

PORTFOLIO OPTIMIZATION UNDER A PARTITIONED-BETA MODEL

RICCARDO BRAMANTE
Catholic University
Largo Gemelli, 1
20123 MILAN (Italy)
riccardo.bramante@unicatt.it
(Corresponding author)

DIEGO ZAPPA
Catholic University
Largo Gemelli, 1
20123 MILAN (Italy)
diego.zappa@unicatt.it

Abstract

Portfolio optimization has been a highly researched area in finance. Since the seminal work of Markowitz (1959) there had been many advances in portfolio analysis, attempting to combine the conceptual world of scholars with the pragmatic view of practitioners and to couple with increased electronic computing power. Among the proposals, the Capital Asset Pricing Model (*CAPM*) is one of the potential solutions to simplify the calculation of optimal portfolios and to directly relate each stock return to the return referred to a market index. *CAPM* assumes that stock riskiness, which are captured by their market beta, are constant over the domain. However, there exists substantial empirical evidence that this assumption may be inaccurate and hazardous in asset allocation decisions, mainly when the relationship between risk and excess returns in “Bear” and “Bull” markets would be modelled separately.

In this paper we propose the use of a mixture of truncated normal distributions in returns modelling. An optimization algorithm has been developed to obtain the best fit both in the univariate and in the bivariate case. Moreover, the procedure permits to decompose the global beta coefficient into local betas referred to specific regions of the market returns domain. Partitioning the domain provides a set of disjoint conditional regions where the local relationship between portfolio components and the benchmark can be slightly different with respect to the one on the domain as a whole. To appreciate how much closed to reality our proposal is, we provide an empirical analysis referred both to Country and Sector data.

Keywords: Portfolio Optimization, Mixtures of Distributions, Capital Asset Pricing Model

JEL classification: C61 – G32

In applied financial literature a relevant issue is the modelling of the empirical distribution of returns since many decision-making and asset pricing models depend on the assumptions related to the stochastic model underlying the data. Research on probability models in finance has given rise to several works (see e.g. Ruppert, 2011) but many questions are open:

- a) Are the models supported by financial markets data?
- b) How are the parameters in these models estimated?
- c) Can the models be simplified?

Quoting the very famous fact by George Box “All models are false but some models are useful”, it can generally be agreed that complex models may be closer to reality but often involve many parameters and are not easy to be interpreted; on the other hand, too simple models may not capture important features of the data and can lead to serious bias.

Starting from these preliminaries, we made some considerations on how and whether the normal distribution might be exploited in financial modelling (Bramante, Zappa, 2012). In particular we have investigated how to improve the fitting to the sample data by exploiting the so called minimum distance approach (*MDA*) (Parr, 1985 , Basu et al, 2011), partitioning the domain and interpolating the observed distribution by a sequence of truncated normal distributions. The main difference of *MDA* with respect to the *ML* approach is that the emphasis in *MDA* is on the accuracy of the fitting, i.e. how much able is the estimated distribution to replicate the empirical one. It is then a method strictly closed to a data-mining context where the aim is to describe the observed data by considering them as if they represent the entire population. By contrast the main focus of *ML* is on searching the estimators of the parameters with good statistical properties based on sample data in order to obtain a distribution able to represent at the best the unknown population.

It is obvious to say that the normal distribution may be considered too simple in fitting returns, but it is far to be excluded in the practice. In fact, many practitioners still make extensive use of normal distribution for returns modelling even if different approaches are described in a very rich literature (McNeil et al., 2005 for an extensive review). Our approach seems to be a good compromise between theory and practice, since encouraging results

have been obtained both in the direction of the accuracy of the fitting and in the easy interpretation of the parameters of the final distribution.

In this paper we extend the proposal in Bramante, Zappa (2012) to a bivariate context, exploring in particular how and when the global beta coefficient may be conditionally decomposed into local betas, allowing a different interpretation of dependence among returns; moreover, we describe how results may be used to set up a quite flexible procedure which permits to optimize portfolio allocation under the common market – neutral strategy assumption to eliminate portfolio’s response to market movements and to allow to potentially benefit from both undervalued and overvalued securities.

I. Methodology

A. Fitting Bivariate Distributions: notation and preliminary exploratory results

Let us fix some preliminary ideas and notation.

Let $\mathbf{Z} = (X, Y)$ be a bivariate random variable and let $\hat{F}_n(\mathbf{z}), F_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\theta})$ be the empirical and the cumulative distribution function (*ECDF* and *CDF*) of (X, Y) respectively. Let $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ be a sample drawn from \mathbf{Z} . If $F_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\theta})$ describes the random nature of \mathbf{Z} , we expect that the bivariate *QQ*-plot of

$$\{\mathbf{z}_i, F_{\mathbf{Z}}^{-1}(\hat{F}_n(\mathbf{z}_i); \boldsymbol{\theta})\} \quad \text{for } i = 1, 2, \dots, n$$

or equivalently the distance

$$\|\mathbf{z}_i - F_{\mathbf{Z}}^{-1}(\hat{F}_n(\mathbf{z}_i); \boldsymbol{\theta})\|_p \quad \text{for } i = 1, 2, \dots, n \quad (1)$$

for some $p > 0$, is closed to zero $\forall i$. If $\{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ comes from a distribution \mathbf{W} different from the one we have chosen we expect that the locus of points is locally different from zero. In an analogous manner, the analysis of the *PP*-plot over the domain $[0, 1]^2$ may be considered, i.e. the plot of

$$\{\hat{F}_n(\mathbf{z}_i), F_{\mathbf{Z}}(\mathbf{z}_i; \boldsymbol{\theta})\} \quad \text{for } i = 1, 2, \dots, n \quad .$$

Parameter estimation is typically based on standard maximum likelihood (*ML*) or by robust estimation procedures, i.e. the median and the median absolute deviation from median (*MAD*) (Ruppert, 2011). Differently from *ML*, in order to let the fitting process as flexible and maximally data dependent as possible, the *MDA* approach (see Basu et al. 2011) is becoming popular in applications also because of its theoretical implications. It consists in solving the general unconstrained problem

$$\min_{\theta} d\left(\hat{F}_n(\mathbf{z}), F_{\mathbf{Z}}(\mathbf{z}; \theta)\right) \quad \mathbf{Z} \in \mathbb{R}^2 \quad (2)$$

where $d(\cdot)$ is an appropriate measure of discrepancy (or loss function)¹. If $F_{\mathbf{Z}}(\mathbf{z}; \theta)$ is the “true” distribution then the unconstrained estimator $\hat{\theta}$ minimizing (2) has been shown to be strongly consistent.

Depending on which $d(\cdot)$ is used, further properties, e.g. robustness to extreme influence values, may be defined in addition. Let $A(\mathbf{z})$ and $B_{\theta}(\mathbf{z})$ be continuous functions. Examples of $d(\cdot)$ are

$$d(A(\mathbf{z}), B_{\theta}(\mathbf{z})) := \begin{cases} KS: & \sup |A(\mathbf{z}) - B_{\theta}(\mathbf{z})| \\ MH: & E[|A(\mathbf{z}) - B_{\theta}(\mathbf{z})|] \\ MQ: & E\left[(A(\mathbf{z}) - B_{\theta}(\mathbf{z}))^2\right] \end{cases} \quad (3)$$

known in the literature as the Kolgomorov, Manhattan, Euclidean (Cramer – von Mises) distances, respectively. To keep this fact into consideration, object (2) can be generalized as follows

$$\min_{\theta} \sum_i d\left([\hat{F}_n(\mathbf{z}_i)]^q, [F_{\mathbf{Z}}(\mathbf{z}_i; \theta)]^q\right)^{\frac{1}{q}} \|\mathbf{z}_i\|_p \quad \mathbf{Z} \in \mathbb{R}^2 \quad (4)$$

where $\mathbf{z}_i = [x_i \ y_i]'$, $q > 0$, $\|\mathbf{z}_i\|_p = (\sum_i |z_i|^p)^{1/p}$ with $p > 0$. That means we may solve (4) by searching also for that powers (q, p) – both for the distance and for the norms – that at the best guarantee a good fit to the distribution. In this paper – to avoid complexity in the comprehension of the proposal – we will use $q=1$ and $p=2$ to reconcile notation to the definitions in (3) and the norm of the weighs to the standard Euclidean distance of the vector \mathbf{z}_i from the origin.

¹ It can be easily noticed that (2) is a transformation of the quantities used in the *PP*-plot.

B. Fitting Bivariate Truncated Normal Distributions

Optimization in (4) can be applied also to a partition of \mathbf{Z} e.g. by assuming $F_{\mathbf{Z}}(\mathbf{z}_i; \boldsymbol{\theta})$ to be represented by a sequence of truncated distributions, also called spliced distributions. If the distribution is locally different from the one fitted over the whole domain, we expect that the sum of the losses measured on each partition is less than the one obtained by using the whole domain. This may be interpreted as an evidence that the underlying process is, e.g., over dispersed or it may be considered as a mixture of distributions (see also Kon, 1984). If the opposite happens, it should be an evidence that locally the process is not different from the unconditional one.

In the case-study of dependence of stock returns on market return the search of spliced distribution to fit the bivariate distribution of market returns (X) and stock returns (Y) in *CAPM* modelling may be done by conditioning on a partition of the return market index. In general we expect that the dependence is different if we consider e.g. the leftmost or the rightmost region of the return space, as it is often heuristically observed. The matter is how to define an “optimal” partition by keeping invariant the structure of the dependence among the stocks. A solution can be found by applying (4) in a two stage process: in the first step we estimate the parameters of the bivariate distribution (X, Y) and then, by keeping fixed these estimates, we look for a partition, if it exists, such that the optimum obtained in the first step is improved².

Let $\mathbf{Z} = (X, Y) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e. \mathbf{Z} distributed as a bivariate normal distribution and let $F_{\mathbf{Z}}(\mathbf{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be its cumulative distribution function where $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ are the mean vector and the and variance-covariance matrix, respectively. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ be the estimates of $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ obtained by solving (4). Let ${}_X tr_i$ be a generic threshold for the marginal X such that, for any ${}_X tr_{i-1} < {}_X tr_i$ and for $i = 1, \dots, K$

$$\begin{aligned} \bigcup_{i=1}^K ({}_X tr_{i-1} \dashv {}_X tr_i) &= \mathbb{R}^2 \\ ({}_X tr_{i-1} \dashv {}_X tr_i) \cap ({}_X tr_{j-1} \dashv {}_X tr_j) &= \emptyset \text{ for } i \neq j \end{aligned} \quad (5)$$

² An alternative solution, which is a fairly common approach, is the fitting a piecewise regression model, searching for those knots that significantly change the estimates of the coefficients or produce an improvement to the considered diagnostics.

with ${}_x tr_0 = -\infty$ and ${}_x tr_K = +\infty$. Let

$${}_T f_Z(\mathbf{z}; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}}; {}_x tr_{i-1}; {}_x tr_i) = \begin{cases} \frac{\exp\left(-\frac{1}{2}(\mathbf{z} - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\right)}{\int \left(\int_{{}_x tr_{i-1}}^{{}_x tr_i} \exp\left(-\frac{1}{2}(\mathbf{z} - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\right) dx \right) dy} & \text{for } \mathbf{z} \in (x \in {}_x tr_{i-1} \rightarrow {}_x tr_i) \cap (y \in \mathbb{R}) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

be the truncated normal pdf and let the CDF in \mathbf{z} be³

$${}_T F_Z(\mathbf{z}; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}}) = \sum_{j=1}^i \left(\iint_{-\infty}^{\mathbf{z}} {}_T f_Z(\mathbf{z}; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}}; {}_x tr_{j-1}; {}_x tr_j) d\mathbf{z} \right) w_j \quad (7)$$

with $w_i > 0 \forall i$ and $\sum_{i=1}^{K-1} w_i = 1 - w_K$

The estimate of the weights and thresholds in (7) is obtained by solving

$$\begin{aligned} & \text{Find } w_1, \dots, w_{K-1} \text{ and } tr_1, \dots, tr_{K-1} : \\ & \sum_i d\left(\hat{F}_n(\mathbf{z}_i), F_Z(\mathbf{z}_i; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}})\right) \|\mathbf{z}_i\|_2 - \min_{w,K} \sum_i d\left(\hat{F}_n(\mathbf{z}_i), {}_T F_Z(\mathbf{z}_i; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}})\right) \|\mathbf{z}_i\|_2 \geq \varepsilon \quad (8) \end{aligned}$$

with $\varepsilon > 0$. It is the same problem stated in (4) but where the unknowns are not the parameters of the bivariate Gaussian distribution but the weights and the thresholds. Observe that we look for the smallest partition such that (8) is fulfilled, since for $K \rightarrow \infty$ the truncated distribution degenerates on the single observation. To exemplify suppose $X =$ ‘‘MSCI World’’ index, $Y =$ ‘‘MSCI Italy’’ index. In Figure 1 the joint distribution has been splitted into 3 truncated distributions (see Table 1 for a comparison of *ML* and *MDA* bivariate normal parameters and Table 2 for thresholds and local betas estimates).

Table 1: MDA results

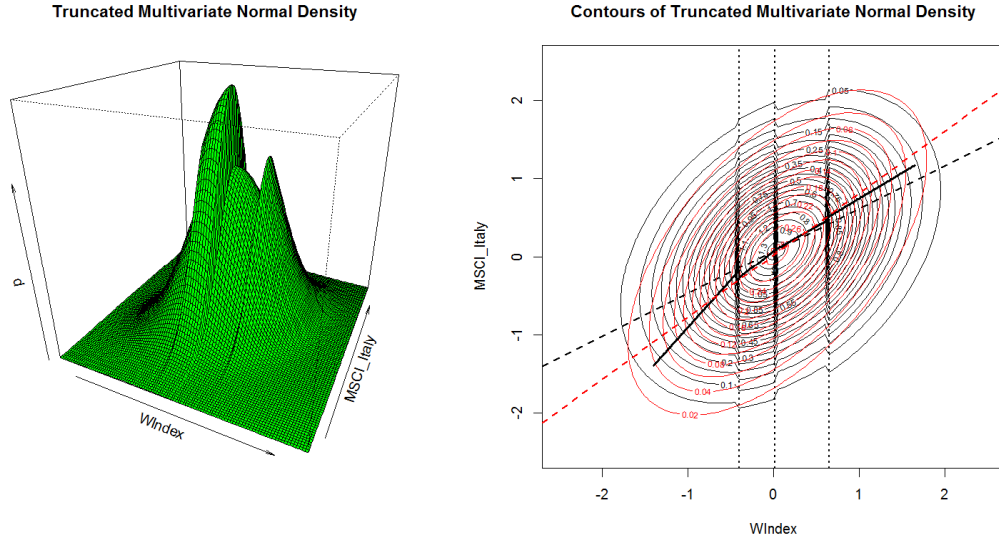
Parameter	MSCI Italy		MSCI World	
	<i>ML</i>	<i>MDA</i>	<i>ML</i>	<i>MDA</i>
Mean	0.14467	0.06226	0.12000	1.08068
Standard Deviation	1.20489	0.67059	0.08387	0.55785
Correlation	0.429542			
Beta	0.771781			

³ From [7], ${}_T F_Z(\mathbf{z}; \hat{\boldsymbol{\mu}}; \hat{\boldsymbol{\Sigma}})$ may also be interpreted as a weighted sum of disjoint truncated distributions.

Table 2: Threshold results

	<i>Alfa</i>	<i>Beta</i>	<i>Freq.</i>
$-\infty \dashv -0.51713$	0.03570	0.69629	15%
$-0.51713 \dashv 0.77926$	0.09797	0.74177	75%
$0.77926 \dashv +\infty$	-0.67063	1.40531	11%

Figure 1: MSCI Italy Index Bivariate Truncated Normal Distribution



The picture on the left reports the contours of the joint distribution and the regression line with parameters estimated by *ML* (red dashed line), and *MDA* (black dashed line) and the regression lines with parameters estimated over each subdomain (black solid line). These lines show that the beta coefficient estimated over the most negative returns is larger than the one estimated using all the data while it is somehow similar when returns are around zero and slightly less than the beta obtained using *ML* over all the data. Similar comments can be made when considering other MSCI indices, too.

The same methodology applied to each component of a portfolio may be exploited to define optimization strategies different from the standard mean-variance optimal criteria in order to improve portfolio risk control even conditional to one or more subregions.

C. Decomposition of the beta coefficient

In the previous Section we have seen that, once the domain X has been partitioned, it is possible to estimate local betas that represent relative risk to the market in each subset.

Consider a portfolio with N assets. According to the Capital Asset Pricing Model (CAPM) the i -th asset return, for $i = 1, 2, \dots, N$, is described by

$$r_i = \alpha_i + \beta_i \cdot r_B + \varepsilon_i$$

where r_i and r_B are the returns of the i -th asset and the benchmark respectively, α_i and β_i are the standard two primary components of investment decisions, i.e. the part of each asset return uncorrelated and correlated with the market. Given a countable partition set of B , the same relation holds conditional to each partition, that is

$$r_{i,k} = \alpha_{i,k} + \beta_{i,k} \cdot r_{B,k} + \varepsilon_{i,k}$$

where $r_{i,k}$ and $r_{B,k}$ are the returns of the i -th asset and the benchmark within the k -th partition, for $k = 1, 2, \dots, K$, and $\alpha_{i,k}$ and $\beta_{i,k}$ are the corresponding alpha and beta parameters. Then the beta coefficient estimated without conditioning over a set of the partition, may be written as follows. Using *OLS*, $\hat{\beta}_i$ estimate can be expressed as the ratio of the codeviance between the i -th asset returns and market index returns and at the denominator the deviance of market index returns.

Let \bar{r}_i, \bar{r}_B be the average returns. Then $\hat{\beta}_i$ may be written as

$$\begin{aligned} \hat{\beta}_i &= \frac{\text{Codev}(r_i, r_B)}{\text{Dev}(r_B)} = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} r_{i,k} \cdot r_{B,k} - N \cdot \bar{r}_i \cdot \bar{r}_B}{\text{Dev}(r_B)} = \\ &= \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} (r_{i,k} \cdot r_{B,k} - \bar{r}_{i,k} \cdot \bar{r}_{B,k}) + \sum_{i=1}^{n_k} (\bar{r}_{i,k} \cdot \bar{r}_{B,k} - \bar{r}_i \cdot \bar{r}_B)}{\text{Dev}(r_B)} \end{aligned}$$

where n_k is the size of the sample and $\bar{r}_{i,k}, \bar{r}_{B,k}$ the corresponding average returns within the k -th partition. Rearranging terms gives:

$$\hat{\beta}_i = \sum_{k=1}^K \hat{\beta}_{i,k} \cdot \frac{\text{Dev}(r_{B,k})}{\text{Dev}(r_B)} + \hat{\beta}_{i,M} \cdot \frac{\text{Dev}(\bar{r}_{B,K})}{\text{Dev}(r_B)}$$

where

$$\begin{aligned} \hat{\beta}_{i,M} &= \frac{\sum_{i=1}^{n_k} (\bar{r}_{i,k} \cdot \bar{r}_{B,k} - \bar{r}_i \cdot \bar{r}_B)}{\text{Dev}(\bar{r}_{B,k})} \\ \text{Dev}(r_B) &= \sum_{k=1}^K \text{Dev}(r_{B,k}) + \text{Dev}(\bar{r}_{B,K}) \end{aligned}$$

The overall beta coefficient $\hat{\beta}_i$ can then be decomposed into a weighted average of all the local betas, with weights given by the fraction of deviance of B within each partition with respect to the overall deviance, plus the estimate of the coefficient $\hat{\beta}_{i,M}$, which is the beta of the linear regression through the means evaluated conditional to the partition, with weight given by the ratio of the deviance between them and the overall deviance. Similarly for the alpha we have

$$\hat{\alpha}_i = \sum_{k=1}^K [\hat{\alpha}_{i,k} + \bar{r}_{i,k}(\hat{\beta}_{i,k} - \hat{\beta}_i)] \cdot \frac{n_k}{N}$$

Recall that the beta of the portfolio (β_p) is generally expressed as a weighted average of the betas of the individual assets in the portfolio (see Elton et al., 2006). When partitions are available, $\hat{\beta}_p$ can be written as follows

$$\hat{\beta}_p = \sum_{j=1}^N \hat{\beta}_j \cdot w_{j,p} = \sum_{j=1}^N \sum_{k=1}^K \hat{\beta}_{j,k} \cdot \frac{Dev(r_{B,k})}{Dev(r_B)} \cdot w_{j,p} + \sum_{j=1}^N \hat{\beta}_{j,M} \cdot w_{j,p} \cdot \frac{Dev(\bar{r}_{B,K})}{Dev(r_B)}$$

where $w_{j,p}$ are the weights assigned to the assets in the portfolio.

This setup allows to have another look at portfolio optimization since the optimization problem can be restated locally for each specific region of the domain or by introducing threshold coefficients dependent on the investor relative risk aversion. In the following section we give some insights to the interpretation of beta coefficients when the benchmark domain is decomposed into local partitions and we provide an example to show advantages of local over global optimization.

3. EMPIRICAL RESULTS

Since a goal of the partitioned beta model is to capture portfolio components dynamics during market turmoil, we provide a portfolio optimization illustration under the common condition of a market neutral strategy. The optimal fraction of wealth invested in each index is determined on the basis of a market – neutral strategy which eliminates portfolio's response

to market movements and allows to potentially benefit from both undervalued and overvalued securities.

Market neutrality is achieved by imposing equivalence of the aggregate beta of long and short components, each being a weighted average of the beta coefficients of individual indices⁴. Assume the N components of a portfolio are subdivided in n_1 long portfolio components and $N - n_1$ short positions. Let $i = 1, \dots, n_1$ the set of subscripts representing only the long portfolio components and $-$ for $i > n_1$ – the subscripts for the others. Given the estimated thresholds and the corresponding bivariate domain partitions referred to all the portfolio components, market neutrality can be achieved in three different ways:

- Over the whole domain, by assuming a global zero-beta

$$\sum_{i=1}^{n_1} w_{i,P} \cdot \beta_i = \sum_{i=n_1+1}^N w_{i,P} \cdot \beta_i$$

- Locally, by assuming a local zero-beta

$$\sum_{i=1}^{n_1} w_{i,P} \cdot \beta_{i,\hat{k}} = \sum_{i=n_1+1}^N w_{i,P} \cdot \beta_{i,\hat{k}}$$

where \hat{k} is the chosen partition.

- Over the whole domain, by weighting each partition on the basis of a coefficient of risk aversion

$$\sum_{i=1}^{n_1} w_{i,P} \cdot \sum_{k=1}^K \gamma_k \cdot \hat{\beta}_{i,k} \cdot \frac{Dev(r_{B,k})}{Dev(r_B)} = \sum_{i=n_1+1}^N w_{i,P} \cdot \sum_{k=1}^K \gamma_k \cdot \hat{\beta}_{i,k} \cdot \frac{Dev(r_{B,k})}{Dev(r_B)}$$

where γ_k ($0 < \gamma_k < 1$) is the k partition risk aversion parameter and

$$\sum_{k=1}^K \gamma_k = 1$$

The second and third optimization conditions aim to compute weights so that market neutrality is achieved within a specific region of the domain in order to manage e.g. an accurate

⁴ There are different interpretations for the term “market-neutral”. In the present paper, the market neutrality always means zero beta

hedging of the downside or – while maintaining global neutrality – to give more / less weight to specific partitions according to the investor aversion to one or more return “areas” – typically the ones referred to negative returns.

Since we are dealing with N assets and since the partition conditioning the estimates of the betas is referred to the benchmark, a generalization of the two step procedure described in §B is required.

In the first step, the optimization tool was used to estimate the vector of the $3 \cdot N + 2$ unknown parameters (means, variances and correlations) simultaneously for the N assets potential candidates for inclusion/addition to the portfolio and the benchmark.

Suppose that $\mathbf{Z}_i = (B, Y_i) \sim N_2(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ and let $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for $i = 1, \dots, N$ where Y_i represents the i -th asset and B the market index (benchmark). The procedure starts by solving

$$\min_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N} \sum_{i=1}^N \sum_{j=1}^{m_i} d\left(\hat{F}_n(\mathbf{z}_{ij}), F_{Z_i}(\mathbf{z}_{ij}; \boldsymbol{\theta}_i)\right) \|\mathbf{z}_{ij}\|_2$$

where m_i is the sample size of returns for the i -th portfolio component.

The criterion function measures the asset cumulative weighted squared distance between the empirical and the bivariate normal distribution. In the second step, the optimal truncation thresholds conditional to the benchmark were computed, i.e. analogously to (8) we have looked for a solution to

$$\sum_i \sum_j d\left(\hat{F}_n(\mathbf{z}_{ij}), F_{Z_i}(\mathbf{z}_{ij}; \hat{\boldsymbol{\theta}}_i)\right) \|\mathbf{z}_{ij}\|_2 - \min_{w, K} \sum_i \sum_j d\left(\hat{F}_n(\mathbf{z}_{ij}), {}_T F_{Z_i}(\mathbf{z}_{ij}; \hat{\boldsymbol{\theta}}_i)\right) \|\mathbf{z}_{ij}\|_2 \geq \varepsilon$$

Observe that we are looking for a partition and a system of weights that will be kept fixed for all the components of the portfolio.

Case studies were performed on two different aggregation levels (by geography and by sector) using the returns of 75 Morgan Stanley Capital International (*MSCI*) indices, provided in the country (separately for developing and emerging markets) and sector sub set, and by assuming the World Index returns to be the benchmark. All the indices are denominated in US dollars and cover the period from January 1996 to December 2012. The entire sample

period was divided into pre specified consecutive intervals with a fixed length of 500 observations, then using an automated three-week calendar rebalancing approach where thresholds and weights are dynamically re determined: in the end, about one thousand optimization runs were completed. In table 3 and 4 results regarding the relative gain in the discrepancy measure, with respect to the bivariate normal case, for the two optimization steps, are reported.

Table 3: Optimization results

Type of Index	% Gain in degree of fit	
	First Opt. Step	Second Opt. Step
Developed Markets Index	45.81	5.16
Emerging Markets Index	46.58	3.86
World Sector Index	43.46	6.14

Table 4: Threshold distribution

Number of Threshold	% of Cases	% Gain in degree of fit	
		First Opt. Step	Second Opt. Step
1	3.38	43.90	1.06
2	32.09	45.80	3.11
3	53.33	45.03	6.05
4	11.20	47.22	6.41

Figures in the two tables suggest that the relative gain in the discrepancy measure by using the *MQ* fit with respect to standard *ML* is particularly relevant in the first optimization step (45.5% on average) and results are quite the same in the three considered aggregation levels; as for the mixture of truncated normal distributions, results are less positive since the relative gain, with respect to the first step, is not so high (5% on average); nevertheless, domain partitioning provides a set of disjoint conditional regions where the local relationship between the index and the benchmark can be slightly different with respect to the one on the domain as a whole.

One interesting result is the reverse in sign of the global beta relation within the domain. Table 5 and 6 report, separately for the three types of the performed optimizations and the resulting threshold distribution, local beta sign inversions (from a global positive/negative beta to at least one negative/positive beta in the partitioned domain): sign inversion occurs on average in 10% of the total case, mainly when the global relation is described by a negative beta and is directly related with the number of the estimated thresholds: this provides evidences that relative risk varies conditionally with the benchmark return regions which are bounded by the computed thresholds.

Table 5: Beta parameters sign inversion (aggregation levels)

Type of Index	% Sign Inversion Local Beta	
	% negative	% positive
Developed Markets Index	6.76	57.63
Emerging Markets Index	17.74	42.65
World Sector Index	1.27	1.12

Table 6: Beta parameters sign inversion (number of thresholds)

Number of Threshold	% Sign Inversion Local Beta	
	% negative	% positive
1	4.95	42.11
2	9.24	43.81
3	10.89	44.21
4	15.99	48.33

As a final step, the optimization algorithm was implemented within the three considered available frameworks to solve the long-short portfolio problem under the condition of market neutrality. The initial portfolio contained only cash, and the algorithm should determine – at each rebalancing period – an optimal investment decision subject to risk constraints. The limit on how large a part of the total portfolio value one single asset can constitute, w_i , was set to $\pm 30\%$ for all i . No limits on the changes in the individual positions were considered and no transaction costs are incurred for buying or selling stocks. The return on cash was set to zero. In order to evaluate portfolio compositions and performances of different strategies with respect to the benchmark, a set of measures (table 7) – that come from the traditional investment world and have been accepted as useful tools to evaluate portfolio risk and return – is considered.

Table 7: Statistical indicators

Annualized excess return	Negative semi-deviation
Frequency of negative returns	Maximum drawdown
Annualized volatility	Skewness
Correlation with MSCI World Index	Excess Kurtosis

Specifically, optimization outcomes – both in the locally and in the risk aversion framework – are compared to the ones obtained in the classical global long-short market neutral schema. Besides the “*Annualized return*” that captures the capability of the portfolio of generating appealing excess returns with respect to the benchmark, all the measures aim at giving evidence of the aptitude of protecting capital. Among these indicators, the “*Negative semi deviation*” is a common solution to measure downside risk and the “*Maximum draw-down*” is often used by practitioners to analyze the downside exposure. Finally, “*Correlation with MSCI World Index*” provides an insight on how each long-short strategy performs, since ex post correlation with the benchmark should be negligible. Results referred to “*risk aversion framework*” are shown graphically (Figure 2 to 5)⁵.

Figure 2: Portfolio optimization results (annualized excess return)

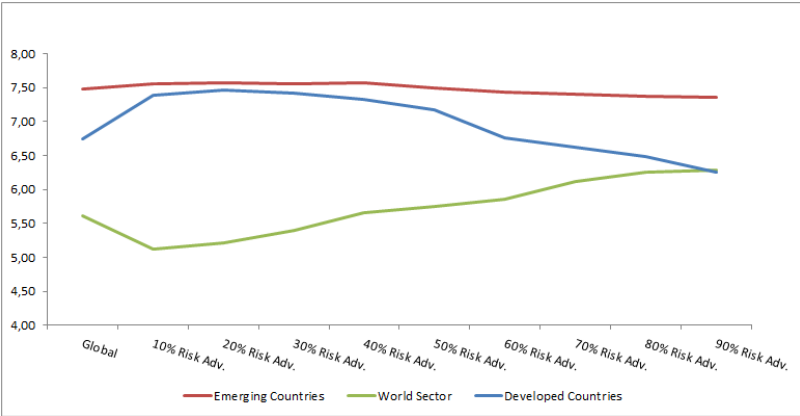
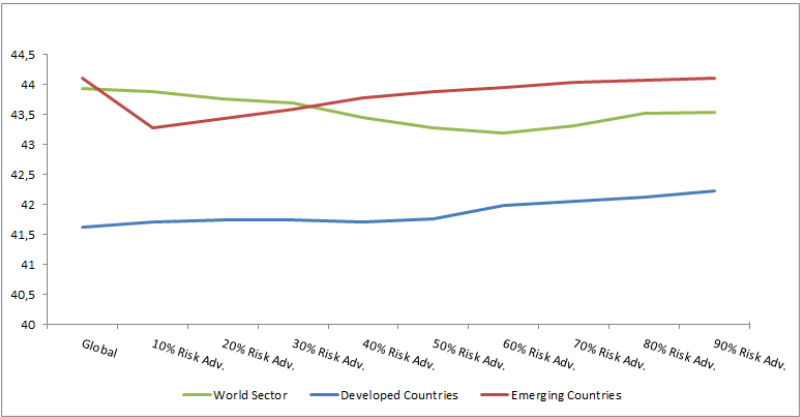


Figure 3: Portfolio optimization results (negative semi-deviation)



⁵ Same considerations hold for the “local optimization” algorithm and so the corresponding results are omitted.

Figure 4: Portfolio optimization results (maximum drawdown)

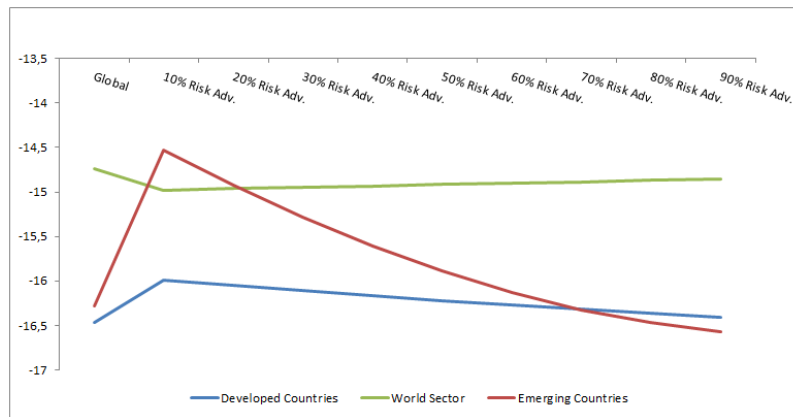
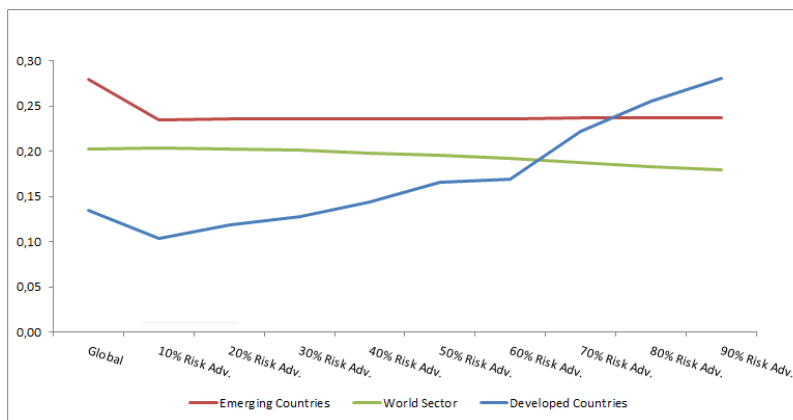


Figure 5: Portfolio optimization results (correlation with MSCI World)



Results show marginal improvement in risk/return performance compared to the one of global risk-neutral optimization. Moreover, whilst correlations with MSCI World values are as expected, it doesn't seem to exist a clear relationship between the "risk aversion parameter" and optimization results. This may be related to the type of time series used which exhibit by themselves intrinsic diversification: the consequence is an alignment of each returns' series to the one of the benchmark over the whole domain. At the same time, there is some evidence that the "risk aversion optimization" permits to control the downside, specifically where the Negative Semi-Deviation indicator is considered.

To sum up, in order to verify more extensively if the proposed optimization techniques are useful in modeling local relations between assets and benchmark, we are planning to do further empirical investigations in the type of data used (individual stocks versus country and sector indices), in the portfolio optimization strategies (active versus passive) and in the optimization function (beta versus Traynor Index).

REFERENCES

- Aas, K. and Hobæk Haff, I., The Generalized Hyperbolic Skew Student's t-distribution. *Journal of Financial Econometrics*, 2006, 4 (2), 275-309.
- Azzalini, A. and Capitanio, A., Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution. *Journal of the Royal Statistical Society B*, 2003, 65, 579-602.
- Basu A., Shioya H., Park C., Statistical Inference: The Minimum Distance Approach, 2011, CRC Press, London
- Bates, D. M. and Watts, D. G., Nonlinear Regression Analysis and its application, 2007, Wiley, New York, NY.
- Black, F. and Scholes M., The pricing of options and corporate liabilities. *Journal of Political Economy*, 1973, 81, 637-659.
- Box, G. E. P. and Draper, N. R., Empirical Model-Building and Response Surfaces, 1987, Wiley, New York, NY.
- Bramante R., Zappa D., 2012, Value at Risk Estimation in a Mixture Normality Framework, Submitted
- Chen, Y., Hardle, W., and Jeong, S., Nonparametric Risk Management With Generalized Hyperbolic Distributions. *Journal of the American Statistical Association*, 2008, 103, 910-923.
- Cont, R., Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 2001, 1, 223-236.
- Elton ,E. J., Gruber, M.J., Brown, S.J. and Goetzmann W. N., Modern Portfolio Theory and Investment Analysis, 2007, John Wiley & Sons, New York.
- Kon, S. J., Models of stock returns – a Comparison. *The Journal of Finance*, 1984, 39, 147-165.
- McNeil, A., Rudiger Frey, J. and Embrechts, P., *Quantitative Risk Management*, 2005, Princeton University Press, Princeton.
- Parr, W. C., Minimum Distance Estimation, *Encyclopedia of Statistical Science*, 1985, 5.
- Ruppert, D., Statistical and data analysis for financial engineering, 2011, Springer, Berlin.
- Wand, M. P., and Jones, M.C., Kernel smoothing, 1995, Chapman & Hall, London.