Risk Attribution and Portfolio Performance Measurement-An Overview

Yongli Zhang

Department of Economics

University of California, Santa Barbara, CA93016, USA

Svetlozar Rachev*

Department of Econometrics and Statistics

University of Karlsruhe, D-76128 Karlsruhe, Germany and

Department of Statistics and Applied Probability

University of California, Santa Barbara, CA93106, USA

September, 2004

Abstract

A major problem associated with risk management is that it is very hard to identify the main resource of risk taken, especially in a large and complex portfolio. This is due to the fact that the risk of individual securities in the portfolio, measured by most of the widely used risk measures such as standard deviation and value-at-risk, don't sum up to the total risk of the portfolio. Although the risk measure of beta in the Capital Asset Pricing Model seems to survive this major deficiency, it suffers too much from other pitfalls to become a satisfactory solution. Risk attribution is a methodology to decompose the total risk of a portfolio into smaller terms. It can be applied to any positive homogeneous risk measures, even free of models. The problem is solved in a way that the smaller decomposed units of the total risk are interpreted as the risk contribution of the corresponding subsets of the portfolio. We present here an overview of the methodology of risk attribution, different risk measures and their properties.

*S. Rachev's research was supported by grants from Division of Mathematical, Life and Physical Sciences, College of Letters and Science, University of California at Santa Barbara and the Deutschen Forschungsgemeinschaft.

1 Introduction

The central task of risk management is to locate the main source of risk and find trading strategies to hedge risk. But the major problem is that it is very hard to identify the main source of risk taken, especially in a large and complex portfolio. This is due to the fact that the risk of individual securities, measured by most of the widely used risk measures such as standard deviation and value-at-risk, don't sum up to the total risk of the portfolio. That is, while the stand-alone risk of an individual asset is very significant, it could contribute little to the overall risk of the portfolio because of its correlations with other securities. It could even act as a hedge instrument that reduces the overall risk. Although the risk measure of beta in the Capital Asset Pricing Model(CAPM) survives the above detrimental shortcoming (i.e. the weighted sum of individual betas equals the portfolio beta), it suffers from the pitfalls of the model on which it is based. Furthermore, beta is neither translation-invariant nor monotonic, which are two key properties possessed by a coherent risk measure that we discuss later in section 4.2. Except for the beta, the stand-alone risk measured by other risk measures provide little information about the composition of the total risk and thus give no hint on how to hedge risk.

Risk attribution is a methodology to decompose the total risk of a portfolio into smaller units, each of which corresponds to each of the individual securities, or each of the subsets of securities in the portfolio.[1] The methodology applies to any risk measures, as long as they are positively homogeneous (see definition 3.2). Risk measures having better properties than beta but not additive are now remedied by risk attribution. The problem is solved in a way that the smaller decomposed units of the total risk can be interpreted as the *risk contribution* of the corresponding sub-portfolios. After the primary source of the risk is identified, active portfolio hedging strategies can be carried out to hedge the significant risk already taken.

It is worthwhile to mention that we assume there exits a portfolio already before the risk attribution analysis is taken. This pre-existing portfolio could be a candidate of or even an optimized portfolio. But this is counterintuitive since the optimized portfolio is supposed to be the "best" already and it

[1] Mina(2002) shows that the decompostion can be taken according to any arbiturary partition of the portfolio. The partition could be made according to active investment decisions, such as sector allocation and security selection. For example, the partition could be made by countries, sectors or industries.

seems unnecessary at all to re-examine it. Why shall we bother? The answer to this question is the heart of *risk hedging*. Investors are typically risk averse and they don't feel comfortable when they find out through risk attribution that the major risk of their optimized portfolios is concentrated on one or few assets. They are willing to spend extra money on buying financial insurance such as put and call options in order to hedge their positions of major risk exposure. Furthermore, Kurth et al. (2002) note that optimal portfolios are quite rare in practice, especially in credit portfolio management. It is impossible to optimize the portfolio in one step, even if the causes of bad performance in a credit portfolio have been located. It is still so in general portfolio management context because of the rapid changes of market environment. Traders and portfolio managers often update portfolio on a daily basis. Successful portfolio management is indeed a process consisting of small steps, which requires detailed risk diagnoses, namely risk attribution. The process of portfolio optimization and risk attribution can be repeatedly performed until the satisfactory result is reached.

There are certainly more than one risk measure that are proposed and used by both the academics and practitioners. There has been a debate about which risk measures are more appropriate ([3] and [4]). While the basic idea of decomposing the risk measure is the same, the methods of estimating and computing the components could differentiate a lot for different risk measures. We give a close look at these risk measures and show their properties. The pros and cons of each risk measure will hopefully be clear.

In the next section, we briefly examine beta in the Capital Asset Pricing Model developed by Sharpe and Tobin. Although beta can be a risk measure whose individual risk components sum up to the total risk, the model in which it lives in is under substantial criticism. The rejection of the questionable beta leads us to seek for new tools of risk attribution. In Section 3, under a general framework without specifying the function forms of the risk measures, the new methodology of differentiating the risk measures is presented . Different risk measures are introduced in Section 4. The methods of calculating derivatives of risk measures follows in Section 5. The last section concludes the paper and gives light to future studies.

2 What is wrong with beta?

The Capital Asset Pricing Model (CAPM) is a corner stone of modern finance. It states that when all investors have the same expectation and the same information about the future, the expected excess return of an asset is proportional to the risk, measured by beta, of the asset. This simple yet powerful model can be expressed mathematically as follows:

$$E[R_i] - R_f = \beta_i(E[R_M] - R_f) \tag{1}$$

where R_i is the random return of the *ith* asset, R_f is the return of the riskless asset and R_M is the random return of *the market portfolio*, which is defined as the portfolio consisting of all the assets in the market. Beta β_i is defined as the ratio of the covariance between asset i and the market portfolio and the variance of the market portfolio.

$$\beta_i = \frac{Cov(R_i, R_M)}{\sigma_M^2} \tag{2}$$

As so defined, beta measures the responsiveness of asset i to the movement of the market portfolio. Under the assumption that agents have homogeneous expectations and information set, all investors will combine some of the riskless asset and some of the market portfolio to form their own optimal portfolios. Therefore beta is a measure of risk in CAPM. It is easy to find out that the weighted sum of individual betas in a portfolio equals the beta of the portfolio, i.e.

$$\beta_P = \sum_{i=1}^{N} w_i \beta_i \tag{3}$$

where w_i is the portfolio weight, i.e. the percentage of wealth invested in asset i. We can identify the main source of risk by examining the values of betas. The largest weighted beta implies that asset contributes most to the total risk of the portfolio. Then risk attribution seems easy in the CAPM world.

However, the CAPM and beta have been criticized for their over simplicity and not being representative about the real world we live in. The famous Roll's critique(1977) asserts that the market portfolio is not observable and the only testable implication about CAPM is the efficiency of the market portfolio. Another representative challenge is the paper by Fama

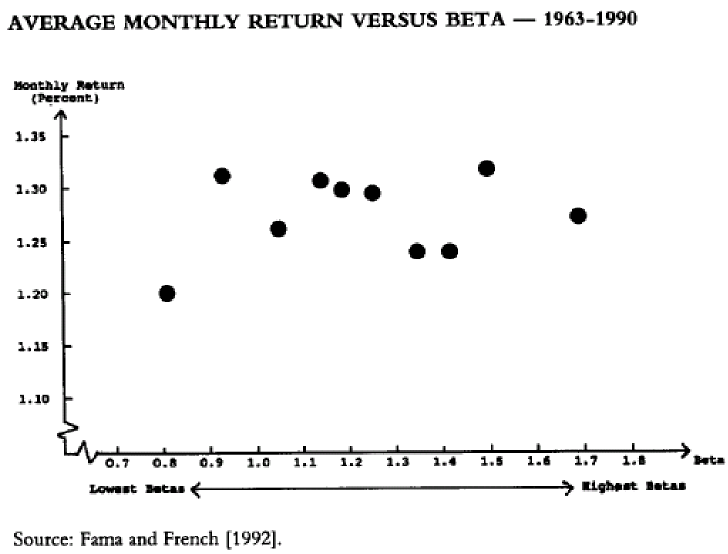


Figure 1: beta and long-run average return are not correlated.

and French(1992). They find some empirical evidence showing that beta and long-run average return are not correlated. (See Figure 1) There are a lot of other similar papers. But we don't attempt to give an extensive survey of the CAPM literature. The point we want to make is that beta, as well as the CAMP that it is based on, is too controversial for us to reply on, especially when there exists an alternative tool to develop risk attribution techniques.

3 Risk Attribution-The Framework

Risk attribution is not a new concept. The term stems from the term *"return attribution"* or *"performance attribution"*, which is a process of attributing the return of a portfolio to different factors according to active investment decisions. The purpose is to evaluate the ability of asset managers. The literature on return attribution or performance attribution started in the 60s,[2] when mutual funds and pension funds were hotly debated. Whereas the literature on risk attribution, which is related to return attribution, didn't start until the mid 90's. Risk attribution differs from return attribution

[2] See Fama (1972) for a short review.

in two major aspects. First, as is clear from their names, the decomposing objectives are risk measures for the former and returns for the latter. Second, the latter uses historical data and thus is ex-post while the former is an ex-ante analysis.

The general framework of decomposing risk measures was first introduced by Litterman (1996), who uses the fact that the volatility, defined as the standard deviation, of a portfolio is linear in position size. His finding is generalized to risk measures that possess this property, which we call later *positive homogeneity*. We shall formally define the general measure of risk.

We assume there are N financial assets in the economy. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\mathcal{G}$ be a set of random variables which are $\mathcal{F}$-measurable. The $N \times 1$ random vector R is the return vector in which the ith element R_i is the random return of asset i, $i = 1, 2, \ldots, N$. Let the $N \times 1$ vector $m \in \mathbb{R}^N$ be the vector of portfolio positions where the ith element m_i is the amount of money invested in asset i, $i = 1, 2, \ldots, N$. A portfolio is represented by the vector m and the portfolio random payoff is $X = m'R \in \mathcal{G}$. All returns belong to some time interval Δt and all positions are assumed to be at the beginning of the time interval.

Definition 3.1 *A* ***risk measure*** *is a mapping* $\rho : \mathcal{G} \to \mathbb{R}$.

Definition 3.2 *A risk measure is* ***homogeneous of degree*** τ *if* $\rho(kX) = k^\tau \rho(X)$, *for* $X \in \mathcal{G}$, $kX \in \mathcal{G}$ *and* $k > 0$.[3]

We are more interested in risk measures which have the property of positive homogeneity because this is one of the properties possessed by the class of *coherent measures*, which we define in the next section. In Litterman's framework, positive homogeneity plays a central role in decomposing risk measures into meaningful components. When all position sizes are multiplied by a common factor $k > 0$, the overall portfolio risk is also multiplied by this common factor. As we can see from the following proposition that each component can be interpreted as the marginal risk contribution of each individual asset or a subset of assets in the portfolio from small changes in the corresponding portfolio position sizes.

[3] Some authors use the term *positive homogeneity* for simplicity to mean the case when $\tau = 1$, i.e. the risk measure is homogeneous of degree one. But this term can cause confusion in some cases thus is not adopted in this paper.

Proposition 3.3 *(Euler's Formula) Let ρ be a homogeneous risk measure of degree τ. If ρ is partially differentiable with respect to m_i, $i = 1, \ldots, N$,*[4] *then*

$$\rho(X) = \frac{1}{\tau}(m_1 \frac{\partial \rho(X)}{\partial m_1} + \cdots + m_N \frac{\partial \rho(X)}{\partial m_N}) \tag{4}$$

Proof. *Consider a mapping $\mu : \mathbb{R}^N \to \mathcal{G}$ and $\mu(m) = m'R$, for $m \in \mathbb{R}^N$. Then $\rho(X) = \rho \circ \mu(m)$, where $\rho \circ \mu$ is a composite mapping: $\mathbb{R}^N \to \mathcal{G} \to \mathbb{R}$. Homogeneity of degree τ implies that for $k > 0$,*

$$\rho(kX) = \rho \circ \mu(km) = k^\tau \rho \circ \mu(m) = k^\tau \rho(X) \tag{5}$$

Taking the first-order derivative with respect to k to equation (5), we have

$$\begin{aligned} \frac{d\rho(kX)}{dk} &= \frac{\partial \rho \circ \mu(km)}{\partial k m_1} m_1 + \cdots + \frac{\partial \rho \circ \mu(km)}{\partial k m_N} m_N \\ &= k^{\tau-1}(\frac{\partial \rho(X)}{\partial m_1} m_1 + \cdots + \frac{\partial \rho(X)}{\partial m_N} m_N) = \tau k^{\tau-1} \rho(X) \end{aligned}$$

Deviding both sides of the above equation by $k^{\tau-1}$ gives the result. ■

In particular, we are more interested in the case when $\tau = 1$, since most widely used risk measures fall into this category. The above proposition (known as the Euler's Formula) is fundamental in risk attribution. It facilitates identifying the main source of risk in a portfolio. Each component $m_i \frac{\partial \rho(X)}{\partial m_i}$, termed as the *risk contribution of asset i*, is the amount of risk contributed to the total risk by investing m_i in asset i. The sum of risk contributions over all securities equals the total risk. If we rescale every risk contribution term by $\frac{1}{\rho(X)}$, we should get the percentage of the total risk contributed by the corresponding asset.

The term $\frac{\partial \rho(X)}{\partial m_i}$ is called *marginal risk* which represents the marginal impact on the overall risk from a small change in the position size of asset i, keeping all other positions fixed. If the sign of marginal risk of one asset is positive, then increasing the position size of the asset by a small amount will increase the total risk; If the sign is negative, then increasing the position

[4] Tasche (1999) has a slightly more general assumption. He shows that if the risk measure is $\tau-$homogeneous, continuous and partially differentiable with respect to m_i, $i = 2, \ldots, N$, then it is also differentiable with respect to m_1.

size of the asset by a small amount will reduce the total risk. Thus the asset with negative marginal risk behaves as a hedging instrument.

But there is one important limitation of this approach ([16]). The decomposition process is only a marginal analysis, which implies that only small changes in position sizes make the risk contribution terms more meaningful. For example, if the risk contribution of asset one–in a portfolio consisting of only two assets–is twice that of asset two, then a small increase in the position of asset one will increase the total risk twice as much as the one caused by the same amount of increase in asset two's position. However, it doesn't imply that removing asset one from the portfolio will reduce the overall risk by 2/3. In fact, the marginal risk and the total risk will both change as the position size of asset one changes. This is because of the definition of marginal risk. *Only if* an increase (say ε_1) in asset one's position size is small enough, the additional risk, or *incremental risk,*[5] can be approximated by $\frac{\partial\rho(X)}{\partial m_1}\varepsilon_1$.[6] The larger ε_1 is, the poorer the approximation would be. Removing asset one entirely represents a large change in the position size and thus the approach is not suitable in this situation.

This limitation casts doubt on the philosophy of risk attribution. The main questions are:

1. Is each of the decomposed terms in (4) really an appropriate representation of the risk contribution of each individual asset?

2. If dropping the asset with most risk contribution won't even help, what do we do in order to reduce the overall risk?

We don't answer the second question for now. It may be found in examining the interaction and relationship between risk attribution and portfolio optimization, which is a topic for future studies.

To answer the first question, two author's work are worth mentioning. Tasche (1999) shows that under a general definition of suitability, the only representation appropriate for performance measurement is the first order partial derivatives of the risk measure with respect to position size, which is exactly the marginal risk we define. Denault (2001) applies game theory

[5] We define the incremental risk as the difference between the total risk after changing the composition of a portfolio and the total risk before the change. Note that some authors (cf. [20]) define the risk contribution $m_i\frac{\partial\rho(X)}{\partial m_i}$ as the incremental risk.

[6] A Taylor series expansion can be performed to yield this result ([11]).

to justify the use of partial derivatives of risk measures as a measure of risk contribution. We hereby briefly discuss the approach by Tasche ([28]).

The approach is more like an axiomatic one, that is, to first define the universally accepted and self-evident properties or principles the object possess, then to look for "candidates" which satisfy the pre-determined criterion.[7] In his attempt to define the criterion, Tasche makes use of the concept of *"Return on Risk-Adjusted Capital"*(RORAC), which has been used a lot in allocating banks' capital. RORAC is defined as the ratio between some certain measure of profit and the bank's internal measure of capital at risk.([19]) In our notation, the *portfolio's RORAC* should be $\frac{E[mR]}{\rho(X)-E[m'R]}$, where $E[mR]$ is the expected payoff of the portfolio and $\rho(X) - E[m'R]$ is the *economic capital*, which is the amount of capital needed to prevent solvency at some confidence level. For every unit of investment (i.e. the position size equals 1), the RORAC of an individual asset i (or *per-unit RORAC of asset i*) can be denoted by $\frac{E[R_i]}{\rho(R_i)-E[R_i]}$, where $\rho(R_i)$ is the risk measure of asset i per unit of position i. The RORAC is very similar to the Sharpe ratio, which measures the return performance per unit of risk. If the RORAC of capital A is greater than that of capital B, then capital A gives a higher return per unit of risk than B and thus has a better performance than B; and vice versa. Tasche defines that the measure suitable for performance measurement should satisfy the following conditions:

i) If, for every unit of investment, the amount $\frac{E[R_i]}{a_i(X)-E[R_i]}$[8] is greater than that of the entire portfolio, then investing a little more in asset i should enhance the performance (measured by RORAC) of the entire portfolio, and reducing the amount invested in asset i should decrease the RORAC of the entire portfolio; In mathematical expressions, this is equivalent to

$$\begin{aligned} \frac{E[R_i]}{a_i(X)-E[R_i]} &> \frac{E[m'R]}{\rho(X)-E[m'R]} \\ \Rightarrow \frac{E[m'R+m_i^\varepsilon R_i]}{\rho(X+m_i^\varepsilon R_i)-E[m'R+m_i^\varepsilon R_i]} &> \frac{E[m'R]}{\rho(X)-E[m'R]} > \\ & \frac{E[m'R-m_i^\varepsilon R_i]}{\rho(X-m_i^\varepsilon R_i)-E[m'R-m_i^\varepsilon R_i]} \end{aligned}$$

[7]Other examples of the axiomatic approach can be found in [4] and [6].

[8]Note that the initial risk measure $\rho(R_i)$ in the per-unit RORAC of asset i is replaced by the candidate of performance measure $a_i(X)$ of asset i.

where $a_i(X)$ is the candidate measure of suitable risk contribution of asset i, $E[R_i]$ is the expected payoff per unit of position i and $0 < m_i^\varepsilon < \varepsilon$, for some small $\varepsilon > 0$.

ii) If, for every unit of investment, the amount $\frac{E[R_i]}{a_i(X)-E[R_i]}$ is smaller than that of the entire portfolio, then investing a little more in asset i should decrease the performance of the entire portfolio, and reducing the amount invested in asset i should enhance the RORAC of the entire portfolio; In mathematical expressions, this is equivalent to

$$\begin{aligned} \frac{E[R_i]}{a_i(X)-E[R_i]} &< \frac{E[m'R]}{\rho(X)-E[m'R]} \\ \Rightarrow \frac{E[m'R+m_i^\varepsilon R_i]}{\rho(X+m_i^\varepsilon R_i)-E[m'R+m_i^\varepsilon R_i]} &< \frac{E[m'R]}{\rho(X)-E[m'R]} < \\ &\quad \frac{E[m'R-m_i^\varepsilon R_i]}{\rho(X-m_i^\varepsilon R_i)-E[m'R-m_i^\varepsilon R_i]} \end{aligned}$$

where $a_i(X)$, $E[R_i]$ and m_i^ε are defined in the same way as in 1.

Tasche shows in Theorem 4.4 ([28]) that the *only* function form fulfill the above requirements is $\frac{\partial\rho(X)}{\partial m_i}$.

We note that the limitation of his criterion is that the use of RORAC is more appropriate to banks than to other financial institutions. So this argument might not be appropriate in general. Nevertheless, the use of $\frac{\partial\rho(X)}{\partial m_i}$ as a measure for risk contribution is justified in the sense that it indicates how the global performance changes if there is a little change locally, given the local performance relationship with the overall portfolio.

4 Risk Measures

We examine three major risk measures, which are the *Standard Deviation(and its variants)*, the *Value-at-Risk* and the *Expected Shortfall*. Their strength and weakness in measuring risk are compared. The criterion of good risk measures, namely *coherent risk measures* are reviewed.

4.1 Standard Deviation and Its Variants

Following Markowitz ([18]), scholars and practitioners has been taking the standard deviation as a "standard" risk measure for decades. Its most popular form of great practical use is called the *Tracking error,* which is defined as the standard deviation (also known as the *volatility*) of the excess return (or payoff) of a portfolio relative to a benchmark[9]. Despite its appealing feature of computational ease, the standard deviation has been criticized for its inefficiency of representing risk. The inherent flaw stems from the definition of the standard deviation: both the fluctuations above the mean and below the mean are taken as contributions to risk. This implies that a rational investor would hate the potential gain to the same degree as the potential loss, if the standard deviation were used as the risk measure when he optimizes his portfolio. Furthermore, the standard deviation underestimates the tail risk of the pay distribution, especially when the distribution is nonsymmetric. The following example shows that, with the presence of option, the payoff distribution of a portfolio is asymmetric and thus the standard deviation fails to capture the tail risk.

To remedy the deficiency of the standard deviation, Markowitz ([18]) proposed a variant of the standard deviation, which emphasizes on the loss part of the distribution. The general form is called *the lower semiα-moment*. It is defined as follows:

$$\rho(X) = \rho(m'R) = \sqrt[\alpha]{E[((m'R - E(m'R))^{-})^{\alpha}]} \tag{6}$$

where $(m'R - E(m'R))^{-} = \begin{cases} -(m'R - E(m'R)) \text{ if } m'R - E(m'R) < 0 \\ 0 \text{ if } m'R - E(m'R) \geq 0 \end{cases}$.

Note that when $\alpha = 2$, $\rho(m'R) = \sqrt{E[((m'R - E(m'R))^{-})^{2}]}$ is called the lower semi-standard deviation, which was proposed by Markowitz.

4.2 Value-at-Risk (VaR)

Value-at-risk, or VaR for short, has been widely accepted as a risk measure in the last decade and has been frequently written into industrial regulations(see [14] for an overview). The main reason is because it is conceptually easy. It is defined as the minimum level of losses at a confidence level of solvency of $1 - \alpha$.(See figure 2). That is, VaR can be interpreted as the minimum amount of capital needed as reserve in order to prevent insolvency which happens with probability α.

[9]Tracking error is sometimes defined as the return difference between a portfolio and a benchmark. We here define it as the risk measure of the standard deviation associated with the excess return, because it is widely accepted by the practioners.

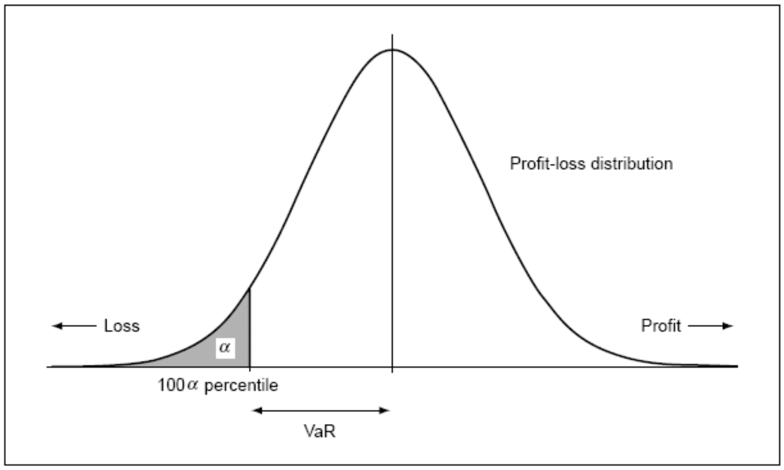


Figure 2: VaR at α level.

Definition 4.1 *The* **VaR** *at confidence level* $(1-\alpha)^{10}$ *is defined as the negative of the lower* α*-quantile of the gain/loss distribution, where* $\alpha \in (0,1)$. *i.e.*

$$VaR_\alpha = VaR_\alpha(X) = -q_\alpha(X) = -\inf\{x|P(X \leq x) \geq \alpha\} \tag{7}$$

where $P(\cdot)$ *is the probability measure.*

An alternative definition of VaR is that $VaR_\alpha(X) = E[X] - q_\alpha(X)$, which is the difference between the expected value of X and the lower α-quantile of X. This relative form of VaR is already used in the performance measurement of the Sharpe ratio in the last section.

Before we introduce the properties of VaR and evaluate how good or bad it is, we have to first introduce the judging rules. Four criterion have been proposed by Artzener et al. (1999).

Axiom 4.2 *A risk measure* $\rho : \mathcal{G} \rightarrow \mathbb{R}$ *is called a* ***coherent risk measure*** *if and only if it satisfies the following properties:*

a Positive homogeneity. (See Definition 3.2)

b Monotonicity: $X \in \mathcal{G},\ X \leq 0 \Rightarrow \rho(X) > 0$.

[10] Typically, α takes the values such as $1\%,\ 5\%$ and 10%.

c Translation invariance: $X \in \mathcal{G},\ c \in \mathbb{R} \Rightarrow \rho(X+c) = \rho(X) - c$

d Subadditivity: $X, Y \in \mathcal{G},\ X+Y \in \mathcal{G} \Rightarrow \rho(X+Y) \leq \rho(X) + \rho(Y)$.

Positive homogeneity makes sense because of liquidity concerns. When all positions are increased by a multiple, risk is also increased by the same multiple because it's getting harder to liquidate larger positions. For monotonicity, it requires that the risk measure should give a "negative" message when the financial asset has a sure loss. The translation invariance property implies that the risk-free asset should reduce the amount of risk by exactly the worth of the risk-free asset. The subadditivity is important because it represents the diversification effect. One can argue that a risk measure without this property may lead to counterintuitive and unrealistic results.[11]

VaR satisfies property a-c but in general fails to satisfy the subadditivity[12], which has been heavily criticized. Another pitfall of VaR is that it only provides a minimum bound for losses and thus ignores any huge potential loss beyond that level. VaR could encourage individual traders to take more unnecessary risk that could expose brokerage firms to potentially huge losses. In the portfolio optimization context, VaR is also under criticism because it is not convex in some cases and may lead to serious problems when being used as a constraint. The following example shows that VaR is not sub-additive(see also [29] for another example).

Example 4.3 *Consider a call(with payoff X) and a put option(with payoff Y) that are both far out-of-money, written by two independent traders. Assume that each individual position leads to a loss in the interval [-4,-2] with probability 3%, i.e.* $P(X<0) = P(Y<0) = 3\% = p$ *and a gain in the interval [1,2] with probability 97%. Thus there is no risk at 5% for each position. But the firm which the two traders belong to may have some loss at 5% level because the probability of loss is now*

$$P(X+Y<0) = \sum_{i=1}^{2} \binom{2}{i} p^i (1-p)^{2-i} = 1 - (1-p)^2 = 5.91\%$$

Therefore $VaR_{5\%}(X+Y) > VaR_{5\%}(X) + VaR_{5\%}(Y) = 0$.

[11] For examle(cf.[4]), an investor could be encouraged to split his or her account into two in order to meet the lower margin requirement; a firm may want to break up into two in order to meet a capital requirement which they would not be able to meet otherwise.

[12] Note that only under the assumption of elliptical distributions is VaR sub-additive([7]). In particular, VaR is sub-additive when $\alpha < .5$ under the Gaussian assumption(cf. [4]).

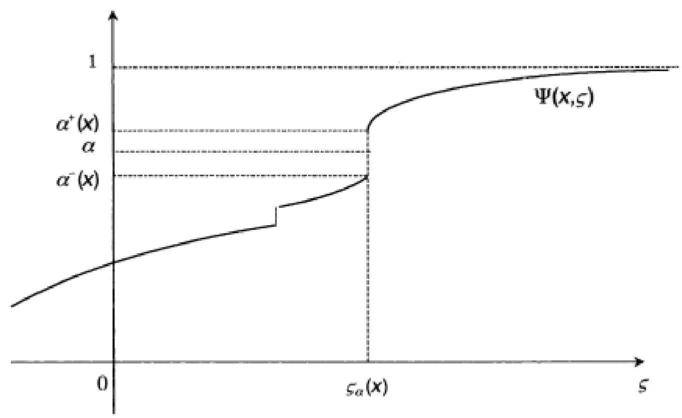


Figure 3: A jump occurs at $\varsigma_\alpha(x) := VaR_\alpha$ of the distribution function $\Psi(x, \varsigma)$ and there are more than one confidence levels $(\alpha^-(x), \alpha^+(x))$ which give the same VaR.

4.3 Conditional Value-at-Risk or Expected Shortfall

While VaR has gained a lot of attention during the late nineties and early this century, that fact that it is not a coherent risk measure casts doubt on any application of VaR. Researchers start looking for alternatives to VaR. A coherent measure, *conditional value-at-risk*(CVaR) or *expected shortfall*(ES) was introduced. Similar concepts were introduced in names of mean excess loss, mean shortfall, worse conditional expectation, tail conditional expectation or tail VaR. The definition varies across different writers. Acerbi and Tasche (2002) clarify all the ambiguity of definitions of the VaR and the CVaR and show the equivalence of the CVaR and the expected shortfall. At the same time independently, Rockafellar and Uryasev (2002) also show the equivalence and generalize their finding in the previous paper to general loss distributions, which incorporate discreteness or discontinuity.

Definition 4.4 *Suppose $E[X^-] < \infty$, the expected shortfall at the level α of*

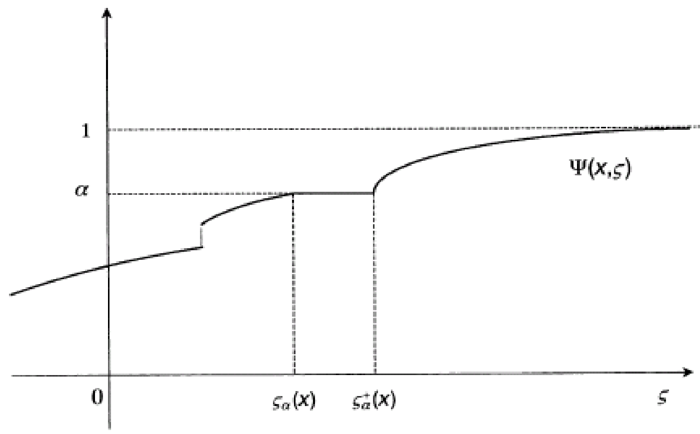


Figure 4: There are more than one candidate $(\varsigma_\alpha(x), \varsigma_\alpha^+(x))$ of the VaR for the same confidence level.

X is defined as

$$\begin{aligned} ES_\alpha &= ES_\alpha(X) \qquad (8) \\ &= -\frac{1}{\alpha}\{E[X\mathbf{1}_{\{X \le -VaR_\alpha\}}] - VaR_\alpha(\alpha - P[X \le -VaR_\alpha])\} \\ &= -\frac{1}{\alpha}\{E[X\mathbf{1}_{\{X \le q_\alpha(X)\}}] + q_\alpha(X)(\alpha - P[X \le q_\alpha(X)])\} \end{aligned}$$

The expected shortfall can be interpreted as the mean of the α-tail of the loss distribution. Rockafellar and Uryasev(2002) define the conditional value-at-risk(CVaR) based on a rescaled probability distribution. Proposition 6 in Rockafellar and Uryasev(2002) confirms that the CVaR is essentially the same as the ES. The subtleness in the definition of ES becomes especially important when the loss distribution has a jump at the point of VaR, which is usually the case in practice. Two cases of jump(or discontinuity) and discreteness of the loss distribution, are illustrated in the figure 3 and figure 4, respectively.

If the loss distribution is continuous, then $\alpha = P[X \le -VaR_\alpha]$ and the expected shortfall defined above will reduce to

$$ES_\alpha(X) = -E[X|X \le -VaR_\alpha]$$

which coincides with the tail conditional expectation defined in Artzener et al. (1999). It is worth mentioning that they show that the tail conditional expectation is generally not subadditive thus not coherent (see also [2]).

We now show that the expected shortfall is a coherent risk measure.

Proposition 4.5 *The expected shortfall(or conditional value-at-risk) defined as (8) satisfies the axiom of coherent risk measures.*
Proof.

i) *Positive homogeneity:*

$$\begin{aligned} & ES_\alpha(tX) \\ = & -\frac{1}{\alpha}\{E[tX\mathbf{1}_{\{tX\le -tVaR_\alpha(X)\}}] - tVaR_\alpha(X)(\alpha - P[tX \le -tVaR_\alpha(X)])\} \\ = & \ tES_\alpha(X) \end{aligned}$$

ii) *Monotonicity:*

$$\begin{aligned} X &\le 0, VaR_\alpha(X) > 0 \Rightarrow \\ ES_\alpha(X) &= -\frac{1}{\alpha}\{\underbrace{E[X\mathbf{1}_{\{X\le -VaR_\alpha\}}]}_{\le 0} - VaR_\alpha(\underbrace{\alpha - P[X \le -VaR_\alpha]}_{=\alpha-1<0})\} \\ &> 0 \end{aligned}$$

iii) *Translation invariance:*

$$\begin{aligned} & ES_\alpha(X+c) \\ = & -\frac{1}{\alpha}\{E[X\mathbf{1}_{\{X\le -VaR_\alpha\}}] + cP[X \le -VaR_\alpha] \\ & -(VaR_\alpha - c)(\alpha - P[X \le -VaR_\alpha])\} \\ = & -\frac{1}{\alpha}\{E[X\mathbf{1}_{\{X\le -VaR_\alpha\}}] - VaR_\alpha(\alpha - P[X \le -VaR_\alpha])\} - c \\ = & \ ES_\alpha(X) - c \end{aligned}$$

iv) *Subadditivity: This proof is based on Acerbi et. al. (2001). They use an indicator function as follows:*

$$1^\alpha_{\{X\le x\}} = \begin{cases} 1_{\{X\le x\}} & if\ P[X=x]=0 \\ 1_{\{X\le x\}} + \frac{\alpha - P[X\le x]}{P[X=x]}1_{\{X=x\}} & if\ P[X=x]>0 \end{cases}$$

It is easy to see that

$$E[1^{\alpha}_{\{X\le x\}}] = \alpha \tag{9}$$

$$1^{\alpha}_{\{X\le q_\alpha\}} \in [0,1] \tag{10}$$

$$\frac{1}{\alpha}E[X1^{\alpha}_{\{X\le q_\alpha\}}] = -ES_\alpha(X) \tag{11}$$

We want to show that $ES_\alpha(X+Y) \le ES_\alpha(X) + ES_\alpha(Y)$*. By 9, 10 and 11*

$$\begin{aligned}
& ES_\alpha(X) + ES_\alpha(Y) - ES_\alpha(X+Y) \\
= & \frac{1}{\alpha}E[(X+Y)1^{\alpha}_{\{X+Y\le q_\alpha(X+Y)\}} - X1^{\alpha}_{\{X\le q_\alpha(X)\}} - Y1^{\alpha}_{\{Y\le q_\alpha(Y)\}}] \\
= & \frac{1}{\alpha}E[X(1^{\alpha}_{\{X+Y\le q_\alpha(X+Y)\}} - 1^{\alpha}_{\{X\le q_\alpha(X)\}}) + \\
& Y(1^{\alpha}_{\{X+Y\le q_\alpha(X+Y)\}} - 1^{\alpha}_{\{Y\le q_\alpha(Y)\}})] \\
\ge & \frac{1}{\alpha}(q_\alpha(X)\underbrace{E[1^{\alpha}_{\{X+Y\le q_\alpha(X+Y)\}} - 1^{\alpha}_{\{X\le q_\alpha(X)\}}]}_{0} + \\
& q_\alpha(Y)\underbrace{E[1^{\alpha}_{\{X+Y\le q_\alpha(X+Y)\}} - 1^{\alpha}_{\{X\le q_\alpha(X)\}}]}_{0}) \\
= & 0
\end{aligned}$$

■

Pfulg(2000) proves that CVaR is coherent by using a different definition of CVaR, which can be represented by an optimization problem(see also [24]).

5 Derivatives of Risk Measures

We are now ready to go one step further to the core of risk attribution analysis, namely calculating the first order partial derivatives of risk measures with respect to positions (recall from (4)). The task is not easy because the objective functions of differentiation of VaR and $CVaR$ are probability functions or quantiles. We introduce here the main results associated with the derivatives.

5.1 Tracking Error

5.1.1 Gaussian Approach

The tracking error is defined as the standard deviation of the excess return(or payoff) of a portfolio relative to a benchmark(see the footnote in section 3.1). It is a well-established result that the standard deviation is differentiable. By assuming Gaussian distributions, Garman (1996, 1997) derives the close form formula for the derivative of VaR[13] from the variance-covariance matrix . Mina (2002) implements the methodology to perform risk attribution, which incorporates the feature of institutional portfolio decision making process in financial institutions.

We first assume Gaussian distributions. Denote by $b = (b_i)_{i=1}^N$the positions of a benchmark. Let $w = (w_i)_{i=1}^N = (m_i - b_i)_{i=1}^N = m - b$ be the excess positions(also called "bet") relative to the benchmark. Then $w'R$ is the excess payoff of the portfolio relative to the benchmark. Let Ω be the variance-covariance matrix of the returns $(r_i)_{i=1}^N$. Then the tracking error is

$$TE = \sqrt{w'\Omega w} \tag{12}$$

The first order derivative with respect to w is

$$\frac{\partial TE}{\partial w} = \frac{\Omega w}{\sqrt{w'\Omega w}} = \nabla \tag{13}$$

which is an $N \times 1$ vector. Therefore the risk contribution of the bet on asset i can be written as

$$w_i \nabla_i = w_i (\frac{\Omega w}{\sqrt{w'\Omega w}})_i \tag{14}$$

The convenience of equation (13) is that we can now play with any arbitrary partition of the portfolio so that the risk contribution of a subset of the portfolio can be calculated as the inner product of ∇ and the corresponding vector of bets. For example, the portfolio can be sorted by industries $I_1, \ldots, I_n$, which are mutually exclusive and jointly exhaustive. The risk contribution of industry I_j is then $\phi_j' \nabla$, where $\phi_j = w \otimes \mathbf{1}_{\{i \in I_j\}}$ and $\mathbf{1}_{\{i \in I_j\}}$ is an $N \times 1$ vector whose's ith elements is one if $i \in I_j$ and zero otherwise, for all $i = 1, \ldots, N$. We can further determine the risk contribution of different sectors in industry I_j in a similar way.

[13] One can thus derive the derivative of the standard deviation from the one of VaR, because under normal distributions(more generally, under elliptical distributions), VaR is a linear function of the standard deviation.

5.1.2 Spherical and Elliptical Distributions

The Gaussian distributions can be generalized to the spherical or more generally, the elliptical distributions so that the tracking error can be still calculated in terms of the variance-covariance matrix. We briefly summarize the facts about spherical and elliptical distributions. See Embrechts et al. (2001) for details.

A random vector R has a *spherical distribution* if for every orthogonal map $U \in \mathbb{R}^{N\times N}$(i.e. $U^T U = UU^T = I_N$, where I_N is the N-dimentional identity matrix), $UX \stackrel{d}{=} X$.The definition implies that the distribution of a spherical random variable is invariant to rotation of the coordinates. The characteristic function of R is $\Phi_R(\theta) = E[\exp(i\theta' R)] = \phi(\theta'\theta) = \phi(\theta_1^2 + \cdots + \theta_N^2)$ for some function $\phi : \mathbb{R}^+ \to \mathbb{R}$, which is called the *characteristic generator* of the spherical distribution and we denote $R\tilde{\ } S_N(\phi)$.Examples of the spherical distributions include the Gaussian distributions, student-t distributions, logistic distributions and etc. The random vector R is spherically distributed($R\tilde{\ } S_N(\phi)$) if and only if there exists a positive random variable D such that

$$R \stackrel{d}{=} D \cdot U \tag{15}$$

where U is a uniformly distributed random vector on the unit hypersphere(or sphere) $S_N = \{s \in \mathbb{R}^N | \, \|s\| = 1\}$.

While the spherical distributions generalize the Gaussian family to the family of symmetrically distributed and uncorrelated random vectors with zero mean, the elliptical distributions are the affine transformation of the spherical distributions. They are defined as follows:

Definition 5.1 *A random vector R has an elliptical distribution, denoted by $R\tilde{\ } E_N(\mu, \Sigma, \phi)$, if there exist $X\tilde{\ } S_K(\phi)$, an $N \times K$ matrix A and $\mu \in \mathbb{R}^N$ such that*

$$R \stackrel{d}{=} AX + \mu$$

where $\Sigma = A^T A$ is a $K \times K$ matrix.

The characteristic function is

$$\begin{aligned} \Phi_R(\theta) &= E[\exp(i\theta'(AX + \mu))] & (16) \\ &= \exp(i\theta'\mu)E[i\theta' AX] = \exp(i\theta'\mu)\phi(\theta^T \Sigma\theta)) & (17) \end{aligned}$$

Thus the characteristic function of $R - \mu$ is $\Phi_{R-\mu}(\theta) = \phi(\theta^T \Sigma \theta)$. Note that the class of elliptical distributions includes the class of spherical distributions. We have $S_N(\phi) = E_N(0, I_N, \phi)$.

The elliptical representation $E_N(\mu, \Sigma, \phi)$ is not unique for the distribution of R. For $R\tilde{}E_N(\mu, \Sigma, \phi) = E_N(\tilde{\mu}, \tilde{\Sigma}, \tilde{\phi})$, we have $\tilde{\mu} = \mu$ and there exists a constant c such that $\tilde{\Sigma} = c\Sigma$ and $\tilde{\phi}(u) = \phi(\frac{u}{c})$. One can choose $\phi(u)$ such that $\Sigma = Cov(R)$, which is the variance-covariance matrix of R. Suppose $R\tilde{}E_N(\mu, \Sigma, \phi)$ and $R \stackrel{d}{=} AX + \mu$, where $X\tilde{}S_K(\phi)$. By (15), $X \stackrel{d}{=} D \cdot U$. Then we have $R \stackrel{d}{=} ADU + \mu$. It follows that $E[R] = \mu$ and $Cov[R] = AA^T E[D^2]/N = \Sigma E[D^2]/N$ since $Cov[U] = I_N/N$. So the characteristic generator can be chosen as $\tilde{\phi}(u) = \phi(u/c)$ such that $\tilde{\Sigma} = Cov(R)$, where $c = N/E[D^2]$. Therefore, the elliptical distribution can be characterized by its mean, variance-covariance matrix and its characteristic generator.

Just like the Gaussian distributions, the elliptical class preserves the property that any affine transformation of an elliptically distributed random vector is also elliptical with the same characteristic generator ϕ. That is, if $R\ \tilde{}E_N(\mu, \Sigma, \phi)$, $B \in \mathbb{R}^{K \times N}$ and $b \in \mathbb{R}^N$ then $BR + b\tilde{}E_K(B\mu + b, B\Sigma B^T, \phi)$.

Applying these results to the portfolio excess payoff $Y = w'R$, we have $X\tilde{}E_1(w'\mu, w'\Sigma w, \phi)$. The tracking error is again $TE = \sqrt{w'\Sigma w}$. The derivative of the tracking error is then similar to the one under the Gaussian case derived in section (5.1.1). We can see that the variance-covariance matrix, under elliptical distributions, plays the same important role of measuring dependence of random variables, as in the Gaussian case. That is why the tracking error can still be express in terms of the variance-covariance matrix.

5.1.3 Stable Approach

It is well-known that portfolio returns don't follow normal distributions. The early work of Mandelbrot (1963) and Fama (1965) built the framework of using *stable* distributions to model financial data. The excessively peaked, heavy-tailed and asymmetric nature of the return distribution made the authors reject the Gaussian hypothesis in favor of the stable distributions, which can incorporate excess kurtosis, fat tails, and skewness.

The class of all stable distributions can be described by four parameters: $(\alpha, \beta, \mu, \sigma)$.The parameter α is the index of stability and must satisfy $0 < \alpha \leq 2$. When $\alpha = 2$ we have the Gaussian distributions. The term *stable Paretian* distributions is to exclude the case of Gaussian distributions ($\alpha = 2$) from the general case. The parameter β, representing skewness of the distribution, is within the range $[-1,1]$. If $\beta = 0$, the distribution is symmetric; If $\beta > 0$, the distribution is skewed to the right and to the left if $\beta < 0$. The location is described by μ and σ is the scale parameter, which measures the dispersion of the distribution corresponding to the standard deviation in Gaussian distributions.

Formally, a random variable X is stable (Paretian stable, α-stable) distributed if for any $a > 0$ and $b > 0$ there exists some constant $c > 0$ and $d \in \mathbb{R}$ such that $aX_1 + bX_2 \overset{d}{=} cX + d$, where X_1 and X_2 are independent copies of X. The stable distributions usually don't have explicit forms of distribution functions or density functions. But they are described by explicit characteristic functions derived through the Fourier transformation. So the alternative definition of a stable random variable X is that for $\alpha \in (0,2]$, $\beta \in [-1,1]$ and $\mu \in \mathbb{R}$, X has the characteristic function of the following form:

$$\Phi_X(t) = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha (1 - i\beta sign(t) \tan\frac{\pi\alpha}{2} + i\mu t)\} & if\ \alpha \neq 1 \\ \exp\{-\sigma|t|(1 + i\beta sign(t)\ln|t| + i\mu t)\} & if\ \alpha = 1 \end{cases} \tag{18}$$

Then the stable random variable X is denoted by $X\tilde{}S_\alpha(\sigma,\beta,\mu)$.In particular, when both the skewness and location parameters β and μ are zero, X is said to be symmetric $\alpha-$stable and denoted by $X\tilde{}S\alpha S$.

A random *vector* R of dimension N is multivariate stable distributed if for any $a > 0$ and $b > 0$ there exists some constant $c > 0$ and a vector $D \in \mathbb{R}^N$ such that $aR_1 + bR_2 \overset{d}{=} cR + D$, where R_1 and R_2 are independent copies of R.The characteristic function now is

$$\Phi_R(\theta) = \begin{cases} \exp\{-\int_{S_N} |\theta^T s|^\alpha (1 - i sign(\theta^T s)\tan\frac{\pi\alpha}{2})\Gamma(ds) + i\theta^T\mu\} & if\ \alpha \neq 1 \\ \exp\{-\int_{S_N} |\theta^T s|^\alpha (1 - i\frac{2}{\pi} sign(\theta^T s)\ln|\theta^T s|)\Gamma(ds) + i\theta^T\mu\} & if\ \alpha = 1 \end{cases} \tag{19}$$

where $\theta \in \mathbb{R}^N$, Γ is a bounded nonnegative measure(also called a *spectral measure*) on the unit sphere $S_N = \{s \in \mathbb{R}^N | \, \|s\| = 1\}$ and μ is the shift vector. In particular, if $R\tilde{}S_\alpha S$, we have $\Phi_R(\theta) = \exp\{-\int_{S_N} |\theta^T s|^\alpha \Gamma(ds)\}$. For an in-depth coverage of properties and applications of stable distributions, we refer to Samorodnitsky and Taqqu (1994) and also Rachev and Mittnik (2000).

As far as risk attribution is concerned, we want to first express the portfolio risk under the stable assumption and then differentiate the measure with respect to the portfolio weight. If the return vector R is multivariate stable Paretian distributed($0 < \alpha < 2$), then all linear combinations of the components of R are stable with the same index α. For $w \in \mathbb{R}^N$, defined as the difference between the portfolio positions and the benchmark positions, the portfolio gain $Y = w'R = \sum_{i=1}^{N} w_i R_i$ is $S_\alpha(\sigma_Y,\ \beta_Y,\ \mu_Y)$. It can be shown that the scale parameter of $S_\alpha(\sigma_Y,\ \beta_Y,\ \mu_Y)$ is

$$\sigma_Y = (\int_{S_N} |w's|^\alpha \Gamma(ds))^{\frac{1}{\alpha}} \tag{20}$$

which is the analog of standard deviation in the Gaussian distribution. This implies that σ_Y is the tracking error under the assumption that asset returns are multivariate stable Paretian distributed. Thus we can use (20) as the measure of portfolio risk. Similarly, the term $\sigma_Y^\alpha = \int_{S_N} |w's|^\alpha \Gamma(ds)$ is the *variation* of the stable Paretian distribution, which is the analog of the variance.

The derivatives of σ_Y with respect to w_i can be calculated for all i.

$$\frac{\partial \sigma_Y}{\partial w_i} = \frac{1}{\alpha}(\int_{S_N} |w's|^\alpha \Gamma(ds))^{\frac{1}{\alpha}-1}(\int_{S_N} \alpha |w's|^{\alpha-1} |s_i| \Gamma(ds)) \tag{21}$$

As a special case of stable distributions, as well as a special case of the elliptical distributions, we look at the *sub-Gaussian* $S_\alpha S$ random vector. A random vector R is sub-Gaussian $S_\alpha S$ distributed if and only if it has the characteristic function $\Phi_Z(\theta) = \exp\{-(\theta' Q \theta)^{\alpha/2} + i\theta'\mu\}$, where Q is a positive definite matrix called the dispersion matrix. By comparing this characteristic function to the one in equation (16), we can see that the distribution of R belongs to the elliptical class. The dispersion matrix is defined by

$$Q = [\frac{R_{i,j}}{2}],\ where\ \frac{R_{i,j}}{2} = [R_i; R_j]_\alpha \, \|R_j\|_\alpha^{2-\alpha} \tag{22}$$

The *covariation* between two symmetric stable Paretian random variables with the same α is defined by

$$[R_i; R_j]_\alpha = \int_{S_2} s_i s_j^{<\alpha-1>} \Gamma(ds)$$

where $x^{<k>} = |x|^k sign(x)$. It can be shown that when $\alpha = 2,\ [R_1; R_2]_\alpha = \frac{1}{2} cov(R_1, R_2)$. The *variation* is defined as $[R_i; R_i]_\alpha = \|R_j\|_\alpha^\alpha$.

Since Z is elliptically distributed, by the results in the last section, the linear combination of its components $w'R \sim E_1(w'\mu, Q, \phi)$ for some characteristic generator ϕ. Then the scale parameter $\sigma_{w'R}$, which is just the tracking

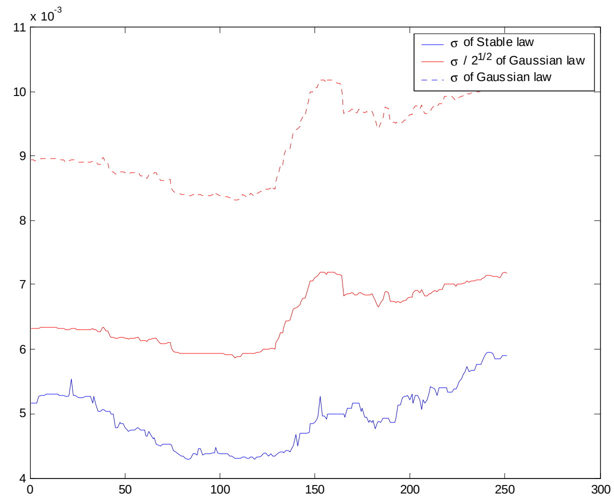


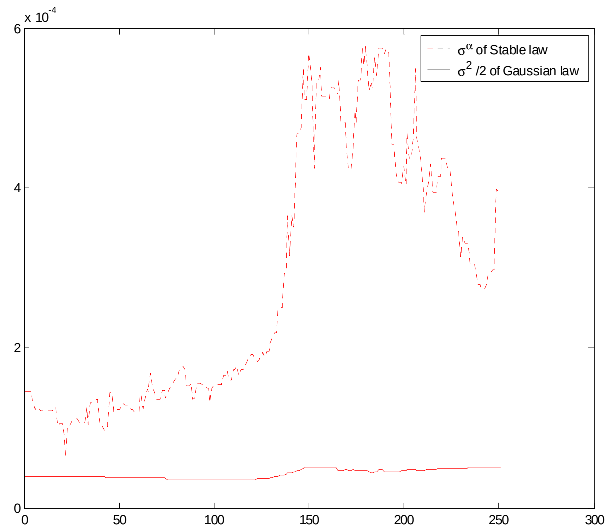
Figure 5: Comparision of the standard deviation and the scale parameter under Gaussian distributions and Stable distributions.

error under this particular case, should be

$$\sigma_{w'R} = \sqrt{w'Qw}$$

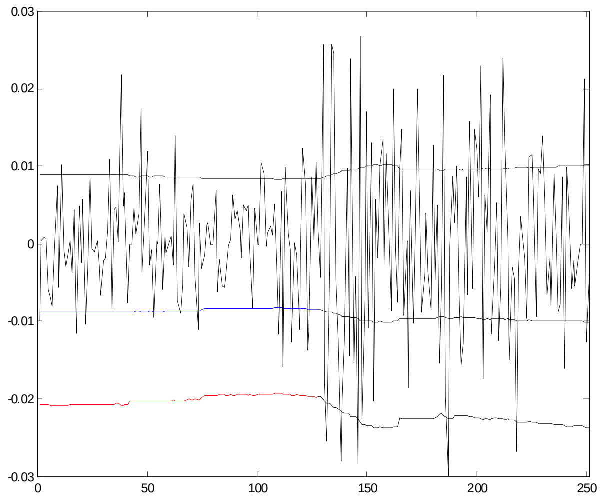
where Q is determined by (22). The derivative of the tracking error is then similar to the one under the Gaussian case derived in section (5.1.1).

For a given portfolio, figure 5 compares the difference of the tracking error modeled under Gaussian distributions and the one under stable distributions(sub-Gaussian $S\alpha S$). Figure 6 compares the variance and variation under different distribution assumptions.

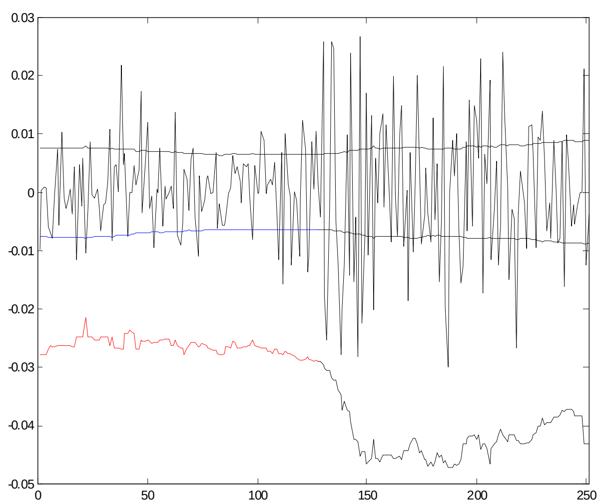


Comparision of the variance and the variation under Gaussian distributions and Stable distributions.

As a result, the stable assumption improves the tracking error's prediction power better than the Gaussian case. This is shown by the results of backtesting, which is designed to test the predicting power of the model consisting of VaR. Since the model we use all belong to the elliptical class, it can be shown that VaR, defined as a quantile of the underlying distribution, is always a linear combination of the standard deviation or the scale parameter(in the stable model), which are both defined as the tracking error. Therefore the efficiency of the tracking error in the stable model can be proved by the backtesting of VaR model. Figure (7) shows that there are 5 exceedings under the Gaussian assumption. The model is rejected because the number of exceedings shouldn't surpass 2.5 were the model correct. To the contrary, figure (8) confirms that the stable model captures the heavy tail and thus is a better one.



The plot of one-year daily excess returns of a portfolio under the Gaussian assumption. The blue curves represent the16-84 quantile-range.The red curve is the Gaussian VaR at 99% confidence interval. There are 5 exceedings and thus the model is rejected.



The plot of one-year daily excess returns of a portfolio under the stable assumption. The blue curves represent the16-84 quantile-range.The red curve is the stable VaR at 99% confidence interval. There is zero exceeding.

5.2 Quantile Derivatives(VaR)

Though VaR is shown to be a problematic risk measure, we still want to exploit its differentiability because the derivative of the expected shortfall depends on the derivative of the quantile measure. Under the assumption of the Gaussian or the more general elliptical distributions, it is not hard to calculate the derivatives of VaR as shown above, which can also be obtained by rescaling the variance-covariance matrix because the VaR is a linear function of the tracking error under the elliptical distributions ([12]). We present here the general case. We assume that there exists a well-defined joint density function for the random vector R.[14]

Proposition 5.2 *Let R be an $N \times 1$ random vector with the joint probability density function of $f(x_{1,} \ldots, x_n)$ satisfying $P[X = VaR_\alpha(X)] \neq 0$ and $\rho(X) = \rho(m'R) = VaR_\alpha(X) = VaR_\alpha$ be the risk measure defined in (7). Then,*

$$\frac{\partial VaR_\alpha(X)}{\partial m_i} = -E[R_i| - X = VaR_\alpha(X)],\ i = 1, \ldots, n \tag{23}$$

Proof. First consider a bivariate case of a random vector (Y, Z) with a smooth $p.d.f.$ $f(y, z)$.Define VaR_α by

$$P[Y + mR\!\!\!/ \geq VaR_\alpha] = \alpha$$

Let $m \neq 0$,

$$\int \int_{(VaR_\alpha - mR\!\!\!/)} f(y, z) dy dz = \alpha$$

Taking the derivative with respect to m, we have

$$\int (\frac{\partial VaR_\alpha}{\partial m} - z) f(y = VaR_\alpha - mz, z) dz = 0$$

Then

$$\frac{\partial VaR_\alpha}{\partial m} \int f(VaR_\alpha - mz, z) dz = \int z f(VaR_\alpha - mz, z) dz$$

[14] Tasche (2000) discusses a slightly more general assumption, where he only assumes the existence of the conditional density function of X_i given $(X_{-i}) := (X_j)_{j \neq i}$. He notes that the existence of joint density implies the existence of the conditional counterpart but not necessarily vice versa.

Since $\int f(VaR_\alpha - mz, z)dz = P[Y + mZ = VaR_\alpha] \neq 0$, we have

$$\frac{\partial VaR_\alpha}{\partial m} = \frac{\int z f(VaR_\alpha - mz, z)dz}{\int f(VaR_\alpha - mz, z)dz} = E[Z|Y + mZ = VaR_\alpha]$$

Now replace $Y = -\sum_{j \neq i}^{n} m_j R_j$ and $m = m_i$ and $Z = -R_i$, we have for all i,

$$\frac{\partial VaR_\alpha}{\partial m_i} = \frac{\partial \rho(X)}{\partial m_i} = -E[R_i| - m'R = VaR_\alpha]$$

■

The risk contribution, defined as $m_i \frac{\partial \rho(X)}{\partial m_i}$ in (4), is $-m_i E[R_i| - m'R = VaR_\alpha]$ in the case of VaR.

5.3 Differentiating The Expected Shortfall

The expected shortfall can be written in terms of an interval of VaR.(cf. [2]) This representation facilitates differentiating the expected shortfall because we already know the derivative of VaR.

Proposition 5.3 *(Tasche 2002) Let X be a random variable, $q_\alpha(X)$ the α-quantile defined in (7) for $\alpha \in (0,1)$ and $f : \mathbb{R} \to [0, \infty)$ a function such that $E|f(X)|^- < \infty$. Then ,*

$$\int_0^a f(q_u(X))du = E[f(X)1_{\{X \leq q_\alpha(X)\}}] + f(q_\alpha(X))(\alpha - P[X \leq q_\alpha(X)]) \quad (24)$$

Proof. Consider a uniformly distributed random variable U on $[0,1]$. We claim that the random variable $Z = q_U(X)$ has the same distribution as X, because $P(Z \leq x) = P(q_U(X) \leq x) = P(F_X^-(U) \leq x) = P(U \leq F_X(x)) = F_X(x)$. Since $u \to q_U(X)$ is non-decreasing we have

$$\{U \leq \alpha\} \subset \{q_U(X) \leq q_\alpha(X)\} \quad (25)$$
$$\{U > \alpha\} \cap \{q_U(X) \leq q_\alpha(X)\} \subset \{q_U(X) = q_\alpha(X)\} \quad (26)$$

(25) implies that $\{U > \alpha\} \cap \{q_U(X) \leq q_\alpha(X)\} + \{U \leq \alpha\} = \{q_U(X) \leq q_\alpha(X)\}$.Then

$$\begin{aligned}\int_0^a f(q_u(X))du &= E_U[f(Z)1_{\{U\leq\alpha\}}] \\ &= E_U[f(Z)1_{\{Z\leq q_\alpha(X)\}}] - E_U[f(Z)1_{\{U>\alpha\}\cap\{Z\leq q_\alpha(X)\}}] \\ &= E[f(X)1_{\{X\leq q_\alpha(X)\}}] + f(q_\alpha(X))P[\{U>\alpha\}\cap\{X\leq q_\alpha(X)\}] \\ &= E[f(X)1_{\{X\leq q_\alpha(X)\}}] + f(q_\alpha(X))(\alpha - P[X\leq q_a(X))\end{aligned}$$

■

Corollary 5.4 *Given the VaR and the expected shortfall defined in (7) and (8), the following is true:*

$$ES_\alpha = \frac{1}{\alpha}\int_0^\alpha VaR_u(X)du \tag{27}$$

Proof. Let $f(X) = X$. By (8) and (24), we have

$$\begin{aligned}\int_0^a f(q_u(X))du &= \int_0^a q_u(X)du \\ &= E[f(X)1_{\{X\leq q_\alpha(X)\}}] + q_\alpha(X)(\alpha - P[X\leq q_\alpha(X)]) = -\alpha ES_\alpha\end{aligned}$$

Then dividing both sides by $-\alpha$ and replacing $-q_u(X)$ by $VaR_u(X)$ yield the result. ■

Proposition 5.5 *The partial derivative of ES_α defined in (8) is*

$$\frac{\partial ES_\alpha}{\partial m_i} = -\frac{1}{a}\{E[R_i 1_{\{X\leq q_\alpha(X)\}}] + E[R_i|X = q_u(X)](\alpha - P[X\leq q_\alpha(X)])\} \tag{28}$$

Proof. Given the representation in (27) and by (23), we have

$$\begin{aligned}\frac{\partial ES_\alpha}{\partial m_i} &= \frac{1}{\alpha}\int_0^\alpha \frac{\partial VaR_u}{\partial m_i}du & (29)\\ &= -\frac{1}{\alpha}\int_0^\alpha E[R_i| -X = VaR_u(X)]du \\ &= -\frac{1}{\alpha}\int_0^\alpha E[R_i|X = q_u(X)]du & (30)\end{aligned}$$

We can apply Proposition 5.3 again. Let $f(x) = E[R_i|X = x]$, then

$$\int_0^\alpha E[R_i|X = q_u(X)]du = E[E[R_i|X]1_{\{X\leq q_\alpha(X)\}}]+E[R_i|X = q_u(X)](\alpha-P[X \leq q_\alpha(X)]) \tag{31}$$

The first term

$$\begin{aligned} E[E[R_i|X]1_{\{X\leq q_\alpha(X)\}}] &= E\{E[R_i|X]|X \leq q_\alpha(X)\} \cdot P[X \leq q_\alpha(X)] \\ &= E[R_i|X \leq q_\alpha(X)\} \cdot P[X \leq q_\alpha(X)] \\ &= E[R_i 1_{\{X\leq q_\alpha(X)\}}] \end{aligned} \tag{32}$$

Then equation (31) becomes

$$\int_0^\alpha E[R_i|X = q_u(X)]du = E[R_i 1_{\{X\leq q_\alpha(X)\}}]+E[R_i|X = q_u(X)](\alpha-P[X \leq q_\alpha(X)])$$

Plugging this into (29) completes the proof. ■

6 Conclusion

We have reviewed the methodology of risk attribution, along with the properties of different risk measures and their calculation of derivatives. The rationale of risk attribution is that risk managers need to know where the major source of risk in their portfolio come from. The stand-alone risk statistics are useless in identifying the source of risk because of the presence of correlations with other assets. The partial derivative of the risk measure is justified to be an appropriate measure of risk contributions and therefore can help locating the major risk.

Having a good measure of risk is critical for risk attribution. A good risk measure should at least satisfy the coherent criterion. The widely accepted measure of volatility could be a poor measure when the distribution is not symmetric. VaR is doomed to be a history because of its non-subadditivity and non-convexity. The conditional VaR or expected shortfall seems promising and is expected to become a dominant risk measure widely adopted in risk management.

Yet there are still some questions that haven't been answered. The statistical methods of estimating the risk contribution terms need to be further studied. Under the more general assumption of the distribution, the risk attribution results might be more accurate. The limitation of risk attribution analysis is that the risk contribution figure is a marginal concept. Risk attribution serves the purpose of hedging the major risk, which is closely related to portfolio optimization. How exactly the information extracted from the risk attribution process can be used in the portfolio optimization process still needs to be exploited. The two processes seem to be interdependent. Their interactions and relationship could be the topic of further studies.

References

[1] Acerbi, C., Nordio, C. and Sirtori, C.(2001). "Expected Shortfall as A Tool for Financial Risk Management." *Working paper.*

[2] Acerbi, C. and Tasche, D.(2002). "On the Coherence of Expected Shortfall." *Journal of Banking & Finance* 26(7), 1487-1503.

[3] Arzener, P., Delban, F., Eber,J.M. and Heath, D. (1997). "Thinking Coherently." *Risk* 10(11), 58-60.

[4] Arzener, P., Delban, F., Eber,J.M. and Heath, D.(1999). "Coherent Measures of Risk." *Mathematical Finance* 9(3), 203-228.

[5] Bradley, B.O. and Taqqu, M.S.(2003). "Financial risk and heavy tails." In: Rachev, S.T.(Ed.), *Handbook of Heavy Tailed Distribution in Finance. Elsevier.*

[6] Denault, M.(1999). "Coherent Allocation of Risk Capital." *Journal of Risk* 4(1), 1-34.

[7] Embrechts, P., McNeil, A.J., Straumann, D.(2001). "Correlation and dependence in risk management: properties and pitfalls." In: Dempster, M., Moffatt, H.K.(Eds.) *Risk Management: Value at Risk and Beyond. Cambridge University Press.*

[8] Fama, E.F.(1965). "The behavior of stock market prices." *Journal of Business* 38(1), 34-105.

[9] Fama, E.F.(1972). "Components of Investment Performance." *Journal of Finance* 27(3), 551-567.

[10] Gourieroux, C., Laurent, J.P. and Scaillet, O.(2000). "Sensitivity Analysis of Value at Risk." *Journal of Empirical Finance* 7, 225-245.

[11] Garman, M.(1996). "Improving on VaR." *Risk* 9(5), 61-63.

[12] Garman, M.(1997). "Taking VaR into Pieces." *Risk 1*0(10), 70-71.

[13] Hallerbach, W.G.(1999). "Decomposing Portfolio Value-at-Risk: A General Analysis." *Discussion paper TI 99-034/2. Tinbergen Institute.*

[14] Khindanova, I.N. and Rachev, S.T.(2000). "Value at Risk: Recent Advances." *Journal of Risk Analysis* 2, 45-76.

[15] Kurth, A. and Tasche, D.(2003). "Contributions to Credit Risk." *Risk* 16(3), 84-88

[16] Litterman, R.(1996). "Hot Spots^{TM} and Hedges." *Journal of Portfolio Management,* Special Issue, 52-75.

[17] Mandelbrot, B.B.(1963). "The variation of certain speculative prices." *Journal of Buisiness* 36, 392-417.

[18] Markowitz, H. M.(1952). "Portfolio Selection." *Journal of Finance,* 7(1), 77-91.

[19] Matten, C. (2000). "Managing Bank Capital: Capital Allocation and Performance Measurement. 2nd Edition." *Wiley, Chichester..*

[20] Mina, J.(2002). "Risk Attribution for Asset Managers." *RiskMetrics Journal* 3(2), 33-56.

[21] Pflug, G.C.(2000). "Some Remarks on The Value-at-Risk and The Conditional Value-at-Risk." In: Uryasev, S.(Ed.), *Probabilistic Constrained Optimization: Methodology and Applications. Kluwer Academic Publishers, Dordrecht.*

[22] Rachev, S.T. and Mittnik, S.(2000). "Stable paretian models in finance." *Wiley, Chichester.*

[23] Rockafellar, R.T. and Uryasev, S.(2000). "Optimization of Conditional Value-at-Risk." *Journal of Risk* 2, 21-41.

[24] Rockafellar, R.T. and Uryasev, S.(2002). "Conditional Value-at-Risk for General Loss Distributions." *Journal of Banking & Finance* 26, 1443-1471

[25] Roll, R.(1977). "A Critique of the Asset Pricing Theory's Tests: Part I." *Journal of Financial Economics,* 4, 129-176.

[26] Samorodnitsky, G. and Taqqu, M.(1994). "Stable non-Gaussian random processes." *Chapman & Hall, New York.*

[27] Sharpe, W.(1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance,* 19(3), 425-442.

[28] Tasche, D.(1999). "Risk Contributions and Performance Measurement." *Working Paper.*

[29] Tasche, D.(2002). "Expected Shortfall and Beyond." *Journal of Banking & Finance* 26, 1519-1533.

[30] Tobin, J.(1958). "Liquidity Preference as Behavior Towards Risk." *Review of Economic Studies,* 18(5), 931-955.