Performance Measures in Financial Markets

Bertrand B. Maillet

Massimiliano Caporin

Francesco Lisi

Second Release – April 2017

We thank here the GRI in Financial Services for support.

2

Measures of Relative Performance

Summary. The relative performance measures can be expressed, in general terms, according to the following *formula*[1]:

$$PM_p = \mathscr{P}(r_p - \tau_1) \times \left[\mathscr{R}(r_p - \tau_2) \times c_p\right]^{-1}, \tag{2.1}$$

where $\mathscr{P}(\cdot)$ is a function that depends upon the observed performance, $\mathscr{R}(\cdot)$ is a risk measure of the investor's portfolio, τ_1 and τ_2 are two specific threshold returns, and c_p is a correction term. More precisely, "relative risk-adjusted performance measures" aim at comparing the expected return of managed portfolios in excess of a threshold, *per* unit of risk. The portfolio performance corresponds to an increasing function of the measured performance and a decreasing function of the estimated risk.
The Sharpe ratio (Sharpe, 1966) is the most representative measure in this class. Directly related to the Mean-Variance model developed by Markowitz (1952), this ratio uses a mean return as a numerator and a standard deviation of portfolio returns as a risk measure. When comparing potential benefits of an investment portfolio, relative to its underlying total risk, practitioners often refer to the Sharpe ratio as the reference performance measure. In some cases, managed portfolios are ranked by means of Sharpe ratios (Ackermann et al., 1999; Liang, 1999; Schneeweis et al., 2002); however, academics and practitioners have proposed

[1] For the sake of simplicity, we do not consider the time subscript for most of the performance measures presented in this book (except when it is strictly necessary).

many other Sharpe-based performance measures in order to overcome some of its well-known limitations (mainly regarding the evaluation of portfolio total risk).

2.1 The Sharpe (1966) Reward-to-variability ratio

This first ratio, also called "Reward-to-variability ratio", was developed by Sharpe (1966). This performance measure equals the expected return of the investor's portfolio in excess of the risk-free rate (a quantity also called "risk *premium*" or "reward") over the standard deviation of returns on the same portfolio. The Sharpe ratio is expressed as follows:

$$S_p = \left[E\left(r_p\right) - r_f\right] \times \left(\sigma_{r_p}\right)^{-1}. \tag{2.2}$$

This measure evaluates the compensation earned by the portfolio manager *per* unit of portfolio total risk, namely both systematic and idiosyncratic risks[2].

Despite some limitations, the Sharpe (1966) ratio is still considered as the reference performance measure. If we derive this ratio within a Markowitz framework, it shares the same drawbacks of the Mean-Variance model, where the representative investor is characterized by a quadratic utility function and/or the portfolio returns are assumed Gaussian. However, it is well-known that financial returns are not Gaussian, also due to investment strategies based on derivatives with time-varying exposures and leverage effects. An incorrect assumption of Gaussianity may lead to an underestimate of the portfolio total risk (see Amin and Kat, 2003; Geman and Kharoubi, 2003) and, thus, to biased investment rankings (a downward biased risk evaluation induces an upward biased Sharpe ratio). Furthermore, the standard deviation equally weights positive and negative excess returns, and it has been shown that this quantity can be subject to manipulations (see Goetzmann et al., 2007) . Another issue of using volatility as a risk measure is related to the liquidity problem of some categories of risky assets (see Getmansky et al., 2004) . Moreover, the Sharpe ratio does not completely reflect the attitude

[2] We should mention that Treynor and Black (1973) propose a performance measure, named "Appraisal ratio" (see, p. 74), corresponding to the Sharpe ratio squared, as noticed by Sharpe (1994) on p. 52.

towards risk for all categories of investors, as it assumes a constant risk-free rate, identical for lending and borrowing. Finally, the validity of this performance measure heavily relies on the accuracy and stability of the first and second moment estimations (Merton, 1981; Engle and Bollerslev, 1986) . Israëlsen (2005) highlights that when the average return is negative, the portfolio with higher standard deviation leads to a better Sharpe ratio, which is counter-intuitive. Thus, he suggests correcting this drawback, when it is negative, by multiplying – instead of dividing – the return of the managed portfolio by its total risk.

2.2 Sharpe-like Performance Measures

The measures presented herein aim at evaluating the relative risk-adjusted performance of investment portfolios. They define risk using quantities that are directly linked to the standard deviation, but with some amendments.

The Jobson-Korkie (1981) T-Stat. Jobson and Korkie (1981) propose a test-statistic, used to verify the null hypothesis of equality between two Sharpe ratios for any couple of managed portfolios

However, this test is not valid when the distribution of the investor's portfolio returns have tails heavier than a Gaussian return distribution (corrections have been proposed by Memmel, 2003). , which is given by:

$$\widehat{JK}_{p_1,p_2} = \left[\widehat{\sigma}_{r_{p_1}} \times \left(\bar{r}_{p_2} - r_f\right) - \widehat{\sigma}_{r_{p_2}} \times \left(\bar{r}_{p_1} - r_f\right)\right] \times \widehat{\theta}^{-\frac{1}{2}}, \tag{2.3}$$

with:

$$\begin{aligned}\widehat{\theta} = N^{-1} \times \Bigg\{ & 2\widehat{\sigma}^2_{r_{p_1}}\widehat{\sigma}^2_{r_{p_2}} - 2\widehat{\sigma}_{r_{p_1}}\widehat{\sigma}_{r_{p_2}}\widehat{\sigma}_{r_{p_1},r_{p_2}} + \frac{1}{2}\bar{r}^2_{p_1}\widehat{\sigma}^2_{r_{p_2}} + \frac{1}{2}\bar{r}^2_{p_2}\widehat{\sigma}^2_{r_{p_1}} \\ & - \left[\left(\bar{r}_{p_2}\bar{r}_{p_1}\right) \times \left(2\widehat{\sigma}_{r_{p_1}}\widehat{\sigma}_{r_{p_2}}\right)^{-1}\right] \times \left[\widehat{\sigma}^2_{r_{p_1},r_{p_2}} + \widehat{\sigma}^2_{r_{p_1}}\widehat{\sigma}^2_{r_{p_2}}\right] \Bigg\},\end{aligned}$$

where $\bar{r}_{p_n}$, with $n = [1,2]$, is the mean return of two portfolios and N is the total number of observations.

Jobson and Korkie (1981) show that, under the null hypothesis of two identical Sharpe ratios and when portfolio returns are i.i.d, the $\widehat{JK}_{p_1,p_2}$ statistic follows a law which is asymptotically distributed as a Gaussian one with a zero mean and an unit standard deviation. Jobson and Korkie (1981) and Jorion (1985) note that the statistical power of this test is low, especially for small sample sizes. Such an outcome could be due to the presence of serial correlation and/or heteroskedasticity in the portfolio returns. A more robust approach for testing the equivalence of Sharpe ratios has been proposed by Ledoit and Wolf (2008).

The Israëlsen (2005) Reward-to-absolute Excess Return Ratio. Israëlsen (2005) proposes the "Reward-to-absolute excess return ratio" in order to correct a drawback of the Sharpe (1966) ratio when the estimated expected return of the managed portfolio is negative. This performance measure is defined as:

$$IS_p = \left[E\left(r_p\right) - r_f\right] \times \left(\sigma_{r_p}^{-1}\right)^{\lambda_p}, \tag{2.4}$$

with $\lambda_p = Sgn\left[E\left(r_p\right) - r_f\right] \times 1$.

Results obtained with the Sharpe (1966) ratio and the Israëlsen (2005) reward-to-absolute excess return ratio differ when the investor's portfolio expected excess return is negative. In this peculiar case, the expected excess return of the manager's portfolio is adjusted (*i.e.*, multiplied) by the total risk, in order to heavily penalize its under-performance while maintaining a coherence in terms of rankings compared to those obtained with the Sharpe (1966) ratio.

The Morey-Vinod (2001) Double Sharpe Ratio. Morey and Vinod (2001) develop a performance measure, named "Double Sharpe ratio", which refines the original Sharpe ratio by taking into account the sampling error. The Double Sharpe ratio is given as:

$$DS_p = \left\{\left[E\left(r_p\right) - r_f\right] \times \sigma_{r_p}^{-1}\right\} \times \left(\sigma_{S_p}\right)^{-1}, \tag{2.5}$$

where σ_{S_p} is the standard deviation of the estimated Sharpe ratios.

In order to obtain an estimate of σ_{S_p}, the authors generate a large number of excess return series by resampling the observed excess return using a simple bootstrap methodology (Memmel, 2003; Ledoit and Wolf, 2008) . Then, they compute Sharpe ratios for each simulated series and calculate the standard deviation of all these Sharpe ratios. Their main innovation is to explicitly consider model risk in the computation of Sharpe ratios, maintaining the interpretation similar to that of the original Sharpe performance measure. However, the use of bootstrap methods for independent data might be a problem.

The Watanabe (2006) Extended Sharpe Ratio. Watanabe (2006) proposes an extension of the original Sharpe (1966) dealing with higher-moments, which is written as:

$$WS_p = S_p + \left[sk_{r_p} \times \left(\kappa_{r_p}\right)^{-1}\right], \tag{2.6}$$

where S_p is the Sharpe (1966) ratio, sk_{r_p} and κ_{r_p} respectively correspond to the skewness and the *kurtosis* of the investor's portfolio return distribution.

Based on a similar approach, Zakamouline-Koekebakker (2009) propose the Adjusted for Skewness Sharpe ratio.

The Zakamouline-Koekebakker (2009) Adjusted for Skewness Sharpe Ratio. Zakamouline and Koekebakker (2009) build a performance measure, named "Adjusted for Skewness Sharpe Ratio" that accounts for the third moment of the portfolio return within an Expected Utility theory framework. The authors introduce the notion of relative preferences to higher moments of return distributions. These are defined as the relation between the absolute preference to the third moment and the absolute preference to the second moment (Pratt, 1964; Arrow, 1971) . In a Mean-Variance-Skewness framework, the proposed measure is given by:

$$ASSR_p = \left\{\left[E\left(r_p\right) - r_f\right] \times \sigma_{r_p}^{-1}\right\} Unexpectedeqnlinespacer$$
$$\times \left\{1 + b_{3,p} \times \left[sk_{r_p} \times (3)^{-1}\right] \times \left\{\left[E\left(r_p\right) - r_f\right] \times \sigma_{r_p}^{-1}\right\}\right\}^{\frac{1}{2}}, Unexpectedeqnlinespacer \tag{2.7}$$

where $b_{3,p}$, corresponding to the investor's relative preferences for skewness, is defined as:

$$\begin{cases} b_{3,p} = (-1)^2 \times \left\{ \left\{ U^{(3)}\left(r_p\right) \times \left[U^{(1)}\left(r_p\right) \right]^{-1} \right\} \times \left(\gamma_p\right)^{-2} \right\} \\ \gamma_p = -U^{(2)}\left(r_p\right) \times \left[U^{(1)}\left(r_p\right) \right]^{-1}, \end{cases}$$

and γ_p is the Arrow-Pratt measure of absolute risk aversion.

In other words, Equation (2.7) corresponds to the original Sharpe ratio corrected by a "skewness adjustment factor". The value of $b_{3,p}$ is directly linked to the investor's utility function. Indeed, a CARA utility function is implied by a coefficient $b_{3,p} = 1$, whilst a CRRA utility function is induced by any coefficient $b_{3,p} > 1$. Zakamouline and Koekebakker (2009) develop an extension which considers the fourth moment (see also Watanabe, 2006) .

The Sharpe (1994) Information Ratio. Since the reference to competitors is a crucial topic in the management industry, Sharpe (1994) proposes a second performance measure, named "Information Ratio" (IR), which takes into account the Tracking Error volatility of an actively managed portfolio. The Information Ratio is given by:

$$IR_p = \left[E\left(r_p - r_B\right)\right] \times \left(TE_{r_p,r_B}\right)^{-1}. \tag{2.8}$$

This measure allows us to track the persistence of over or under-performance between the investor's portfolio and the chosen benchmark. In other words, it provides an assessment of the "quality" of investment choices made by the portfolio manager. Similar to the correction proposed for the Sharpe ratio, Israëlsen (2005) suggests a modified Information Ratio where the average return of the investor's portfolio is negative. In this case, the excess return of the investor's portfolio is multiplied by its Tracking Error volatility..

The Gillet-Moussavou (2000) Information Ratio. In order to only take into account the negative deviations, Gillet and Moussavou (2000) suggest another corrected Information Ratio by replacing the Tracking Error with the semi-variance of the excess return of the managed portfolio. This measure thus reads:

$$GMIR_p = \left[E\left(r_p - r_B\right)\right] \times \left[GLPM_{r_p - r_B, E(r_p - r_B), 0, 2}\right]^{-\frac{1}{2}}, \tag{2.9}$$

where $GLPM_{r_p - r_B, E(r_p - r_B), 0, 2}$ corresponds to the semi-variance of the managed portfolio returns in excess of the benchmark portfolio returns.

The Israëlsen (2005) Information Ratio. In the same vein as the previous measure, Israëlsen (2005) also suggests a corrected Information Ratio (Sharpe, 1994) , specifically when the estimated expected return of the investor's portfolio is negative. This measure reads:

$$IIR_p = \left[E\left(r_p - r_B\right)\right] \times \left(TE^{-1}_{r_p, r_B}\right)^{\lambda_p}, \tag{2.10}$$

where $\lambda_p = Sgn\left[E\left(r_p\right) - r_f\right] \times 1$.

Results obtained with the Sharpe (1994) and the Israëlsen (2005) information ratios only vary when the investor's portfolio expected excess return is negative. In this case, the expected excess return of the manager's portfolio is adjusted by the tracking error in order to penalize its under-performance.

The Treynor (1965) Reward-to-volatility Ratios. Treynor (1965) proposes two "Reward-to-volatility ratios" that are based on the systematic risk sensitivity of the managed portfolio returns. This risk measure is directly linked to the non-diversifiable risk, related to the so-called "*beta*", and obtained from a single index regression. The first ratio, known as the "Treynor ratio", is written as:

$$T_{1,p} = \left[E\left(r_p\right) - r_f\right] \times \left(\beta_{r_p, r_m}\right)^{-1}, \tag{2.11}$$

and the second one, also referred to as the Appraisal Ratio[3] , is formulated as:

$$T_{2,p} = \left\{\left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p, r_m}\right\} \times \left(\beta_{r_p, r_m}\right)^{-1}. \tag{2.12}$$

[3] See also Smith and Tito (1969) for a discussion about the Treynor (1965) ratios. The second ranking measure $T_{2,p}$ proposed by Treynor (1965) - see p. 75 - is sometimes referred to a few authors as the Treynor-Black (1973) Appraisal ratio.

Assuming a perfectly diversified portfolio (with a null specific risk), $T_{1,p}$ evaluates the compensation earned by the manager, with respect to the market risk exposure of portfolio p, while $T_{2,p}$ assesses the *extra* performance realized by the manager over the risk *premium* offered by markets. This second ratio illustrates the investor's abilities to over-perform the market portfolio, when investors hold an amount of private information, potentially not reflected yet in the current market prices.

Several authors propose close variants of the Treynor ratio $T_{1,p}$. Bacon (008b) suggests multiplying the systematic risk sensitivity of the investor's portfolio return by the total risk of the market portfolio, while Srivastava and Essayyad (1994) replace the traditional *beta* in Equation (2.11) with a modified version defined as the ratio between partial moments. In the context of the Arbitrage Pricing Theory developed by Ross (1976), Hübner (2005) proposes a generalized version of $T_{2,p}$ including the effects of risk factors.

The Bacon (008b) modified Treynor Ratio. Bacon (008b) introduces another modified version of the original Treynor (1965) ratio by also considering in the denominator the total risk of the market portfolio such as:

$$mT_{1,p} = \left[E\left(r_p\right) - r_f\right] \times \left(\beta_{r_p,r_m} \times \sigma_{r_m}\right)^{-1}, \tag{2.13}$$

where σ_{r_m} is the total risk of the market portfolio.

The measures related to the Sharpe ratio have been subjected to different criticisms. In order to correct the potential estimation error when the Sharpe ratio is obtained from historical data, Morey and Vinod (2001) propose a simple bootstrap procedure. Other, more efficient, approaches have been proposed for dealing with model risk. However, the exact characterization of a return distribution by moments is a difficult problem (Kimball, 1993) . In addition, estimations are usually based on Conventional-moments, known to be highly responsive to sampling variability (see Hosking, 1990). Sharpe (1994) proposed using the Tracking Error. Still, this complementary risk measure does not differentiate positive from negative variations, and the definition of the benchmark is crucial. Treynor (1965) focuses on the systematic risk sensitivity of the investor's portfolio returns to those of the market portfolio. Nevertheless, this supposes a unique risk factor economy and a portfolio exposure to the systematic risk being stable over time. Finally, the idiosyncratic risk is just neglected, which cannot be without consequences in the context of performance measurement.

2.3 Relative Performance Measures based on Other Risk Measures

The set of performance measures, presented hereafter, adjusts the reward earned by the portfolio manager according to alternative risk measures. Some examples are given by Semivolatility, Linear-moment of order 2, Mean Absolute Deviation, Value-at-Risk and range ratios.

The Sortino-Satchell (2001) Reward-to-Lower Partial Moment Ratio. Sortino and Satchell (2001) develop a performance measure based on Lower Partial Moments. The "Reward-to-Lower Partial Moment ratio" is formulated as:

$$RLPM_p = \left[E\left(r_p\right) - \tau\right] \times \left(GLPM_{r_p,\tau,\tau,o}\right)^{-\frac{1}{o}}, \quad (2.14)$$

where o is a positive constant and τ is a generic threshold. Sortino and Meer (1991) introduce a close variant, named the "Sortino ratio", which is based on the square root of an *LPM* of order 2 (*i.e.* with $o = 2$) and on a threshold equals to a MAR. Watanabe (2006) proposed an extension of the Sortino ratio by adding a term related to the skewness and *kurtosis* of the portfolio returns. Kaplan and Knowles (2004) develop the so-called "Kappa 3 ratio"which uses $o = 3$. A further case is the "Downside Risk Sharpe ratio" suggested by Ziemba (2005) (see also Gergaud and Ziemba, 2012) , which is obtained by considering the square root of an *LPM* of order 2 (*i.e.* $o = 2$) multiplied by 2 in the denominator of Equation (2.14) and for a threshold equal to the expected return $E(r_p)$. The Sortino-Satchell (2001) index evaluates the portfolio managers' performance by considering their risk profile. For example, a high (low) threshold level characterizes managers with a high (low) degree of risk aversion.

The Watanabe (2006) Extended Sortino Ratio. In the same vein, Watanabe (2006) also suggests an extension of the Sortino-Meer (1991) ratio, by adding a term that takes into account the third and fourth moments of the underlying return distribution. It is given by:

$$WRLPM_p = So_p + \left[sk_{r_p} \times \left(\kappa_{r_p} \right)^{-1} \right], \tag{2.15}$$

with:

$$So_p = \left[E\left(r_p \right) - MAR \right] \times \left(GLPM_{r_p,MAR,MAR,2} \right)^{-\frac{1}{2}},$$

where *MAR* denotes the *Minimum* Acceptable Return.

The Dowd (2000) Reward-to-Value-at-Risk Ratio. Based on a quantile model, Dowd (2000) introduces a performance measure that adjusts the expected excess return of the investor's portfolio by the *a*-Value-at-Risk (*a*-VaR) of the portfolio return distributions (see Chapter 1). The "Reward-to-Value-at-Risk ratio" is written as:

$$RVaR_p = \left[E\left(r_p \right) - r_f \right] \times |VaR_{r_p,a}|^{-1}. \tag{2.16}$$

This ratio allows the investor to gauge the performance of the managed portfolio rescaled by a measure of extreme risk, instead of total risk. Note that the Value-at-Risk is widely used in finance and insurance for capital and risk management. However, in recent years, it has been criticized following Artzner et al. (1999) who showed that *VaR* does not have, theoretically, all the four coherence properties (translation invariance, monotonicity, sub-additivity, positive homogeneity). Those properties are required for any "good" risk measure. In particular, *VaR* does not respect the sub-additivity principle.

With the same reasoning as that of the *RVaR* (see also Alexander and Baptista, 2003) , a few authors favor substitute risk measures, such as the modified *VaR* (Favre and Galeano, 2002) , the Conditional *VaR* (Martin et al., 2003) and the MiniMax measure (Young, 1998) .

The Alexander-Baptista (2003) modified Reward-to-Value-at-Risk Ratio. Strongly inspired by Dowd (2000), Alexander and Baptista (2003) proposes to add the risk-free rate at the denominator of the original Dowd (2000) Reward-to-Value-at-Risk ratio such as:

$$mRVaR_p = \left[E\left(r_p\right) - r_f\right] \times \left[|VaR_{r_p,a}| + r_f\right]^{-1}, \tag{2.17}$$

where $VaR_{r_p,a}$ is the a-Value-at-Risk of the investor's portfolio return distribution.

The Favre-Galeano (2002) Reward-to-modified Value-at-Risk. Favre and Galeano (2002) introduce a measure named the "Reward-to-modified Value-at-Risk ratio" (RmVaR for short) considering higher-order moments in order to gauge the non-Gaussianity of the investor's portfolio returns. Based on the Cornish-Fisher expansion (1938) , the authors propose to use a modified Value-at-Risk, *i.e.* $mVaR_{r_p,a}$, as:

$$RmVaR_p = \left[E\left(r_p\right) - r_f\right] \times \left(mVaR_{r_p,a}\right)^{-1}, \tag{2.18}$$

with:

$$\begin{cases} mVaR_{r_p,a} = -\left\{E\left(r_p\right) + \sigma_{r_p} \times \left[q_a + \left(q_a^2 - 1\right) \times sk_{r_p}/6 \right.\right. \\ \qquad\qquad + \left(q_a^3 - 3 \times q_a\right) \times \kappa_{r_p}/24 - \left(2 \times q_a^3 - 5 \times q_a\right) \\ \qquad\qquad \left.\left. \times \left(sk_{r_p}\right)^2/36\right]\right\}, \end{cases}$$

where q_a is the a-quantile of the portfolio returns.

Rankings obtained with the measures developed by Dowd (2000) and Favre and Galeano (2002) are identical when considering Gaussian portfolio return distributions.

The Martin-Rachev-Siboulet (2003) Reward-to-Conditional Value-at-Risk. Martin et al. (2003) suggest the "STARR" (standing for Stable Tail Adjusted Return Ratio) which uses the Conditional Value-at-Risk or CVaR (see Artzner et al., 1999) . This risk measure aims to give information about the average amount of potential losses suffered by the portfolio manager, contrary to VaR which only provides a probability of potential losses defined from a threshold. In order to keep consistent with our previous notations, we rename this measure the "Reward-to-Conditional Value-at-Risk ratio" (RCVaR for short) which is given by:

$$RCVaR_p = \left[E\left(r_p\right) - E\left(r_B\right)\right] \times \left(CVaR_{r_p,a}\right)^{-1}, \tag{2.19}$$

with:

$$CVaR_{r_p,a} = -E\left[\left(r_p - r_B\right)|r_p - r_B \leq VaR_{r_p,a}\right] = -ES_{r_p,r_B,\tau_3},$$

where $\tau_3 = VaR_{r_p,a}$ and $CVaR_{r_p,a}$ is the Conditional Value-at-Risk of returns of the p portfolio in excess of the benchmark portfolio returns r_B, which is equivalent to the Expected Shortfall ES.

CVaR is suggested when investors want to evaluate the performance of their portfolio according to the average amount of their worst potential losses, defined beyond a given a-Value-at-Risk level.

The Young (1998) MiniMax Ratio. Young (1998) develops a similar approach to the Mean-Variance portfolio selection rule (Markowitz, 1952) based on a linear programming problem. This principle uses *minimum* return of the investor portfolio rather than variance as a measure of risk to set the optimal portfolio, also denoted as the "MiniMax" portfolio. More precisely, the chosen portfolio is the one that minimizes the *maximum* loss over all past observation periods, for a given level of return. Formally, the optimal portfolio is the solution of the linear program proposed by Young as:

$$\widehat{MiniMax}_t = \underset{w_j \in [0,\ldots,1]}{\arg\max} \left(\widehat{M}_t\right), \tag{2.20}$$

where the estimator of the *minimum* investor's portfolio return is:

$$\widehat{M}_t = \min_{i \in [1,\ldots,t]} \left(\sum_{i=1}^{I} w_j r_{j,i}\right).$$

In fact, given a vector of weights, $\widehat{M}_t$ is the estimated *minimum* return over time. Moreover, the difference between the returns of the portfolio over the period of interest and the *minimum* return will either be positive or zero. From the solution presented above, Young

(1998) introduces a further performance measure:

$$Yo_p = \left[E\left(r_p\right) - r_f\right] \times (MiniMax)^{-1}. \tag{2.21}$$

We should mention that the MiniMax measure can also be interpreted as the investor's portfolio return associated to a $VaR_{r_p,100\%}$. Given historical or simulated future returns data on a collection of assets, an optimal "MiniMax portfolio" can be constructed using such a linear programming technique. Under weak conditions, the MiniMax principle corresponds approximately to an expected utility maximizing principle, with the implied utility function representing an extreme form of risk aversion.

The Yitzhaki (1982) Gini Ratio. Yitzhaki (1982) suggests an alternative method to the Mean-Variance approach for comparing uncertain prospects based on the Gini coefficient. He proposes the following performance measure, called the Gini ratio:

$$Yi_p = \left[E\left(r_p\right) - r_f\right] \times \left(G_p\right)^{-1}, \tag{2.22}$$

where the Gini coefficient G_p (see Yitzhaki, 1982) , is defined as:

$$G_p = \frac{1}{2} E\left(\mid r_p - r_f \mid\right).$$

The Gini coefficient is used as a statistical measure of dispersion; one of its advantages is that it ranges from 0 to 1.

The Darolles-Gouri éroux-Jasiak (2009) "L-performance" Measure. Darolles et al. (2009) develop a measure, called "L-performance", similar to the previous Gini ratio. Indeed, the L-performance is defined as the ratio between the first and the second order of Trimmed L-moments (Elamir and Seheult, 2003) , which can be seen as truncated versions of traditional Linear-moments (L-moments). The L-moment of order 1 is equal to the mean and the L-moment of order 2 is defined as:

$$L_p^2 = \frac{1}{2} \times E\left(r_{p,[2:2]} - r_{p,[1:2]}\right), \tag{2.23}$$

where $r_{p,[1:2]}$ and $r_{p,[2:2]}$ are order statistics corresponding, respectively, to the smallest and largest returns in a sample composed by two observations.

L-moments are more robust to outliers than Conventional-moments. Indeed, their linear structure makes them less sensitive on the effects of sampling variability. [4]. As mentioned in Darolles et al. (2009), Trimmed L-moments can be associated to different values of quantiles, or equivalently to *VaR*, defined according to several confidence levels. The trimming parameter allows the assigning of more weight to neighbourhoods of some risk levels of interest. By increasing the sample size of returns from n to $(2n+1)$, thus underweighting the outliers of the return distribution, the authors propose a set of generalized L-performance ratios using, for evaluating the dispersion measure in the denominator, the difference between the $(1-a)$-*VaR* and the a-*VaR*, such as:

$$L_{p,a} = VaR_{r_p,0.5} \times \left(VaR_{-r_p,a} - VaR_{r_p,a}\right)^{-1}. \tag{2.24}$$

The Yitzhaki (1982) Gini ratio and the Darolles et al. (2009) Linear-performance have an interpretation similar to the one of the Sharpe ratio: they normalize the expected return of the managed portfolio by a measure that evaluates the portfolio total risk through a measure of dispersion. Based on the same approach, some authors have proposed additional performance measures using different quantities to assess risk, such as the Mean Absolute Deviation (Konno and Yamazaki, 1991) and the Range ratio (Caporin and Lisi, 2011) .

The Konno-Yamazaki (1991) Mean Absolute Deviation Ratio. Konno and Yamazaki (1991) present another way of measuring performance by defining the Mean Absolute Deviation (Konno , 1988 and 1990) as the risk measure, a more robust estimator of the scale compared to the standard deviation and more resilient to outliers in a data set. When computing the standard deviation, distances between portfolio returns from their mean are squared; so, on average, large deviations are weighted more heavily and, thus, outliers can strongly

[4] See Sillitto (1951) and Hosking (1989 and 1990) for more details on L-moments.

influence results. Thus, the proposed ratio, known as the "Mean Absolute Deviation ratio" is defined as:

$$KY_p = \left[E\left(r_p\right) - r_f\right] \times \left(MAD_{r_p}\right)^{-1}, \tag{2.25}$$

with the following usual estimator for the Mean Absolute Deviation of the managed p portfolio returns denoted MAD_{r_p}:

$$\widehat{MAD}_{r_p} = \sum_{i=1}^{I} \left[\left| r_{p,i} - \sum_{i=1}^{I} r_{p,i} \right|\right].$$

The interpretation of this measure is similar to the Sharpe (1966) ratio, except that it penalizes more heavily the performance of portfolios whose returns strongly vary negatively or positively around their mean.

The Caporin-Lisi (2011) Expected Return over Range Ratio. Grounded on the *maximum* magnitude between the investor's portfolio returns, Caporin and Lisi (2011) design a performance measure, named "Expected Return over Range ratio" (ERR for short), which is given by:

$$ERR_p = \left[E\left(r_p\right) - r_f\right] \times \left(RG_{r_p}\right)^{-1}. \tag{2.26}$$

Note that the Range of the managed portfolio p is estimated as:

$$\widehat{RG}_{r_p} = \left\{ \max_{i \in [1,\ldots,t]} \left(r_{p,i}\right) - \min_{i \in [1,\ldots,t]} \left(r_{p,i}\right) \right\},$$

where $\max(\cdot)$ and $\min\,(\cdot)$ are respectively the largest and the smallest investor's portfolio returns, denoted $r_{p,i}$, over the time period $i = [1,\ldots,t]$.

This ratio gauges the direct impact of market shocks on the performance of the studied portfolio. For instance, a high value of the range measure will imply a strong sensitivity of the investor's portfolio returns to the market turbulences and *vice-versa*.

The Martin-McCann (1989) Ulcer Index Performance. Martin and McCann (1989) present a performance measure based on the Mean-Downside Deviation, the "Ulcer Index Performance", which compares the expected excess return with the square root of its average squared weekly Drawdowns. This risk measure is called the "Ulcer Index", due to ulcers (and sleepless nights) caused by high losses. Formally, it is given by:

$$UIP_p = \left[E\left(r_p\right) - r_f\right] \times \left(UI_p\right)^{-1}, \tag{2.27}$$

where an estimation of the Ulcer Index is:

$$\widehat{UI}_{p,t} = \left[\frac{1}{t}\sum_{i=1}^{t}\left(\widehat{D}_{p,i}\right)^2\right]^{\frac{1}{2}},$$

and:

$$\widehat{D}_{p,i} = \frac{\max_{\tau\in[1,\ldots,i]}\left(NAV_\tau\right) - NAV_i}{\max_{\tau\in[1,\ldots,i]}\left(NAV_\tau\right)},$$

where $\widehat{D}_{p,i}$ are the Drawdowns of the investor's p portfolio returns at time $i = [1,\ldots,t]$ computed from weekly Net Asset Values (NAV).

Another way to define Drawdowns (see Caporin and Lisi, 2011) is:

$$\widehat{D}_{p,i} = \min\left(\widehat{D}_{p,i-1} + r_{p,i}, 0\right), \tag{2.28}$$

where $D_{p,0} = 0$ and $r_{p,i}$ are the investor's portfolio returns at time $i \in [0,\ldots,t]$.

In practice, Drawdowns of the portfolio returns are computed over at least a five-year time period. Several authors introduced similar performance measures with alternative Drawdown measures, such as the Total Squared Drawdowns (Burke, 1994) , the *Maximum* Drawdown (Young, 1991) and the Average Annual *Maximum* Drawdown (Kestner, 1996) . These indices based on Drawdowns are well-suited for evaluating the appropriateness of portfolio manager's investment choices when focusing on extreme losses, and for comparing the effectiveness of market timing strategies in reducing risk and avoiding large market downturns.

The Burke (1994) Sharper Ratio. Burke (1994) proposes the "Sharper ratio" by modifying the Ulcer Index as the Total Squared Drawdowns of investor portfolio:

$$B_p = \left[E\left(r_p\right) - r_f\right] \times \left(TSD_{r_p}\right)^{-1}, \tag{2.29}$$

where the Total Squared Drawdowns is estimated as:

$$\widehat{TSD}_{r_p} = \left[\sum_{i=1}^{t} \left(\widehat{D}_{r_p,i} \right)^2 \right]^{\frac{1}{2}}.$$

Burke (1994) computes the Drawdowns of the investor's portfolio NAVs over one year from monthly quotes.

Morever, Young (1991) suggests the "CALMAR Ratio"by focusing on the *Maximum* Drawdown. We thus have:

$$C_p = \left[E\left(r_p\right) - r_f \right] \times \left(MD_{r_p} \right)^{-1}, \tag{2.30}$$

with the following estimation of *Maximum* Drawdown:

$$\widehat{MD}_{r_p} = \max_{i \in [1,\ldots,t]} \left(\widehat{D}_{r_p,i} \right).$$

The term "CALMAR" given by Young (1991) to his performance measure corresponds to the initials of his company's name, *i.e.* CALifornia Managed Accounts Reports. Young (1991) recommends calculating the *Maximum* Drawdown over (at least) a three-year time period from monthly NAVs.

Furthermore, Kestner (1996) presents the "Sterling ratio", a slightly modified version of the Young (1991) CALMAR ratio, which uses the average annual *Maximum* Drawdown over several years to which he adds 10%, as follows:

$$St_p = \left[E\left(r_p\right) - r_f \right] \times \left[E\left(MD_{r_p} \right) + 10.00\% \right]^{-1}, \tag{2.31}$$

where $E\left(MD_{r_p}\right)$ is estimated by:

$$\widehat{E\left(MD_{r_p}\right)} = \left(\frac{1}{Y} \sum_{y=1}^{Y} \widehat{MD}_{r_p,t}^{(y)} \right),$$

where $\widehat{MD}_{r_p,t}^{(y)}$ is the annual *Maximum* Drawdown of the investor's p portfolio returns for the year y such as $y = [1,\ldots,Y]$.

This performance measure was originally developed by Deane Sterling Jones. Kestner (1996) also introduces the "K-ratio" which is computed as the slope of the regression line of the logarithmic investor's portfolio NAVs divided by the standard error of the slope, normalized by the root square of the number of observations. The recommended time period is 5 years. For computing the average annual *Maximum* Drawdown, Kestner (1996) suggests using monthly quotes over a three-year period of time.

The indices based on the Drawdown are well-suited for evaluating the appropriateness of the portfolio manager's investment choices when focusing on extreme losses, and for comparing the effectiveness of market timing strategies in reducing risk and avoiding large market downturns.

The Hübner (2005) Generalized Treynor Ratio. In the context of the Arbitrage Pricing Theory developed by Ross (1976), Hübner (2005) proposes a generalized version of the second ranking measure proposed by Treynor (1965) - denoted $T_{2,p}$ previously - which is defined as the abnormal return of a portfolio (*Cf.* Jensen, 1968) *per* unit of weighted-average systematic risk sensitivity, normalized by the *premium*-weighted average systematic risk sensitivity of the reference portfolio. The Generalized Treynor ratio is thus defined as:

$$\begin{aligned} GT_p = & \left\{\left[E\left(r_p\right)-r_f\right]-\left[E\left(r_m\right)-r_f\right]\times\beta_{r_p,r_m}\right\} \\ & \times\left\{\left\{\sum_{k=1}^{K}\left[E\left(F_k\right)\right]\times\beta_{r_p,F_k}\right\}\times\left\{\sum_{k=1}^{K}\left[E\left(F_k\right)\right]\times\beta_{r_m,F_k}\right\}^{-1}\right\}^{-1}. \end{aligned} \tag{2.32}$$

This generalized measure is equivalent to a weighted Jensen (1968) *alpha* which penalizes the performance of the portfolio manager when the exposition of his portfolio to each risk loading is superior to the exposition of the market portfolio to these same loading factors. In other words, the higher β_{r_p,F_k} compared to β_{r_m,F_k}, the lower the *extra*-performance of the investor's portfolio. The Generalized Treynor ratio is reduced to its original version $T_{2,p}$ when the number of risk loadings is equal to one (in a single index model).

The Scholz-Wilkens (05 b) Investor-Specific Performance Measure. Combining the Sharpe (1966) ratio and the first ranking measure introduced by Treynor (1965), Scholz and Wilkens (05 b) suggest a performance measure, labelled the "Investor-Specific performance

Measure", based on a Stochastic Dominance approach. The authors assume that an investor holds a position in the portfolio p_1 (not liquid). Furthermore, the investor wants to allocate some cash to a different fund, say portfolio i, which is the object of analysis, and will also borrow or lend at the risk-free rate. Scholz and Wilkens (05 b) associate this additional investment with a second portfolio denoted p_2. The aggregation of these two portfolios, p_1 and p_2, represents the portfolio p. It will be necessary to build as many indexes as the number of funds to be analyzed and/or compared. Assuming that the portfolio p_1 is the market portfolio m, we have:

$$\begin{aligned} ISM_i = & -w_{p_2} \times \left\{ \left[E\left(r_{p_2}\right) - r_f \right] \times \sigma_{r_m}^{-1} \right\}^2 \times S_i^{-2} \\ & -2 \times \left(1 - w_{p_2}\right) \times \left\{ \left[E\left(r_{p_2}\right) - r_f \right] \times T_{1,i}^{-1} \right\}, \end{aligned} \tag{2.33}$$

where $w_{p_2} \in [0,1]$ is the proportion fixed by the investor to allocate to the second portfolio p_2, $1 - w_{p_2}$ is the proportion invested in the market portfolio p_1, r_{p_2} and r_p are the returns of the portfolio p_2 and the global portfolio p, while S_i and $T_{1,i}$ are defined in (2) and (6), respectively.

The use of these different substitute risk measures presents some advantages but also drawbacks. Variants of the standard deviation (semi-volatility, L-moment of order 2, Mean Absolute Deviation, Range Ratio among others) might not be theoretically justified for non-Gaussian distributions. The Sortino-Satchell (2001) measure, and other derived ratios (for instance, Sortino and van der Meer, 1991; Kaplan and Knowles, 2004) , refer to LPM only to account for negative excess returns. However, small changes on the threshold, which is related to the investor's risk profile, can largely influence the LPM values. Additionally, in a dynamic framework, an investor can be misguided by focusing on LPM when a speculative bubble *phenomenon* occurs. Measures based on Value-at-Risk to evaluate risk of losses depend upon the estimation methodology of *VaR* (parametric, non-parametric or semi-parametric) and implicitly consider specific forms of utility functions (Kingston, 1989) . Finally, regarding the measures of Drawdowns, inference and estimation become complex, because of the uncertainty around the knowledge of the proper distribution of those quantities.

3

Measures of Absolute Performance

Summary. The second family presents the absolute PM, which can be formulated, in general terms, as follows:

$$PM_p = \Pi \left[\mathscr{P} \left(r_p - \tau_1 \right), \mathscr{P}^{th} \left(r_p - \tau_2 \,|\Omega \right) \right], \tag{3.1}$$

where $\mathscr{P}^{th} (\cdot\,|\cdot)$ is a function related to the theoretical performance of the investor's portfolio, conditionally to a set of information denoted by Ω. Most of the absolute Performance Measures can be, however, expressed in a straightforward manner, in which the $\Pi(\cdot,\cdot)$ function is linear, such as:

$$PM_p = \mathscr{P} \left(r_p - \tau_1 \right) - \mathscr{P}^{th} \left(r_p - \tau_2 \,|\Omega \right). \tag{3.2}$$

Absolute performance measures are positively influenced by an increase of the observed portfolio performance and by a decrease of the theoretical one and, in standard practices, are expressed in basis points.

The most influential measure is the Jensen *alpha* Jensen (1968). Based on the Capital Asset Pricing Model (CAPM), this measure aims at evaluating the portfolio manager's stock picking abilities. Many alternative versions have been proposed since the 70s. We describe the main ones in the next three subsections. Firstly, we introduce the Jensen (1968) *alpha*, then we present the main related measures and, finally, we group together miscellaneous absolute measures, which adopt other approaches for gauging absolute investor management skills.

3.1 The Jensen (1968) *Alpha*

Jensen (1968) proposes a performance measure, named "*alpha*", defined as the expected return of the investor's portfolio in excess of that predicted by the CAPM. The Jensen *alpha* is formulated as follows:

$$\alpha_p^J = \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m}. \tag{3.3}$$

This measure evaluates the contribution of manager's choices to the performance of his portfolio. The better the manager's stock picking skills, the higher the value of the Jensen *alpha*. The systematic risk sensitivity of the managed portfolio returns, denoted β_{r_p,r_m}, might refer to a *proxy* of the market portfolio, since investments realized by the investor may only represent a portion of the market.

Despite the wide use of the Jensen *alpha* for gauging the selectivity skills of managers, a number of criticisms have been advanced for this measure. The main one is related to the definition of the market *proxy*. Indeed, a misspecification can significantly modify the final ranking (see Roll, 1978 and 1979) of alternative investments. Furthermore, managers may change their portfolio's sensitivity to the systematic risk, according to their outlooks on market variations. In this case, the Jensen *alpha*, based on a constant single risk factor model, may counter-intuitively yield negative values (Treynor and Mazuy, 1966) . Finally, as for the Sharpe ratio, the characteristics of the risk-free rate (unique and constant) are not realistic.

3.2 Jensen-type Absolute Measures of Performance

The performance measures, presented below, are close variants of the Jensen (1968) *alpha*. However, they overcome some of its limits by attempting, with different approaches, to separately gauge the selectivity and market timing abilities of portfolio managers. This subsection is organized as follows. First, we present Jensen-type measures of performance based on a single-factor model as, for instance, the Black (1972) zero-*beta* version as well as the Fama (1972) index. Secondly, we discuss the multi-factor models. The main ones are the Connor-Korajczyk (1986) model and the Ferson-Schadt (1996) conditional *alpha*. Thirdly, we present the performance measures capturing, specifically, the investor's market timing abilities, such as, the Treynor-Mazuy (1966) and the Henriksson-Merton (1981) parametric models.

The Black (1972) Zero-*beta* CAPM. Black (1972) proposes the "zero-*beta* CAPM" by exploring the nature of capital market *equilibrium* under two restrictive assumptions. First, he assumes the absence of a riskless asset, but considers different risk-free borrowing and lending. Secondly, he supposes the existence of a riskless asset and the possibility to only take long positions on it (short positions are not allowed). In both cases, the investor can take unlimited long or short positions in risky assets. Then, the *extra* performance of the Black (1972) zero-*beta* CAPM is given by:

$$\alpha_p^{ZB} = \left[E\left(r_p\right) - E\left(r_z\right)\right] - \left[E\left(r_m\right) - E\left(r_z\right)\right] \times \beta_{r_p,r_m}, \tag{3.4}$$

where r_z is the return of the zero-*beta* portfolio z.

The interpretation of this performance measure is quite similar to the Jensen (1968) *alpha*, since it only differs by considering the expected return of a zero-*beta* portfolio instead of the risk-free rate. In this framework, the risky portion of the managed portfolio is a weighted combination of the market *proxy* and the *minimum*-variance zero-*beta* portfolio.

Another similar model is the one suggested by Brennan (1970) who proposes to incorporate the individual taxation when evaluating the portfolio performance. Finally, Leland (1999) suggests a corrected *beta* in order to take into account the investor's preferences for skewness.

The Fama (1972) Net Selectivity Index. In order to quantify the selectivity abilities of portfolio managers, Fama (1972) proposes a global performance measure, named the "Overall Performance index", which is written as:

$$OP_p = \left[E\left(r_p\right) - \bar{r}\left(\beta_{r_p,r_m}\right)\right] + \left[\bar{r}\left(\beta_{r_p,r_m}\right) - r_f\right], \tag{3.5}$$

where $\bar{r}\left(\beta_{r_p,r_m}\right)$ is given by:

$$\bar{r}\left(\beta_{r_p,r_m}\right) = r_f + \left[\left(r_m - r_f\right) \times \left(\sigma_{r_m}\right)^{-1}\right] \times \beta_{r_p,r_m}.$$

The Overall Performance index, given in Equation (3.5), is thus decomposed into two parts. The first term, $E\left(r_p\right) - \bar{r}(\beta_{r_p,r_m})$, is the selectivity reward component while the second term, $\bar{r}(\beta_{r_p,r_m}) - r_f$, corresponds to the risk reward component. From this statement, Fama (1972) suggests a second measure, labeled "Net Selectivity index", which is given by:

$$NS_p = \left[E\left(r_p\right) - r_f\right] - \left\{\left[E\left(r_m\right) - r_f\right] \times \left(\sigma_{r_m}\right)^{-1}\right\} \times \sigma_{r_p}. \tag{3.6}$$

The term $E\left(r_p\right) - r_f$ describes the compensation earned by the manager for bearing his portfolio total risk, whilst $\left\{\left[E\left(r_m\right) - r_f\right] \times \left(\sigma_{r_m}\right)^{-1}\right\} \times \sigma_{r_p}$ just reflects the remuneration that the manager would have obtained if his portfolio specific risk was rewarded as the systematic risk. The Net Selectivity index thus evaluates the stock picking abilities of the manager by assessing the *extra* return gained by the portfolio manager, simply comparing its excess return to that he would have earned, if he had been totally invested in the market portfolio only - for the same risk[1].

The Leland (1999) Measure. Leland (1999) develops a performance measure, based on the Rubinstein (1976) asset pricing model, which supposes lognormal market portfolio returns and a representative investor characterized by a Power Utility function - the third derivative of this utility function is positive, which implies a skewness preference. The Leland (1999) measure is formally given by:

$$\alpha_p^{LB} = \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta^*_{r_p,r_m}, \tag{3.7}$$

with:

$$\beta^*_{p,m} = \sigma_{r_p,\left[-(1+r_m)^{-b}\right]} \times \left\{\sigma_{r_m,\left[-(1+r_m)^{-b}\right]}\right\}^{-1},$$

and:

[1] For similar approaches, see also the measures proposed by Chauveau and Maillet (1997), Morey and Morey (1999), Cantaluppi and Hug (2000), Chauveau (2004), Briec et al. (2007), and, Briec and Kerstens (2010).

$$b = \left[E\left(1+r_m\right) - \left(1+r_f\right)\right] \times \sigma_{1+r_m}^{-2}.$$

Leland (1980) shows that an investor wants a portfolio insurance, if his risk tolerance grows with wealth more quickly than the average investor's risk tolerance. On the contrary, the investor will prefer a rebalancing strategy (concavity), if his risk tolerance grows less quickly than the markets. Moreover, risk tolerance grows more quickly when the investor has a preference for greater skewness. If the returns distributions are Gaussian, the results of the Leland (1999) measure and of the Jensen (1968) *alpha* are similar. The Leland (1999) measure gauges the contribution of the manager's investment choices to the performance of his portfolio. The better the portfolio manager's stock picking abilities, the higher the Leland (1999) measure and the portfolio performance.

The Smith-Tito (1969) Modified Jensen Measure. Based on the Jensen (1968) *alpha*, Smith and Tito (1969) propose a measure, originally named the "Modified Jensen" measure, that is written as:

$$\alpha_p^{ST} = \left\{\left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m}\right\} \times \left(\beta_{r_p,r_m}\right)^{-1}, \qquad (3.8)$$

where β_{r_p,r_m} is the systematic risk sensitivity of the investor's portfolio returns.

The Ang-Chua (1979) Excess Return Index. From the three-moment CAPM (see Rubinstein, 1973; Kraus and Litzenberger, 1976) , Ang and Chua (1979) propose a two-factor CAPM, whose peculiarity is the introduction of the asymmetry coefficient of the portfolio return distribution, which is written as:

$$\alpha_p^{AC} = \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{1,r_p,r_m} - \left[E\left(r_z\right) - r_f\right] \times \beta_{2,r_p,r_z}. \qquad (3.9)$$

This measure assesses the expected return of the investor's portfolio relative to the market portfolio and the *minimum*-variance zero-*beta* portfolio, which is assumed to have a non-null skewness coefficient. Moreover, the risk *premium* on the zero-*beta* portfolio and its associated sensitivity coefficients, β_{2,r_p,r_z}, should always have opposite signs, as long as a positive skewness is preferred by the manager. Thus, the more the portfolio return distribution is positively skewed, the better the performance.

The Treynor-Mazuy (1966) Market Timing Model. In order to assess the investor's market timing abilities, Treynor and Mazuy (1966) develop a bivariate regression model. The additional term, compared with the mono-factorial model previously introduced, corresponds to the squared market *proxy premium* and captures the convexity of the managed portfolio return function of the market return. The function can be evaluated with market data referring to the following regression model:

$$r_{p,t} - r_f = \alpha_p^{TM} + \left(r_{m,t} - r_f\right) \times \beta_{1,r_p,r_m} + \left(r_{m,t} - r_f\right)^2 \times \beta_{2,r_p,r_m} + \varepsilon_t. \qquad (3.10)$$

where α_p^{TM} is the Treynor-Mazuy *alpha*, the t subscripts denote the time index, β_{1,r_p,r_m} is the systematic risk sensitivity of returns on the investor's portfolio, while β_{2,r_p,r_m} is the corresponding market timing coefficient.

This regression model provides rankings that are directly linked to the manager's abilities to forecast the sign of future returns of the market *proxy*, and thus to adjust his market exposure. If we interpret *ex post* this model, a positive value of the estimate $\widehat{\beta}_{2,r_p,r_m}$ implies some market timing abilities. A few authors (see, for instance, Goetzmann et al., 2007; Hübner, 2011) reinterpret the original formulation developed by Treynor and Mazuy (1966) with investment strategies including derivatives. Indeed, we can compute the *extra* performance of a market timer, defined in equation (3.10), as the difference between the *ex post* returns of a portfolio actively managed by an investor and those of a replicating portfolio, consisting of a combination of options (put or call, depending on the directional and quadratic exposure, respectively $\widehat{\beta}_{1,r_p,r_m}$ and $\widehat{\beta}_{2,r_p,r_m}$) and the risk-free rate. In other words, we can interpret the term $\left(r_{m,t} - r_f\right) \times \widehat{\beta}_{1,r_p,r_m} - \left(r_{m,t} - r_f\right)^2 \times \widehat{\beta}_{2,r_p,r_m}$ as the estimated cost (*premium*) saved by the portfolio manager from buying options.

The Henriksson-Merton (1981) Parametric Market Timing Model. Henriksson and Merton (1981) - see also Merton (1981) and Henriksson (1984) - suggest a bivariate regression model for market timing inspired by option theory. The main idea is to introduce a put option pay-off, interacting with the evolution of portfolio market returns, in order to separately identify investors' micro- and macro-forecasting abilities. These abilities correspond to, respectively, selectivity and market timing skills. The suggested regression model is:

$$r_{p,t} - r_f = \alpha_p^{HM} + \left(r_{m,t} - r_f\right) \times \beta_{1,r_p,r_m} + \max\left(r_f - r_{m,t}, 0\right) \times \beta_{2,r_p,r_m} + \varepsilon_t, \tag{3.11}$$

and β_{2,r_p,r_m} is the market timing coefficient, while α_p^{HM} is the Henriksson-Merton *alpha*.

As mentioned by Henriksson and Merton (1981), it is crucial to view this model by considering both the performance measured by the *alpha*, reflecting the investor's abilities in terms of stock selections, and the second *beta* associated with the agent's market timing skills. Considering this parametric model as an *ex post* measure, the authors interpret the market timing ability or "macro-forecasting" skills, expressed by the estimated term $\max\left(r_f - r_{m,t}, 0\right) \times \widehat{\beta}_{2,r_p,r_m}$, as the ability to implement a hedge strategy with some put options written on the market portfolio. More precisely, this term is equivalent to an investment strategy consisting of buying the market portfolio protected by $\widehat{\beta}_{2,r_p,r_m}$ put options, where the strike is equal to the risk-free rate. This strategy will pay off if the *ex post* market portfolio return is lower than the risk-free rate and the balance is invested (borrowed), if positive (negative) at the risk-free rate. The coefficient $\widehat{\beta}_{2,r_p,r_m}$ thus represents the number of options that have been spared by the portfolio manager thanks to his market timing capacity for having the same overall *ex post* return as the protected strategy. For alternative ways to measure the contribution of market timing to active returns, see Coles et al. (2006) , Bollen and Busse (2001 and 2005), Comer et al. (2009) and Hübner (2011).

The Morey-Morey (1999) Mutual Fund Performance Appraisals. Morey and Morey (1999) present two efficiency measures, called "Radial Expansion" and "Radial Contraction" for identifying managed portfolios that are strictly dominated by others. Firstly, they focus on the expected return and, then, on the total risk level. The two performance indexes are given by:

$$RE_p = \arg\max_{\theta \geq 1} \left(\theta E(r_p) \leq E(r_B) \middle| \sigma_{r_B}^2 \leq \sigma_{r_p}^2 \right), \tag{3.12}$$

and:

$$RC_p = \underset{0<\theta\leq 1}{\arg\min}\left(\sigma^2_{r_B} \leq \theta\sigma^2_{r_p}\middle| E(r_B) \geq E(r_p)\right). \tag{3.13}$$

These two measures correspond to optimization, in some directions, to the *ex post* efficiency frontier in the risk – expected return plane. The focus of the “Radial Expansion” approach is to expand the mean return of the managed portfolio (while keeping the total risk constant). Conversely, the “Radial Contraction” approach is to contract the total risk of the studied portfolio (for a same expected return). The value of the resulting θ will thus be the loss in terms of returns or the over-risk given by the portfolio when compared to the efficient ones (with a θ above the unit in the first approach and below the unit in the second).

There also exist other efficiency measures built on similar methodologies; for example, Cantaluppi and Hug (2000) propose to evaluate the level of under-optimization of the investor’s portfolio simply by dividing its Sharpe ratio by that of the optimal portfolio – in a Mean-Variance plane – when total risks are identical for both compared portfolios. Chauveau and Maillet (1997), and Chauveau (2004) share the same line of reasoning with borrowing constraints, while Briec et al. (2007) and Briec and Kerstens (2010), respectively, add a skewness and *kurtosis* parameter to the analysis framework.

The Hwang-Satchell (1998) Higher-Moment CAPM. According to a Mean-Variance-Skewness-*Kurtosis* approach, Hwang and Satchell (1998) consider the incremental value of higher moments in modeling a CAPM for emerging markets. The main objective is to know if the performance of the investor’s portfolio invested in emerging markets can be better explained by taking into account the effect of the skewness and *kurtosis* of studied portfolio return distributions. The model yields to the following form:

$$\alpha_p^{HS} = \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \lambda_{1,p,m} - \left(\beta_{r_p,r_m} - sk_{r_p,r_m}\right) \times \lambda_{2,m}, \tag{3.14}$$

with:

$$\begin{cases} \lambda_{1,r_p,r_m} = \left[(sk_{r_m})^2 \times sk_{r_p,r_m} - (\kappa_{r_m} - 1) \times \beta_{p,m}\right] \\ \qquad \times \left[(sk_{r_m})^2 - (\kappa_{r_m} - 1)\right]^{-1}, \\ \lambda_{2,r_m} = (sk_{r_m} \times \sigma_{r_m}) \times \left[(sk_{r_m})^2 - [\kappa_{r_m} - 1]\right]^{-1}, \end{cases}$$

and:

$$\begin{cases} \beta_{r_p,r_m} = E\left\{\left[r_p - E\left(r_p\right)\right] \times \left[r_m - E\left(r_m\right)\right]\right\} \times E\left\{\left[r_m - E\left(r_m\right)\right]^2\right\}^{-1} \\ sk_{r_p,r_m} = E\left\{\left[r_p - E\left(r_p\right)\right] \times \left[r_m - E\left(r_m\right)\right]^2\right\} \times E\left\{\left[r_m - E\left(r_m\right)\right]^3\right\}^{-1} \end{cases},$$

where sk_{r_p,r_m} is the co-skewness between the investor's p portfolio and the m market portfolio returns.

We can notice that if the market portfolio return distribution is Gaussian, the first sensitivity coefficient, λ_{1,r_p,r_m}, is equal to the systematic risk sensitivity of the returns on the investor's portfolio β_{r_p,r_m}, and the second one, λ_{2,r_m}, is null, leading to the Jensen (1968) *alpha.* From the Sears-Wei (1985) Four-Moment CAPM, Hwang and Satchell (1999) also propose a Higher-Moment CAPM with the introduction of the co-*kurtosis* and a quadratic return generating process.

The Brennan (1970) Tax-based Model. Based on the Farrar and Selwyn (1967) works about individual taxation, Brennan (1970) develops a tax-based CAPM which considers the average taxation rates applied to dividends and capital gains earned by the portfolio manager. Thus, the Brennan (1970) tax-based model gives the following *extra*-performance:

$$\begin{aligned} \alpha_p^B = & \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m} \\ & - \left[\left(\varphi_p - r_f\right) - \left(\varphi_m - r_f\right) \times \beta_{r_p,r_m}\right] \times \eta_p, \end{aligned} \tag{3.15}$$

with:

$$\eta_p = \left(\eta_1 - \eta_2\right) \times \left(1 - \eta_2\right)^{-1},$$

where η_p is the marginal tax rate, η_1 and η_2 are, respectively, the average taxation rates for dividends and capital gains, φ_p and φ_m are the dividend yields of the investor's portfolio p and of the market portfolio m.

This model gauges the investor's portfolio *extra*-performance from the arbitrage made by the investor between the exposure to the systematic risk of his portfolio returns and the dividend yields perceived from the holding of risky assets. The first term, $\left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m}$, represents the expected return of the investor's portfolio in excess of the market portfolio, and corresponds to the Jensen (1968) *alpha*. The second component, $\left[\left(\varphi_p - r_f\right) - \left(\varphi_m - r_f\right) \times \beta_{r_p,r_m}\right] \times \eta_p$, evaluates the dividend yield, adjusted to the investor's marginal tax rate, obtained from the holding of risky assets in the managed portfolio in excess of those he would have obtained, if he had held the market portfolio.

The McDonald (1973) Total Performance Measure. Within an international framework, McDonald (1973) introduces a measure, *alias* "total performance", which evaluates the performance of portfolios holding risky assets in two different markets which are assumed totally independent. The author defines a portfolio p held by a manager who wishes to invest in the French and the US stock markets, denoted m_1 and m_2. The McDonald (1973) total performance measure is thus:

$$\alpha_p^{McD} = \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_{m_1}\right) - r_f\right] \times \beta^*_{r_p,r_{m_1}} - \left[E\left(r_{m_2}\right) - r_f\right] \times \beta^*_{r_p,r_{m_2}}, \tag{3.16}$$

where:

$$\begin{cases} \beta^*_{r_p,r_{m_1}} = w_{p,m_1} \times \left[\left(\sigma_{r_p,r_{m_2}} \times \sigma_{r_{m_1},r_{m_2}}\right) \right. \\ \qquad\qquad \left. - \left(\sigma^2_{r_{m_2}} \times \sigma_{r_p,r_{m_1}}\right)\right] \\ \qquad\qquad \times \left[\left(\sigma_{r_{m_1},r_{m_2}}\right)^2 - \left(\sigma^2_{r_{m_1}} \times \sigma^2_{r_{m_2}}\right)\right]^{-1}, \\ \beta^*_{r_p,r_{m_2}} = w_{p,m_2} \times \left\{\left[\sigma_{r_p,r_{m_1}} \times \sigma_{r_{m_1},r_{m_2}}\right] \right. \\ \qquad\qquad \left. - \left(\sigma^2_{r_{m_1}} \times \sigma_{r_p,r_{m_2}}\right)\right\} \\ \qquad\qquad \times \left[\left(\sigma_{r_{m_1},r_{m_2}}\right)^2 - \left(\sigma^2_{r_{m_1}} \times \sigma^2_{r_{m_2}}\right)\right]^{-1}, \end{cases}$$

where r_{m_1} and r_{m_2} are the returns of the French stock market portfolio m_1 and the US stock market portfolio m_2, $\beta^*_{r_p,r_{m_1}}$ and $\beta^*_{r_p,r_{m_2}}$ are the weighted systematic risk sensitivities of portfolio p, $w_{p,m_1} = 1 - w_{p,m_2}$ are the corresponding portfolio weights.

This measure assumes a world of segmented markets, where the non diversifiable risk sensitivity is defined relative to the domestic market alone. The French risk-free rate is used as a benchmark. If we consider a French investor, he will need to figure out the best combinations in terms of investments and systematic risk sensitivities to domestic and foreign markets to get the higher excess return. Pogue et al. (1973) generalize the McDonald (1973) total performance measure to a portfolio containing two asset classes, say bonds and securities, and where risky assets are negotiated in several international stock markets, with no limit about the number of countries. These two performance measures evaluate the portfolio manager's ability to choose the most promising markets and his skill in selecting the best stocks in each market.

The Fama-French (1993) Three-factor CAPM. Fama and French (1993) propose another version of the three-factor CAPM (*Cf.* Rubinstein, 1973; Kraus and Litzenberger, 1976) by assuming that two asset classes outperform the market: small-cap companies and stocks with a high book-to-market ratio (customarily termed "value stocks", and their opposites "growth stocks"). Building on this intuition, they propose to introduce to the traditional Jensen (1968) *alpha* two additional factors: *SMB*, detecting the *extra*-performance of Small companies with respect to the Big ones (Small-*Minus*-Big factor) and HML (High-*Minus*-Low) that is associated with the *extra*-return given by companies with a high book-to-market ratio compared to those with a low book-to-market ratio. In this model, the *extra*-performance is given as:

$$\begin{aligned}\alpha_p^{FF} = & \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m} \\ & - \left[E\left(r_{SMB}\right)\right] \times \beta_{r_p,r_{SMB}} - \left[E\left(r_{HML}\right)\right] \times \beta_{r_p,r_{HML}},\end{aligned} \tag{3.17}$$

where r_{SMB} and r_{HML} are the returns of the SMB and HML risk factors, while $\beta_{r_p,r_{SMB}}$ and $\beta_{r_p,r_{HML}}$ are the corresponding sensitivities.

An extension is proposed by Carhart (1997), who suggests a four-factor CAPM by adding a fourth sensitivity factor to assess the persistence of studied risky assets over the time period of interest and then takes into account the *momentum* effect (Jegadeesh and Titman, 1993) . This model has the following form:

$$\begin{aligned}\alpha_p^{C} = & \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m} \\ & - \left[E\left(r_{SMB}\right)\right] \times \beta_{r_p,r_{SMB}} - \left[E\left(r_{HML}\right)\right] \times \beta_{r_p,r_{HML}} \\ & - \left[E\left(r_{PR1YR}\right)\right] \times \beta_{r_p,r_{PR1YR}},\end{aligned} \tag{3.18}$$

where the factor $PR1YR$ (stands for "PRevious 1 YeaR") corresponds to the one-month lagged eleven-month past returns.

These two multi-factor models allow investors to have a more appropriated decomposition of their portfolio returns, where the sensitivity coefficients and risk *premia* on the factor-mimicking portfolios indicate the proportion of mean return associated with three (four) main investment strategies: high *versus* low *beta* stocks, large *versus* small market capitalization stocks, value *versus* growth stocks (and a one-year return *momentum versus* contrarian stocks).

The Connor-Korajczyk (1986) Model. Connor and Korajczyk (1986) present a generalization of the CAPM model. Their proposal explores the effects of several risk factors in the analysis of investment performance. This *alpha*-index implied by the model is:

$$\alpha_p^{CK} = \left[E\left(r_p\right) - r_f\right] - \sum_{k=1}^{K} \left[E\left(F_k\right)\right] \times \beta_{r_p,F_k}, \tag{3.19}$$

where $E\left(F_k\right)$ is the expected risk *premium* of the k-th factor and β_{r_p,F_k} is the sensitivity of the investor's portfolio returns to the k-th factor.

Connor and Korajczyk (1986) determine the relevant risk factors by using traditional model specification techniques. If the value of the expected risk *premium* of the k-th risk factor is significantly positive, this factor is kept as a rewarding factor, and disregarded otherwise. The two-step analysis is carried out again with the remaining factors.

The Ferson-Schadt (1996) Conditional Performance Measure. A refinement of the Connor-Korajczyk (1986) model is developed by Ferson and Schadt (1996) in order to explain the evolution of the systematic risk sensitivity of the managed portfolio over time. The main idea is to use time-varying conditional expected returns and conditional *betas*. This conditional performance measure is defined as:

$$\alpha_p^{FS} = \left[E\left(r_p\right) - r_f\right] - \sum_{k=1}^{K} E\left(F_k\right) \times \beta_{r_p,F_k,t}, \tag{3.20}$$

where $\beta_{r_p,F_k,t}$ are time-varying coefficients and represent the risk sensitivity of the managed portfolio returns p to the k-th risk factor at time t.

Time-variation of managed portfolio *betas* may come from three distinct sources: the sensitivities of the underlying assets may change over time; the weights of underlying assets in the benchmark will vary as relative value changes; and a manager is paid for actively managing weights of his portfolio. The Ferson and Schadt (1996) approach shows that a negative value of a Jensen (1968) *alpha* might be due, in reality, to a time-varying *beta*.

The Aftalion-Poncet (1991) Index. Proposing a new definition of the market price of risk, Aftalion and Poncet (1991) build an index which compares the expected return of the investor's portfolio in excess of the expected return of its reference portfolio, to the return that should have been reached according to the total risk of the managed portfolio. The Aftalion and Poncet (1991) index is defined such as:

$$AP_p = \left[E\left(r_p\right) - E\left(r_B\right)\right] - PXR\left(\sigma_{r_p} - \sigma_{r_B}\right), \tag{3.21}$$

with:

$$PXR(\sigma_{r_p} - \sigma_{r_B}) = \left[E\left(r_p\right) - E\left(r_B\right)\right] \times \left(\sigma_{r_p} - \sigma_{r_B}\right)^{-1},$$

where $PXR(\cdot)$ is the market price of risk of the portfolio p.

The expected excess return of portfolio p contributes positively to this index, while the excess total risk contributes negatively. The market price of risk, which has the same dimension as an expected excess return divided by an excess total risk, ensures that the two terms in the index have the same scale. The Aftalion and Poncet (1991) index represents the *extra* return that investors require on average for each additional point of risk. Authors illustrate their measure by doing a comparison between a portfolio invested in cash over 30 years and a diversified portfolio invested in the French stock market. The expected returns of these two portfolios are, respectively, 8.00% and 11.00%, while the corresponding volatilities are equal to 3.00% and 18.00%. In this specific case, the authors conclude that an *extra* total risk equal to 15.00% is compensated for an *extra* return of 3.00%. The market price of risk is thus equal to .20, *i.e.* 3.00% over 15.00%.

The Sortino-Price (1994) Fouse Index. Combining a Mean-Downside risk computation and the Expected Utility Theory, Sortino and Price (1994) propose a measure, entitled the "Fouse index", which is computed from a Generalized Lower Partial Moment of order 2 of the investor's portfolio returns, adjusted for the coefficient of risk aversion of the manager (see Pratt, 1964; Arrow, 1971). The Fouse index is formulated as:

$$\alpha_p^F = E\left(r_p\right) - \left(A \times GLPM_{r_p,\underline{r},\underline{r},2}\right). \tag{3.22}$$

This model determines the performance of the investor's portfolio as the difference between its expected return and a "subjective" measure of its downside risk. This function of the LPM of order 2 of the investor's portfolio returns may thus vary widely according to the manager's risk profile.

The Melnikoff (1998) Index. In the same vein as Sortino and Price (1994), Melnikoff (1998) proposes to combine a Mean-Downside risk computation and the Prospect Theory (Kahneman and Tversky, 1979) . Indeed, the proposed measure corresponds to the difference between the investor's portfolio return in excess of a "subjective" measure of its downside risk that is given by:

$$\alpha_p^M = E\left(r_p\right) - \left[(\varpi - 1) \times GLPM_{r_p,r_f,r_f,1}\right], \tag{3.23}$$

where ϖ is a constant reflecting the loss-to-gain aversion weight.

The interpretation of this model is quite similar to the one of the Sortino-Price (1994) Fouse Index.

The Lobosco (1999) Style Risk-Adjusted Performance Measure. Directly based on the Modigliani and Modigliani (1997) works, Lobosco (1999) suggests a measure, named "Style Risk-Adjusted Performance" , which compares the performance of the managed portfolio to its benchmark with respect to the investor's management style (Sharpe, 1992) . The Lobosco

(1999) Style Risk-Adjusted Performance measure is equal to:

$$\begin{aligned} SRAP_p = & \left[E\left(r_p\right) \times \left(1+\lambda_{1,r_p,r_m}\right) - r_f \times \lambda_{1,r_p,r_m}\right] \\ & - \left[E\left(r_B\right) \times \left(1+\lambda_{2,r_p,r_B}\right) - r_f \times \lambda_{2,r_p,r_B}\right], \end{aligned} \tag{3.24}$$

where $\lambda_{1,r_p,r_m} = \sigma_{r_m} \times \left(\sigma_{r_p}\right)^{-1}$ and $\lambda_{2,r_p,r_B} = \sigma_{r_m} \times \left(\sigma_{r_B}\right)^{-1}$.

Lobosco (1999) considers that portfolios characterized by high Risk-Adjusted Performances (Modigliani and Modigliani, 1997) are deemed to have better managers than those with low results. While this may often be the case, it is not always true since higher results can only be the consequence of a style mandate, not of the manager's skills. The proposed measure is an attempt to compensate this drawback. In other words, the main objective is to show that the underperformance of a portfolio may be perceived as a warning about the investment style forced by the company, beyond the personal abilities of the fund manager.

The Graham-Harvey (1997) Measures. Graham and Harvey (1997) build two performance measures based on market timing advice, provided by investment newsletters, which recommend increasing investments in risky assets before market appreciation and decrease before market shocks. The Graham and Harvey (1997) measures are written as:

$$GH_{1,p} = E\left(r_p\right) - E\left(r_m\right) \times \lambda_{r_m}, \tag{3.25}$$

and:

$$GH_{2,p} = E\left(r_p\right) \times \lambda_{r_p} - E\left(r_m\right), \tag{3.26}$$

where $\lambda_{r_p} \in \mathbb{R}_+$ and $\lambda_{r_m} \in \mathbb{R}_+$ are the (un-)leverage factors of the investor's portfolio p and the market *proxy* m.

The first performance measure only deals with the (un)leverage of the market *proxy,* while the second focuses on the (un)leverage of the investor portfolio. In both cases, the (un)leverage factor is determined by supposing that the two elements entering the performance measures have the same total risk.

The Briec-Kerstens-Jokung (2007) Overall Efficiency Index. Briec et al. (2007) introduce a performance measure, based on a Mean-Variance-Skewness (MVS) approach and using a shortage function, which can be seen as an extension of the Briec et al. (2004) work. The shortage function aims at looking for possible increases in return and skewness and decreases in variance. The Briec et al. (2007) overall efficiency index is computed as:

$$OE_p = \underset{\theta \in \mathbb{R}_+}{\arg\max} \left[\phi \left(r_p \right) + \theta g \leq \phi \left(r_B \right) \right], \tag{3.27}$$

where $\phi(\cdot) = [E(\cdot), -\sigma(\cdot), Sk(\cdot)]$ is a function which represents for a given portfolio its expected return, variance and skewness, and $g = (g_E, g_V, g_{Sk})$ is a directional vector for each of the first three Conventional moments.

This model implies a simultaneous maximization of all the three moments, namely mean, (*minus*) variance and skewness, until finding the best (more efficient) portfolio. For instance, assuming that an investor decides to invest all his cash in a single risky asset, he obtains in his study a value of 92.00% by computing the Briec et al. (2007) overall efficiency. This result implies that it is possible to do better in terms of overall efficiency compared to this risky asset if we apply the optimal portfolio weight, which will improve return and skewness and reduce risk of this same asset by 92.00%. This example can be easily extended to a portfolio containing several risky assets.

These absolute performance measures, directly based on the Jensen *alpha*, display some peculiar problems, regardless of the different proposed improvements. Fama (1972) analyzes the *ex post* compensation earned by the manager, compared to that he would have received if his portfolio had been completely exposed to the systematic risk. Also in this case, the definition of the market *proxy* plays a crucial role. The efficiency measures (Chauveau and Maillet, 1997; Morey and Morey, 1999; Cantaluppi and Hug, 2000) aim at determining the potential under-performance experienced by the manager, in comparison to the performance he could have reached by holding the *ex post* optimal portfolio. Extensions to the third (Briec et al., 2007) and to the fourth moments (Briec and Kerstens, 2010) also exist. Yet, they are subject to issues related to the characterization of return densities. Treynor and Mazuy (1966) study the investor's market timing abilities by analyzing the convexity of the return function linking the portfolio to the market return. In the same vein, Henriksson and Merton (1981) propose using call options for *ex post* reproducing this function. Here, the critical point is that exact rankings rely upon the precise characteristics of derivatives that are used for replication (strike, time to maturity, Greeks...). Black (1972) defines a zero-*beta* portfolio as the benchmark, assuming the absence of a riskless asset. However, it still relies on CAPM assumptions. Finally, Connor and Korajczyk (1986) extend these approaches within a multi-factorial framework, while Ferson and Schadt (1996) suggest a correction by considering time-varying *betas*, that explicitly highlights, as in Treynor and Mazuy (1966), the potential mixture between market timing and selectivity qualities of portfolio managers, that definitely cannot be considered as orthogonal. However, since the role of a portfolio manager is precisely to adjust his positions according to expected behaviors of factors and market conditions, we may end up in a paradoxical situation in which the performance of a "good" manager is totally explained, and thus null.

3.3 Other Miscellaneous Absolute Performance Measures

Miscellaneous measures, described hereafter, are based on a similar approach, when comparing the observed performance and the theoretical one. However, contrary to the Jensen-type measures, the link between the observed and the theoretical expected returns is much less straightforward. In other words, the way these measures are built is more complex, and they might not be expressed in return terms.

The Moses-Cheyney-Veit (1987) Measure. Moses et al. (1987) develop a measure that considers both the expected active return and the diversification level of risky asset portfolio held by an informed investor. The main idea is to evaluate, in terms of performance, the arbitrage made by the manager between his performance and his idiosyncratic risk, assumed to be priced by the market. Formally, the measure is defined as:

$$MCV_p = \left\{ \left[E\left(r_p\right) - r_f\right] - \left[E\left(r_m\right) - r_f\right] \times \beta_{r_p,r_m} \right\} \times \lambda_p^{-1}, \tag{3.28}$$

with:

$$\lambda_p = \left[E\left(r_m\right) - r_f\right] \times \left[\left(\sigma_{r_p} \times \sigma_{r_m}^{-1}\right) - \beta_{r_p,r_m}\right].$$

This measure corresponds to a weighted Jensen *alpha* such as the *extra* performance of the investor's portfolio is adjusted by the market risk *premium* associated with its *extra* portfolio total risk (see Moses et al., 1987) .

Other performance measures have been proposed with the aim of taking into account management skills: Muralidhar (2001) develops the Correlation-Adjusted Performance to compare the management skills of several investors within a peer group, with respect to a target Tracking Error; Muralidhar (2002) refines the approach allowing a comparison between portfolios with different inception dates, for instance; Scholz and Wilkens (005a) generalize Muralidhar (2002) focusing on the systematic risk sensitivity of portfolio returns.

The Modigliani-Modigliani (1997) Risk-Adjusted Performance Measure.
Studying the impact of leverage strategies on the portfolio performance, Modigliani and Modigliani (1997) develop a close variant of the measure proposed by Statman (1987) named "eSDAR" (standing for "excess Standard Deviation Adjusted Return). The authors called it "M² [2]" or "Risk-Adjusted Performance" and the aim is to scale down (up) the expected return of the investor's portfolio with respect to its total risk, when the latter is higher (lower) than the market portfolio total risk. The proposed measure is written as:

$$RAP_p = \gamma_p \times \left[E\left(r_p\right) - r_f\right] + r_f, \tag{3.29}$$

where the leverage parameter γ_p is defined as $\gamma_p = \left(\sigma_{r_m} \times \sigma_{r_p}^{-1}\right)$.

This measure evaluates the expected return of a portfolio composed with risky and risk-free assets *per* unit of market total risk. In other words, it allows the investor to assess the compensation earned for the *extra* total risk with respect to that of the market portfolio risk.

Scholz and Wilkens (005a) propose a similar performance measure, the "Market Risk-Adjusted Performance" by considering the systematic risk sensitivity of the managed portfolio returns instead of its total risk. The Scholz and Wilkens (005a) Market Risk-Adjusted

Performance measure is expressed as:

$$MRAP_p = T_{1,p} + r_f, \tag{3.30}$$

where $T_{1,p}$ is the Treynor (1965) ratio and $r_f > 0$ is the risk-free rate.

Differently, Graham and Harvey (1997) build two measures by making the comparison between the unleveraged (respectively leveraged) investor's portfolio expected return and the leveraged (resp. unleveraged) market portfolio expected return, assuming an identical total risk for both portfolios. Finally, Lobosco (1999) develops the "Style Risk-Adjusted Performance", which compares the performance of the managed portfolio to its benchmark with respect to the investor's management style (Sharpe, 1994) .

The Grinblatt-Titman (1989) Positive Period Weighting Measure. In response to the timing-related biases of the Jensen (1968) *alpha*, Grinblatt and Titman (1989) propose a measure, named "Positive Period Weighting measure" (PPW in short), which attributes a positive weight to the excess return of the portfolio of a real (*versus* lucky) market timer over the studied time horizon. This measure is computed as:

$$PPW_{p,t} = \sum_{i=1}^{t} w_i \times \left(r_{p,i} - r_f\right), \tag{3.31}$$

with weights satisfying:

$$\begin{cases} \sum_{i=1}^{t} w_i \times \left(r_{m,i} - r_f\right) = 0 \\ \sum_{i=1}^{t} w_i = 1. \end{cases}$$

The weight vector is chosen to have non-negative values that make the weighted sum of the excess returns of the benchmark portfolio equal to zero. We suppose, here, that the benchmark/market portfolio is not always above the risk-free rate, otherwise the structure of weights cannot respect the first condition. This measure is thus both benchmark and sample period specific. There are many sets of weights that satisfy the conditions above. Following Grinblatt and Titman (1994), the weights are derived from marginal utilities of an uninformed investor characterized by a power utility function (see Grinblatt and Titman, 1994 , for additional details). In this context, a negative value of the Jensen (1968) *alpha* attributed to a real market timer can thus be explained in terms of (negative) marginal utilities, when considering that the portfolio return level reached by the investor exceeds the satiation point defined by its utility function. Consequently, the PPW measure assigns positive weights to the portfolio excess returns of informed investors having stock picking and/or market timing abilities and a null performance otherwise. A similar measure, based on a two-period comparison, was introduced by Cornell (1979).

The Henriksson-Merton (1981) Non parametric Market Timing Model. Henriksson and Merton (1981) suggest a non parametric model for gauging market timing abilities of portfolio managers. Under some assumptions detailed in Merton (1981) and Henriksson and Merton (1981), the authors show that the sum of two conditional probabilities of a correct forecast is a sufficient statistic for the evaluation of the portfolio manager's market timing abilities. More precisely, the non parametric market timing model is written as:

$$HM_p = \text{ Prob }_1 + \text{ Prob }_2 - 1, \tag{3.32}$$

with:

$$\begin{cases} \text{Prob }_1 = \text{ Prob } \left[\delta = 0 | \, r_m \leq r_f \right] \\ \text{Prob }_2 = \text{ Prob } \left[\delta = 1 | \, r_m > r_f \right], \end{cases}$$

where δ is the market timer's forecast variable, and $\delta = 0$ ($\delta = 1$) corresponds to an incorrect (correct) forecast of the market direction.

The aim of this measure is to focus on the correctness of forecasts of a portfolio manager regarding the evolution of portfolio market returns. It is, however, assumed that the conditional probabilities Prob_1 and Prob_2 do not depend on the magnitude of the expected portfolio market return in excess of the risk-free rate. In other words, it is implicitly assumed that a better forecasting ability on average leads to a better performance.

The Statman (1987) excess Standard Deviation-Adjusted Return. Inspired by the works of Markowitz (1952), Statman (1987) proposes the following measure:

$$eSDAR_p = \gamma_p \times \left[E\left(r_p\right) - r_f\right] + r_f - E\left(r_m\right), \tag{3.33}$$

where the leverage parameter γ_p is defined as $\gamma_p = \left(\sigma_{r_m} \times \sigma_{r_p}^{-1}\right)$.

The Cantaluppi-Hug (2000) Efficiency Ratio. Cantaluppi and Hug (2000) propose directly evaluating the performance of the investor's portfolio by reference to its *maximum* potential performance. They suggest using the quotient between the standard Sharpe ratio of a portfolio and its optimal Sharpe ratio, assuming an identical portfolio total risk. The Cantaluppi and Hug (2000) efficiency ratio writes:

$$RE_p = \left\{\left[E\left(r_p\right) - r_f\right] \times \left(\sigma_{r_p}\right)^{-1}\right\} \times \left\{\left[E\left(r_B\right) - r_f\right] \times \left(\sigma_{r_B}\right)^{-1}\right\}^{-1}. \tag{3.34}$$

In the Cantaluppi and Hug measure, the reference portfolio (the benchmark) is the tangency portfolio obtained in the unconstrained Mean-Variance framework of Markowitz (1952). The Cantaluppi and Hug (2000) efficiency ratio then considers the *maximum* potential return of the portfolio under consideration. A result strictly inferior to one means that the studied portfolio is currently below its *maximum* potential return, whereas a ratio equal to one implies that the portfolio reached its higher return level, according to the given investor's risk profile. Then, the closer to one the Cantaluppi and Hug (2000) efficiency ratio, the closer to the *maximum* performance the portfolio under analysis.

The Muralidhar (2001) Correlation-Adjusted Performance. In order to compare the management skills of several investors within a *peer* group, Muralidhar (2001) develops a performance measure, called "M^3" or CAP (which stands for Correlation-Adjusted Performance), which considers their common objectives, both in terms of portfolio total risk and Tracking Error. The author assumes an investor who splits his portfolio p between an investment fund, p_1, a benchmark and the risk-free rate. The M^3 is given as:

$$CAP_p = \left[E\left(r_{p_1}\right) \times w_{p_1}\right] + \left[E\left(r_B\right) \times w_B\right] + \left[r_f \times \left(1 - w_{p_1} - w_B\right)\right], \tag{3.35}$$

with:

$$\begin{cases} w_{p_1} = \left[\sigma_{r_B} \times \left(\sigma_{r_{p_1}}\right)^{-1}\right] \times \left\{\left[1-\rho_\tau^2\right] \times \left[1-\rho_{r_{p_1},r_B}^2\right]^{-1}\right\}^{\frac{1}{2}} \\ w_B = \rho_\tau - w_{p_1} \times \left\{\left[\sigma_{r_{p_1}} \times \sigma_{r_B}^{-1}\right] \times \rho_{r_{p_1},r_B}\right\}, \end{cases}$$

and:

$$\begin{cases} \rho_\tau = 1 - \left[TE_{r_p,r_B} \times \left(2 \times \sigma_{r_B}^2\right)^{-1}\right] \\ \rho_{r_{p_1},r_B} = 1 - \left[TE_{r_{p_1},r_B} \times \left(2 \times \sigma_{r_B}^2\right)^{-1}\right], \end{cases}$$

where $w_{p_1} = [0,1]$ and $w_B = [0,1]$ are, respectively, the weights of the investment fund and the benchmark. Furthermore, $\rho_{r_{p_1},r_B} = [-1,1]$ is the correlation coefficient between the investment fund and the benchmark, and $\rho_\tau = [-1,1]$ is the target correlation coefficient.

Muralidhar (2001) argues that this model enables us to ensure that rankings based on Correlation-Adjusted Performance are identical to those founded on management abilities. While w_{p_2} and $(1 - w_{p_1} - w_{p_2})$ may be greater than or less than zero (negative coefficients being equivalent to shortening the futures contract relating to the benchmark and borrowing at the risk-free rate), w_{p_1} is constrained to be positive. Active and passive management approaches can be analyzed through the amount invested in the benchmark: the higher the value of w_{p_2}, the more active the investment strategy. Then, the calculation of optimal proportions will allow us to determine a Correlation-Adjusted Portfolio characterized by the higher expected return, caused by a low volatility and a high correlation with the benchmark, given the fixed target, and a weak set of correlations with other managed portfolios. A corrected version of Muralidhar (2001) is developed by Muralidhar (2002), named “Skill, History And Risk-ADjusted” which aims to take into account the differences in terms of data history between the studied portfolios. The Skill, History And Risk-ADjusted measure writes:

$$SHARAD_p = CAP_p \times C(S), \tag{3.36}$$

where CAP_p is the Muralidhar (2001) Correlation-Adjusted Portfolio, the degree of “confidence” in the manager’s skills is $C(S) \in [0,1]$ and S is a function of the Information Ratio defined as:

$$S = H^{2^{-1}} \times \left\{IR_p - \left[\left(\sigma_{r_p}^2 - \sigma_{r_B}^2\right) \times \left(2 \times TE_{r_p,r_B}^{1/2}\right)^{-1}\right]\right\}, \tag{3.37}$$

where H is the *maximum* number of returns for a common time horizon when comparing the performance of several portfolios with different lengths of data history, IR_p is the Information Ratio of the portfolio p.

This performance measure shares all the properties of the Muralidhar (2001) "M ^{3} " measure and, in addition, accounts for the data period in such a way that it is more consistent with the manager's skill evaluation. The author defines the "confidence" in the manager's skill, denoted $C(S)$, as the cumulative probability of a return distribution assumed Gaussian for a given quantile S. As an illustration, we can read from the standard normal table that $C(S) = 84.13\%$ when considering the quantile S equal to 1.

The measures reported in this subsection differ from Jensen-type ones, since they cannot be directly expressed as a comparison between observed and theoretical performances. The contributions of these approaches are numerous, but their application displays other limits. Henriksson and Merton (1981) develop a free-parameter model by considering managers' correct and incorrect forecasts. The accuracy of this methodology relies on the exact knowledge of all positions of the portfolio manager. Similarly, Fama (1972), Moses et al. (1987) compare the compensations earned by the manager, given his exposure to the systematic and specific risks. Nevertheless, the structure of the measure is rather *ad-hoc*, and it disregards the correlation level between the managed and market portfolio returns (Muralidhar, 2001) . Grinblatt and Titman (1989) propose to adjust the investor's portfolio excess returns by time period-varying weights to reflect the behaviour of market timers. Yet, rankings are sensitive to the methodology adopted for recovering the weighting scheme. Cornell (1979) compares the performance of managed portfolios between two disjointed time intervals. However, the choice of periods is fundamental (date, frequency and length) and this implies that we are able to recover the prices of all assets in both time windows. Like others, Modigliani and Modigliani (1997) compare the performance reached by the manager to that of a riskless asset, both adjusted by the *surplus* of portfolio total risk, relative to the market *proxy*. However, this methodology heavily relies on the assumptions of the Mean-Variance framework.

4

General Performance Measures explicitly based on the Return Distribution

Summary. Developed from the end of the 90s, the third family includes measures based on some general features of the return distribution. In general terms, this class of performance measures can be written in the following form:

$$PM_p = \mathscr{P}^+\left(r_p\right) \times \left[\mathscr{P}^-\left(r_p\right)\right]^{-1}, \tag{4.1}$$

where $\mathscr{P}^+(\cdot)$ and $\mathscr{P}^-(\cdot)$ relate to a specific (respectively right and left) part of the support of the returns density. Measures that belong to this third family are based on features of the return distribution other than the first two moments or some quantiles. However, some of them might be viewed as simple extensions of relative measures presented in the first section. Most of the following density-based Performance Measures can be defined as the ratio of two Generalized Partial Moments (GPM for short). We thus introduce, hereafter, a new generalized function, denoted $\mathscr{H}_p(.)$, corresponding to the ratio of a GHPM onto a GLPM, both raised to two different powers, namely $1/k_1$ and $1/k_2$, adjusted with a term in which appears the cumulative density function of the returns:

$$\begin{aligned} PM_p &= \mathscr{H}_p\left(r_p, \tau_1, \tau_2, \tau_3, \tau_4, o_1, o_2, k_1, k_2, k_3, k_4\right) \\ &= \frac{\left[1 - F_{r_p}(\tau_3)\right]^{k_3}}{F_{r_p}(\tau_4)^{k_4}} \times \left[E\left(\left|\tau_1 - r_p\right|^{o_1} \middle| r_p > \tau_3\right)\right]^{(k_1)^{-1}} \\ &\qquad \times \left\{\left[E\left(\left|\tau_2 - r_p\right|^{o_2} \middle| r_p < \tau_4\right)\right]^{(k_2)^{-1}}\right\}^{-1}, \end{aligned} \tag{4.2}$$

Sign - has been suppressed in both terms of the H function. The cumulative density function is here to respect the conditionnal expectation definition. where τ_1 corresponds to a threshold (a reserve return $\underline{r}$, a MAR, the null return or r_f...) for computing gains and τ_2 for calculating losses or risk, $\tau_3 = VaR_{-r_p,a_1}$ is another threshold – related to a given confidence level a_1 – specifying the right part of the support of the return density under study (*i.e.* gains), $\tau_4 = VaR_{r_p,a_2}$ is a last threshold – linked to another given confidence level a_2 – associated with the left part (*i.e.* losses), o_1 and o_2 are intensification constants reflecting the investor's attitude towards gains and losses, k_1,k_2,k_3, and k_4 are normalizing constants strictly positive. This notation can be easily applied to several relative performance measures. *The examples seem wrong.These ratios can't be put under H(....) format because they could be negative whereas $\mathscr{H}_p(.)$ is strictly positive, So I've deleted this footnote.* [1] Generalized performance measures are evaluated by taking ratios of some ranges of positive and negative portfolio excess returns. In other words, compared to relative performance measures, the following quantities differ both in their numerator (expected return, as in the previous case) and in their denominator (risk measure). The first measure we consider is the Bernardo-Ledoit (2000) Gain-Loss ratio, which penalizes the over-performance (gains – in the numerator) realized by the manager by his under-performance (losses – in the denominator). Another well-known performance measure is the Keating-Shadwick (2002) *Omega* measure that extends the Bernardo-Ledoit (2000) measure with a unique general threshold. The introduction of these two density-based measures is the starting point of many further developments. After introducing the two previously cited measures, we focus on specific performance measures that are directly based on a comparison between some upside and downside sequences of portfolio returns (measured by Partial Moments). Finally, we present the most generalized measures of this family, namely the Biglova et al. (2004) Generalized Rachev Ratio.

[1] For instance, the Sharpe (1966) ratio, the Reward-to-CVaR (2003) ratio, the Sortino-Meer (1991) ratio and the Information (1994) Ratio can be written as: $S_p = \mathscr{H}_p\left(r_p, r_f, -\infty, E\left(r_p\right), +\infty, 1, 1, 2, 2\right)$, $RCVaR_p = \mathscr{H}_p\left(r_p, r_f, -\infty, 0, VaR_{r_p,a}, 1, 1, 1, 1\right)$, $SM_p = \mathscr{H}_p\left(r_p, MAR, -\infty, MAR, MAR, 1, 2, 1, 2\right)$ and $IR_p = \mathscr{H}_p\left(r_p, r_B, -\infty, r_B, +\infty, 1, 2, 1, 2\right)$.

4.1 The Bernardo-Ledoit (2000) Gain-Loss Ratio and the Keating-Shadwick (2002) Measure

Bernardo and Ledoit (2000) build a performance measure, the "Gain-Loss ratio", which aims at evaluating the attractiveness of an investment opportunity by considering specific parts of the portfolio return density. This measure is defined as the expected positive excess return of a portfolio divided by (the opposite of) its expected negative excess return. Using Generalized Higher/Lower Partial Moments, the Gain-Loss ratio writes as:

$$\begin{aligned} GL_p &= \left[GHPM_{r_p,r_f,r_f,1}\right] \times \left[GLPM_{r_p,r_f,r_f,1}\right]^{-1} \\ &= \mathscr{H}_p\left(r_p, r_f, r_f, r_f, r_f, 1, 1, 1, 1, 1, 1\right). \end{aligned} \tag{4.3}$$

The formula has been changed to fit $\mathscr{H}_p(.)$ new definition. The approach of Bernardo and Ledoit (2000) was later generalized by Keating and Shadwick (2002) with their *Omega* measure. The *Omega* is a simple generalization of the Gain-Loss ratio, and is obtained by relaxing the threshold, which is not constrained to be equal to the risk-free return, such as:

$$\begin{aligned} O_p &= \left[GHPM_{r_p,\tau,\tau,1}\right] \times \left[GLPM_{r_p,\tau,\tau,1}\right]^{-1} \\ &= \mathscr{H}_p\left(r_p, \tau, \tau, \tau, \tau, 1, 1, 1, 1, 1, 1\right). \end{aligned} \tag{4.4}$$

The *Omega* ratio separately gauges favourable and unfavourable excess returns, with respect to a threshold τ that has to be defined.

The performance measurement approaches just mentioned go beyond the study of the first moments, and consider additional features of the return distribution. However, this methodology raises a few issues. Evaluations of rewards and losses are directly linked to the characteristics of underlying densities. They are, thus, subject to usual misspecification and estimation problems. For instance, the Gain-Loss ratio is very sensitive to the presence of outliers because it focuses, by definition, on the extreme positive and negative excess returns. Moreover, this performance measure assumes a constant return reference, generally associated to a target return. However, as mentioned before, an error on the chosen threshold can significantly influence the accuracy of resulting values. Furthermore, the impact of volatility on the final rankings based on these measures is rather uncertain in some cases.

4.2 Upside / Downside-related Performance Measures

These performance measures are directly based on the two representative ratios of this family, namely the Gain-Loss ratio and the *Omega* measure, although they can be seen as improved forms.

The Gemmill-Hwang-Salmon (2006) Loss-Averse Performance Measures. Gemmill et al. (2006) suggest two Loss-Averse Performances, respectively denoted by "LAP_p^S" ("S" stands for "simple") and by "LAP_p^H" (where "H" refers to the "House-money" effect). The "House-money" effect refers to the behavior of investors who are often willing to take more risk when they have experienced gains in the past. These two performance measures are derived within a behavioral finance framework. Indeed, the authors define their model according to the utility (or value) function introduced by Kahneman and Tversky (1979). According to our general function $\mathscr{H}_x(.)$, these two Loss-Averse Performance measures can be simply summarized as:

$$\begin{aligned} LAP_p &= \gamma_p \times \left[GHPM_{r_p,\tau,\tau,o_1}\right] \times \left[GLPM_{r_p,\tau,\tau,o_2}\right]^{-1} \\ &= \gamma_p \times \mathscr{H}_p\left(r_p,\tau,\tau,\tau,\tau,o_1,o_2,1,1,1,1\right), \end{aligned} \tag{4.5}$$

where $\gamma_p = 1$ for LAP_p^S and γ_p is a positive number for LAP_p^H.

The interpretation of the first proposed measure LAP_p^S, when the loss aversion coefficient L_p is set to 1, is similar to the *Omega* measure (exactly equal if the density is symmetrical and the aversion to loss for the investor for this portfolio is equal to 1). Regarding the second one, denoted by LAP_p^H, the time-varying loss aversion coefficient allows the introduction of the "House-money" behavioral effect described in Thaler and Johnson (1990) and Barberis et al. (2001); in fact, past losses realized by the portfolio manager, with respect to the benchmark, affect the current performance assessment of the portfolio in a negative way.

Working within the Prospect Theory framework, Watanabe (2006) proposes a similar measure called "Prospect Ratio" that is defined such as:

$$PR_p = \left\{ \left[E\left(r_p \left| r_p \geq 0\right.\right) + 2.25 \times E\left(r_p \left| r_p \leq 0\right.\right)\right] - MAR \right\} \times E\left[\left(r_p - MAR\right)^2 \middle| r_p - MAR < 0 \right]^{-1/2}. \tag{4.6}$$

The Sortino-Meer-Plantinga (1999) Upside-Potential Ratio. From the use of a Generalized Lower Partial Moment of order 2 for estimating downside risk, Sortino et al. (1999) propose the "Upside-Potential ratio" which compares the managed portfolio returns in excess of a MAR to the root square of the downside deviation of the investor's portfolio returns. This performance measure is written as:

$$\begin{aligned} UP_p &= \left[GHPM_{r_p,MAR,MAR,1}\right] \times \left[GLPM_{r_p,MAR,MAR,2}\right]^{-1/2} \\ &= \mathscr{H}_p\left(r_p, MAR, MAR, MAR, MAR, 1, 2, 1, 2, 1, 1/2\right). \end{aligned} \tag{4.7}$$

Farinelli et al. (2008) provide a generalized performance measure which is defined as the ratio between a Generalized Higher Partial Moment of order o_1 and a Generalized Lower Partial Moment of order o_2. In our notation, the proposed measure is given by:

$$\begin{aligned} FT_p &= \left[GHPM_{r_p,\underline{r},\underline{r},o_1}\right]^{1/o_1} \times \left[GLPM_{r_p,\underline{r},\underline{r},o_2}\right]^{-1/o_2} \\ &= \mathscr{H}_p\left(r_p, \underline{r}, \underline{r}, \underline{r}, \underline{r}, o_1, o_2, o_1, o_2, 1/o_1, 1/o_2\right). \end{aligned} \tag{4.8}$$

This performance measure is a generalized version of the upside *versus* downside deviation measure. The choice of the parameters o_1 and o_2 depends on the investor's preferences: high values of o_1 imply a strong preference for gains, while large values of o_2 correspond to an increase in the aversion to losses; in other words, o_1 (respectively o_2) reflects the investor's feelings about the consequences of being above (respectively below) the threshold defined as a reserve return. If the investor's main concern is simply to fall below this threshold without particular regard to the magnitude, then small values of o_1 and o_2 are appropriate. On the contrary, large values of o_2 are required to penalize strong negative deviations from the reserve return.

The Biglova-Ortobelli-Rachev-Stoyanov (2004) Ratios. Biglova et al. (2004) develop two performance measures, named "Rachev ratio" and "Generalized Rachev ratio", which are directly linked to the Bernardo-Ledoit (2000) Gain-Loss ratio by focusing on two specific parts of the return density. These two measures aim at proposing a robust *criterion* when the investor's portfolio return distribution is heavy-tailed. The Rachev Ratio is defined as:

$$\begin{aligned} RR_p &= \left[-ES_{(-r_p),r_f,\tau_3}\right] \times \left[-ES_{r_p,r_f,\tau_4}\right]^{-1} \\ &= \mathscr{H}_p\left(r_p,r_f,r_f,r_f+\tau_3,r_f+\tau_4,o_1,o_2,1,1,0,0\right), \end{aligned} \tag{4.9}$$

whilst the Generalized Rachev Ratio[2] is given by:

$$\begin{aligned} GRR_p &= \left[-PES_{(-r_p),r_f,\tau_3,o_1}\right]^{1/o_1} \times \left[-PES_{r_p,r_f,\tau_4,o_2}\right]^{-1/o_2} \\ &= \mathscr{H}_p\left(r_p,r_f,r_f,r_f+\tau_3,r_f+\tau_4,o_1,o_2,o_1,o_2,0,0\right). \end{aligned} \tag{4.10}$$

These two performance measures allow investors to assess the frequency and impact of extreme events and then incorporate this risk feature when gauging the portfolio performance. The Rachev ratio can be seen as a special case of the Generalized Rachev ratio which proposes to characterize, through the coefficients o_1 and o_2, the investor's attitude towards risk. Following the same analysis, Ortobelli et al. (2010) propose two other performance measures based on Drawups-Drawdowns, where Drawups are defined similarly to Drawdowns, but focusing on positive returns. The first ratio, called the "Rachev Average Drawup-Drawdown ratio", is computed as the average Drawup of the investor's portfolio returns over its average Drawdown. The second ratio is the "Rachev *Maximum* Drawup-Drawdown ratio", using the

[2] The original definition of the Generalized Ratchev Ratio is $GRR_p = \{E[max(r_p - r_f,0)^{o_1}|r_p - r_f > \tau_3]\}^{o_1^{-1}} \times \{\{E[max(r_f - r_p,0)^{o_2}|r_f - r_p < \tau_4]\}^{o_1^{-1}}\}^{-1}$. We assume here that τ_3 is a positive number, typically $VaR_{95\%}$ and τ_4 is a negative number, typically $VaR_{5\%}$, which allows us to write the GRR as in Equation (4.9) , since the (double-)conditioning is now useless (because $\tau_3 > 0$ and $\tau_4 < 0$).

maximum operator instead of the average to compute the portfolio performance.

The Lavinio (2000) d-ratio. Similar to the Bernardo-Ledoit (2000) Gain-Loss ratio and the Keating-Shadwick (2002) *Omega* measure, Lavinio (2000) proposes a performance measure that compares the positive returns of an investor's portfolio with the negative ones. The d-ratio thus reads:

$$dR_p = \left| n^u_{r_p} \sum_{t=1}^{T} \max\left(r_{p,t}, 0\right) \times \left[n^d_{r_p} \sum_{t=1}^{T} \min\left(r_{p,t}, 0\right) \right]^{-1} \right|, \tag{4.11}$$

where $n^u_{r_p}$ and $n^d_{r_p}$ are the numbers of returns, respectively, superior and inferior to zero.

The Kazemi-Schneeweis-Gupta (2004) Sharpe-*Omega* Ratio. Combining, in a sense, the *Omega* measure with the Sharpe ratio, Kazemi et al. (2004) introduce a performance measure, the "Sharpe-*Omega* ratio", whose peculiarity is to compare an expected excess return - as the Sharpe ratio - to the expected value of a put option - as the Omega ratio. More formally, the Kazemi et al. (2004) Sharpe-*Omega* ratio is given by:

$$SO_p = \left[E\left(r_p\right) - \underline{r}\right] \times \left(GLPM_{r_p,\underline{r},\underline{r},1}\right)^{-1}. \tag{4.12}$$

This ratio focuses on the shape of the investor's portfolio return distribution below the threshold $\underline{r}$. The price of the put option is associated with the cost of protecting portfolio returns from large negative deviations (when these are inferior to the threshold). We can consider two main cases. If the investor's portfolio excess return is negative (then $SO_p < 0$), the higher the put price and the better the portfolio performance. Indeed, high volatility will increase the put price and the value of the Sharpe-*Omega* ratio. If the result is positive, the higher the put price and the worse the portfolio performance. Unlike the first case, high volatility will increase the put price and reduce the score of the Sharpe-*Omega* ratio.

Among these different improvements which aim at overcoming the main limits of the two representative ratios, namely the Gain-Loss ratio (Bernardo and Ledoit, 2000) and the *Omega* measure (Keating and Shadwick, 2002) , Farinelli et al. (2008) present the most generalized expression when varying the thresholds and the powers applied to the managed portfolio excess returns. Nevertheless, they share some of the same limits as other measures also based on *VaR*. Biglova et al. (2004) respectively estimate rewards and losses through GHPM of the order o_1 and GLPM of the order o_2. However, these parameters should be linked to the investor's attitude towards performance and risk, which have to be further established and are difficult to estimate.

5

Performance Measures directly derived from Utility Functions

Summary. The main feature of measures in this fourth family is to be straightly obtained from the study of general and explicit utility functions. This kind of PM can be summarized using the following form:

$$PM_p = \mathscr{G}\left\{E\left[U\left(r_p - \tau\right)\right]\right\}, \tag{5.1}$$

where $U(\cdot)$ is a value (or utility) function and $\mathscr{G}(\cdot)$ is a specific function that depends upon the performance of the investor's portfolio. Then, the main purpose of these measures is to explicitly incorporate the investor's preferences and risk profiles, through representative utility functions. This fourth family of measures, expressed *per* unit of marginal utility, is represented by the Morningstar (2002) Risk-Adjusted Return. This measure aims at assessing the portfolio performance by considering a fixed investor's risk-aversion coefficient. Two distinct sub-functions can be defined based on a Power Utility Function that depends upon the value of this risk-aversion coefficient. Following the same approach adopted for the MRAR, several measures (see, for instance, Stutzer, 2000; Kaplan, 2005; Goetzmann et al., 2007) have been proposed in the literature. Hereafter, we first make a brief description of the most used utility-based performance measure, namely the Morningstar (2002) Risk-Adjusted Return. Secondly, we group together measures derived from the same principle.

5.1 The Morningstar (2002) Risk-Adjusted Return

Morningstar (2002) develops a performance measure, named “Morningstar Risk-Adjusted Return” (MRAR), which is built on the Expected Utility theory and considers a Power Utility function. The MRAR is defined as the expected value of the certainty equivalent annualized geometric return on a given horizon[1]. The Morningstar (2002) Risk-Adjusted Return is given by:

$$MRAR_{A,p} = \begin{cases} E\left[\left(1+r_p\right)^{-A}\right]^{\frac{-12}{A}} - 1, & A > -1, A \neq 0 \\ \exp\left\{E\left[\ln\left(1+r_p\right)\right]\right\} - 1, & A = 0 \end{cases} \tag{5.2}$$

where in the Morningstar rating system, portfolio returns are adjusted for management fees, taxes and are expressed in deviation from the risk-free rate.

We can notice that the risk aversion coefficient is equal to 2 by Morningstar. Sharma (2004) suggests a risk aversion coefficient of 3, while Aït-Sahalia et al. (2004) estimate a coefficient of 2.20 for Ultra High Net Worth individuals, raising some doubts on the value used by Morningstar. This is particularly relevant given that the Morningstar rating system is mostly used by retail (non Ultra High Net Worth) investors[2].

The MRAR aims at “predicting” over- and under-performing funds. The fund ranking published by Morningstar is represented by “stars”, from one to five, with a five star evaluation being the best.

Most of the measures collected in the first three families consider, in some ways, the investor’s utility function to evaluate performance, but the link is sometimes not explicit. On the contrary, this is the case for the last family of measures. To our knowledge, the MRAR is probably one of the most followed by investors (see also Stutzer, 2005 , for a detailed study). However, it presents two major limits. The author refers to a Power Utility Function for characterizing the behaviour of investors, which displays an unrealistic Constant Relative Risk Aversion coefficient over time. Furthermore, in the Morningstar rating system this measure assumes a risk aversion coefficient equal to 2. Yet, this value is questioned since it is shown to vary according to the main investor’s risk profiles (see Lisi and Caporin, 2012) and to market

[1] See also Pézier (2010 and 2012) for more details about the use of the certainty equivalent when evaluating the performance of investment portfolios.

[2] see Lisi and Caporin (2012) for further comments on this aspect.

conditions (see Li, 2007; Coudert and Gex, 2008) . Lastly, in a dynamic setting, an investor using the MRAR *criterion*, can expose his portfolio to a simple *momentum* effect.

5.2 Other Utility-based Performance Measures

We present hereafter measures that are similar to the Morningstar (2002) Risk-Adjusted Return, but rely on different families of utility functions to characterize the behaviour of the final investor.

The Stutzer (2000) Performance Index. Moving from a behavioral analysis framework, Stutzer (2000) develops a measure, named "Performance Index" which considers the investor's degree of risk aversion when they are characterized by an Exponential Utility function. The proposed index is thus written:

$$PI_p = -\log\left\{\exp\left[-A \times E\left(r_p - r_B\right)\right]\right\}. \tag{5.3}$$

The Performance Index can be interpreted as the (decay) rate at which the probability that the managed portfolio underperforms his benchmark declines over time.

We can notice that when portfolio return distributions are i.i.d., Stutzer (2000) defines the loss probability of a portfolio manager such as:

$$Prob\left(\overline{r_{p,t}} - \overline{r_{B,t}} \leq 0\right) \simeq \left(A \times T^{-.5}\right) \times \exp\left(-PI_p \times t\right), \tag{5.4}$$

where T is the number of observations and PI_p is the decay rate, which corresponds to the "Performance Index".

In other words, the higher the value of the Performance Index, the closer to the null value is the loss probability of the studied portfolio. When the returns distributions is Gaussian, Stutzer (2000) shows that this performance measure is equivalent to the half of the squared Sharpe ratio. In the absence of normality, rankings obtained with the proposed index will take into account the degree of preference for positive skewness.

The Kaplan (2005)*Lambda* Measure. Combining the features of the Kaplan-Knowles (2004) *Kappa* measure and the Stutzer (2000) Performance Index, Kaplan (2005) presents a performance measure, named "*Lambda*", which is based on a generalized Fishburn utility function and a loss penalty function. Kaplan (2005) defines the Generalized Fishburn Utility Function as:

$$U\left(r_p - r_B\right) = \left(r_p - r_B\right) - l\left[\max\left(r_B - r_p, 0\right)\right], \tag{5.5}$$

where $l(X)$ designs the loss penalty function defined in equation (5.7).

Another version of the penalty function is suggested by Watanabe (2006) who defines it as:

$$l\left(r_p\right) = E\left[\max\left(r_p, 0\right) + 2.25 \times \min\left(r_p, 0\right)\right]. \tag{5.6}$$

Moreover, Kaplan (2005) defines a new class of utility functions named "Proportional Risk Aversion" in which the investor-specific risk aversion parameter multiplies the active return in the utility function. The Kaplan (2005) *Lambda* measure can be obtained by solving the following maximization problem:

$$\Lambda_p = \left[-A \times E\left(r_p - r_B\right)\right] - l\left(\max\left(r_B - r_p, 0\right)\right), \tag{5.7}$$

where $l(X) = \exp(X) - X - 1$ is the loss penalty function.

The main innovation of this performance measure is to propose a variant of the Stutzer (2000) Performance Index, penalizing the negative excess returns through $l_p(\max(r_B - r_p, 0))$. In other words, the Kaplan (2005) *Lambda* measure can be associated to a downside version of the Stutzer (2000) Performance Index, whose main objective is to correct the positive excess return of the manager's portfolio by past drawdowns.

The Goetzmann-Ingersoll-Spiegel-Welch (2007) Manipulation-Proof Performance Measure. Goetzmann et al. (2007) develop an "ungamable" performance measure, named "Manipulation-Proof Performance Measure" (MPPM for short), to gauge the performance of an active manager[3]. It is based on the risk aversion of investors, and is given by:

[3] This can be viewed as a generalization of the Morningstar (2002) Risk-Adjusted Return since $\Theta_{A,p} = ln\left[1 + MRAR_{A-1,p}\right]$.

$$\Theta_{A,p} = [(1-A)\,\Delta t]^{-1} \times \ln\left\{E\left\{\left[\left(1+r_p\right)\times\left(1+r_f\right)^{-1}\right]^{1-A}\right\}\right\}, \tag{5.8}$$

where the risk aversion satisfies $A > 1$ and Δt is the observation frequency expressed on a yearly basis.

Goetzmann et al. (2007) define manipulation (or gaming) of a performance measure as an action taken to increase a fund's performance measure that does not actually add value for the investors. This measure severely penalizes negative excess returns as the risk aversion coefficient increases. Brown et al. (2010) introduce a derived performance measure, called "Doubt Ratio", which is given by:

$$DR_p = \left[\Theta_{2,p} \times \left(\Theta_{2,p} - \Theta_{3,p}\right)^{-1}\right] + 2. \tag{5.9}$$

The Doubt Ratio is based on differences between MPPM when considering two risk aversion coefficients. Beyond a certain degree of manipulation of portfolio returns, values of MPPMs will tend to be (almost) equal, for any risk aversion coefficients, *i.e.* whatever the type of investor risk profiles. In other words, this *phenomenon* means that the portfolio will have a similar MPPM ranking for all investors, which is doubtedly right.

Brown et al. (2010) thus propose to make the difference between MPPMs computed for two different risk aversion coefficients in order to reflect the possibility of manipulation. In the presence of manipulated return distributions, values of $\Theta_{2,p}$ and $\Theta_{3,p}$ will be close, and thus, the Doubt Ratio will be high, which could be suspicious.

More recently, Joenväärä et al. (2013) propose a conditional version of the MPPM that takes macro-economic information into account.

The Billio-Jannin-Maillet-Pelizzon (2014) Generalized Utility-based N-moment Measure. Following Billio et al. (2011), Billio et al. (2014) propose a flexible measure of performance, named "Generalized Utility-based N-moment measure" (GUN for short), relying on a characterization of the whole return distribution, which is hardly gamable. More precisely, through a Taylor expansion, it takes into account the first four moments of the return distribution and the associated sensitivities of a representative investor, reflecting his preferences and risk profile. It is written such as:

$$GUN_{N,i,p} = \sum_{n=1}^{N} \lambda_{n,i,p} \times m_{n,p}\left(r_p\right), \tag{5.10}$$

where $\lambda_{n,i,p}$ are the sensitivities of a representative investor i regarding the n-th moment defined as:

$$\lambda_{n,i,p} = (-1)^{n-1}(n!)^{-1}\,\omega_{n,i} \times g_{n,i}\left[m_{1,p}(r_p)\right]^{\tau_n}, \tag{5.11}$$

where $n = [1,\ldots,N]$, $n!$ is the n-factorial, τ_n is a power, $\omega_{n,i}$ is a weight and $g_{n,i}(\cdot)$ is a function of the first moment $m_{1,p}(.)$ of the underlying return distribution.

The Smetters-Zhang (2013) General Ranking Measure. Smetters and Zhang (2013) develop a performance measure, named "General Ranking Measure" (GRM for short), which is a specific case of the GUN measure since the sensitivities are defined by the unit roots. More precisely, the GRM is given by:

$$GRM_{N,i,p} = \sum_{n=1}^{N} (n!)^{-1}\lambda^*_{n,i,p} \times C_{n,p} \times (z_N)^n, \tag{5.12}$$

where z_N is the smallest absolute real root z that solves $\sum_{n=1}^{N}[(n-1)!]^{-1}\lambda^*_{n,i,p} \times C_{n,p} \times z^{n-1} = 0$, with $\lambda^*_{n,i,p} = 1$ for $n = 1$, or $\lambda^*_{n,i,p} = (A)\ldots(A+n-2)$ for $n \geq 2$ and, where A denotes the risk aversion coefficient.

As exhibited by the different measures previously mentioned, several improvements have been proposed to solve the main limits of the MRAR, but they are still exposed to some drawbacks. Stutzer (2000) proposes to use an Exponential Utility function since it exhibits a positive relative risk aversion. Kaplan (2005) adds a Utility Function with a penalty to punish the positive excess return of the manager's portfolio by its past losses. Again, these two measures are dependent on the definition of the benchmark and on a raw approximation of the investor's risk aversion coefficient. Finally, Goetzmann et al. (2007) propose a manipulation-proof measure, but it happens in some cases that this measure greatly relies on the first moment of the underlying return distribution.