partial u} \right),
\]  

(3.2.10)

where \(x^*_t, t = 1, \ldots, T\), are the observations of (opposite) primitive asset returns reordered according to the order statistics of (opposite) portfolio returns \((y^*_t(a))\).\(^4\) For instance, the estimator for delta-TVaR can be written as:

\[
\frac{\partial \hat{TVaR}_T}{\partial a}(p; a) = \frac{1}{T_p} \sum_{t=\lceil T(1-p) \rceil}^{T} x^*_t,
\]

where \([\cdot]\) denotes integer part. The empirical estimator is easier to compute and will also converge at the standard parametric rate \(\sqrt{T}\).

### 3.2.3 Asymptotic properties of the delta-DRM estimators

The accuracy of the estimated risk measures and their sensitivities are often overlooked by practitioners. However, this potential estimation error can put investors in a very risky position. This issue has been categorized as the estimation risk in the risk management terminology of the Basle Committee. In this section, we derive the asymptotic distributions of both nonparametric estimators of the delta-DRM.

#### I. Asymptotic distributions of the estimators

\(^3\)In some special cases (e.g. \(H(u; p) = u^p\), for \(p < 1\)), the estimator can be unbounded since the empirical distortion density goes to infinity when \(t = T\). This problem can be solved by replacing \(T\) by \(T + 1\) in the distortion density.

\(^4\)This estimator is a L-statistics constructed on two dependent series \(x_t\) and \(y_t(a)\) (see e.g. Shorack and Wellner, 1986, Chapter 19).
The kernel estimator of the delta-DRM is defined as an integral of Nadaraya-Watson estimator of the delta-VaR, which is known to be asymptotically Gaussian. The empirical estimator of the delta-DRM is defined as a sample average. Thus, with i.i.d. observations \( x_t, t = 1, \ldots, T \), the Central Limit Theorem applies directly. Thus, both estimators will converge to Gaussian processes. These results are summarized in the proposition below (see Appendices C.1–C.4).

**Proposition 12.** For an i.i.d. sequence \( x_t = (x_{1t}, \ldots, x_{dt}) \), \( t = 1, \ldots, T \), following a joint distribution \( F(X) = C(F^1(X^1), \ldots, F^d(X^d)) \) with copula function \( C \) and marginal cdf \( F^i(X^i) \), if \( H \) is differentiable, we have:

\[
\sqrt{T} \left( \frac{\partial \hat{\Pi}}{\partial a}(H, a) - \frac{\partial \Pi}{\partial a}(H, a, F) \right) \\
\Rightarrow \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}} \frac{\partial}{\partial y} (E[X|a'X = y]) \nabla H(y; a) 1_{a'x \leq y} dy \right\} dK(F^1(x^1), \ldots, F^d(x^d)) - \int_{\mathbb{R}^d} \left\{ x - E[X|a'X = a'x] \right\} \nabla H(a'x; a) dK(F^1(x^1), \ldots, F^d(x^d)).
\]

If \( H \) is twice differentiable, we have:

\[
\sqrt{T} \left( \frac{\partial \hat{\Pi}}{\partial a}(H, a) - \frac{\partial \Pi}{\partial a}(H, a, F) \right) \\
\Rightarrow \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \frac{\partial^2}{\partial u^2} H(1 - G(a'z; a)) \right\} dK(F^1(x^1), \ldots, F^d(x^d)) + \int_{\mathbb{R}^d} \left\{ x - E[X|a'X = a'x] \right\} \nabla H(a'x; a) dK(F^1(x^1), \ldots, F^d(x^d)).
\]

where \( \Rightarrow \) denotes the weak convergence of processes,\(^6\) \( \nabla H(y; a) = \frac{\partial H(1-G(y,a))}{\partial u} \), and \( K \) is a

---

\(^5\)We assume that the distortion function is continuous and differentiable. The asymptotic variances associated with the VaR and Tail-VaR are provided in Appendix A.3.

\(^6\)The processes are indexed by allocation \( a \). When the distortion function depends on some pessimism parameter, the results can be extended to processes doubly indexed by the allocation and the pessimism parameter (see Gourieroux and Liu, 2006 for the analysis of sensitivity w.r.t. pessimism parameter).
multi-dimensional Brownian bridge on \([0, 1]^d\).

The i.i.d. assumption is crucial for obtaining the proposition above. This assumption is implicitly used in the historical simulation method suggested by Basel committee. On the other hand, a similar result can be derived when process \((x_t)\) is stationary, in which case, Central Limit Theorem for stationary processes must be applied (see Gourieroux and Liu, 2006 Appendix A.3 for a brief discussion).

II. Estimation of the asymptotic variance

The variance-covariance matrices of both delta-DRM estimators corresponding to portfolio allocation \(\mathbf{a}\) are given by (see Appendix A.3):

\[
\Omega(\mathbf{a}, \mathbf{a}) = \lim_{T \to \infty} V \left[ \sqrt{T} \left( \frac{\partial \Pi_T}{\partial \mathbf{a}}(H, \mathbf{a}) - \frac{\partial \Pi}{\partial \mathbf{a}}(H, \mathbf{a}, F) \right) \right] = V \left[ A(\mathbf{a}' \mathbf{X}, H) - (\mathbf{X} - E[\mathbf{X} | \mathbf{a}' \mathbf{X}]) \nabla H(\mathbf{a}' \mathbf{X}; \mathbf{a}) \right]
\]

\[
\Sigma(\mathbf{a}, \mathbf{a}) = \lim_{T \to \infty} V \left[ \sqrt{T} \left( \frac{\partial \hat{\Pi}_T}{\partial \mathbf{a}}(H, \mathbf{a}) - \frac{\partial \Pi}{\partial \mathbf{a}}(H, \mathbf{a}) \right) \right] = V \left[ \mathbf{X} \nabla H(\mathbf{a}' \mathbf{X}; \mathbf{a}) - A(\mathbf{a}' \mathbf{X}, H) \right],
\]

where,

\[
A(y, H) = \int z 1_{(y \leq \mathbf{a}' z)} \frac{\partial^2 H(1 - G(\mathbf{a}' z; \mathbf{a}))}{\partial u^2} dF(z)
\]

\[
= E \left[ \mathbf{X}^* 1_{(\mathbf{a}' \mathbf{X}^* \leq y)} \frac{\partial^2 H(1 - G(\mathbf{a}' \mathbf{X}^*; \mathbf{a}))}{\partial u^2} \right],
\]

and \(\mathbf{X}^*\) is independent of \(\mathbf{X}\) with the same distribution. These variance-covariance matrices cannot be ordered and neither estimator of delta-DRM dominates the other one. However, a better estimator may be obtained by combining these two estimators, that is, by considering

\[
\alpha \frac{\partial \Pi_T}{\partial \mathbf{a}}(H, \mathbf{a}) + (1 - \alpha) \frac{\partial \hat{\Pi}_T}{\partial \mathbf{a}}(H, \mathbf{a}),\]

where \(\alpha\) is the solution of

\[
\min_{\alpha} \alpha^2 \Omega(\mathbf{a}, \mathbf{a}) + 2\alpha (1 - \alpha) \text{COV}(\mathbf{a}, \mathbf{a}) + (1 - \alpha)^2 \Sigma(\mathbf{a}, \mathbf{a}),
\]
Estimators of the variance-covariance matrices can be derived as follows: let us define the pseudo-observations of the components within the variance-covariance matrices as

\[ \hat{S}_1^t = x^*_t \frac{\partial H(1-t/T)}{\partial a}, \]

\[ \hat{S}_2^t = x^*_t \frac{\partial Q_T(1-t/T; a)}{\partial a} \frac{\partial H(1-t/T)}{\partial u}, \]

\[ \hat{S}_3^t = \frac{1}{T} \sum_{i=1}^{T} x^*_i \frac{\partial^2 H(1-t/T)}{\partial u^2}. \]

The variance-covariance matrices can be consistently estimated by their pseudo sample counterparts:

\[ \hat{\Omega}(a, a) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{S}_1^t - \hat{S}_2^t \right)^\prime \left( \hat{S}_1^t - \hat{S}_2^t \right) - \frac{1}{T} \sum_{t=1}^{T} \left( \hat{S}_3^t - \hat{S}_2^t \right)^\prime \left( \hat{S}_3^t - \hat{S}_2^t \right), \]

\[ \hat{\Sigma}(a, a) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{S}_1^t - \hat{S}_3^t \right)^\prime \left( \hat{S}_1^t - \hat{S}_3^t \right) - \frac{1}{T} \sum_{t=1}^{T} \left( \hat{S}_2^t - \hat{S}_3^t \right)^\prime \left( \hat{S}_2^t - \hat{S}_3^t \right), \]  

(3.2.11)

where \( x^*_i \) is defined as before [see e.g. Genest et al. (1995) for related results and regularity conditions, in the framework of copula estimation].

3.2.4 Finite sample properties of the estimators of the delta-DRM

The empirical estimator is less smooth than the kernel estimator with respect to allocation, especially if the distortion functions are discontinuous. Figure 16 illustrates this feature by considering simulated portfolios including two primitive assets. The two series of (opposite) returns are simulated from a bivariate Gaussian white noise. The covariance matrix is estimated from returns of the 1st and 5th deciles (in term of capitalization) of NYSE equity portfolios provided by CRSP. We use 250 observations corresponding to year 2000. The allocation \((a^1)\) associated with asset 1 (resp. \(a^2 = 1 - a^1\) associated with asset 2) is chosen to vary from 0.01 to 0.99 and the delta-DRM is estimated by both the kernel (solid lines) and empirical (dashed lines) approaches.\(^7\) We consider two risk measures, that are the Tail-VaR (TVaR in short with \(H(u; p) = (u/p) \wedge 1\)) and the Proportional Hazard risk measure (PH in short with \(H(u; p) = u^p\)).

\(^7\)Following Scaillet (2004), we choose the bandwidth parameter \(h = 0.5s_yT^{-1/5}\), where \(s_y\) is the estimated standard deviation of \(y_t\).
TVaR is designed to measure the extreme losses (e.g. \( p = 0.05 \)) with weighting function:

\[
\nabla H(a'X; a) = \frac{1}{p} \mathbf{1}(a'X \geq Q(1-p; a))
\]

which is not continuous in \( a \). In the left graph of Figure 16, we observe that the kernel estimator (solid line) of the delta-TVaR is much smoother than the empirical estimator (dashed line). In fact, the empirical estimator is not appropriate for estimating the sensitivity associated with the Tail-VaR (see also Scaillet, 2004). However, the right graph of Figure 16 shows that both estimators of the delta-PH are sufficiently smooth. Indeed, the PH risk measure assigns weights, which are continuous functions of \( a \), at all loss levels.

Let us now study the accuracy of both estimators in finite sample. We replicate 1000 times the previous simulation and compute the associated simulated variances. The finite sample variances of the kernel estimators (solid lines) and of the empirical estimators (dashed lines) for both the delta-TVaR and delta-PH are provided in Figure 17. The finite sample accuracies of both estimators are of similar magnitudes. More precisely, the variance of the empirical estimator of the delta-TVaR is up to 20% higher than that of the corresponding kernel estimator. When considering delta-PH, the kernel estimator is less accurate, but the difference is only 5%.

### 3.3 Efficient portfolio

The mean-variance approach has been prevalent in portfolio management for the past decades (see e.g. Markowitz, 1952). In this framework, efficient portfolios are constructed by minimizing the variance while keeping a desired level of expected return and satisfying a budget constraint. However, this approach can be misleading if the return distributions deviate from Gaussian. Indeed, the variance may not be a proper measure of risk. A few alternatives have been introduced to analyze the portfolio decision based on a DRM, such as VaR under

---

8 The weighting function associated with PH is \( \nabla H(a'X) = p(1 - G(a'X; a))^{p-1} \) (see e.g. Gourieroux and Liu, 2006).
the mean and budget constraints [Basak and Shapiro (2001)] and TailVaR [Lemus (1999), Rockafellar and Uryasev (1999, 2000), Yamai and Yoshiba (2002), Bassett et al. (2004)]. In a recent work, Gourieroux and Monfort (2005) provide a complete analysis of expected utility based efficient portfolios, including the estimation of efficient portfolios and tests for portfolio efficiency. Their approach corresponds to the so-called direct approach of stochastic dominance. In this section, we revisit this issue of portfolio choices for the dual approach of risk comparison based on the optimization of a DRM under DRM and budget constraints (see e.g. Wang and Young, 1998 for a comparison of direct and dual approaches of stochastic dominance). Such a careful analysis is required if more complicated DRM than VaR and TailVaR are suggested by the regulators. Indeed, the portfolio managers will adjust their portfolio management according to the new measures.

3.3.1 The efficiency frontier

i) The optimization problem

Let us consider a market with a risk-free asset and \( d \) primitive risky assets. Without loss of generality, the risk-free rate is set equal to zero; the current prices of risky assets are set equal to 1; we consider an agent facing a one-period investment decision on the allocation \( \mathbf{a} = [a^1, ..., a^d] \) of his/her initial wealth \( W_0 \). We impose no positivity restriction on the allocations, that is, short sales are allowed. Let us denote by \( \mathbf{x} = [x^1, ..., x^d] \) the vector of one-period (opposite) returns on the \( d \) risky assets. Similar to the mean-variance problem, the agent can solve the following DRM-DRM problem:

\[
\min_{a^0, \mathbf{a}} \Pi(H_1, (a^0, \mathbf{a}), F_0), \quad \text{s.t.} \quad \Pi(H_0, (a^0, \mathbf{a}), F_0) = \Pi_0^* \quad \text{and} \quad a^0 + \sum_{i=1}^{d} a^i = W_0, \quad (3.3.1)
\]

where \( a^0 \) is the allocation in the risk-free asset and \( \Pi_0^* \) is a predetermined value. This includes several familiar optimization problems as special cases. For instance, by letting \( H_0(u) = u \) and \( H_1(u) = 1_{(u \geq p)} \) (resp. \( H_1(u) = (u/p) \land 1 \)), we get the mean-VaR (resp. mean-TailVaR) optimization problem. By defining \( H_1(u) = (u/p) \land 1 \) and \( H_0(u) = 1_{(u \geq p)} \), we get the following
optimization problem in the VaR-Tail VaR space:

$$\min_{a^0,a} TVaR(p), \text{ s.t. } VaR(p) = \Pi_0^* \text{ and } a^0 + \sum_{i=1}^{d} a^i = W_0.$$  \hspace{1cm} (3.3.2)

The optimization problem (3.3.2) matches well the current regulation, in which the reserves are invested in a risk-free asset and used to satisfy exactly a VaR(5%) constraint. In this framework, the constraint is an inequality constraint $VaR(5\%) \leq \Pi_0^*$, which is binding at the optimum.

The convexity properties of the DRM with respect to portfolio allocation are useful to ensure a unique solution to the DRM-DRM optimization problem and sufficient first-order conditions. Such conditions are satisfied for the DRM when the distortion function is concave (see Denneberg, 1994).

Since a DRM is drift invariant,\(^9\) the solution of the DRM-DRM problem depends on $W_0$ only by means of the allocation in risk-free asset. For the allocations in risky assets, the optimization is equivalent to:

$$\min_a \Pi(H_1,a,F_0), \text{ s.t. } \Pi(H_0,a,F_0) = \Pi_0,$$

where $\Pi_0 = \Pi_0^* - W_0$ and $F_0$ is the joint distribution of excess returns.

ii) The first-order conditions

Let $\lambda$ denote the Lagrange multiplier associated with the DRM constraint. The optimization of the Lagrangian implies the following first-order conditions:

$$\frac{\partial \Pi}{\partial a}(H_1,a^*,F_0) - \lambda^* \frac{\partial \Pi}{\partial a}(H_0,a^*,F_0) = 0,$$

$$\Pi(H_0,a^*,F_0) = \Pi_0,$$

whose solution $a^*$, $\lambda^*$ are the efficient portfolio allocation and optimal Lagrange multiplier, respectively.\(^{10}\) Since the DRM is homogeneous of degree one with respect to portfolio allocation,

\(^9\)A function $\rho$ is drift invariant if $\rho(X + c) = \rho(X) + c$, for any constant $c$.

\(^{10}\)The solutions $a^*$, $\lambda^*$ depend on $H_0, H_1, \Pi_0, F_0$. This dependence will be mentioned when necessary.
the system (3.3.4) implies:

\[(\alpha^*)' \partial \Pi(H_1, \alpha^*, F_0) - \lambda^* (\alpha^*)' \partial \Pi(H_0, \alpha^*, F_0) = 0\]

\[\Leftrightarrow \Pi(H_1, \alpha^*, F_0) - \lambda^* \Pi(H_0, \alpha^*, F_0) = 0\]

\[\Leftrightarrow \lambda^* = \frac{\Pi(H_1, \alpha^*, F_0)}{\Pi(H_0, \alpha^*, F_0)} = \frac{\Pi(H_1, \alpha^*, F_0)}{\Pi_0}.\] (3.3.5)

As usual the expression of the multiplier depends on the different inputs of the optimization problem; these inputs, including the income and the DRM level \(\Pi_0\), have to satisfy some inequality restrictions to ensure that the Lagrange multiplier has the right sign.

**iii) The two funds separation theorem**

The efficient allocation is characterized by:

\[\frac{\partial \Pi}{\partial a}(H_1, \alpha^*, F_0) - \frac{\Pi(H_1, \alpha^*, F_0)}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, \alpha^*, F_0) = 0,\] (3.3.6a)

\[\Pi(H_0, \alpha^*, F_0) = \Pi_0.\] (3.3.6b)

The efficient allocation \(\alpha^*\) depends on the level \(\Pi_0\) of the DRM constraint. Since the DRM are homogeneous function of degree one in the allocation, the equation (3.3.6a) is equivalent to:

\[\frac{\partial \Pi}{\partial a}(H_1, \frac{\alpha^*}{\Pi_0}, F_0) - \frac{\Pi(H_1, \frac{\alpha^*}{\Pi_0}, F_0)}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, \frac{\alpha^*}{\Pi_0}, F_0) = 0.\]

Thus, the optimal allocation associated with level \(\Pi_0\) is \(\Pi_0\) times the optimal allocation associated with level 1, and all efficient portfolios are proportional to a same portfolio when \(\Pi_0\) varies. This is the well-known two-funds separation theorem (see e.g. Cass and Stiglitz, 1970, Ross, 1978), which is extended here to DRM measures of risk. The two-funds separation theorem is clearly a consequence of the homogeneity property. The efficiency frontier can be represented in the DRM-DRM space. The drift invariance and homogeneity properties imply that the efficiency frontier is a half-line.
iv) The optimization of the performance

The optimal Lagrange multiplier extends the standard notion of Sharpe ratio introduced in the mean-variance framework (see Sharpe, 1966) to the case of DRM. For instance if the DRM defining the constraint is the VaR(p) and the DRM to be optimized is TVaR(p), the Lagrange multiplier is simply the amplifying factor considered in Gourieroux and Liu (2006). Moreover, the extended Sharpe ratio:

$$\lambda(H_0, H_1, a, F_0) = \frac{\Pi(H_1, a, F_0)}{\Pi(H_0, a, F_0)}, \tag{3.3.7}$$

is homogeneous of degree zero in the allocation. As a consequence the efficiency frontier can also be derived by minimizing the extended Sharpe ratio. Thus, an efficient allocation will also satisfy the first-order condition:

$$\frac{\partial \lambda}{\partial a}(H_0, H_1, a^*, F_0) = \frac{1}{\Pi(H_0, a^*, F_0)} \left[ \frac{\partial \Pi}{\partial a}(H_1, a^*, F_0) - \lambda(H_0, H_1, a^*, F_0) \frac{\partial \Pi}{\partial a}(H_0, a^*, F_0) \right] = 0. \tag{3.3.8}$$

The solution $a^*$ of this first-order condition (3.3.8) is unique up to a multiplicative factor.

The solution becomes unique under the additional constraint $\Pi(H_0, a^*, F_0) = \Pi_0$, or if we consider the modified first-order condition:

$$\frac{\partial \Pi}{\partial a}(H_1, a^*, F_0) - \frac{\Pi(H_1, a^*, F_0)}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, a^*, F_0) = 0. \tag{3.3.9}$$

Indeed, by pre-multiplying (3.3.9) by $(a^*)'$ and applying the homogeneity property, we get:

$$\Pi(H_1, a^*, F_0) - \frac{\Pi(H_1, a^*, F_0)}{\Pi_0} \Pi(H_0, a^*, F_0) = 0$$

$$\Leftrightarrow \Pi(H_0, a^*, F_0) = \Pi_0,$$

thus, the additional constraint is automatically satisfied. Note that the theoretical results derived in Section 3.3.1 are also valid in a dynamic framework with serially dependent returns, by considering the conditional distribution instead of the unconditional one.
v) Interpretation in terms of preferences

The economic literature on choices under risk has introduced criteria of the type \( \int u(y)d\tilde{H}(y) \), where \( u \) is a utility function and \( \tilde{H} \) a probability distribution representing the subjective view of the investor about loss frequencies. When \( \tilde{H} \) is equal to the objective probability, we get the standard theory based on expected utility. When \( u(y) = y \), we get the dual theory of risk introduced by Yaari (1987). A DRM can be written as \( \Pi(H, a; F) = \int yd\tilde{H}[F(y)] \), which is the basic criterion of dual theory of risk. Thus, the DRM-DRM optimization can be thought as maximizing the criterion of dual theory subject to a regulatory constraint. A similar analysis is performed in Gourieroux and Monfort (2005) under the expected utility theory.

As shown in Yaari (1987) and Wang and Young (1998), the stochastic dominance defined under the dual theory is equivalent to the standard one up to the second order. Thus, a solution portfolio selected from the DRM-DRM criterion is automatically not stochastically dominated up to a certain order implied by the objective risk measure \( \Pi(H, a) \). For example, using VaR (resp. TailVaR) as objective DRM is sufficient to ensure that the portfolio is not dominated by others at the first (resp. second) order.

3.3.2 Empirical estimator of the efficient allocation

Nonparametric estimators of the optimal allocation \( a^* \) and extended Sharpe ratio \( \lambda^* \) can be defined in two different ways according to the selected nonparametric estimator of the delta-DRM. For expository purpose, we consider below the empirical estimator, but a similar analysis can be done with the kernel estimator. The empirical estimators \( \hat{a}_T, \hat{\lambda}_T \) are solutions of the empirical counterparts of first-order conditions (3.3.4):

\[
\begin{align*}
\frac{\partial \Pi}{\partial a}(H_1, \hat{a}_T, \hat{F}_T) - \hat{\lambda}_T \frac{\partial \Pi}{\partial a}(H_0, \hat{a}_T, \hat{F}_T) &= 0, \\
\Pi(H_0, \hat{a}_T, \hat{F}_T) &= \Pi_0.
\end{align*}
\]  

(3.3.10)
The system above provides jointly an estimator of the efficient allocation and of the extended Sharpe performance. This system is equivalent to:

\[
\begin{cases}
\frac{\partial \Pi}{\partial a}(H_1, \hat{a}_T, \hat{F}_T) - \frac{\Pi(H_1, \hat{a}_T, \hat{F}_T)}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, \hat{a}_T, \hat{F}_T) = 0, \\
\hat{\lambda}_T = \frac{\Pi(H_1, \hat{a}_T, \hat{F}_T)}{\Pi_0},
\end{cases}
\]

(3.3.11)

which corresponds to the optimization of the extended Sharpe performance.

### 3.3.3 Asymptotic properties of the estimated efficient allocation and extended Sharpe performance

The estimators \( \hat{F}_T, \hat{a}_T, \) and \( \hat{\lambda}_T \) converge to their theoretical counterparts, \( F_0, a^* \) and \( \lambda^* \), when \( T \to \infty \). Thus, the first-order conditions (3.3.11) can be expanded when \( \hat{F}_T, \hat{a}_T \) and \( \hat{\lambda}_T \) are close to \( F_0, a^*, \lambda^* \), respectively. The expansions are performed in A.4 and summarized below.

**Proposition 13.**

i) \( \sqrt{T}(\hat{a}_T - a^*) = \left[ -\frac{\partial^2 \Pi}{\partial a \partial a'}(H_1 - \lambda^*H_0, a^*, F_0) + \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, a^*, F_0) \frac{\partial \Pi}{\partial a'}(H_1, a^*, F_0) \right]^{-1} \)

\[
\sqrt{T} \left( \frac{\partial \Pi}{\partial a}(H_1 - \lambda^*H_0, a^*, F_0) - \frac{\partial \Pi}{\partial a}(H_1 - \lambda^*H_0, a^*, F_0) \right)
\]

\[ - \left[ \frac{\partial^2 \Pi}{\partial a \partial a'}(H_1 - \lambda^*H_0, a^*, F_0) + \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, a^*, F_0) \frac{\partial \Pi}{\partial a'}(H_1, a^*, F_0) \right]^{-1} \]

\[
\frac{(a^*)'}{\Pi_0} \frac{\partial \Pi}{\partial a}(H_0, a^*, F_0) \sqrt{T} \left[ \frac{\partial \Pi}{\partial a}(H_1, a^*, F_0) - \frac{\partial \Pi}{\partial a}(H_1, a^*, F_0) \right] + o_p(1);
\]

ii) \( \sqrt{T}(\hat{\lambda}_T - \lambda^*) = \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a'}(H_1, a^*, F_0) \sqrt{T}(\hat{a}_T - a^*) + \frac{1}{\Pi_0} (a^*)' \sqrt{T} \left[ \frac{\partial \Pi}{\partial a'}(H_1, a^*, F_0) - \frac{\partial \Pi}{\partial a'}(H_1, a^*, F_0) \right] + o_p(1). \)

Thus, the joint asymptotic properties of \( \hat{a}_T, \hat{\lambda}_T \) can be derived from the asymptotic properties of the estimated DRM sensitivities (see Proposition 12).

If the DRM is twice differentiable, we can explicitly evaluate its second-order derivative by
differentiating (3.2.5) with respect to $a$. More precisely, we have:

$$\frac{\partial^2 \Pi}{\partial a \partial a'} (H, a, F) = \frac{\partial}{\partial a'} E \left[ X \frac{\partial H(1 - G(a'X; a))}{\partial u} \right]$$

$$= -E \left[ X \frac{\partial^2 H(1 - G(a'X; a))}{\partial u^2} \right] \left\{ g(a'X; a)X' + \frac{\partial G(a'X; a)}{\partial a'} \right\}$$

$$= -E \left[ X \frac{\partial^2 H(1 - G(a'X; a))}{\partial u^2} \right] g(a'X; a) + \frac{\partial G(a'X; a)}{\partial a'}$$

since $\frac{\partial G(Y; a)}{\partial a} = -E[X|Y]g(Y; a)$. Then, by applying the iterated expectation theorem, we get,

$$\frac{\partial^2 \Pi}{\partial a \partial a'} (H, a, F) = -E \left[ V(a'X|a'X) \frac{\partial^2 H(1 - G(a'X; a))}{\partial u^2} \right] g(a'X; a). \quad (3.3.12)$$

Since $V(a'X|a'X) = 0$, we see that $a' \frac{\partial^2 \Pi}{\partial a \partial a'} (H, a, F) a = 0$, and the Hessian of the DRM has a diminished rank $d - 1$. This is a consequence of the homogeneity property of the DRM. If the distortion function is concave, $\frac{\partial^2 H}{\partial u^2}$ is negative and the matrix $\frac{\partial^2 \Pi}{\partial a \partial a'}$ is positive semi-definite. This provides an alternative proof of the result by Denneberg (1994) mentioned above.

The approach developed above seems easier to implement and more general than the approach for estimating pessimistic portfolio introduced in Bassett et al. (2004). First, it can be applied to any pair of DRMs, not only to a pair (DRM, expected return). Second, by applying the functional theorems, we know how to choose the bandwidth to get the asymptotic results (whereas the weights $\nu_k$, $k = 1, \ldots, m$ in Bassett et al. (2004), p 489, are difficult to choose in an appropriate way). Finally, our approach allows for the derivation of the joint asymptotic distribution of the optimal allocations and of the Lagrange Multiplier, needed for the efficiency tests presented in Section 3.4.

### 3.4 Efficiency test of a given portfolio

The asymptotic expressions of the delta-DRM estimators and the estimated efficient allocations can be used to construct different tests of the efficiency of a given portfolio $a_0$, say.
As usual three types of test statistics can be considered by analogy with the Wald, Lagrange Multiplier and Likelihood Ratio tests introduced in the statistical literature. For expository purpose, we provide the results associated with the empirical estimators of the delta-DRMs. A similar analysis can be performed for the kernel estimator.

3.4.1 The null hypothesis

Let us consider a given allocation in risky assets $a_0$, say. The hypothesis to be tested is the DRM-DRM efficiency of $a_0$ for given risk measures, that are given distortion functions $H_0, H_1$. If $a^*(H_0, H_1, \Pi_0, F_0)$ denotes the efficient allocation associated with constraint level $\Pi_0$, the null hypothesis can be written as:

$$
\mathcal{H}_0 = \left\{ \exists \Pi_0 \text{ s.t.: } a^*(H_0, H_1, \Pi_0, F_0) = a_0 \right\}
$$

Alternatively, the efficiency hypothesis can also be defined from the extended Sharpe ratio by:

$$
\mathcal{H}_0 = \left\{ a_0 \text{ optimizes } \lambda(H_0, H_1, a, F_0) \right\}.
$$

The null hypothesis involves $d - 1$ independent restrictions on the true distribution $F_0$, and we can expect test procedures with $d - 1$ degree of freedom.

3.4.2 The constrained and unconstrained estimators

The general results of Section 3.3.2, 3.3.3 cannot be used directly. Indeed, the estimator $\hat{a}_T = \hat{a}_T(H_0, H_1, \Pi_0)$ assumes a known constraint level $\Pi_0$. This is not the case when considering the tests for efficiency. However, we can define the realized constrained level under the null:

$$
\hat{\Pi}_{0T} = \Pi(H_0, a_0, \hat{F}_T),
$$
and look for the efficient allocation and Lagrange multiplier corresponding to this level. The unconstrained estimators are denoted by \( \hat{\mathbf{a}}_T, \hat{\lambda}_T \) and satisfy the first-order conditions:

\[
\begin{align*}
\frac{\partial \Pi}{\partial \mathbf{a}}(H_1, \hat{\mathbf{a}}_T, \hat{F}_T) - \frac{\Pi(H_1, \hat{\mathbf{a}}_T, \hat{F}_T)}{\Pi(0)} \frac{\partial \Pi}{\partial \mathbf{a}}(H_0, \hat{\mathbf{a}}_T, \hat{F}_T) &= 0, \\
\hat{\lambda}_T &= \frac{\Pi(H_1, \hat{\mathbf{a}}_T, \hat{F}_T)}{\Pi(0)}. 
\end{align*}
\]

The estimators constrained by the null hypothesis are \( \hat{\mathbf{a}}_{0T} = \hat{\mu}_T \mathbf{a}_0 \) and \( \hat{\lambda}_{0T} \), say. The estimator of the constrained allocation satisfies the first-order condition:

\[
\frac{\partial \Pi}{\partial \mathbf{a}}(H_1, \hat{\mu}_T \mathbf{a}_0, \hat{F}_T) - \frac{\Pi(H_1, \hat{\mu}_T \mathbf{a}_0, \hat{F}_T)}{\Pi(0)} \frac{\partial \Pi}{\partial \mathbf{a}}(H_0, \hat{\mu}_T \mathbf{a}_0, \hat{F}_T) = 0. 
\]

By using the homogeneity property of the DRM, we get:

\[
\frac{\partial \Pi}{\partial \mathbf{a}}(H_1, \mathbf{a}_0, \hat{F}_T) - \frac{\Pi(H_1, \mathbf{a}_0, \hat{F}_T)}{\Pi(0)} \frac{\partial \Pi}{\partial \mathbf{a}}(H_0, \mathbf{a}_0, \hat{F}_T) = 0. 
\]

Then, pre-multiplying the system by \( \mathbf{a}_0' \) and using the Euler condition, we get: \( \hat{\mu}_T = 1 \). We deduce that the constrained estimators are:

\[
\hat{\mathbf{a}}_{0T} = \mathbf{a}_0, \quad \hat{\lambda}_{0T} = \frac{\Pi(H_1, \mathbf{a}_0, \hat{F}_T)}{\Pi(H_0, \mathbf{a}_0, \hat{F}_T)},
\]

where \( \hat{\lambda}_{0T} \) is the realized extended Sharpe (RES) ratio of portfolio \( \mathbf{a}_0 \).

### 3.4.3 Asymptotic expansion of the unconstrained estimated allocation

The asymptotic expansion of the difference between the unconstrained and constrained efficient allocations \( \hat{\mathbf{a}}_T - \mathbf{a}_0 \) is derived in A.5.
Proposition 14. We get:

\[
\sqrt{T}(\hat{a}_T - a_0) = B^{-1}A\sqrt{T} \left[ \frac{\partial \Pi_T}{\partial a}(H_1 - \lambda_0H_0, a_0, F_0) - \frac{\partial \Pi}{\partial a}(H_1 - \lambda_0H_0, a_0, F_0) \right] + o_p(1),
\]

\[
= B^{-1}\sqrt{T}\frac{\partial \Pi}{\partial a}(H_1 - \hat{\lambda}_0T, a_0, \hat{F}_T) + o_p(1),
\]

where

\[
A = Id - \frac{1}{\Pi(H_0, a_0, F_0)} \frac{\partial \Pi}{\partial a}(H_0, a_0, F_0)a_0',
\]

\[
B = - \frac{\partial^2 \Pi}{\partial a \partial a'}(H_1 - \lambda_0H_0, a_0, F_0) + \frac{1}{\Pi(H_0, a_0, F_0)} \frac{\partial \Pi}{\partial a'}(H_1, a_0, F_0).
\]

The matrix \(A\) is such that:

\[
a_0' A = a_0' \left( Id - \frac{1}{\Pi(H_0, a_0, F_0)} \frac{\partial \Pi}{\partial a}(H_0, a_0, F_0)a_0' \right)
\]

\[
= a_0' - \frac{a_0' \frac{\partial \Pi}{\partial a}(H_0, a_0, F_0)}{\Pi(H_0, a_0, F_0)} a_0'
\]

\[
= a_0' - a_0' = 0, \quad \text{by the Euler condition.}
\]

Thus, matrix \(A\) has rank \(d - 1\). If \(\Sigma\) denotes the asymptotic variance-covariance matrix of

\[
\sqrt{T} \left[ \frac{\partial \Pi_T}{\partial a}(H_1 - \lambda_0H_0, a_0, F_0) - \frac{\partial \Pi}{\partial a}(H_1 - \lambda_0H_0, a_0, F_0) \right],
\]

we have asymptotically

\[
\sqrt{T}(\hat{a}_T - a_0) \xrightarrow{d} N\left(0, B^{-1}A\Sigma A'(B')^{-1}\right). \quad (3.4.5)
\]

The variance-covariance matrix of \(\sqrt{T}(\hat{a}_T - a_0)\) has also a diminished rank \(d - 1\), since the difference between estimated efficient allocations satisfies the restriction \(\frac{\partial \Pi}{\partial a}(H_0, a_0, F_0)\sqrt{T}(\hat{a}_T - a_0) = o_p(1)\) due to the DRM constraint.

3.4.4 The test statistics

The test statistics for the efficiency hypothesis are the following.
i) The Wald statistic

The Wald statistic is based on a comparison of the estimated efficient allocation with the given allocation $a_0$. It is defined as:

$$
\xi_W = \sqrt{T}(\hat{a}_T - a_0)' \left[ \hat{V}((\hat{a}_T) \right]^{-1} \sqrt{T}(\hat{a}_T - a_0),
$$

where $[\hat{V}(\hat{a}_T)]^{-1} = \hat{B}'(\hat{A}\hat{\Sigma}\hat{A})^{-1}\hat{B}$ is a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{a}_T - a_0)$, and $[\cdot]^{-1}$ denotes a generalized inverse. Indeed, we know that $V(\hat{a}_T)$ has the reduced rank $d - 1$.

ii) The delta-DRM statistic

The delta-DRM statistics is given by:

$$
\xi_{\text{delta-DRM}} = T \frac{\partial \Pi}{\partial a'} \left( H_1 - \hat{\lambda}_0 H_0, a_0, \hat{F}_T \right) \left( \hat{A}\hat{\Sigma}\hat{A}' \right)^{-1} \frac{\partial \Pi}{\partial a} \left( H_1 - \hat{\lambda}_0 H_0, a_0, \hat{F}_T \right).
$$

From Proposition 14, we deduce the result below.

**Proposition 15.** Under the null hypothesis of efficiency, the Wald and delta-DRM statistics are asymptotically equivalent and follow asymptotically a chi-square distribution with $d - 1$ degree of freedom.

iii) The RES based statistic

Finally, let us consider the statistic

$$
\xi_{\lambda} = T \left( \hat{\lambda}_T - \hat{\lambda}_{0,T} \right),
$$

based on the comparison of RES ratios. This is the analogue of the standard likelihood ratio
test encountered in maximum likelihood theory. Under the null hypothesis, we have:

\[
T \left( \hat{\lambda}_T - \hat{\lambda}_{0T} \right) = T \left[ \lambda(H_0, H_1, \hat{a}_T, \hat{F}_T) - \lambda(H_0, H_1, a_0, \hat{F}_T) \right]
\]

\[
= \sqrt{T} \left[ \sqrt{T} \left( \hat{a}_T - a_0 \right) \right]' \frac{\partial \lambda}{\partial a} (H_0, H_1, a_0, \hat{F}_T) + o_p(1)
\]

\[
= \sqrt{T} \left( \hat{a}_T - a_0 \right)' \frac{\partial^2 \lambda}{\partial a \partial F} (H_0, H_1, a_0, F_0) \sqrt{T} (\hat{F}_T - F_0) + o_p(1), \tag{3.4.9}
\]

since under the null \( \frac{\partial \lambda}{\partial a} (H_0, H_1, a_0, F_0) = 0. \)

Moreover, the unconstrained efficient allocation \( \hat{a}_T \) satisfies the first-order conditions:

\[
\frac{\partial \lambda}{\partial a} (H_0, H_1, \hat{a}_T, \hat{F}_T) = 0;
\]

these conditions can be expanded to get:

\[
\frac{\partial^2 \lambda}{\partial a \partial a} (H_0, H_1, a_0, F_0) \sqrt{T} (\hat{a}_T - a_0) + \frac{\partial^2 \lambda}{\partial a \partial F} (H_0, H_1, a_0, F_0) \sqrt{T} (\hat{F}_T - F_0) + o_p(1) = 0.
\]

By substituting in (3.4.9), we get the following proposition:

**Proposition 16.** Under the null hypothesis of efficiency,

\[
T \left( \hat{\lambda}_T - \hat{\lambda}_{0T} \right) = \sqrt{T} (\hat{a}_T - a_0)' \left[ -\frac{\partial^2 \lambda}{\partial a \partial a} (H_1 - \lambda_0 H_0, a_0, F_0) \right] \sqrt{T} (\hat{a}_T - a_0) + o_p(1)
\]

\[
= \sqrt{T} (\hat{a}_T - a_0)' \left[ -\frac{\partial^2 \Pi}{\partial a \partial a} (H_1 - \lambda_0 H_0, a_0, F_0) \right] \sqrt{T} (\hat{a}_T - a_0) + o_p(1).
\]

In general \( -\frac{\partial^2 \Pi}{\partial a \partial a} (H_1 - \lambda_0 H_0, a_0, F_0) \) is not a generalized inverse of \( V(\hat{a}_T) \), and the RES based statistic is not equivalent to the two other statistics. This result is compatible with the general theory of asymptotic tests when the objective function cannot be interpreted as a log-likelihood function (see Gourieroux and Monfort, 1995, Vol. 2, Chapter 18).
3.5 Finite sample properties

We will now discuss the finite sample properties of the empirical estimators of the efficient allocation and extended Sharpe ratio for a TVaR-VaR optimization problem. By considering Gaussian returns in the Monte-Carlo, we will also compare the nonparametric estimators with the parametric estimators computed using the Gaussian assumption. Let us first describe the DRM-DRM problem in a Gaussian framework.

3.5.1 The DRM-DRM problem in a Gaussian framework

Let us assume Gaussian (opposite) excess returns: $X \sim N(-m, \Omega)$. The quantile function associated with the (negative) portfolio return $y(a) = a'x$ is given by:

$$Q(u; a) = -a'm + \Phi^{-1}(u)(a'\Omega a)^{1/2}.$$ (3.5.1)

We deduce the expression of a DRM in the Gaussian framework:

$$\Pi(H, a, \Phi) = \int_0^1 \left[-a'm + \Phi^{-1}(u)(a'\Omega a)^{1/2}\right]dH(u)$$

$$= -a'm + (a'\Omega a)^{1/2}\beta(H),$$ (3.5.2)

where $\beta(H) = \int_0^1 \Phi^{-1}(1 - u)dH(u)$. Thus, the DRM is an affine combination of the expected (negative) portfolio return and its standard error.

Let us now consider a DRM-DRM problem such as:

$$\min_a -a'm + (a'\Omega a)^{1/2}\beta(H_1), \text{ s.t. } -a'm + (a'\Omega a)^{1/2}\beta(H_0) = \Pi_0.$$ (3.5.3)

The first-order conditions imply that the efficient allocation is proportional to the efficient mean-variance allocation $a^*_0 = \Omega^{-1}m$. In particular, it is easily checked that the set of efficient
allocations is given by:

\[ a^* = \Omega^{-1} m \frac{\Pi_0}{m'\Omega^{-1}m} \left[ \frac{(m'\Omega^{-1}m)^{1/2}}{-(m'\Omega^{-1}m)^{1/2} + \beta(H_0)} \right], \]

which is generated by the vector \( a^*_0 \) and only depend on the choice of \( (\Pi_0, H_0) \).\(^{11}\) We can note that \( a^* \) is equivalent to mean-variance efficient allocation when the expected return is equal to \( \Pi_0 (m'\Omega^{-1}m)^{1/2} / [-(m'\Omega^{-1}m)^{1/2} + \beta(H_0)] \). The associated extended Sharpe ratio is:

\[ \lambda^* = \frac{-a^* m + (a^* \Omega a^*)^{1/2} \beta(H_1)}{-a^* m + (a^* \Omega a^*)^{1/2} \beta(H_0)} = \frac{-(m'\Omega^{-1}m)^{1/2} + \beta(H_1)}{-(m'\Omega^{-1}m)^{1/2} + \beta(H_0)}. \] (3.5.4)

The extended Sharpe ratio is in a one-to-one relationship with the standard Sharpe ratio, that is \( m'\Omega^{-1}m \), but depends on both selected distortion measures.

3.5.2 The Monte-Carlo study

Let us consider two risky assets with excess returns independently drawn in a Gaussian distribution with parameters \( m = (0.00044, 0.007)' \), \( \Omega = \begin{pmatrix} 0.0031 & 0.0028 \\ 0.0028 & 0.0064 \end{pmatrix} \). These parameters are estimated using the same data set as in Section 3.2.4.\(^{12}\) The corresponding mean-variance efficient portfolio is \( a^*_0 = (-1.415, 1.722)' \), and can be normalized as \( \tilde{a}_0 = (-4.6, 5.6)' \) to get a sum of the components equal to 1.

Then, we introduce a TVaR-VaR problem for risk level \( p = 5\%, 10\% \), respectively. For a given risk level and a given set of \( T = 501 \) return observations, we compute:

i) the empirical estimate of the efficient allocation normalized by \( \hat{a}_T^1 + \hat{a}_T^2 = 1 \);

ii) the empirical estimate of the extended Sharpe ratio;

iii) the parametric estimate of the normalized efficient allocation \( \hat{\Omega}_T^{-1}\hat{m}_T / (\hat{\Omega}_T^{-1}\hat{m}_T) \), where

\(^{11}\)As long as the convexity of the Lagrangian objective function is ensured.

\(^{12}\)The returns are multiplied by 10 before calculating these parameters. This is to ensure the numerical invertibility of the variance-covariance matrix from the simulated data.
\( \widehat{m}_T \) and \( \widehat{\Omega}_T \) are the realized mean and volatility, \( \iota \) is a \( d \times 1 \) vector of ones;

iv) the parametric estimate of the standard Sharpe ratio \( \widehat{m}_T \widehat{\Omega}_T^{-1} \widehat{m}_T \);

v) the implied Sharpe ratio from the empirical estimate of extended Sharpe ratio

\[
\tilde{\lambda}_T = \left( \frac{\widehat{\lambda}_T \beta(H_0) - \beta(H_1)}{\widehat{\lambda}_T - 1} \right)^2.
\]

These computations are replicated \( S = 400 \) times to get the finite sample distributions.

The finite sample distributions of the normalized estimated allocation in asset 1 \((a^1)^*\) are provided in Figure 18. The solid line corresponds to the parametric estimator, the dashed line (resp. shorter dashed line) to the empirical estimator when \( p = 0.05 \) (resp. \( p = 0.1 \)) and the vertical (thick) line indicates the true efficient allocation \( \tilde{a}_1^0 \). These curves are produced for the range \([-100, 100] \), but more simulated values are outside this range for the parametric estimator than for the empirical estimators. This is a consequence of the lack of robustness of the parametric mean-variance approach to extreme returns. It is observed that all estimators feature bias in finite sample. The bias is larger for the maximum likelihood parametric estimator than for the empirical estimator when \( p = 0.05 \), and that for the empirical estimator when \( p = 0.1 \). The average biases are 7, 6.26 and 5.46, respectively. The order of their finite sample variances are the same as that of their biases. The parametric estimator can be very noisy (the variance is equal to 1656.32). Between the empirical estimators, the one with \( p = 0.05 \) has slightly larger variance than that with \( p = 0.1 \) (139.02 v.s. 117.76).

The distributions of implied Sharpe ratios are plotted in Figure 19. The vertical (thicker) line represents the true value (i.e. 0.107). The parametric estimator (solid line) exhibits the smallest bias (the bias is 0.013) and finite sample variance (0.002). As expected, due the shortage of observations, the empirical estimator with \( p = 0.05 \) (dashed line) is less accurate as compared to the one with \( p = 0.1 \) (shorter dashed line), in the sense that it has a larger bias (the average biases of the implied Sharpe ratio are 0.219 and 0.13, respectively) and a larger variance (0.094 v.s. 0.038).
3.6 Conclusion

In this paper, we first introduce two nonparametric estimators of the delta-DRM, defined as the sensitivity of DRM with respect to portfolio allocations, and derive their asymptotic distributions. Then, these estimators are used to calculate the efficient portfolio allocations assuming a DRM objective function and a DRM based constraint. We show that the limiting behaviors of the estimators of efficient allocations only depend on the asymptotic properties of the delta-DRM. Three test statistics are proposed to test the efficiency hypothesis for a given portfolio. These test statistics are analogous to the standard Wald, LM and LR test statistics in the Maximum Likelihood context. Finally, a Monte-Carlo study is implemented to compare the finite sample properties of the nonparametric estimators and the parametric estimators in a Gaussian framework. We find that, when estimating the efficient allocation in finite sample, the empirical estimator may be preferred to its parametric counterpart even though the parametric model is well specified. While considering the implied Sharpe ratio, the parametric estimator dominates. Of course, the main advantage of the empirical estimator is to be consistent even if the parametric model is misspecified.

Appendices

A.1 Preliminary lemma

Lemma 2 (Scaillet (2004), Proof of Proposition 3.1). Let us consider a d-dimensional random vector $X$ with a continuous joint distribution $F$. For any given $d \times 1$ real vector $a$, any continuous function $\Psi : \mathbb{R} \to \mathbb{R}$, any continuous function $\varphi : \mathbb{R}^d \to \mathbb{R}$, any value $\xi$ and any kernel $k$ such that $\int k(u)du = 1$ and $\int u k(u)du = 0$, we have, as $T \to \infty$: 
(i). \[
\frac{1}{h_T} \int_{\mathbb{R}} x k\left(\frac{a'x - y}{h_T}\right) \Psi(y) dy = -x\Psi(a'x) + O(h_T^2),
\]
\[
\frac{1}{h_T} \int_{\mathbb{R}} E[X|a'X = y] k\left(\frac{a'x - y}{h_T}\right) \Psi(y) dy = -E[X|a'X = a'x]\Psi(a'x) + O(h_T^2);
\]

(ii). \[
\frac{1}{h_T} E\left[\varphi(X)k\left(\frac{a'X - \xi}{h_T}\right)\right] = E[\varphi(X)|a'X = \xi] g(\xi; a) + O(h_T^2),
\]
\[
\frac{1}{h_T} E\left[(\varphi(X)k\left(\frac{a'X - \xi}{h_T}\right))^2\right] = E[\varphi(X)\varphi(X)'|a'X = \xi] g(\xi; a) \int k(u)^2 du + O(h_T),
\]

where \( h_T \) is the bandwidth satisfying \( h_T \to 0 \) and \( T h_T \to \infty \) as \( T \to \infty \), and \( g(\cdot; a) \) is the pdf of \( a'X \).

**A.2 Comovement interpretation of the delta-DRM**

**Proof.** If \( H \) is continuous and differentiable in \( u \), the \( i \)th right hand side of equation (3.2.4) can be written as:

\[
RHS^i = \int_{0}^{1} \int_{\mathbb{R}^{d-1}} \frac{x^i}{a^j} f\left(x^1, \ldots, \frac{Q(1-u; a) - \sum_{l \neq j} a^l x^l}{a^j}, \ldots, x^d\right) \prod_{l \neq j} dx^l \frac{\partial H(u)}{\partial u} du.
\]

By applying the change of variables \( y = Q(1-u; a) \), this expression becomes:

\[
RHS^i = \int_{\mathbb{R}} \frac{x^i}{a^j} f\left(x^1, \ldots, \frac{y - \sum_{l \neq j} a^l x^l}{a^j}, \ldots, x^d\right) \prod_{l \neq j} dx^l \frac{\partial H(1 - (G(y; a))}{\partial u} (1 - G(y; a))
\]
\[
= \int_{\mathbb{R}} \frac{x^i}{a^j} f\left(x^1, \ldots, \frac{y - \sum_{l \neq j} a^l x^l}{a^j}, \ldots, x^d\right) \prod_{l \neq j} dx^l dy.
\]

Since \( \frac{1}{a^j} f\left(x^1, \ldots, \frac{y - \sum_{l \neq j} a^l x^l}{a^j}, \ldots, x^d\right) \) is the joint density of distribution of \((X^l, l \neq j)\) and \( a'X \), we deduce the result. \( \square \)
A.3 Asymptotic expansions of the estimators

Regularity conditions

A sufficient set of regularity conditions for deriving the asymptotic expansions of the random estimators and applying functional limit theorems is given below.

1. Conditions on the return process and portfolio allocation:

**Assumption A. 1.** The process \((x_t)\) is a strong white noise, that is the \(x_t\)'s are i.i.d. random vectors.

**Assumption A. 2.** The distribution of \((X)\) is continuous, with a continuous strictly positive density.

**Assumption A. 3.** \(E[||X||^\gamma] < \infty\) for some \(\gamma > 0\).

**Assumption A. 4.** The true portfolio allocation belongs to a bounded set \(A = \prod_{i=1}^d (a_i, \bar{a}_i),\) say.

Assumption A.2 avoids nonstandard behavior, when the distribution features some point masses (see e.g. Laurent (2003) for a discussion of this problem), and Assumption A.3 imposes a uniform tail behavior for the portfolio returns. The value of \(\gamma\) has to be sufficiently large to give a meaning to the DRM of interest.

**Assumption A. 5.** The distribution of \(X\) given \(a'X = y\) is continuous on \(\{a'X = y\}\) with a continuous strictly positive density. The conditional moments \(E[||X||^\gamma|a'X = y]\) exist for any \(y\) and any \(a \in A\).

2. Conditions on the distortion function:

**Assumption A. 6 (Smoothness).** The distortion function \(H\) is increasing on \((0, 1]\), twice continuously differentiable.
Assumption A.7 (Boundedness). [see Shorack (1972), or Shorack and Wellner (1986), Chapter 19] We have \( \left| \frac{\partial H(u)}{\partial u} \right| \leq cu^{-\alpha}(1-u)^{-\beta} \), for \( \alpha < 1/2 - 1/\gamma \), \( \beta < 1/2 - 1/\gamma \), \( \gamma > 0 \) and \( c < \infty \).

Assumption A.7 ensures that the weighting function do not attribute too much weight on extreme risks. It is used jointly with Assumption A.3 on the tail behavior of the return distribution.

Assumption A.8. The discretized distortion function (resp. first, second order derivative of the distortion function) tends uniformly to the true distortion function (resp first, second order derivative of the distortion function).

This assumption is introduced to get a negligible discretization error.

3. Conditions on the kernel:

Assumption A.9. [see Fermanian and Scaillet (2005)] \( k \) is a strictly symmetric Parzen kernel of order 2 on \( \mathbb{R} \) such that \( \lim_{u \to \infty} uk(u) = 0 \), \( \int |u|k(u)du < \infty \), \( \int k(u)du = 1 \) and \( \int uk(u)du = 0 \). \( k \) is three times differentiable, \( k' \) and \( k'' \) are integrable and \( k''' \) is bounded.

Assumption A.9 is satisfied by standard kernels such as the Gaussian kernel. It is not satisfied by some optimal kernels, such as the Epanechnikov kernel.

4. Conditions on the bandwidth:

Assumption A.10. \( Th_5^5 \to 0 \) and \( Th_2^{7/2} \to \infty \), when \( T \) tends to infinity.

Assumption A.10 removes the bias from the kernel estimator (see Yatchew (2003)).

Under this set of conditions, we can apply the multivariate Functional Central Limit Theorem to the sample cdf.
Multivariate Functional Limit Theorem. Let $x_1, \ldots, x_T$ be i.i.d. random observations in $\mathbb{R}^d$ with continuous marginal cdfs $F^j(x^j)$ with $j = 1, \ldots, d$, $u = (u^1, \ldots, u^d) = (F^1(x^1), \ldots, F^d(x^d))$ a $1 \times d$ vector of uniformly distributed ranks and $\tilde{F}_T(x)$ its joint sample cdf, we have:

$$\sqrt{T} (\tilde{F}_T(x) - F(x)) \Rightarrow \mathcal{K}(u), \quad (3-A-1)$$

where $\mathcal{K}$ is a multivariate Brownian bridge on $[0, 1]$, which is Gaussian with zero mean and covariance,

$$\text{COV}(\mathcal{K}(u), \mathcal{K}(u')) = C(u \wedge u') - C(u) C(u'), \quad (3-A-2)$$

where $u \wedge u' = \{\min(u^1, (u')^1), \ldots, \min(u^d, (u')^d)\}$.

Note that, for any real function $\Psi$ such that $E[\Psi(X)^2] < \infty$, we have:

$$E \left[ \int \Psi(x) d\mathcal{K} \left( F^1(x^1), \ldots, F^d(x^d) \right) \right] = 0,$$

$$E \left[ \left( \int \Psi(x) d\mathcal{K} \left( F^1(x^1), \ldots, F^d(x^d) \right) \right)^2 \right] = V(\Psi(x)).$$

Kernel estimator

Let us consider the expansion of the kernel estimator of the delta-VaR (3.2.6). We have:

$$\sqrt{Th_T} \left( \frac{\partial \tilde{Q}_T}{\partial a}(\cdot; a) - \frac{\partial Q}{\partial a}(\cdot; a) \right)$$

$$= \sqrt{Th_T} \left( \frac{\partial \tilde{Q}_T}{\partial a}(\cdot; a) - E[X|a'X = Q(\cdot; a)] \right)$$

$$= \sqrt{Th_T} \left( \frac{1}{Th_T} \sum_{t=1}^T \mathbf{x}_t k \left( \frac{a'x_t - \tilde{Q}_T(\cdot; a)}{h_T} \right) - \frac{1}{Th_T} \sum_{t=1}^T \mathbf{x}_t k \left( \frac{a'x_t - Q(\cdot; a)}{h_T} \right) \right) + \sqrt{Th_T} \left( \frac{1}{Th_T} \sum_{t=1}^T \mathbf{x}_t k \left( \frac{a'x_t - Q(\cdot; a)}{h_T} \right) - \frac{1}{g(Q(\cdot; a); a)} \int_{\mathbb{R}^d} \mathbf{x} k \left( \frac{a'x - Q(\cdot; a)}{h_T} \right) \sqrt{d} (\tilde{F}_T(x) - F(x)) \right)$$

$$= \nabla \mathcal{K}_T(Q(\cdot; a); a) \sqrt{Th_T} (\tilde{Q}_T(\cdot; a) - Q(\cdot; a)) + \frac{1}{g(Q(\cdot; a); a)} \frac{1}{\sqrt{h_T}} \int_{\mathbb{R}^d} \mathbf{x} k \left( \frac{a'x - Q(\cdot; a)}{h_T} \right) \sqrt{d} (\tilde{F}_T(x) - F(x))$$
By a change of variables, standard parametric rate and

where:

\[
\nabla K_T(Q(:; a); a) = \frac{\partial}{\partial Q(:; a)} \left( \frac{1}{h_T} \sum_{t=1}^{T} x_t k \left( \frac{a'x_t - Q(:; a)}{h_T} \right) \right),
\]

and \( \hat{F}_T(x) = \frac{1}{T} \sum_{t=1}^{T} \mathbbm{1}(x_t \leq x) \) denotes the sample multivariate cdf and we use the standard expansion for the Nadaraya-Watson estimator. Thus, we get:

\[
\sqrt{T} h_T \left( \frac{\partial \hat{\Pi}_T}{\partial a} (H, a) - \frac{\partial \Pi}{\partial a} (H, a) \right)
= \int_0^1 \nabla K_T(Q(1 - u; a); a) \sqrt{T} \left( \hat{Q}_T(1 - u; a) - Q(1 - u; a) \right) d H(u)
+ \int_0^1 \frac{1}{g(Q(1 - u; a); a)} \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} x k \left( \frac{a'x - Q(1 - u; a)}{h_T} \right) \sqrt{T} d \left[ \hat{F}_T(x) - F(x) \right] \right\} d H(u)
- \int_0^1 \mathbb{E}_{X|a'X = Q(1 - u; a)} \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} k \left( \frac{a'x - Q(1 - u; a)}{h_T} \right) \sqrt{T} d \left[ \hat{F}_T(x) - F(x) \right] \right\} d H(u) + o_p(\sqrt{h_T}).
\]

Due to its integral expression, the kernel estimator of the delta-DRM will converge at the standard parametric rate \( \sqrt{T} \). Indeed, we have:

\[
\sqrt{T} \left( \frac{\partial \hat{\Pi}_T}{\partial a} (H, a) - \frac{\partial \Pi}{\partial a} (H, a) \right)
= \int_{\mathbb{R}} \nabla K_T(y; a) \sqrt{T} \left[ \hat{Q}_T(1 - G(y; a); a) - y \right] d H(1 - G(y; a))
- \int_{\mathbb{R}} \frac{1}{g(y; a)} \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} x k \left( \frac{a'x - y}{h_T} \right) \sqrt{T} d \left[ \hat{F}_T(x) - F(x) \right] \right\} d H(1 - G(y; a))
+ \int_{\mathbb{R}} \mathbb{E}_{X|a'X = y} \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} k \left( \frac{a'x - y}{h_T} \right) \sqrt{T} d \left[ \hat{F}_T(x) - F(x) \right] \right\} d H(1 - G(y; a)) + o_p(1).
\]
When $H$ is first-order differentiable, we get:

$$\sqrt{T} \left( \frac{\partial \tilde{\Pi}_T}{\partial a}(H, a) - \frac{\partial \Pi}{\partial a}(H, a) \right)$$

$$= \int \nabla K_T(y; a) \sqrt{T} \left[ \hat{Q}_T(G(y; a); a) - y \right] \frac{\partial H(1 - G(y; a))}{\partial u} g(y; a) dy$$

$$+ \int \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} x k \left( \frac{a'x - y}{h_T} \right) \sqrt{T} [\hat{F}_T(x) - F(x)] \right\} \frac{\partial H(1 - G(y; a))}{\partial u} dy$$

$$- \int E[X|a'X = y] \left\{ \frac{1}{h_T} \int_{\mathbb{R}^d} x k \left( \frac{a'x - y}{h_T} \right) \sqrt{T} [\hat{F}_T(x) - F(x)] \right\} \frac{\partial H(1 - G(y; a))}{\partial u} dy + o_p(1)$$

where the second equality is a consequence of the Bahadur representation of the nonparametric quantile estimator:

$$\sqrt{T} \left( \hat{Q}_T(G(y; a); a) - y \right) = - \frac{1}{g(y; a)} \int_{\mathbb{R}^d} 1_{a'x \leq y} \sqrt{T} [\hat{F}_T(x) - F(x)] + o_p(1).$$

By commuting the integrations with respect to $x$ and $y$ and applying Lemma 2 i) in A.1, we deduce the limiting behavior of the kernel estimator:

$$\sqrt{T} \left( \frac{\partial \tilde{\Pi}_T}{\partial a}(H, a) - \frac{\partial \Pi}{\partial a}(H, a) \right)$$

$$\Rightarrow \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \frac{\partial}{\partial y} (E[X|a'X = y]) \Psi(y; a) 1_{a'x \leq y} dy \right\} dK(F^1(x^1), \ldots, F^d(x^d))$$

$$- \int_{\mathbb{R}^d} x \Psi(a'x; a) dK(F^1(x^1), \ldots, F^d(x^d)) + \int_{\mathbb{R}^d} E[X|a'X = a'x] \Psi(a'x; a) dK(F^1(x^1), \ldots, F^d(x^d))$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{\partial}{\partial y} (E[X|a'X = y]) \Psi(y; a) 1_{a'x \leq y} dy dK(F^1(x^1), \ldots, F^d(x^d))$$

$$- \int_{\mathbb{R}^d} \left\{ x - E[X|a'X = a'x] \right\} \Psi(a'x; a) dK(F^1(x^1), \ldots, F^d(x^d)) + o_p(1).$$

(3-A-3)
Empirical estimator

Let us now consider the expansion of the empirical estimator (3.2.9). When \( H \) is twice differentiable, we have:

\[
\sqrt{T} \left( \frac{\partial \hat{\Pi}_T}{\partial a}(H, a) - \frac{\partial \Pi}{\partial a}(H, a) \right)
\]

\[
= \sqrt{T} \left( \int_{\mathbb{R}^d} x \frac{\partial H(1 - \hat{G}_T(a'; x; a))}{\partial u} d \hat{F}_T(x) - \int_{\mathbb{R}^d} x \frac{\partial H(1 - G(a'x; a))}{\partial u} d F(x) \right)
\]

\[
+ \sqrt{T} \left( \int_{\mathbb{R}^d} x \frac{\partial H(1 - G(a'x; a))}{\partial u} d \hat{F}_T(x) - \int_{\mathbb{R}^d} x \frac{\partial H(1 - G(a'x; a))}{\partial u} d F(x) \right)
\]

\[
= - \int_{\mathbb{R}^d} x \frac{\partial^2 H(1 - G(a'x; a))}{\partial u^2} \sqrt{T} [\hat{G}_T(a'; x; a) - G(a'x; a)] d F(x)
\]

\[+ \int_{\mathbb{R}^d} x \frac{\partial H(1 - G(a'x; a))}{\partial u} \sqrt{T} d [\hat{F}_T(x) - F(x)] + o_p(1). \]

Since \( G(a'z; a) = P[a'X \leq a'z] = \int_{\mathbb{R}^d} 1_{(a'x \leq a'z)} d F(x) \), and \( \hat{G}_T(a'z; a) = \int_{\mathbb{R}^d} 1_{(a'x \leq a'z)} d \hat{F}_T(x) \),
the right hand side of the expansion can be rewritten as:

\[
\int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} -z \frac{\partial^2 H(1 - G(a'z; a))}{\partial u^2} 1_{(a'x \leq a'z)} d F(z) + x \frac{\partial H(1 - G(a'x; a))}{\partial u} \right\} \sqrt{T} d [\hat{F}_T(x) - F(x)] + o_p(1),
\]

which weakly converges to:

\[
\int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} -z \frac{\partial^2 H(1 - G(a'z; a))}{\partial u^2} 1_{(a'x \leq a'z)} d F(z) + x \frac{\partial H(1 - G(a'x; a))}{\partial u} \right\} d K(F^1(x^1), \ldots, F^d(x^d)).
\]

Alternative expansion of the kernel estimator

If the distortion function \( H \) is twice differentiable, we can integrate by part the first component in the asymptotic expression of the kernel estimator given in (3-A-3). Let us denote by
Z a variable with the same distribution as \( X \). We have:

\[
\int_{\mathbb{R}} \frac{\partial}{\partial y} (E[Z^i | a'Z = y]) \nabla H(y; a) 1_{a'x \leq y} dy
\]

\[
= 1_{a'x \leq y} \nabla H(y; a) E[Z^i | a'Z = y] \bigg|_{-\infty}^{\infty} - \int_{\mathbb{R}} E[Z^i | a'Z = y] 1_{a'x \leq y} d\nabla H(y)
\]

\[
= \nabla H(\infty; a) E[Z^i | a'Z = \infty] + \int_{\mathbb{R}} 1_{a'x \leq y} \frac{\partial^2 H(1 - G(y; a); a)}{\partial u^2} \frac{1}{a^2} f \left( \frac{z_1, \ldots, \frac{y - \sum_{i=1}^{j} a_i z_i}{a_i}, \ldots, z_d}{a} \right) \prod_{l \neq j} dz^l g(y)
\]

Since \( g(y; a) = \int_{\mathbb{R}^{d-1}} \frac{1}{a^2} f \left( \frac{z_1, \ldots, \frac{y - \sum_{i=1}^{j} a_i z_i}{a_i}, \ldots, z_d}{a} \right) \prod_{l \neq j} dz^l \) and \( \int_{\mathbb{R}^d} dK(F_1(x^1), \ldots, F_d(x^d)) = 0 \), we get:

\[
\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{\partial}{\partial y} (E[Z^i | a'Z = y]) \nabla H(y; a) 1_{a'x \leq y} dy dK(F_1(x^1), \ldots, F_d(x^d))
\]

\[
= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_{a'x \leq a'z} \frac{\partial^2 H(1 - G(a'z; a); a)}{\partial u^2} \frac{1}{a^2} f \left( \frac{z_1, \ldots, z_d}{a} \right) dz^1 \ldots dz^d dK(F_1(x^1), \ldots, F_d(x^d))
\]

\[
+ \nabla H(\infty) E[Z^i | a'Z = \infty] \int_{\mathbb{R}^d} dK(F_1(x^1), \ldots, F_d(x^d))
\]

\[
= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_{a'x \leq a'z} \frac{\partial^2 H(1 - G(a'z; a); a)}{\partial u^2} \frac{1}{a^2} f \left( \frac{z_1, \ldots, z_d}{a} \right) dz^1 \ldots dz^d dK(F_1(x^1), \ldots, F_d(x^d)).
\]

The above equation holds for any asset \( i = 1, \ldots, d \). Thus, we deduce that:

\[
\sqrt{T} \left( \frac{\partial \Pi_T}{\partial a} (H, a) - \frac{\partial \Pi}{\partial a} (H, a) \right) \Rightarrow \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left[ \frac{\partial^2 H}{\partial u^2} \left[ 1 - G(a'z; a) \right] \right] dz \frac{1}{a^2} f \left( \frac{z_1, \ldots, z_d}{a} \right) dK(F_1(x^1), \ldots, F_d(x^d))
\]

\[
- \int_{\mathbb{R}^d} \left[ x - E[X | a'X = a'x] \right] \nabla H(a'x; a) dK(F_1(x^1), \ldots, F_d(x^d)). \quad (3-A-4)
\]

**Asymptotic variances**

The asymptotic variance-covariances of both estimators are obtained by using (see A.3):

\[
V \left( \int_{\mathbb{R}^d} \psi(x) dK(F_1(x^1), \ldots, F_d(x^d)) \right) = V(\psi(x)).
\]
a) If the distortion function $H$ is twice differentiable, the variance-covariance matrices of the limiting processes are:

$$
\Omega(a, a) = \mathbb{V} \left[ \int z_{1(a'X \leq a'z)} \frac{\partial^2 H(1 - G(a'z; a))}{\partial u^2} \partial F(z) - (X - E[X|a'X]) \nabla H(a'X; a) \right],
$$

$$
\Sigma(a, a) = \mathbb{V} \left[ X \nabla H(a'X; a) - \int z_{1(a'X \leq a'z)} \frac{\partial^2 H(1 - G(a'z; a))}{\partial u^2} \partial F(z) \right].
$$

b) However, the distortion functions $H$ associated with the VaR and Tail-VaR are not twice differentiable. The variance-covariance matrices of their kernel estimators are given below.

b.i) For the delta-VaR, we have:

$$
\lim_{T \to \infty} V \left( \sqrt{T h_T} \left[ \frac{\partial \hat{\text{VaR}}_{T,k}(p, a)}{\partial a} - \frac{\partial \text{VaR}(p, a)}{\partial a} \right] \right)
$$

$$
= \frac{1}{g(Q(1 - p; a); a)} \int k^2(u) d u \left[ E[k^2(a'X = Q(1 - p; a))] - \left( E[k(a'X = Q(1 - p; a))] \right)^2 \right].
$$

Proof. Let us denote $g = g(Q(1 - p; a); a)$; we get:

$$
V \left( \sqrt{T h_T} \left[ \frac{\partial \hat{Q}_T(\cdot, a)}{\partial a} - \frac{\partial Q(\cdot, a)}{\partial a} \right] \right)
$$

$$
= \frac{1}{g^2 h_T} \left\{ E \left[ X^k k^2 \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) \right] - E \left[ k X^2 \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) \right]^2 \right\}
$$

$$
+ \frac{(E[X|a'X = Q(\cdot; a)])^2}{g^2 h_T} \frac{1}{h_T} \left[ E k^2 \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) - \left( E k \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) \right)^2 \right]
$$

$$
- \frac{2(E[X|a'X = Q(\cdot; a)])}{g^2 h_T} \frac{1}{h_T} \left\{ E \left[ X k^2 \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) \right] - E \left[ k \left( \frac{a'X - Q(\cdot; a)}{h_T} \right) \right] \right\}.
$$

By applying Lemma 2 ii) in A.1 for $\varphi(x) = 1, x, x^2$, respectively, we deduce:

$$
\lim_{T \to \infty} V \left( \sqrt{T h_T} \frac{\partial \hat{\text{VaR}}_T(p, a)}{\partial a} - \frac{\partial \text{VaR}(p, a)}{\partial a} \right)
$$
\[
= \frac{1}{g} \int k^2(u) d u \left[ E[X^2|a'X = Q(1 - p; a)] - \left( E[X|a'X = Q(1 - p; a)] \right)^2 \right] + o(1).
\]

\(\square\)

ii) For the delta-TVaR, we have:

\[
\Omega_{TVaR(p)} = V \left( -\frac{1}{p} E[X|a'X = Q(1 - p; a)] 1_{a'X \leq Q(1 - p; a)} - \frac{1}{p} X 1_{a'X \geq Q(1 - p; a)} \right) + \frac{1}{p} E[X|a'X = Q(1 - p; a)] 1_{a'X \geq Q(1 - p; a)}.
\]

A.4 Expansions of \(\hat{a}_T\) and \(\hat{\lambda}_T\)

i) Let us consider the first equation of system (3.3.10). We get:

\[
\frac{\partial^2 \Pi}{\partial a \partial \alpha^*} (H_1 - \lambda^* H_0, a^*, F_0) - \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a} (H_0, a^*, F_0) \frac{\partial \Pi}{\partial \alpha^*} (H_1, a^*, F_0) \cdot \sqrt{T} (\hat{a}_T - a^*) + \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a} (H_0, a^*, F_0) \cdot \sqrt{T} (\hat{F}_T - F_0) = o_p(1),
\]

(3-A-5)

where \(\frac{\partial \Pi}{\partial a^*}\) stands for Hadamard derivative. Equivalently, we have:

\[
\left[ - \frac{\partial^2 \Pi}{\partial a \partial \alpha^*} (H_1 - \lambda^* H_0, a^*, F_0) + \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a} (H_0, a^*, F_0) \frac{\partial \Pi}{\partial \alpha^*} (H_1, a^*, F_0) \right] \cdot \sqrt{T} (\hat{a}_T - a^*) = \sqrt{T} \left[ \frac{\partial \Pi}{\partial a} (H_1 - \lambda^* H_0, a^*, F_0) - \frac{\partial \Pi}{\partial a} (H_1 - \lambda^* H_0, a^*, F_0) \right] - \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a} (H_0, a^*, F_0) \cdot \sqrt{T} \left[ \Pi_T (H_1, a^*, F_0) - \Pi (H_1, a^*, F_0) \right] + o_p(1).
\]

The result of Proposition 13 i) follows by noting that \(\hat{\Pi}_T (H_1, a^*, F_0) = (a^*)' \frac{\partial \Pi}{\partial a} (H_1, a^*, F_0)\).

ii) Let us now consider the expansion of \(\hat{\lambda}_T\), we get:

\[
\sqrt{T} (\hat{\lambda}_T - \lambda^*) = \frac{1}{\Pi_0} \frac{\partial \Pi}{\partial a} (H_1, a^*, F_0) \cdot \sqrt{T} (\hat{a}_T - a^*) + \frac{1}{\Pi_0} \sqrt{T} \left[ \Pi_T (H_1, a^*, F_0) - \Pi (H_1, a^*, F_0) \right] + o_p(1).
\]
A.5 Expansion of $\hat{a}_T$

The expansion of the first equation of (3.4.4) provides:

\[ B\sqrt{T}(\hat{a}_T - a_0) = \sqrt{T} \left[ \frac{\partial \Pi}{\partial a}(H_1, a_0, \hat{F}_T) - \frac{\Pi(H_1, a_0, \hat{F}_T)}{\Pi(H_0, a_0, \hat{F}_T)} \frac{\partial \Pi}{\partial a}(H_0, a_0, \hat{F}_T) \right] + o_p(1) \]

\[ = \sqrt{T} \left[ \frac{\partial \Pi}{\partial a}(H_1, a_0, \hat{F}_T) - \frac{a_0' \frac{\partial \Pi}{\partial a}(H_1, a_0, \hat{F}_T)}{a_0' \frac{\partial \Pi}{\partial a}(H_0, a_0, \hat{F}_T)} \frac{\partial \Pi}{\partial a}(H_0, a_0, \hat{F}_T) \right] + o_p(1) \]

\[ = \left[ Id - \frac{1}{\Pi(H_0, a_0, F_0)} \frac{\partial \Pi}{\partial a}(H_0, a_0, F_0) a_0' \right] \sqrt{T} \left[ \frac{\partial \hat{\Pi}_T}{\partial a}(H_1 - \lambda_0 H_0, a_0, F_0) - \frac{\partial \Pi}{\partial a}(H_1 - \lambda_0 H_0, a_0, F_0) \right] \]

Note finally that the first equality can be written as:

\[ B\sqrt{T}(\hat{a}_T - a_0) = \sqrt{T} \frac{\partial \Pi}{\partial a} \left( H_1 - \lambda_0 T H_0, a_0, \hat{F}_T \right) + o_p(1). \]
Chapter 4

Control and Out-of-Sample Validation of Dependent Risks

4.1 Introduction

The current regulation introduced in Finance by Basel I and Basel II (or in Insurance by Solvency II) has at least three limitations: i) The first one is the choice of the so-called Value-at-Risk (VaR) to measure the risk and to fix the required capital. Indeed, the VaR accounts for the probability of a loss, but not for the expected magnitude of this loss (as the TailVaR), or for the possibility of extreme losses (as distortion risk measures). ii) The second limitation is the lack of coherency between the suggested computation of the VaR as a conditional quantile and the ex-post validation of this computation, which uses unconditional (historical) hedging errors. iii) Finally, the computation of the VaRs are performed independently on different lines of financial products, without taking into account the possible dependence of the risks between lines.

The aim of this paper is to propose solutions to the two last limitations by focusing on the (implicit) decision criterion of the internal or external validator. For expository purpose, we follow the practice of measuring the risk by means of the VaR. In Section 4.2, we focus on a single line of financial products, that is on a single risk. We introduce the decision criterion which
underlies the choice of a VaR and discuss the criterion value when a suboptimal control variate is selected. The criterion value can be decomposed into its minimal attainable value plus a term measuring the lack of optimality. In a perspective of measuring the out-of-sample performances of a proposed control variate, we analyze the joint dynamic of these two components. The case of several lines of financial products is considered in Section 3. We introduce a decision criterion function, which extends naturally the criterion considered for a single budget line. By optimizing this criterion, with or without constraint on the global risk, we get the allocations of reserves by budget lines. Section 4.4 concludes. Proofs are gathered in appendices.

4.2 Conditional risk control

It is emphasized in the recent literature [see e.g. Hansen (2004), Giacomini and White (2006)] that the “out-of-sample comparison of predictive ability has to be based on inference about conditional expectations of forecast and forecast errors rather than their unconditional expectations”. The same remark applies for the validation by the regulators of the approaches followed by banks and insurance companies to fix their reserves. This section discusses this principle for the computation of the VaR and the analysis of its out-of-sample performances.

4.2.1 The underlying criterion

The approach is based on the underlying criterion followed by regulators to derive the VaR, that is a conditional quantile of the future portfolio value. Let us denote by $y_t$ the opposite of the portfolio value at date $t$, by $I_t$ the information available at time $t$, which includes at least the current and lagged values of $y$ and by $z_t$ the level of reserve, which can be considered as a control variate. This level is proposed by the bank according to the information $I_t$. The standard decision criterion is:

$$
\Psi(t, z_t) = E_t \left[ \alpha (y_{t+1} - z_t)^+ + (1 - \alpha) (y_{t+1} - z_t)^- \right],
$$

(4.2.1)
where \( E_t \) denotes the conditional expectation given \( I_t \), \( y^+ = \max(y, 0) \), \( y^- = \max(-y, 0) \), and \( \alpha \) is the announced risk level. This decision criterion corresponds to an asymmetric loss function (written in \$). Two types of errors can arise. The amount of reserve \( z_t \) can be too small, implying a loss \( y_{t+1} - z_t \); it can also be too large, which implies an amount of money \( z_t - y_{t+1} \) not used for profitable investment. The criterion is weighting both types of errors, with more weights to avoid the first type of error from the prudential (regulator) point of view.\(^1\)

Let us denote \( f_t \) and \( F_t \) the conditional pdf and cdf of \( y_{t+1} \), respectively. It is well-known \[\text{see e.g. Koenker (2005)}\], that the optimal value of the control variate:

\[
\hat{z}_t = \arg \min_{z_t \in I_t} E_t \left[ \alpha(y_{t+1} - z_t)^+ + (1 - \alpha)(y_{t+1} - z_t)^- \right], \tag{4.2.2}
\]

is the conditional quantile at risk level \( \alpha \):

\[
\hat{z}_t = F_t^{-1}(\alpha). \tag{4.2.3}
\]

However, there exists a large number of decision criteria providing the conditional quantile as the optimal control variate. It has been shown in Gourieroux and Monfort (1995), Vol 1, Section 8.5.9, that these criteria can be written as:

\[
\Psi(t, z_t) = E_t \left[ \alpha \left( A(y_{t+1}) - A(z_t) \right)^+ + (1 - \alpha) \left( A(y_{t+1}) - A(z_t) \right)^- \right], \tag{4.2.4}
\]

where \( A \) is an increasing function. In practice, it is important to choose among them, a criterion with clear financial interpretation, that is weighting appropriately the loss magnitude.\(^2\)

Note also that, although the optimal control variate is clearly defined, this control variate is difficult to approximate in practice due to the conditioning which involves a lot of conditioning variables. A nonparametric approach for approximating the conditional quantile is not very

---

\(^1\)Since the quantile concerns the loss and profit variable, \( \alpha \) is large in practice. Typically, it is equal to 95%, i.e. to a level of 5% when the problem is written in terms of profit and loss.

\(^2\)In this respect, the decision criterion introduced in Hansen (2004), Example 2, neglects the magnitude of the loss.
accurate, whereas a parametric approach has a significant risk of misspecification. In the rest of the paper, the optimal control will be a benchmark useful to discuss some properties and interpretations. However, the control variates proposed by banks (or insurance companies) differ generally from this optimal control and their suboptimal methodologies can vary in time.

For any control variate, the criterion value can be decomposed as:

\[ \Psi(t, z_t) = \Psi(t, \hat{z}_t) + \left[ \Psi(t, z_t) - \Psi(t, \hat{z}_t) \right], \quad (4.2.5) \]

where the second component of the right hand side measures the lack of optimality. For the standard objective function, where \( A \) is the identity function, this measure is equal to [see A.1]:

\[ c(t, z_t) = \int_{z_t}^{\hat{z}_t} \left[ F_t(\hat{z}_t) - F_t(y) \right] dy \quad (4.2.6) \]

\[ = \begin{cases} 
E_t \left[ (y_{t+1} - z_t) 1_{(z_t < y_{t+1} < \hat{z}_t)} \right], & \text{if } \hat{z}_t > z_t, \\
E_t \left[ (z_t - y_{t+1}) 1_{(\hat{z}_t < y_{t+1} < z_t)} \right], & \text{if } \hat{z}_t \leq z_t. 
\end{cases} \quad (4.2.7) \]

In particular, when the selected control variate is close to the optimal one \( z_t \simeq \hat{z}_t \), we get the expansion:

\[ c(t, z_t) \simeq \frac{1}{2} f_t(\hat{z}_t)(\hat{z}_t - z_t)^2 = \frac{1}{2} f_t \left[ F_t^{-1}(\alpha) \right] (\hat{z}_t - z_t)^2. \quad (4.2.8) \]

The first component of decomposition (4.2.5) is equal to:

\[ \Psi(t, \hat{z}_t) = \alpha P_t [y_{t+1} > \hat{z}_t] E_t \left[ (y_{t+1} - \hat{z}_t)^+ | y_{t+1} > \hat{z}_t \right] \\
+ (1 - \alpha) P_t [y_{t+1} < \hat{z}_t] E_t \left[ (y_{t+1} - \hat{z}_t)^- | y_{t+1} < \hat{z}_t \right] \\
= \alpha (1 - \alpha) \left[ TailVaR^+ (\alpha) + TailVaR^- (\alpha) \right], \quad (4.2.9) \]

where the TailVaR measures the "expected loss" when a loss occurs. This expression highlights the importance of the TailVaR in the validating criterion even if it is not the optimal level of the control variate. Thus, if the optimal control is applied, the regulator needs to follow up
both the VaR for the control variate, and the TailVaR, for the objective function.

### 4.2.2 Dynamic of the decision criterion

It is now possible to study the consequence on the criterion of any proposed methodology to compute the control variate. More precisely, for any methodology $z_t$, we can consider the joint dynamics of $\Psi(t, \hat{z}_t)$ and $c(t, z_t)$.

As an illustration, let us assume that the dynamics of $y_{t+1}$ is such that:

$$y_{t+1} = m_t + \sigma_t u_{t+1}, \quad (4.2.10)$$

where $m_t, \sigma_t$ are functions of the past and $(u_t)$ is a sequence of i.i.d. error terms. This specification includes the ARMA-GARCH model, where the information set corresponds to the current and lagged values of $y$, but also the stochastic mean and volatility model, where the information set also includes the current and lagged values of the factors driving the stochastic parameters. Let us denote by $Q$ the quantile function of the standardized error $u$, $H(x) = E(u - x)^+$, and $E(u) = 0$. The conditional quantile is:

$$\hat{z}_t = m_t + \sigma_t Q(\alpha). \quad (4.2.11)$$

Thus, we deduce:

$$\Psi(t, \hat{z}_t) = \alpha E_t \left[ (y_{t+1} - \hat{z}_t)^+ \right] + (1 - \alpha) E_t \left[ (y_{t+1} - \hat{z}_t)^- \right]$$

$$= \alpha \sigma_t E_t \left[ (u_{t+1} - Q(\alpha))^+ \right] + (1 - \alpha) \sigma_t E_t \left[ (u_{t+1} - Q(\alpha))^- \right]$$

$$= \sigma_t \delta(\alpha), \quad (4.2.12)$$

where

$$\delta(\alpha) = H \left[ Q(\alpha) \right] + (1 - \alpha) Q(\alpha). \quad (4.2.13)$$

Let us now assume that the VaR proposed by the bank is computed by historical simulation
(on a large window), that is, by considering the unconditional quantile. Then, we get $z_t = z$, independent of the past, and the value of the criterion becomes:

$$\Psi(t, z) = \alpha E_t \left[(y_{t+1} - z)^+\right] + (1 - \alpha) E_t \left[(y_{t+1} - z)^-\right]$$

$$= E_t \left[(y_{t+1} - z)^+\right] - (1 - \alpha) \left(E_t[y_{t+1}] - z\right)$$

$$= \sigma_t H \left[\frac{z - m_t}{\sigma_t}\right] - (1 - \alpha)(m_t - z). \quad (4.2.14)$$

The correcting component is:

$$c(t, z_t) = \sigma_t \left[H \left(\frac{z - m_t}{\sigma_t}\right) - H \left(Q(\alpha)\right)\right] - (1 - \alpha) \left[m_t + \sigma_t Q(\alpha) - z\right], \quad (4.2.15)$$

and its local expansion becomes (see A.2):

$$c(t, z_t) = \frac{1}{2\sigma_t} g \left[Q(\alpha)\right] \left(m_t + \sigma_t Q(\alpha) - z\right)^2, \quad (4.2.16)$$

where $g$ is the pdf of the error term $u$.

This example shows that:

1. The minimal attainable criterion value is a stochastic function of the scale factor $\sigma_t$ only;

2. The measure of the lack of optimality of the unconditional VaR depends on both the location and scale factors $m_t, \sigma_t$; therefore, the two components of the decision criterion have dependent joint dynamics;

3. The same remark applies for the relative lack of optimality $c(t, z_t)/\Psi(t, \hat{z}_t)$;

4. The optimal criterion value is highly volatile, since it is proportional to $\sigma_t$. This is a consequence of the usual choice of a fixed risk level $\alpha$ by validators instead of a path dependent risk level $\alpha_t$. This feature has already been noted when discussing how to take into account the cycle effect in the regulation, that is the point-in-time (PIT) versus the through-the-cycle (TTC) methodologies. In this framework, it has been suggested to
adjust the risk level along the cycle. Similarly, the risk level could be adjusted to the volatility level.

As an illustration, let us consider the stochastic volatility model:

$$r_t = -y_t = 0.5\sigma^2_t + \sigma_t \varepsilon_t,$$

(4.2.17)

where the volatility $\sigma^2_t$ follows an autoregressive gamma (ARG) process with parameters $\rho = 0.96$, $\delta = 3.6$, $c = 7.34 \times 10^{-3}$ [see Gourieroux and Jasiak (2006a)]. The historical parameters $\rho, \delta, c$ are such that the ARG volatility process matches the stationary mean, variance and first-order autocorrelation of the discretely sampled Cox-Ingersoll-Ross process estimated by Andersen et al. (2002). The dynamic model is an extension of the standard Ball and Roma (1994) model, including a risk premium. We provide in Figure 20 a joint simulated path of $(r_t, \sigma^2_t)$, $t = 1, \ldots, 250$, corresponding to solid line and dashed line, respectively.

This path is used to estimate the historical (unconditional) distribution of the return and to deduce its (estimated) unconditional quantile at $1 - \alpha = 5\%$. This unconditional quantile is equal to $\hat{z} = 2.31$. Figure 21 displays the evolution of the conditional quantile $\hat{z}_t$ (solid line) and compares with the unconditional quantile (dashed line) $\hat{z}$. The simulated path of $r_t$ (shorter dashed line) is also plotted to reveal the relative locations of the estimated quantiles. As expected the unconditional quantile is above the average of the conditional quantiles. This represents the cost of not taking into account the largest information when evaluating the risk.

We observe the clustering of reserve levels deduced from the underlying volatility clustering. Moreover, the observations exceed the conditional and unconditional VaRs by a similar number of times, even though the VaR estimates are quite different during large and small volatility periods. This suggests the insufficiency of the current validation based on the number of exceedances criterion. Figure 22 provides the joint dynamics of the value of the optimal criterion.

---

3The conditional quantile is computed with the value of the parameters used in the simulation study. In practice these values are unknown and have to be estimated, which induces an additional uncertainty.
(solid line), the value of the suboptimal criterion (shorter dashed line) and the measure of suboptimality (dashed line), evaluated when unconditional quantile is used. By definition, the criterion value computed with the unconditional quantile is above the criterion value computed with the conditional one. The optimal criterion value exhibits similar pattern to that of conditional VaR in Figure 21. This is due to the fact that both processes are proportional to the volatility.

[Insert Figure 22: Joint Evolution of $Ψ(t, \hat{z}_t), Ψ(t, z), \text{and } c(t, z_i)$].

For a suboptimal control, both components of the objective function have to be analyzed jointly. We provide in Table 6 the historical summary statistics of the series, that are their mean, variance, skewness, kurtosis.

[Insert Table 6: Historical Summary Statistics].

The distribution of the loss function $Ψ(t, z)$, when the suboptimal unconditional quantile is used, features strong positive skewness and very fat tails. The fat right tail arises since the unconditional VaR estimate tends to underestimate the TailVaR when the volatility is large. This underestimation of TailVaR is further amplified by the volatility level as specified by equation (4.2.14). This effect is significantly smaller when the conditional quantile is used, and the kurtosis is four times smaller.

Cross-autocorrelations of the series and their squares are provided in Table 7. $ρ(k)$ denotes the kth order autocorrelation matrix. For instance,

$$ρ(3) = corr \left( [ψ(t, \hat{z}_t), c(t, z_t), ψ(t, z)]', [ψ(t − 3, \hat{z}_{t−3}), c(t − 3, z_{t−3}), ψ(t − 3, z)]' \right).$$

By construction, the serial correlations on the optimal criterion value $Ψ(t, \hat{z}_t)$ and their squares are equal to the serial correlations of $σ_t$ and $σ_t^2$. Thus, it is not surprising to observe the long memory effect corresponding to the high persistence of volatility. A smaller persistence can be observed on the criterion value and its squares, when unconditional quantile is used as control variate.
4.2.3 Comparing the methodologies proposed to compute the reserves

The aim of this section is to compare different methodologies $z_{1,t}, z_{2,t}$, say, used to compute the reserve corresponding to a same risky portfolio and a same announced risk level. First of all, two methodologies cannot be ranked unambiguously in general; indeed, methodology 1 can be preferred in some environments (such as expansion periods, or high volatility periods), whereas methodology 2 can be preferable in other environments (such as recession periods, or low volatility periods).

However, it is possible to develop tractable diagnostic concerning properties valid for all environments, for instance to test the hypothesis:

$$H_0: (\text{methodology 1 is better than methodology 2 for any environment.})$$

The null hypothesis is:

$$H_0: [\Psi(t, z_{1,t}) \leq \Psi(t, z_{2,t}), \text{ for any value of the conditioning variable}]$$

and involves an infinite number of conditional inequality moment restrictions. This hypothesis can be written equivalently in terms of unconditional inequality moment restrictions by introducing instruments $g_t$ as:

$$H_0: \left[ E \left\{ g_t \left[ \alpha(y_{t+1} - z_{1,t})^+ + (1 - \alpha)(y_{t+1} - z_{1,t})^- \right] \right\} \leq E \left\{ g_t \left[ \alpha(y_{t+1} - z_{2,t})^+ + (1 - \alpha)(y_{t+1} - z_{2,t})^- \right] \right\} , \right.$$  

for any positive function $g_t$ depending on information $I_t$.

At this step, it is important to mention that the standard diagnostics, proposed in the literature [see e.g. Diebold and Mariano (1995), Harvey et al. (1997, 1998), Harvey and Newbold (2000),...
or the survey by Stock and Watson (2005) or used by practitioners, validators, or regulators, are based on the unconditional loss function only, that is, select \( g_t = 1 \) as the single instrumental variable.\(^4\) The recent literature has emphasized the importance of selecting a larger number of appropriate instruments. For instance, Gourieroux and Jasiak (2006b) considered lagged values of the loss function, such as:

\[
g_{j,k,t} = \alpha(y_{t+1-k} - z_{j,t-k})^+ + (1 - \alpha)(y_{t+1-k} - z_{j,t-k})^-; j = 1, 2.
\]

By considering the single unitary instrument, we focus on the historical average loss. By choosing lagged values of the objective function as instruments, we consider also the possibility of loss clustering, that is of a large cumulated loss on consecutive periods of time. When a finite number of instruments \( g_{1,t}, \ldots, g_{K,t} \), say, are selected, the hypothesis concerns \( K \) positivity restrictions on unconditional moments. It can be tested by Wald type procedures using the results established in Kudô (1963), Gourieroux et al. (1980, 1982), Wolak (1987), and Wolak (1991).

The procedure is as follows:

1. Approximate each unconditional moment by its sample counterpart. Namely, the following moment:

\[
\theta_k = \mathbb{E} \left[ g_{k,t} \left\{ \alpha \left[ (y_{t+1} - z_{2,t})^+ - (y_{t+1} - z_{1,t})^+ \right] \\
+ (1 - \alpha) \left[ (y_{t+1} - z_{2,t})^- - (y_{t+1} - z_{1,t})^- \right] \right\} \right], k = 1, \ldots, K,
\]

is estimated by:

\[
\hat{\theta}_{k,T} = \frac{1}{T} \sum_{t=1}^{T-1} g_{k,t} \left\{ \alpha \left[ (y_{t+1} - z_{2,t})^+ - (y_{t+1} - z_{1,t})^+ \right] \\
+ (1 - \alpha) \left[ (y_{t+1} - z_{2,t})^- - (y_{t+1} - z_{1,t})^- \right] \right\}
\]

\(^4\)Direct comparisons of the proposed and optimal VaR have the same drawback [see e.g. Mancini and Trojani (2005)].
2. Estimate the asymptotic variance-covariance matrix $\hat{\Omega}_T$ of $\hat{\theta}_T = (\hat{\theta}_{1,T}, \ldots, \hat{\theta}_{K,T})'$, by the sample variance of the vector of sample averages, given by,

$$\hat{\Omega}_T = \hat{\Gamma}_T(0) + \sum_{\tau=-L}^{L} \hat{\Gamma}_T(\tau),$$

where $\hat{\Gamma}_T(\tau)$ is the sample counterpart of the autocovariance matrix at lag $\tau$ of the multivariate process $g_t \{\alpha \left[ (y_{t+1} - z_{2,t})^+ - (y_{t+1} - z_{1,t})^+ \right] + (1 - \alpha) \left[ (y_{t+1} - z_{2,t})^- - (y_{t+1} - z_{1,t})^- \right]\}$ and $L$ defines the bandwidth.

3. Compute the inequality constrained estimator $\hat{\theta}_0^0$ of $\theta$ as the solution of:

$$\hat{\theta}_0^0 = \arg \min_{\theta: \theta_k \geq 0, k=1,\ldots,K} \left( \hat{\theta}_T - \theta \right)' \hat{\Omega}_T^{-1} \left( \hat{\theta}_T - \theta \right).$$

4. Compute the test statistics as the standardized sum of squared residuals:

$$SSR_T = \left( \hat{\theta}_T - \hat{\theta}_0^0 \right)' \hat{\Omega}_T^{-1} \left( \hat{\theta}_T - \hat{\theta}_0^0 \right).$$

5. Reject the null hypothesis $H_0$ if $SSR_T > c$, where the critical value with type-I error 5% is the 95% quantile of a mixture of chi-square distributions $\sum_{k=0}^{K} \pi_k \chi^2(k)$, where the weights depend on the pattern of the $\Omega$ matrix. For instance, if $K = 2$, we have

$$\pi_0 = \frac{1}{2\pi} \cos^{-1} \left( \rho(\theta_1, \theta_2) \right), \quad \pi_1 = \frac{1}{2}, \quad \text{and} \quad \pi_2 = \frac{1}{2} - \pi_0,$$

where $\rho(\theta_1, \theta_2)$ is the correlation between $\theta_1$ and $\theta_2$.

As an illustration, let us continue the example of Section 4.2.2, with the suboptimal unconditional quantile $z_{1t} = \hat{z}_t$, and the optimal conditional quantile $z_{2t} = \hat{z}_t$. The test procedure can be applied with the following instruments:
Case I: $g_{1t} = 1$;

Case II: $g_{1t} = 1, g_{2t} = |r_t - \hat{z}|$;

Case III: $g_{1t} = 1, g_{3t} = |r_t - \hat{z}_{t-1}|$;

Case IV: $g_{1t} = 1, g_{2t} = |r_t - \hat{z}|, g_{3t} = |r_t - \hat{z}_{t-1}|$.

Case I is the standard practice of historical averaging. Case II (resp. Case III) includes also an instrument to capture clustering effect for the suboptimal control (resp. optimal control). Case IV gathers all instruments. The testing procedures are applied for $T = 250$ (i.e. one year of daily data) and $\hat{\Omega}_T$ is computed assuming $L = 5$. The results of the test are given in Table 8, where $Y$ means that the test concludes that the estimated optimal procedure $\hat{z}_t$ is better than the suboptimal procedure $\hat{z}$. When $K \leq 2$, the critical values $c$ are computed by using the closed form expressions provided by Gourieroux et al. (1982). For case IV, the weights $\pi_k$ are obtained through a simulation procedure [see e.g. Wolak (1987)]. As expected, the optimal procedure is proved to be better than the suboptimal procedure in all cases. The comparison of the methodologies based on conditional and unconditional VaR is less clear, if we follow the standard validation practice, which counts the number of exceedances. Indeed, the numbers of times that the observed returns exceed the estimated VaRs are similar in our example for both models.

[Insert Table 8: Test results].

The example above has just been given to illustrate the power of the testing procedure, since in practice the competing control variates $z_{1t}, z_{2t}$ have generally no optimal properties and the validator has only partial knowledge on the way they have been computed.

## 4.3 Dependent risks

This section considers the joint analysis of two lines of financial products. The method has to account for the possible dependence between the extreme risks associated with the two lines. First, we consider this question by focusing on an appropriate criterion function, and , as a
by-product, derives a notion of bidimensional quantile. Then, we discuss the problem of reserve allocation when the global reserve is fixed. The validation approach introduced in Section 4.2 can be easily extended to this framework.

4.3.1 The decision criterion

Let us now consider two lines of financial products with loss and profit values $y_{j,t}, j = 1, 2$, and denote $z_{j,t}, j = 1, 2$ the associated amount of reserves. The criterion function can be generalized to:

$$
\Psi(t, z_{1,t}, z_{2,t}) = E_t \left[ \alpha_{11}(y_{1,t+1} - z_{1,t})^+ (y_{2,t+1} - z_{2,t})^- + \alpha_{12}(y_{1,t+1} - z_{1,t})^+ (y_{2,t+1} - z_{2,t})^- + \alpha_{21}(y_{1,t+1} - z_{1,t})^- (y_{2,t+1} - z_{2,t})^+ + \alpha_{22}(y_{1,t+1} - z_{1,t})^- (y_{2,t+1} - z_{2,t})^- \right],
$$

(4.3.1)

where $\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$ are positive weights. By introducing the cross-products, the loss function accounts for all types of joint losses on the two product lines.

If $\alpha_{11} = \alpha^2, \alpha_{12} = \alpha_{21} = \alpha(1 - \alpha), \alpha_{22} = (1 - \alpha)^2$, the criterion becomes:

$$
\Psi(t, z_{1,t}, z_{2,t}) = E_t \left\{ \alpha(y_{1,t+1} - z_{1,t})^+ + (1 - \alpha)(y_{1,t+1} - z_{1,t})^- \right\} \left\{ \alpha(y_{2,t+1} - z_{2,t})^+ + (1 - \alpha)(y_{2,t+1} - z_{2,t})^- \right\}.
$$

(4.3.2)

or, equivalently,

$$
\Psi(t, z_{1,t}, z_{2,t}) = E_t \left[ \alpha(y_{1,t+1} - z_{1,t})^+ + (1 - \alpha)(y_{1,t+1} - z_{1,t})^- \right] E_t \left[ \alpha(y_{2,t+1} - z_{2,t})^+ + (1 - \alpha)(y_{2,t+1} - z_{2,t})^- \right]

+ Cov_t \left[ \alpha(y_{1,t+1} - z_{1,t})^+ + (1 - \alpha)(y_{1,t+1} - z_{1,t})^- \right] \left[ \alpha(y_{2,t+1} - z_{2,t})^+ + (1 - \alpha)(y_{2,t+1} - z_{2,t})^- \right].
$$

(4.3.3)

The first component is the product of criterion values for a separate analysis of the two product
lines, whereas the second term emphasizes the effect of the possible dependence between risks.

**Remark 1.** As in Section 4.2, alternative decision criteria can be introduced by weighting differently the losses according to their magnitudes. For instance, we can consider the criterion:

$$
\Psi(t, z_{1,t}, z_{2,t}; A_1, A_2) = E_t \left\{ \alpha_{11} \left[ A_1(y_{1,t+1}) - A_1(z_{1,t}) \right]^+ \left[ A_2(y_{2,t+1}) - A_2(z_{2,t}) \right]^+ \\
+ \alpha_{12} \left[ A_1(y_{1,t+1}) - A_1(z_{1,t}) \right]^+ \left[ A_2(y_{2,t+1}) - A_2(z_{2,t}) \right]^-
+ \alpha_{21} \left[ A_1(y_{1,t+1}) - A_1(z_{1,t}) \right]^- \left[ A_2(y_{2,t+1}) - A_2(z_{2,t}) \right]^+ \\
+ \alpha_{22} \left[ A_1(y_{1,t+1}) - A_1(z_{1,t}) \right]^- \left[ A_2(y_{2,t+1}) - A_2(z_{2,t}) \right]^-ight\},
$$

where $A_1, A_2$ are two increasing functions.

### 4.3.2 The optimal control variates

The optimal control variates are solutions of the first-order conditions (note that the objective function is convex); we get:

$$
\begin{align*}
\frac{\partial \Psi}{\partial z_1}(t, \hat{z}_{1,t}, \hat{z}_{2,t}) &= 0, \\
\frac{\partial \Psi}{\partial z_2}(t, \hat{z}_{1,t}, \hat{z}_{2,t}) &= 0,
\end{align*}
$$

(4.3.4)

where:

$$
\frac{\partial \Psi}{\partial z_1}(t, z_{1,t}, z_{2,t}) = E_t \left\{ -1_{y_{1,t+1}>z_{1,t}} \left[ \alpha_{11}(y_{2,t+1} - z_{2,t})^+ + \alpha_{12}(y_{2,t+1} - z_{2,t})^- \right] \right\}
+ E_t \left\{ 1_{y_{1,t+1}<z_{1,t}} \left[ \alpha_{21}(y_{2,t+1} - z_{2,t})^+ + \alpha_{22}(y_{2,t+1} - z_{2,t})^- \right] \right\},
$$

and a similar expression for $\frac{\partial \Psi}{\partial z_2}(t, z_{1,t}, z_{2,t})$.

The system of first-order conditions has no closed form solution in general. However, it can be easily solved recursively when $\alpha_{11} = \alpha^2, \alpha_{12} = \alpha_{21} = \alpha(1 - \alpha), \alpha_{22} = (1 - \alpha)^2$, since the
criterion function is easily concentrated with respect to $z_{1,t}$ (resp. $z_{2,t}$). Indeed, let us assume that $z_{2,t}$ is given. We have:

$$
\Psi(t, z_{1,t}, z_{2,t}) = E_t \left\{ (y_{1,t+1} - z_{1,t})^+ E \left[ \alpha^2 (y_{2,t+1} - z_{2,t})^+ + \alpha(1 - \alpha)(y_{2,t+1} - z_{2,t})^- \middle| y_{1,t+1}, I_t \right] \right. \\
+ (y_{1,t+1} - z_{1,t})^- E \left[ \alpha(1 - \alpha)(y_{2,t+1} - z_{2,t})^+ + (1 - \alpha)^2(y_{2,t+1} - z_{2,t})^- \middle| y_{1,t+1}, I_t \right] \right\} \\
= E_t \left\{ \delta(y_{1,t+1}, t) \left[ \alpha(y_{1,t+1} - z_{1,t})^+ + (1 - \alpha)(y_{1,t+1} - z_{1,t})^- \right] \right\},
$$

where

$$
\delta(y_{1,t+1}, t) = E_t \left[ \alpha(y_{2,t+1} - z_{2,t})^+ + (1 - \alpha)(y_{2,t+1} - z_{2,t})^- \middle| y_{1,t+1}, I_t \right].
$$

Thus, $\Psi(t, z_{1,t}, z_{2,t})$ is proportional to

$$
E_t^{Q_1} \left[ \alpha(y_{1,t+1} - z_{1,t})^+ + (1 - \alpha)(y_{1,t+1} - z_{1,t})^- \right],
$$

where $Q_1^t$ denotes the probability deduced from the historical one by applying a change of density proportional to $\delta(y_{1,t+1}, t)$. This modified probability depends on the methodology $z_{2,t}$.

Since a similar approach can be used to concentrate with respect to $z_{2,t}$, the optimal strategies can be derived recursively. At step $p$, the strategies are $z_{1,t}^{(p)}$, $z_{2,t}^{(p)}$, say. They are used to derive the two modified probabilities $Q_1^{1,p}$ and $Q_2^{2,p}$. Then $z_{1,t}^{(p+1)}$, $z_{2,t}^{(p+1)}$ are the $\alpha$-quantiles of the modified conditional probabilities $Q_1^{1,p}$ and $Q_2^{2,p}$.

Thus, even for this situation when the weights are restricted by $\alpha_{11} = \alpha^2$, $\alpha_{12} = \alpha_{21} = \alpha(1 - \alpha)$, $\alpha_{22} = (1 - \alpha)^2$, the distributions have to be changed for computing the quantile, and the optimal strategy does not correspond to the standard choice of VaR performed separately for each product line.
4.3.3 Reserve allocation under global constraint

The external regulator controls the risk on some aggregated budget line. For internal control, the management of the bank has to allocate this total reserve among the different sublines. We will consider a global line composed of two sublines $y_{1t}, y_{2t}$. The disaggregate reserves are $z_{1,t}, z_{2,t}$, whereas the total reserve is $z_t = z_{1,t} + z_{2,t}$. If we retain the criterion function (4.3.1), the optimization problem becomes:

$$
\begin{align*}
\min_{z_{1,t}, z_{2,t}} & \Psi(t, z_{1,t}, z_{2,t}), \\
\text{s.t.} & \quad z_{1,t} + z_{2,t} = z_t.
\end{align*}
$$

(4.3.6)

The restriction on the total reserve can be solved to get the new optimization problem:

$$
\begin{align*}
\min_{z_{1,t}} E_t \left[ \alpha_{11} \left( y_{1,t+1} - z_{1,t} \right)^+ + \alpha_{12} \left( y_{2,t+1} - z_{1,t} \right)^+ + \alpha_{21} \left( y_{1,t+1} - z_{1,t} \right)^- + \alpha_{22} \left( y_{2,t+1} - z_{1,t} \right)^- \right]. \\
\end{align*}
$$

(4.3.7)

The solutions of this problem will be denoted by $\hat{z}_{j,t}(z_t, \alpha), j = 1, 2$, where $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$. Thus, $\hat{z}_{j,t}(z_t, \alpha)$ solves:

$$
\begin{align*}
\frac{\partial \Psi}{\partial z_1} \left( t, \hat{z}_{1,t}(z_t, \alpha), z_t - \hat{z}_{1,t}(z_t, \alpha) \right) &= 0, \\
\hat{z}_{2,t}(z_t, \alpha) &= z_t - \hat{z}_{1,t}(z_t, \alpha),
\end{align*}
$$

(4.3.8)

where:

$$
\frac{\partial \Psi}{\partial z_1} \left( t, z_{1,t}, z_t - z_{1,t} \right)
= E_t \left[ y_{1,t+1} > z_{1,t} \frac{1}{z_{1,t} - z_{1,t}} \alpha_{11} - y_{1,t+1} > z_{1,t} \frac{1}{z_{1,t} - z_{1,t}} \alpha_{12} - y_{2,t+1} < z_{1,t} \frac{1}{z_{1,t} - z_{1,t}} \alpha_{21} + y_{2,t+1} < z_{1,t} \frac{1}{z_{1,t} - z_{1,t}} \alpha_{22} \right] (y_{1,t+1} - y_{2,t+1} + z_t - 2z_{1,t}).
$$
4.3.4 Comparison of the methodologies

By comparing the standard approach and the new methodologies introduced in Sections 4.3.2, 4.3.3, we can distinguish three ways of fixing the reserves $z_{1,t}, z_{2,t}$.

Method 1: Fix a risk level $\alpha^*$ and compute the disaggregate reserve levels by $\hat{z}_{j,t}(\alpha^*) = F_{j,t}^{-1}(\alpha^*)$, where $F_{j,t}$ is the conditional cdf of $y_{j,t}$.

Method 2: Apply the optimization problem (4.3.1) without restrictions on $z_{1,t}, z_{2,t}$. We get solutions $\hat{\hat{z}}_{j,t}(\alpha), j = 1, 2$, say, which depend on $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$.

Method 3: Apply a two-step approach. First fix a risk level $\alpha^*$ for the aggregated line and compute $\hat{z}_t(\alpha^*) = F_t^{-1}(\alpha^*)$, where $F_t$ is the conditional cdf of the aggregate line $y_t = y_{1,t} + y_{2,t}$. Then, look for solutions of the constrained optimization problem, that are $\hat{\hat{z}}_{j,t}[\hat{z}_t(\alpha^*), \alpha], j = 1, 2$.

The three approaches above depend on different risk level parameters, that are $\alpha^*$ for method 1, $\alpha$ for method 2, and $(\alpha^*, \alpha)$ for method 3. They are difficult to compare in a general setting. We discuss below the reserve allocation for specific weighting functions and return distributions.

i) Unconstrained methods

As mentioned above, methods 1 and 2 coincide if $\alpha_{11} = (\alpha^*)^2, \alpha_{22} = (1 - \alpha^*)^2, \alpha_{12} = \alpha_{21} = \alpha^*(1 - \alpha^*)$, and if the disaggregate risks are independent. In the sequel, we select these weights, but do not assume independent risks. To highlight the differences between methods 1 and 2, let us consider the case of i.i.d. Gaussian variables $(y_{1,t}, y_{2,t})' \sim N\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}\right)$. It is equivalent to write $y_{j,t} = m_j + \sigma_j u_{j,t}, j = 1, 2$, where $(u_{1,t}, u_{2,t})' \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. It is easily checked that the reserves computed with $y_{j,t}$ for both methods 1 and 2 are deduced from the reserves computed with $u_{j,t}, j = 1, 2$ by the same affine transformations $u_j \rightarrow m_j + \sigma_j u_j$. Thus, without loss of generality, we can assume $m_1 = m_2 = 0, \sigma_1^2 = \sigma_2^2 = 1$; then, the solutions satisfy $z_{1,t}(\alpha) = z_{2,t}(\alpha)$ for both methods. We provide in Figure 23 the pattern of...
the disaggregate reserve level as function of correlation $\rho$, for $\alpha^* = 95\%$. The reserve level corresponding to $\rho = 0$ corresponds to the standard derivation line per line. The horizontal line is 1.645, representing the 95th percentile of the standard normal distribution. In order to reduce the effect of sampling errors, we choose sample size to be 20000 when performing the simulation and computation. We refer this as the true pattern and it is displayed by the solid curve. For comparison, finite sample patterns are also provided with $T = 100, 250, 500, 1000$.

[Figure 23: Disaggregate reserve level, function of $\rho$]

The true pattern suggests that, when estimating jointly, one tends to get higher reserve levels than that estimated independently if $\rho \neq 0$. This is due to the behavior of the weighting function defined (4.3.5). As illustrated in Figure 24, for a non zero correlation $\rho$, the weighting function takes large values at both tails of $y_1$ and small values in the middle. As a consequence, the modified cdf has fatter tails than the marginal cdf. This is illustrated by the graph at the right side of Figure 24, where the cdf of $N(0, 1)$ (the shorter dashed line), and the modified cdf with $\rho = \pm 0.9$ (dashed and solid lines) are plotted.

Errors due to small sample size can be substantial. It is seen in Figure 23 that, for sample sizes less than 500 (i.e. 2 years), the reserve level could be largely underestimated. This underestimation is diminished when sample size rises up to 1000 (i.e. 4 years).

ii) Constrained method

Let us now consider Method 3. Due to the two-step approach, the reserves computed with $y_{j,t}, j = 1, 2$ are no longer deduced by linear transformations of the reserves computed with $u_{j,t}, j = 1, 2$. To highlight this effect, we consider returns $(y_{1,t}, y_{2,t})' \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho \sigma \\ \rho \sigma & 1 \end{pmatrix}\right]$, with a distribution depending on the correlation $\rho$ and the ratio of variances $\sigma^2$.

[Figure 25: Disaggregate reserve level as function of $\rho$ with aggregate reserve constraint]

Figure 25 shows a symmetric situation (i.e. $\sigma = 1$), where the objective function values are plotted against the proportion $\hat{z}_{t}[\alpha^*(\alpha), \alpha]/\hat{z}_{t}(\alpha^*)$, for $\rho = 0, 0.5, 0.9$. The reserve level allocated to asset 1 is roughly one half of the aggregate reserve level for modest correlation
(say, $\rho < 0.6$). However, if correlation is large, it is optimal to assign more reserve to one asset than the other. To illustrate this formally, let us consider the limiting case where $\rho = 1$ (i.e. $y_{1,t} = y_{2,t}$ for all $t = 1, \ldots, T$). The FOC (4.3.8) is satisfied if $z_{1,t}^* = z_{2,t}^* = z_t/2$. Evaluating at the same value of $z_{j,t}^*, j = 1, 2$, the second order condition (SOC) is given by:

$$
\frac{\partial^2 \Psi}{\partial z_1^2}(t, z_{1,t}^*, z_t - z_{1,t}^*) = -2E_t \left\{ \left[ 1_{y_{1,t+1} > z_{1,t}^*} + 1_{y_{2,t+1} > z_t - z_{1,t}^*} \alpha_{11} + 1_{y_{1,t+1} < z_{1,t}^*} + 1_{y_{2,t+1} < z_t - z_{1,t}^*} \alpha_{22} \right] \right\}
$$

$$
= -2 \left( \alpha_{11} Pr \left[ y_{1,t} > \frac{z_t}{2} \right] + \alpha_{22} Pr \left[ y_{1,t} < \frac{z_t}{2} \right] \right) < 0. \quad (4.3.9)
$$

The FOC and SOC imply that a local maximum of the criterion value is achieved when $z_{1,t}^*/z_t = 0.5$. Thus, the minimal value of the criterion function, if exist, must be a different solution ($\tilde{z}_{j,t}^*, j = 1, 2$) to the FOC (4.3.8), where $\tilde{z}_{1,t}^* \neq \tilde{z}_{2,t}^* = z_t - z_{1,t}^*$. If $\tilde{z}_{1,t}^* < \tilde{z}_{2,t}^*$, this implies:

$$
\alpha_{11} Pr \left[ y_{1,t} > \frac{\tilde{z}_{2,t}^*}{2} \right] - \alpha_{12} Pr \left[ \tilde{z}_{1,t}^* \leq y_{1,t} \leq \frac{\tilde{z}_{2,t}^*}{2} \right] + \alpha_{22} Pr \left[ y_{1,t} < \frac{\tilde{z}_{1,t}^*}{2} \right] = 0. \quad (4.3.10)
$$

Given condition (4.3.10), the SOC evaluated at $(\tilde{z}_{j,t}^*, j = 1, 2)$ is given by:

$$
\frac{\partial^2 \Psi}{\partial z_1^2}(t, \tilde{z}_{1,t}^*, z_t - \tilde{z}_{1,t}^*) = (z_t - 2\tilde{z}_{1,t}^*) \left[ \alpha_{22} f(\tilde{z}_{1,t}^*) - \alpha_{11} f(z_t - \tilde{z}_{1,t}^*) \right], \quad (4.3.11)
$$

where $f$ denotes the pdf of $y_1$ (resp. $y_2$). In general, a positive SOC requires the following relationship of $(\alpha_{11}, \alpha_{22}, f, \text{ and } z_t)$:

$$
\alpha_{22} f(\tilde{z}_{1,t}^*) - \alpha_{11} f(z_t - \tilde{z}_{1,t}^*) > 0.
$$

Since $z_{1,t}$ and $z_{2,t}$ are interchangeable, the minimal point is not unique.

We report, in Figure 26, the relative optimal allocation $\tilde{z}_{1,t}/\tilde{z}_t$ as function of $\rho$ and $\sigma^2$. The data are generated with $\rho$ ranging from -0.9 to 0.9 and $\sigma^2 = \{1.5, 2, 2.5, 3\}$. When the value of correlation is small (say less than 0.5), we observe larger reserves are allocated to riskier business line. However, the reverse is true for strongly and positively correlated business lines. In both cases, the relative reserve allocation to riskier business line increases with the risk level.
4.4 Conclusion

The validation of reserve levels proposed by a bank often requires investigation of models’ forecasting accuracies of the VaR. A simple criterion consists in comparing the observed number of exceedances of the estimated VaR with the desired level. However, this criterion is not sufficient since the VaRs estimated by two completely different models may be exceeded by similar number of times in a finite sample. In this paper, we propose an alternative criterion which takes the level of loss beyond VaR into account. In the univariate case, this criterion coincides with the objective function of the standard quantile regression. We show that both the optimal criterion values and the values of lack of optimality can vary with the asset return volatility. A natural extension is conducted from the single line analysis to the risks with multiple lines. We explain how to allocate reserves to individual lines of risk with or without a constraint on global reserve.

Appendices

A.1 The component for lack of optimality

i) We have:

$$
\Psi(t, z_t) - \Psi(t, \hat{z}_t) = \alpha(\hat{z}_t - z_t) + E_t \left[ (z_t - y_{t+1})1_{z_t > y_{t+1}} \right] - E_t \left[ (\hat{z}_t - y_{t+1})1_{\hat{z}_t > y_{t+1}} \right].
$$

Let us consider the case $\hat{z}_t > z_t$. We get:

$$
\Psi(t, z_t) - \Psi(t, \hat{z}_t) = \alpha(\hat{z}_t - z_t) + E_t \left[ (z_t - y_{t+1})1_{z_t > y_{t+1}} - (\hat{z}_t - y_{t+1})1_{\hat{z}_t > y_{t+1}} \right] - E_t \left[ (\hat{z}_t - y_{t+1})1_{\hat{z}_t < y_{t+1} < z_t} \right].
$$

$$
= \alpha(\hat{z}_t - z_t) + (z_t - \hat{z}_t)P_t [z_t > y_{t+1}] - E_t \left[ (\hat{z}_t - y_{t+1})1_{\hat{z}_t < y_{t+1} < z_t} \right]
$$

$$
= (\hat{z}_t - z_t)E_t \left[ 1_{z_t < y_{t+1} < \hat{z}_t} \right] - E_t \left[ (\hat{z}_t - y_{t+1})1_{\hat{z}_t < y_{t+1} < z_t} \right]
$$

$$
= E_t \left[ (y_{t+1} - z_t)1_{z_t < y_{t+1} < \hat{z}_t} \right].
$$
A similar computation can be done when \( z_t > \hat{z}_t \).

ii) We have:

\[
E_t \left[ (y_{t+1} - z_t) \mathbb{1}_{z_t < y_{t+1} < \hat{z}_t} \right] \\
= \int_{z_t}^{\hat{z}_t} (y - z_t) d F_t(y) \\
= \left( y - z_t \right) F_t(y) \bigg|_{z_t}^{\hat{z}_t} - \int_{z_t}^{\hat{z}_t} F_t(y) dy \\
= \left( \hat{z}_t - z_t \right) F_t(\hat{z}_t) - \int_{z_t}^{\hat{z}_t} F_t(y) dy \\
= \int_{z_t}^{\hat{z}_t} \left[ F_t(\hat{z}_t) - F_t(y) \right] dy.
\]

iii) Finally, if \( z_t \simeq \hat{z}_t \), we get:

\[
\int_{z_t}^{\hat{z}_t} \left[ F_t(\hat{z}_t) - F_t(y) \right] dy \simeq \int_{z_t}^{\hat{z}_t} f_t(\hat{z}_t)(\hat{z}_t - y) dy \simeq \frac{1}{2} f_t(\hat{z}_t)(\hat{z}_t - z_t)^2.
\]

### A.2 Weight correction

Let us assume that \( y_{t+1} = m_t + \sigma_t u_{t+1} \), where the error term has a path independent distribution with pdf \( g \) and cdf \( G \). The conditional pdf and cdf of \( y_{t+1} \) are respectively given by:

\[
f_t(y) = \frac{1}{\sigma_t} g \left( \frac{y - m_t}{\sigma_t} \right), \quad F_t(y) = G \left( \frac{y - m_t}{\sigma_t} \right).
\]

We deduce:

\[
f_t \left[ F_t^{-1}(\alpha) \right] = f_t \left[ m_t + \sigma_t G^{-1}(\alpha) \right] = \frac{1}{\sigma_t} g \left[ G^{-1}(\alpha) \right].
\]
Bibliography


Politis, editors, *Recent Advances and Trends in Nonparametric Statistics*, Elsevier Science, Chapter 4, 123–137.


Chapter 5

Tables and Figures
Table 1: Relationship between VaR and Tail-VaR

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$U(a, b)$</th>
<th>$\gamma(1, \lambda)$</th>
<th>Pareto$(a, b)$</th>
<th>$N(0, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{TVaR(p)}{VaR(p)}$</td>
<td>$\frac{b(2-p)+ap}{2(b(1-p)+ap)}$</td>
<td>$1 - \frac{1}{\log(p)}$</td>
<td>$\frac{a}{a-1}$</td>
<td>$\frac{\phi\left(\Phi^{-1}(1-p)\right)}{p\Phi^{-1}(1-p)}$</td>
</tr>
<tr>
<td>$g(p)$</td>
<td>$p/2$</td>
<td>$p/e$</td>
<td>$\left(\frac{a-1}{a}\right)^a p$</td>
<td>$1 - \Phi\left[\frac{1}{p}\phi\left(\Phi^{-1}(1-p)\right)\right]$</td>
</tr>
</tbody>
</table>

The first row gives the distribution specification. The second row provides the ratio between $TVaR(p)$ and $VaR(p)$. In the third row, we provide the value of $p^* = g(p)$ such that $TVaR(p) = VaR(p^*)$. $\Phi$ and $\phi$ denote the cdf and pdf of standard normal distribution, respectively. The parameter $a$ of the Pareto distribution is strictly larger than 1 to ensure that the $TVaR$ exists.

Table 2: Sensitivity of the VaR and Tail-VaR

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$U[a, b]$</th>
<th>$\gamma(1, \lambda)$</th>
<th>Pareto$(a, b)$</th>
<th>$N(0, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\partial VaR(p)}{\partial p}$</td>
<td>$a - b$</td>
<td>$-\frac{1}{\lambda_p}$</td>
<td>$-\frac{b}{a} p^{-(a+1)/a}$</td>
<td>$-\frac{1}{\phi\left(\Phi^{-1}(1-p)\right)}$</td>
</tr>
<tr>
<td>$\frac{\partial TVaR(p)}{\partial p}$</td>
<td>$(a - b)/2$</td>
<td>$-\frac{1}{\lambda_p}$</td>
<td>$-\frac{b}{a-1} p^{-(a+1)/a}$</td>
<td>$\frac{-1}{p^2} \phi\left(\Phi^{-1}(1-p)\right) + \frac{1}{p} \Phi^{-1}(1-p)$</td>
</tr>
</tbody>
</table>

The first row gives the distribution specification. The second and third rows provide the sensitivity of Value-at-Risk and Tail-VaR, respectively. $\Phi$ and $\phi$ denote the cdf and pdf of standard normal distribution, respectively.
Table 3: Asymptotic variance and relative accuracy of the estimated VaR($p$)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$VaR(p)(Q_0(1-p))$</th>
<th>Variance</th>
<th>$\sigma_{VaR(p)}/VaR(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform $U[a, b]$</td>
<td>$b(1-p) + ap$</td>
<td>$(b-a)^2 p(1-p)$</td>
<td>$(b-a)\sqrt{p(1-p)}/b(1-p)+ap$</td>
</tr>
<tr>
<td>Exponential $\gamma(1, \lambda)$</td>
<td>$-\log(p)/\lambda$</td>
<td>$1/\lambda^2 (1-p)$</td>
<td>$-\sqrt{1-p}/\log(p)$</td>
</tr>
<tr>
<td>Pareto $Pareto(a, b)$</td>
<td>$b p^{-1/a}$</td>
<td>$\left(\frac{b}{a}\right)^2 \frac{1-p}{p}$</td>
<td>$\frac{1}{\eta} p^{1/a} \sqrt{\frac{1-p}{1-\frac{1-p}{a}}}$</td>
</tr>
<tr>
<td>Gaussian $N(\mu, \sigma^2)$</td>
<td>$\mu + \sigma \Phi^{-1}(1-p)$</td>
<td>$\sigma^2 \phi\left(\Phi^{-1}(1-p)\right) \sigma^2 p(1-p)$</td>
<td>$\frac{\sigma \sqrt{p(1-p)}}{\phi\left(\Phi^{-1}(1-p)\right) \phi\left(\Phi^{-1}(1-p)\right)}$</td>
</tr>
<tr>
<td>Lévy $Lévy(c)$</td>
<td>$e^{c \Phi^{-1}(1-p)}$</td>
<td>$2\pi^2 p(1-p) \frac{\pi}{\Phi^{-1}(1+2c)} \exp\left(\frac{1}{\Phi^{-1}(1+2c)} \left(\frac{1+2c}{2}\right)^2\right)$</td>
<td>$\frac{\sqrt{p(1-p)}}{\phi\left(\Phi^{-1}(1+2c)\right) \phi\left(\Phi^{-1}(1+2c)\right)}$</td>
</tr>
<tr>
<td>Cauchy $Cauchy(m, b)$</td>
<td>$m + b \tan\left(\pi \left(\frac{1}{2} - p\right)\right)$</td>
<td>$(b\pi)^2 \left{\sec\left(\pi \left(\frac{1}{2} - p\right)\right)\right}^4 p(1-p)$</td>
<td>$\frac{b\pi \left{\sec\left(\pi \left(\frac{1}{2} - p\right)\right)\right}^4 \sqrt{p(1-p)}}{m + b \tan\left(\pi \left(\frac{1}{2} - p\right)\right)}$</td>
</tr>
</tbody>
</table>

The first column gives the distribution specifications with their parameters. The second column lists the quantile functions and the third column provides the asymptotic variances for the nonparametric estimators of VaR. Relative accuracy measured by the ratio between the standard deviation and the $VaR(p)$ are listed in column 4. $\phi$ (resp. $\Phi$) is the pdf (resp. cdf) of standard normal distribution. $\sec$ is the secant function $\sec[x] = 1/\cos[x]$. The variance is defined for the estimator scaled by $\sqrt{T}$. 

Table 4: Asymptotic variance and relative accuracy of the estimated \( TVaR(p) \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( TVaR(p) )</th>
<th>Variance</th>
<th>( \sigma(\widehat{TVaR}_T(p))/TVaR(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform ( U[a, b] )</td>
<td>( \frac{b(2-p)+ap}{2} )</td>
<td>( \frac{(a-b)^2p(4-3p)}{12} )</td>
<td>( \frac{(b-a)\sqrt{2(1-3p)}}{2b+(a-b)p} )</td>
</tr>
<tr>
<td>Exponential ( \gamma(1, \lambda) )</td>
<td>( \frac{1-\log(p)}{\lambda} )</td>
<td>( \frac{2-p}{\lambda^2p} )</td>
<td>( \sqrt{\frac{2-p}{p}}\left(1 - \log(p)\right) )</td>
</tr>
<tr>
<td>Pareto ( Pareto(a, b) )</td>
<td>( \frac{ab}{a-1}p^{-1/a} )</td>
<td>( b^2p^{-2+\frac{a}{a-1}} \left{ \frac{a}{a-1} - \frac{a^2}{(a-1)^2} + (1 - p) \left[ \frac{a}{a-1} - 1 \right]^2 \right} )</td>
<td>( \frac{(a-1)p^{-1/2}}{\sqrt{\frac{a^2 - (a-1)^2}{a}} + (1 - p)\left[ \frac{a}{a-1} - 1 \right]^2} )</td>
</tr>
<tr>
<td>Gaussian ( N(0, 1) )</td>
<td>( \frac{1}{p} \phi\left( \Phi^{-1} \right) )</td>
<td>( \frac{1}{p^2} \left[ p + (2p - 1)\phi\left( \Phi^{-1} \right)\Phi^{-1} + (1 - p)p\left[ \Phi^{-1} \right]^2 - \left[ \phi\left( \Phi^{-1} \right) \right]^2 \right] )</td>
<td>( \frac{\left[ p + (2p - 1)\phi\left( \Phi^{-1} \right)\Phi^{-1} + (1 - p)p\left[ \Phi^{-1} \right]^2 - \left[ \phi\left( \Phi^{-1} \right) \right]^2 \right]}{\phi\left( \Phi^{-1} \right)} )</td>
</tr>
</tbody>
</table>

The first column gives the distribution specifications with their parameters. The second column provides \( TVaR(p) \) and the third column lists the asymptotic variance of the nonparametric estimator of \( TVaR(p) \). The relative accuracy \( (\sigma(\widehat{TVaR}_T(p))/TVaR(p)) \) is provided in column 4. \( \Phi \) and \( \phi \) denote the cdf and pdf of standard normal distribution, respectively. For shortening the expressions associated with the standard normal distribution, we denote \( \Phi^{-1}(1 - p) \) by \( \Phi^{-1} \). The variance is defined for the estimator scaled by \( \sqrt{T} \).
Table 5: Summary statistics of currency portfolio returns

<table>
<thead>
<tr>
<th>Basic</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Excess Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US$</td>
<td>0.000002</td>
<td>0.000049</td>
<td>-0.680470*</td>
<td>5.195544*</td>
</tr>
<tr>
<td>SIN$</td>
<td>-0.000008</td>
<td>0.000042</td>
<td>-1.195269*</td>
<td>12.987396*</td>
</tr>
<tr>
<td>20 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US$</td>
<td>0.000204</td>
<td>0.000968</td>
<td>-0.981831*</td>
<td>2.356374*</td>
</tr>
<tr>
<td>SIN$</td>
<td>-0.000014</td>
<td>0.000697</td>
<td>-0.483116*</td>
<td>0.142152</td>
</tr>
</tbody>
</table>

The star (*) introduced for mean and skewness indicates significant results and for excess kurtosis result significantly different from 3, that is the kurtosis of a standard normal distribution.

Table 6: Summary statistics

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi(t, z_t)$</td>
<td>0.0926</td>
<td>0.0005</td>
<td>0.0762</td>
<td>2.4029</td>
</tr>
<tr>
<td>$c(t, z_t)$</td>
<td>0.0178</td>
<td>0.0003</td>
<td>1.0883</td>
<td>3.5836</td>
</tr>
<tr>
<td>$\psi(t, z)$</td>
<td>0.1104</td>
<td>0.0003</td>
<td>2.6492</td>
<td>10.8276</td>
</tr>
<tr>
<td>Series</td>
<td>$\psi(t, \hat{z}_t)$</td>
<td>$c(t, z_t)$</td>
<td>$\psi(t, z)$</td>
<td>Series</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------</td>
</tr>
<tr>
<td>$\rho(0)$</td>
<td>1.0000</td>
<td>-0.7024</td>
<td>0.6730</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(1)$</td>
<td>0.9562</td>
<td>-0.6681</td>
<td>0.6465</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(2)$</td>
<td>0.8637</td>
<td>-0.6006</td>
<td>0.5877</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(3)$</td>
<td>0.7988</td>
<td>-0.5512</td>
<td>0.5822</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(4)$</td>
<td>0.6082</td>
<td>-0.4021</td>
<td>0.4263</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(5)$</td>
<td>0.4237</td>
<td>-0.2383</td>
<td>0.3302</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(10)$</td>
<td>0.0286</td>
<td>-0.0163</td>
<td>0.0215</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
<tr>
<td>$\rho(25)$</td>
<td>-0.1145</td>
<td>0.0907</td>
<td>-0.0685</td>
<td>$\psi(t, \hat{z}_t)^2$</td>
</tr>
</tbody>
</table>
Table 8: Test results

<table>
<thead>
<tr>
<th>Cases</th>
<th>$SSR_T$</th>
<th>$c$</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>13.0939</td>
<td>2.7056</td>
<td>Y</td>
</tr>
<tr>
<td>II</td>
<td>14.7373</td>
<td>4.9471</td>
<td>Y</td>
</tr>
<tr>
<td>III</td>
<td>14.4964</td>
<td>4.7363</td>
<td>Y</td>
</tr>
<tr>
<td>IV</td>
<td>15.1493</td>
<td>5.9920</td>
<td>Y</td>
</tr>
</tbody>
</table>

The test statistics $SSR_T$ are provided in column 2. Column 3 lists the critical values. When $K \leq 2$, we obtain the values of $c$ by using the closed form expressions of the distribution of $SSR_T$ and a grid search numerical algorithm. Otherwise, the weights $\pi_k$ are obtained through a simulation procedure. The letter "Y" means that the test concludes that the estimated optimal procedure $\hat{z}_t$ is better than the suboptimal procedure $\hat{z}$. 
Figure 1: $\frac{TVaR(p)}{VaR(p)}$ as function of $p$. 

(a) $X \sim U(0,1)$

(b) $X \sim \gamma(1,2)$

(c) $X \sim N(0,1)$
Figure 2: $\frac{\text{TVaR}(p)}{\text{VaR}(p)}$ as function of $a$ when $X \sim \text{Pareto}(a, b)$.

Figure 3: $g(p)$ when $X \sim N(0, 1)$.

Figure 4: $\alpha$ as function of $a$ when $X \sim \text{Pareto}(a, b)$.
Figure 5: Plot of the weights $w(u, p)$ as a function of $u$ in the sensitivity of $PH(p)$.

Figure 6: Plot of the weights $w(u, p)$ as a function of $u$ in the sensitivity of $EX(p)$. 
(a) Variance
Figure 7: Variance and relative accuracy for $\tilde{VaR}_T(p)$. 

(a) Variance

(b) Relative accuracy
(a) Variance
Figure 8: Variance and relative accuracy for $\overline{TVaR}_T(p)$. 

- $U(0, 1)$
- $\gamma(1, \lambda)$
- Pareto$(5, 0.3)$
- $N(0, 1)$
$U(0, 1)$

$\gamma(1, \lambda)$

$\text{Pareto}(5, 0.3)$

$N(0, 1)$

(a) Variance
Figure 9: Variance and relative accuracy for $\widetilde{PH_T}(p)$. 
Horizon = 1 day
Horizon = 20 days
(a) Base: US$
(b) Base: SIN$

Figure 10: Value-at-Risk of the currency portfolio.
Horizon = 1 day  
Horizon = 20 days

(a) Base: US$

(b) Base: SIN$

Figure 11: Tail-VaR of the currency portfolio.
Figure 12: $\frac{TVaR(p)}{VaR(p)}$ of the equally weighted currency portfolio.
Figure 13: Estimated link $g(p)$ for the equally weighted currency portfolio.
Figure 14: Estimated link $g(p)$ for the currency portfolios and varying allocations.
Figure 15: Term structure of slope parameter $\alpha$

(a) base: US$

(b) base: SIN$

Figure 15: Term structure of slope parameter $\alpha$
Figure 16: Estimators of delta-TVaR and delta-PH
The estimates are obtained from a sample size 250. The delta-TVaR and delta-PH are estimated for $\alpha$ varying from 0.1 to 0.9 and the risk/pessimistic parameters are equal to 0.05 and 0.7, respectively. The kernel estimators are represented by solid line and the empirical estimators by dashed line.

Figure 17: Finite sample variances of estimators of delta-TVaR and delta-PH
The variances are calculated from a sample size 250 and a simulation size 1000. The delta-TVaR and delta-PH are estimated for $\alpha$ varying from 0.1 to 0.9 and the risk/pessimistic parameters are equal to 0.05 and 0.7, respectively. The kernel estimators are represented by solid line and the empirical estimators by dashed line.
Figure 18: Finite sample distribution of estimators of \((a^1)^*\)
The densities are estimated from 400 simulated values of the normalized allocation on asset one using kernel method. The curves are provided for the range \([-100, 100]\). The true value (vertical line) is -4.6. The densities plotted are for i) the parametric estimator (solid line); ii) the empirical estimator when \(p = 0.05\) (dashed line); and iii) the empirical estimator when \(p = 0.1\) (shorter dashed line). The averages calculated from the simulated estimates are 2.4, 1.66 and 0.86, respectively. The finite sample variances are 1656.32 (parametric), 139.02 (empirical when \(p = 0.05\)) and 117.76 (empirical when \(p = 0.1\)).
Figure 19: Finite sample distribution of estimators of standard Sharpe ratio

The densities are estimated from 400 simulated values of the standard Sharpe ratio using kernel method. The true value (vertical line) is 0.107. The densities plotted are for i) the parametric estimator (solid line); ii) the empirical estimator when $p = 0.05$ (dashed line); and iii) the empirical estimator when $p = 0.1$ (shorter dashed line). The averages calculated from the simulated estimates are 0.12, 0.326 and 0.237, respectively. The finite sample variances are 0.002 (parametric), 0.094 (empirical when $p = 0.05$) and 0.038 (empirical when $p = 0.1$).
Figure 20: Joint Simulated Path of \((r_t, \sigma_t^2)\)

Figure 21: Conditional and Unconditional Quantiles
Figure 22: Joint Evolution of $\Psi(t, \hat{z}_t)$, $\Psi(t, z)$ and $c(t, z_t)$

Figure 23: Disaggregate reserve level, function of $\rho$
Figure 24: Value of weights as function of $(\rho, y_1)$ and modified cdf when $\rho = \pm 0.9$

Figure 25: Disaggregate reserve level as function of $\rho$ with aggregate reserve constraint
Figure 26: Relative allocation as function of $\rho$ and $\sigma^2$