Smart Beta: Managing Diversification of Minimum Variance Portfolios*

Jean-Charles Richard
Quantitative Research
Lyxor Asset Management, Paris
jean-charles.richard@lyxor.com

Thierry Roncalli
Quantitative Research
Lyxor Asset Management, Paris
thierry.roncalli@lyxor.com

March 2015

Abstract

In this article, we consider a new framework to understand risk-based portfolios (GMV, EW, ERC and MDP). This framework is similar to the constrained minimum variance model of Jurczenko *et al.* (2013), but with another definition of the diversification constraint. The corresponding optimization problem can then be solved using the CCD algorithm. This allows us to extend the results of Cazalet *et al.* (2014) and to better understand the trade-off relationships between volatility reduction, tracking error and risk diversification. In particular, we show that the smart beta portfolios differ because they implicitly target different levels of volatility reduction. We also develop new smart beta strategies by managing the level of volatility reduction and show that they present appealing properties compared with the traditional risk-based portfolios.

Keywords: Smart beta, risk-based allocation, minimum variance portfolio, GMV, EW, ERC, MDP, portfolio optimization, CCD algorithm.

JEL classification: C61, G11.

1 Introduction

The capital asset pricing model (CAPM) of Sharpe (1964) and the empirical study of Jensen (1969) have been the backbone of passive management based on capitalization-weighted (CW) portfolios. In this approach, there is a single market risk premium, measured by the beta, and this risk premium compensates investors for holding non-diversifiable risk. In the CAPM theory, an investor can capture the market risk premium by holding the market portfolio. Applied to the universe of stocks, this justifies the strong development of capitalization-weighed equity indices. But since CAPM was introduced, academic research has put forward convincing evidence that CW portfolios are poorly diversified (1) and there are systematic sources of return in the equity markets other than simply the market beta (2). This justifies the strong development of smart beta in recent years. In fact, the term smart beta refers to two different approaches. First, it includes alternative-weighted portfolios, whose purpose is to be more diversified than CW portfolios. This smart beta approach

*This article has been written for the book *Risk-based and Factor Investing* edited by Emmanuel Jurczenko.

corresponds to the criticism (1) and it is also known as risk-based investing. Second, it refers to portfolios that are designed intentionally to capture alternative risk premia other the market risk premium, such as value, size, momentum, low beta or quality. This second approach corresponds to the criticism (2) and it is also known as factor-based investing.

In this paper, we focus on risk-based portfolios[1]. The main objective of this smart beta approach is to manage the risk more effectively than a CW index, and achieve a better performance. At first sight, risk-based portfolios seem to be heterogeneous because there are several notions of risk and each method considers one specific aspect of diversification. However, we can show that both approaches aim to reduce the volatility compared with the CW portfolio. This means that they are the solution to a minimum variance optimization problem, but with a different weight constraint. Our paper highlights, therefore, the central role of the minimum variance portfolio. Nevertheless, it is impossible to define a unique minimum variance portfolio. In fact, there are many minimum variance portfolios as there are smart beta products. In this situation, it is essential to have some metrics in order to understand their differences. Using a global optimization program, we can measure the different trade-off relationships between volatility reduction, tracking error, weight diversification and risk concentration. In particular, we can show that these minimum variance portfolios behave differently because they do not target the same volatility reduction. Some of them are very aggressive whereas others are closer to the CW portfolio. But once we impose the same level of volatility reduction, the differences between smart beta portfolios vanish even if they consider different weight constraints.

In risk-based investing, the key variable is then the level of volatility reduction. Because the objective of the investor is (almost) always to obtain a better performance, the choice of this parameter is crucial. This is why we also investigate how the performance of the portfolio is related to the volatility reduction. We show that this relationship depends strongly on the level of the market risk premium. Using this result, we can then build minimum variance strategies by targeting a time-varying volatility reduction, which depends on the market conditions.

The article is organized as follows. In section two, we show how the different risk-based portfolios can be cast in a minimum variance problem. In section three, we propose a unique optimization program in order to compare the diversification profile of smart beta strategies. We then analyze their behavior and propose new smart beta strategies by dynamically managing the objective of volatility reduction. Section five offers some concluding remarks.

2 Risk-based investing and variance minimization

Risk-based investing is generally associated with the concept of diversification. Because diversification can not be measured by a single number, practitioners consider different approaches. The most popular are the equally-weighted (EW) portfolio, the equal risk contribution (ERC) portfolio and the most diversified portfolio (MDP). Each of these portfolios maximizes a diversification measure. For instance, the MDP uses the diversification ratio. The EW portfolio minimizes the concentration in terms of weights whereas the ERC portfolio minimizes the concentration in terms of risk contributions.

[1] Even if the boundary between risk-based investing and factor-based investing is blurred in practice.

Besides these three risk-based approaches, practitioners also consider the minimum variance (MV) portfolio[2]. In this case, the goal is to explicitly manage the volatility rather than the diversification of the portfolio. But, as shown by Maillard *et al.* (2010), the EW and ERC portfolios can also be interpreted as constrained MV portfolios. Jurczenko *et al.* (2013) proposed a similar approach, which also encompasses the MDP. In particular, they consider the following optimization problem:

$$\begin{aligned} x^\star(\delta,\gamma) &= \arg\min \frac{1}{2} x^\top \Sigma x \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \sum_{i=1}^n \sigma_i^\delta \left(x_i^{1-\gamma} - 1 \right) \geq c \\ \mathbf{1}^\top x = 1 \end{array} \right. \end{aligned}$$

where Σ is the covariance matrix of asset returns and σ_i is the volatility of asset i. In this optimization program, $\delta \geq 0$ and $\gamma \geq 0$ are two given parameters and c is a scalar to be determined. They obtain the following correspondence between the parameters (δ,γ) and the risk-based portfolios $x^\star(\delta,\gamma)$:

Portfolio	GMV	EW	ERC	MDP
δ	0	0	0	1
γ	0	∞	1	0

In what follows, we consider an extension of the original optimization problem of Maillard *et al.* (2010). Our model is related to the approach of Cazalet *et al.* (2014) and helps to understand that risk-based portfolios are in fact minimum variance portfolios with a diversification constraint. The goal of risk-based portfolios is then to reach a lower volatility than the volatility of the capitalized-weighted portfolio. However, because each approach consider a specific definition of the diversification, there is a trade-off between these different measures of diversification.

2.1 MV portfolio

Global minimum variance (GMV) portfolios are never used by practitioners, because they correspond to mathematical corner solutions that are concentrated in a few number of assets. This is why minimum variance portfolios are always implemented by considering a constrained optimization problem:

$$\begin{aligned} x^\star &= \arg\min \frac{1}{2} x^\top \Sigma x \\ &\text{u.c.} \quad \left\{ \begin{array}{l} x \in \mathcal{C} \\ \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{1}$$

The constraints $x \geq \mathbf{0}$ and $\mathbf{1}^\top x = 1$ imply that the portfolio is long-only. The management of the weight concentration is specified by the constraint $x \in \mathcal{C}$. There are of course different ways to specify $\mathcal{C}$. One of the popular approaches consists in using the Herfindahl index defined by:

$$\mathcal{H}(x) = \sum_{i=1}^n x_i^2$$

[2] We note GMV the long-only global (or unconstrained) MV portfolio. This portfolio plays a special role in limit cases of portfolio optimization.

$\mathcal{H}(x)$ takes the value 1 if the portfolio is perfectly concentrated in one asset. Conversely, $\mathcal{H}(x)$ takes the value $1/n$ if the portfolio is equally-weighted. We can therefore define the weight diversification as:

$$\begin{aligned}\mathcal{D}_w(x) &= \frac{\mathcal{H}^{-1}(x)}{n} \\ &= \frac{1}{n\sum_{i=1}^n x_i^2}\end{aligned}$$

Using this diversification definition, the previous optimization problem becomes:

$$\begin{aligned}x^\star(c) &= \arg\min \frac{1}{2}x^\top \Sigma x \\ \text{u.c.} &\quad \left\{\begin{array}{l}\mathcal{D}_w(x) \geq c \\ \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0}\end{array}\right.\end{aligned} \tag{2}$$

with $c \in [1/n, 1]$. We have $x^\star(1/n) = x_{\mathrm{gmv}}$ and $x^\star(1) = x_{\mathrm{ew}}$. Because $\sigma(x^\star(c))$ is an increasing function of the parameter c, we deduce that:

$$\sigma(x_{\mathrm{gmv}}) \leq \sigma(x^\star(c)) \leq \sigma(x_{\mathrm{ew}})$$

Remark 1 *We notice that the optimization program (2) is equivalent to solving this Lagrange problem:*

$$\begin{aligned}y^\star(\lambda) &= \frac{1}{2}y^\top \Sigma y + \lambda y^\top y \\ u.c. &\quad \left\{\begin{array}{l}\mathbf{1}^\top x = 1 \\ x \geq \mathbf{0}\end{array}\right.\end{aligned} \tag{3}$$

with $\lambda \geq 0$. In this case, the optimal solution $x^\star(c)$ is equal to $y^\star(\lambda)$ with the following relationship:

$$c = \frac{1}{n\sum_{i=1}^n y_i^\star(\lambda)^2}$$

If $c \leq c_{\mathrm{gmv}} = \left(n x_{\mathrm{gmv}}^\top x_{\mathrm{gmv}}\right)^{-1}$, $x^\star(c) = x_{\mathrm{gmv}}$.

2.2 ERC portfolio

Let $\sigma(x) = \sqrt{x^\top \Sigma x}$ be the portfolio volatility. The risk contribution of asset i is defined by:

$$\mathcal{RC}_i = x_i \cdot \frac{\partial\, \sigma(x)}{\partial\, x_i}$$

These risk contributions are key when performing risk allocation, because the sum of risk contributions is exactly equal to the portfolio volatility:

$$\sum_{i=1}^n \mathcal{RC}_i = \sigma(x)$$

In the ERC portfolio, the risk contributions are the same for all assets:

$$\mathcal{RC}_i = \mathcal{RC}_j$$

Maillard *et al.* (2010) show that the ERC portfolio can be found by using the following optimization program:

$$\begin{aligned} y^\star(c') &= \arg\min \frac{1}{2} y^\top \Sigma y \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \sum_{i=1}^n \ln y_i \geq c' \\ y \geq \mathbf{0} \end{array} \right. \end{aligned}$$

where c' is a scalar. The ERC portfolio is then equal to the normalized portfolio $y^\star(c')$:

$$x_{\mathrm{erc}} = \frac{y^\star(c')}{\mathbf{1}^\top y^\star(c')}$$

Let us now consider this second optimization program:

$$\begin{aligned} x^\star(c) &= \arg\min \frac{1}{2} x^\top \Sigma x \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \sum_{i=1}^n \ln x_i \geq c \\ \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{4}$$

where $c \in \left]-\infty, n \ln n\right]$. Maillard *et al.* (2010) demonstrated that there exists a value of c such that the optimized portfolio is the ERC portfolio. In this case, we have the following relationship:

$$c_{\mathrm{erc}} = c' - n \ln \sum_{i=1}^n y_i^\star(c')$$

Roncalli (2013) also deduces that the optimized portfolio $y^\star(c')$ is a leveraged version of the ERC portfolio:

$$y^\star(c') = \exp\left(\frac{c' - c_{\mathrm{erc}}}{n}\right) \cdot x_{\mathrm{erc}}$$

Because $\sigma(x^\star(c))$ is an increasing function of the parameter c, we obtain the same inequality as in the case of constrained minimum variance portfolios:

$$\sigma(x_{\mathrm{gmv}}) \leq \sigma(x^\star(c)) \leq \sigma(x_{\mathrm{ew}})$$

We deduce that:

$$\sigma(x_{\mathrm{gmv}}) \leq \sigma(x_{\mathrm{erc}}) \leq \sigma(x_{\mathrm{ew}})$$

Remark 2 *The Lagrange formulation of the optimization problem (4) is:*

$$\begin{aligned} y^\star(\lambda) &= \frac{1}{2} y^\top \Sigma y - \lambda \sum_{i=1}^n \ln y_i \\ &\textit{u.c.} \quad \left\{ \begin{array}{l} \mathbf{1}^\top y = 1 \\ y \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{5}$$

with $\lambda \geq 0$. In this case, the optimal solution $x^\star(c)$ corresponds to the portfolio $y^\star(\lambda)$ with:

$$c = \sum_{i=1}^n \ln y_i^\star(\lambda)$$

According to this framework, a natural way to measure the diversification is to consider the Herfindahl index applied to risk contributions:

$$\mathcal{D}_{rc}\left(x\right) = \frac{1}{n\sum_{i=1}^{n}\mathcal{RC}_i^2\left(x\right)}$$

Let us consider portfolios with positive risk contributions[3]. It follows that $\mathcal{D}_{rc}\left(x\right) \in [1/n, 1]$ and we have $\mathcal{D}_{rc}\left(x_{\mathrm{erc}}\right) = 1$. This means that the ERC portfolio is then the one that maximizes the risk diversification.

2.3 Most diversified portfolio

Choueifaty and Coignard (2008) introduce the concept of diversification ratio, which corresponds to the following expression:

$$\mathcal{DR}\left(x\right) = \frac{x^\top \sigma}{\sqrt{x^\top \Sigma x}}$$

By construction, $\mathcal{DR}\left(x\right)$ is equal to one if the portfolio is fully invested in one asset or if the correlations $\rho_{i,j}$ are all equal to one. In the other cases, we have $\mathcal{D}_{\sigma}\left(x\right) > 1$. The MDP is then the portfolio which maximizes the diversification ratio:

$$\begin{aligned} x_{\mathrm{mdp}} &= \arg\max \mathcal{DR}\left(x\right) \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{6}$$

2.3.1 A first route toward variance minimization

Let ρ be the correlation matrix deduced from Σ. We note $x_{\mathrm{gmv}}\left(\rho\right)$ the long-only minimum variance portfolio based only on the correlation matrix. The MDP is then a rescaled version of the GMV portfolio:

$$x_{\mathrm{mdp},i} \propto \frac{x_{\mathrm{gmv},i}\left(\rho\right)}{\sigma_i}$$

2.3.2 A second route

Let us consider the following optimization problem:

$$\begin{aligned} y^\star\left(c'\right) &= \arg\min \frac{1}{2} y^\top \Sigma y \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \sum_{i=1}^{n} y_i \sigma_i \geq c' \\ y \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{7}$$

with $c' > 0$. We can demonstrate that the MDP corresponds to the normalized portfolio[4]:

$$x_{\mathrm{mdp}} = \frac{y^\star\left(c'\right)}{\mathbf{1}^\top y^\star\left(c'\right)}$$

[3] It is always the case if the cross-correlations $\rho_{i,j}$ are positive.

[4] Because we have the following property:

$$\frac{y^\star\left(c'\right)}{c'} = \frac{y^\star\left(c''\right)}{c''}$$

It follows that the MDP is the solution of the following optimization program for a specific value of c:

$$\begin{aligned} x^\star(c) &= \arg\min \frac{1}{2} x^\top \Sigma x \\ &\text{u.c.} \quad \left\{ \begin{array}{l} \sum_{i=1}^n x_i \sigma_i \geq c \\ \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{8}$$

where $c \in [0, \max_i \sigma_i]$. Indeed, we have:

$$c_{\mathrm{mdp}} = \frac{c'}{\sum_{i=1}^n y_i^\star(c')}$$

It follows that:

$$\sigma(x_{\mathrm{gmv}}) \leq \sigma(x_{\mathrm{mdp}}) \leq \max_i \sigma_i$$

Remark 3 *The Lagrange formulation of the optimization problem (8) is:*

$$\begin{aligned} y^\star(\lambda) &= \frac{1}{2} y^\top \Sigma y - \lambda y^\top \sigma \\ &u.c. \quad \left\{ \begin{array}{l} \mathbf{1}^\top y = 1 \\ y \geq \mathbf{0} \end{array} \right. \end{aligned} \tag{9}$$

with $\lambda \geq 0$. In this case, the optimal solution $x^\star(c)$ corresponds to the portfolio $y^\star(\lambda)$ with:

$$c = \sum_{i=1}^n y_i^\star(\lambda) \sigma_i$$

If we delete the constraint $\mathbf{1}^\top y = 1$, we obtain the solution $y^\star(c')$ given by the optimization program (7). Let us consider the restricted universe of invested assets, that is the assets i such that $x_{\mathrm{mdp},i} > 0$. It follows that the MDP weights of this restricted universe are:

$$\tilde{x}_{\mathrm{mdp}} = \frac{\tilde{\Sigma}^{-1} \tilde{\sigma}}{\mathbf{1}^\top \tilde{\Sigma}^{-1} \tilde{\sigma}}$$

where $\tilde{\Sigma}$ is the covariance matrix of the invested assets.

2.4 Comparing the trade-off relationships

Following Cazalet *et al.* (2014), we compare the different optimization programs (2), (4) and (8) by changing the value of c. We consider the Eurostoxx 50 index and the one-year empirical covariance matrix estimated in February 2013. The results are reported in Figures 1, 2 and 3. In each figure, the first panel represents the tracking error volatility $\sigma(x \mid x_{\mathrm{cw}})$ with respect to the volatility reduction $\mathcal{VR}(x \mid x_{\mathrm{cw}})$ defined by:

$$\mathcal{VR}(x \mid x_{\mathrm{cw}}) = \frac{\sigma(x_{\mathrm{cw}}) - \sigma(x)}{\sigma(x_{\mathrm{cw}})}$$

In the second panel, we consider the beta $\beta(x \mid x_{\mathrm{cw}})$ of the portfolio with respect to the capitalization-weighted portfolio. The three panels at the bottom show the impact of the volatility reduction on the diversification measures[5]. These results show that investors have to puzzle out the trade-off between volatility, tracking error and diversification. However, we notice that the trade-off relationships are very similar when comparing MV and ERC portfolios (Figures 1 and 2), which is not the case when considering MDP (Figure 3).

[5]The diversification measure $D_\rho(x)$ is the ratio between the diversification ratio $\mathcal{DR}(x)$ of the portfolio and the diversification ratio $\mathcal{DR}(x_{\mathrm{mdp}})$ of the MDP.

Figure 1: Trade-off relationships of Problem (2) (MV)

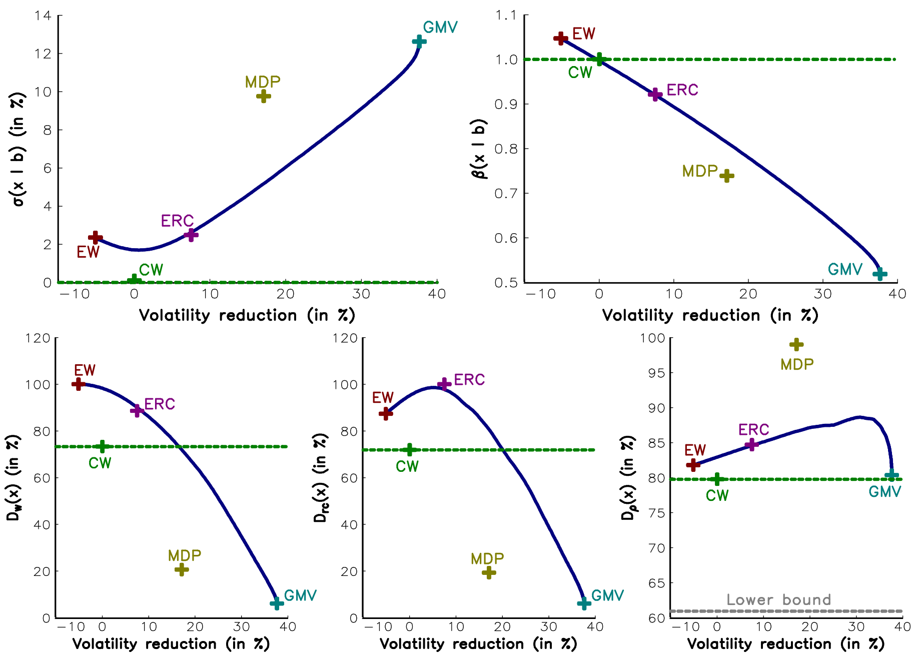


Figure 2: Trade-off relationships of Problem (4) (ERC)

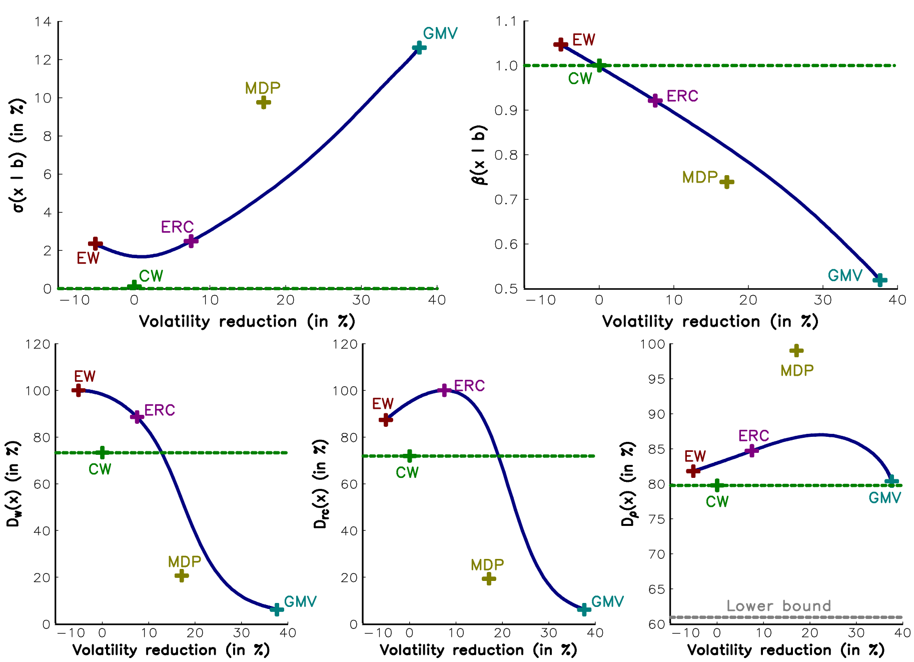
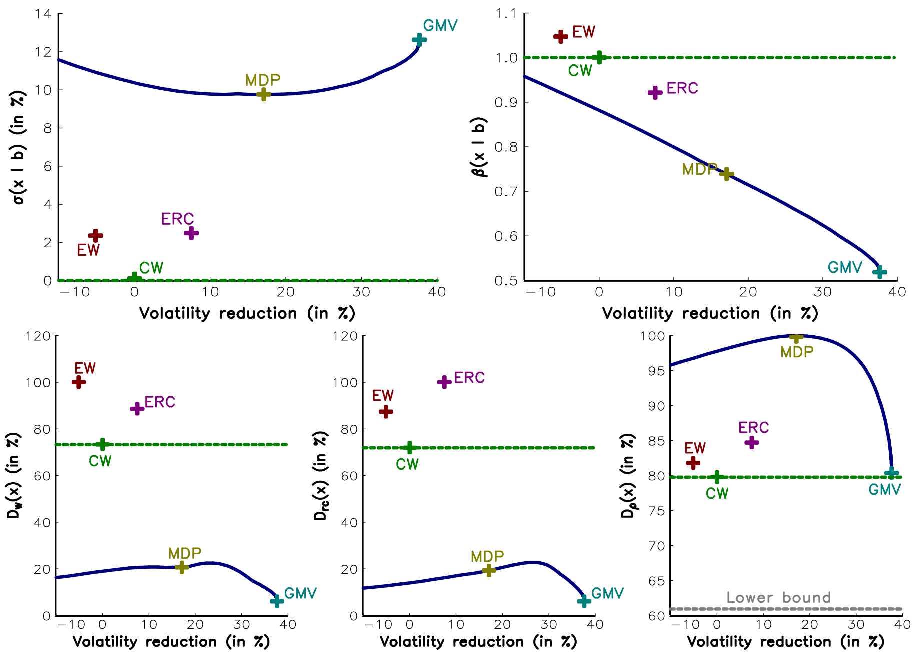


Figure 3: Trade-off relationships of Problem (8) (MDP)



3 Managing the diversification

3.1 Mixing the constraints

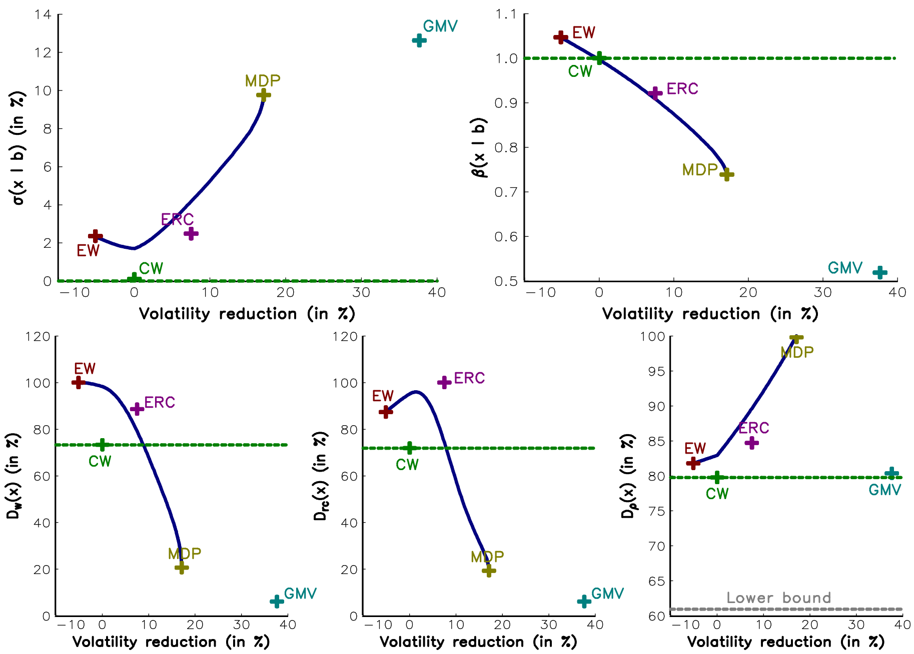
When we consider Figure 3, we observe that solutions are not very interesting because we cannot manage the diversification in terms of weights or risk contributions. This is why we can introduce these constraints into Problem (8). For instance, the MDP optimization problem with the weight diversification becomes:

$$
\begin{aligned}
x^{\star}(c_1, c_2) &= \arg\min \frac{1}{2} x^{\top} \Sigma x \\
&\text{u.c.} \quad \begin{cases} \sum_{i=1}^{n} x_i \sigma_i \geq c_1 \\ \mathcal{D}_w(x) \geq c_2 \\ \mathbf{1}^{\top} x = 1 \\ x \geq \mathbf{0} \end{cases}
\end{aligned}
\tag{10}
$$

In this case, we can build smart beta portfolios between the MDP ($c_1 = c_{\mathrm{mdp}}$ and $c_2 = 0$) and the EW portfolio ($c_1 = c_{\mathrm{mdp}}$ and $c_2 = 1$). An example is given in Figure 4 by setting $c_1 = c_{\mathrm{mdp}}$. If we prefer to consider the risk diversification, we obtain:

$$
\begin{aligned}
x^{\star}(c_1, c_2) &= \arg\min \frac{1}{2} x^{\top} \Sigma x \\
&\text{u.c.} \quad \begin{cases} \sum_{i=1}^{n} x_i \sigma_i \geq c_1 \\ \sum_{i=1}^{n} \ln x_i \geq c_2 \\ \mathbf{1}^{\top} x = 1 \\ x \geq \mathbf{0} \end{cases}
\end{aligned}
\tag{11}
$$

Figure 4: Trade-off relationships of Problem (10) with $c_1 = c_{\mathrm{mdp}}$



3.2 A unified optimization framework

In fact, we can combine these different constraints in a unique variance minimization problem with the following set of constraints:

$$\begin{cases} \mathbf{1}^\top x = 1 \\ \sum_{i=1}^n x_i^2 \leq c_1 \\ \sum_{i=1}^n \ln x_i \geq c_2 \\ \sum_{i=1}^n x_i \sigma_i \geq c_3 \end{cases}$$

The first and fourth constraints allow the GMV portfolio and the MDP respectively to be obtained. The second and third constraints manage the diversification in terms of weights (using the Herfindahl index) and risk contributions. Therefore, we can write the constrained problem using Lagrange multipliers:

$$\begin{aligned} x^\star &= \arg\min \frac{1}{2} x^\top \Sigma x - \\ & \quad \lambda_{\mathrm{gmv}} \left(\sum_{i=1}^n x_i \right) + \lambda_{\mathrm{h}} \left(\sum_{i=1}^n x_i^2 \right) - \\ & \quad \lambda_{\mathrm{erc}} \left(\sum_{i=1}^n \ln x_i \right) - \lambda_{\mathrm{mdp}} \left(\sum_{i=1}^n x_i \sigma_i \right) \\ \text{u.c.} & \quad x \geq \mathbf{0} \end{aligned} \tag{12}$$

with $\lambda_{\mathrm{h}} \geq 0$ and $\lambda_{\mathrm{erc}} \geq 0$. From a technical point of view, there are no restrictions on λ_{gmv} and λ_{mdp} even if some cases are more relevant ($\lambda_{\mathrm{gmv}} \geq 0$ and $\lambda_{\mathrm{mdp}} \geq 0$).

Remark 4 *The previous framework can be extended by replacing the variance minimization problem by the tracking error minimization problem. In Appendix A.1, we show that it is* *equivalent to introducing a constraint in the form $x^\top \Sigma x_{\mathrm{cw}} \geq c_4$. In this case, Problem (12) must include a new penalty function which is equal to:*

$$-\lambda_{\mathrm{te}} \left(\sum_{i=1}^{n} x_i \left(\Sigma x_{\mathrm{cw}} \right)_i \right)$$

The benefits of using the formulation (12) are twofold. First, this optimization problem is very easy to solve using the CCD algorithm. This numerical method was used by Griveau-Billion *et al.* (2013) to find the solution of the ERC portfolio. In Appendix A.2, we extend this analysis to the general problem (12). The second interest lies in the explicit trade-off relationships contained in the optimization problem. If the aim is to emphasize one specific diversification measure, we have to use a larger value for the corresponding Lagrange coefficient, but it is not possible to match all the different diversification constraints. This means that even if there is no restriction between the Lagrange multipliers, only a subset of them is interesting from a financial point of view. This is equivalent to imposing a structure between the different constraints. Let us consider this specific problem for instance:

$$\begin{aligned} x^\star &= \arg\min \frac{1}{2} x^\top \Sigma x \\ \text{u.c.} &\quad \begin{cases} \mathcal{D}(x;\gamma) \geq c_1 \\ \mathcal{B}(x;\delta) = c_2 \\ x \geq \mathbf{0} \end{cases} \end{aligned} \tag{13}$$

where $\mathcal{D}(x;\gamma) = \gamma \sum_{i=1}^{n} \ln x_i - (1-\gamma) \sum_{i=1}^{n} x_i^2$ is a diversification constraint and $\mathcal{B}(x;\delta) = \delta \sum_{i=1}^{n} x_i + (1-\delta) \sum_{i=1}^{n} x_i \sigma_i$ is a budget constraint. The parameter $\gamma \in [0,1]$ controls the trade-off between weighs and risk diversification whereas the parameter $\delta \in [0,1]$ controls the budget allocation. We can then restrict (c_1, c_2) by considering this optimization problem:

$$\begin{aligned} x^\star(\lambda,\gamma,\delta) &= \arg\min \frac{1}{2} x^\top \Sigma x - \lambda \mathcal{D}(x;\gamma) + (\lambda - 1) \mathcal{B}(x;\delta) \\ \text{u.c.} &\quad x \geq \mathbf{0} \end{aligned} \tag{14}$$

where $\lambda \geq 0$ controls the impact on the diversification. Problem (14) is a spacial case of Problem (12), but it is wide enough to include most of the solutions[6].

Remark 5 *If we include a tracking error constraint, the budget constraint becomes $\mathcal{B}(x;\delta,\kappa) = \sum_{i=1}^{n} x_i \left(\delta + \kappa \left(\Sigma x_{\mathrm{cw}}\right)_i + (1-\delta-\kappa)\sigma_i\right)$ with $0 \leq \kappa + \gamma \leq 1$.*

In Table 1 we indicate the parameters that give the different smart beta portfolios (see Appendix A.2.3 for the definition of RP and BP portfolios). For instance $(\lambda,\gamma,\delta,\kappa) = (1,1,0,0)$ gives the ERC portfolio while $(\lambda,\gamma,\delta,\kappa) = (0,1,1,0)$ gives the GMV portfolio.

Table 1: Limits of the smart beta portfolio $x^\star(\lambda,\gamma,\delta,\kappa)$

Parameters	GMV	EW	ERC	MDP	RP	BP	CW
λ	0	$+\infty$	1	0	$+\infty$	$+\infty$	0
γ		0/1	1		1	1	1
δ	1	1		0	1	0	
κ	0	0		0	0	1	1

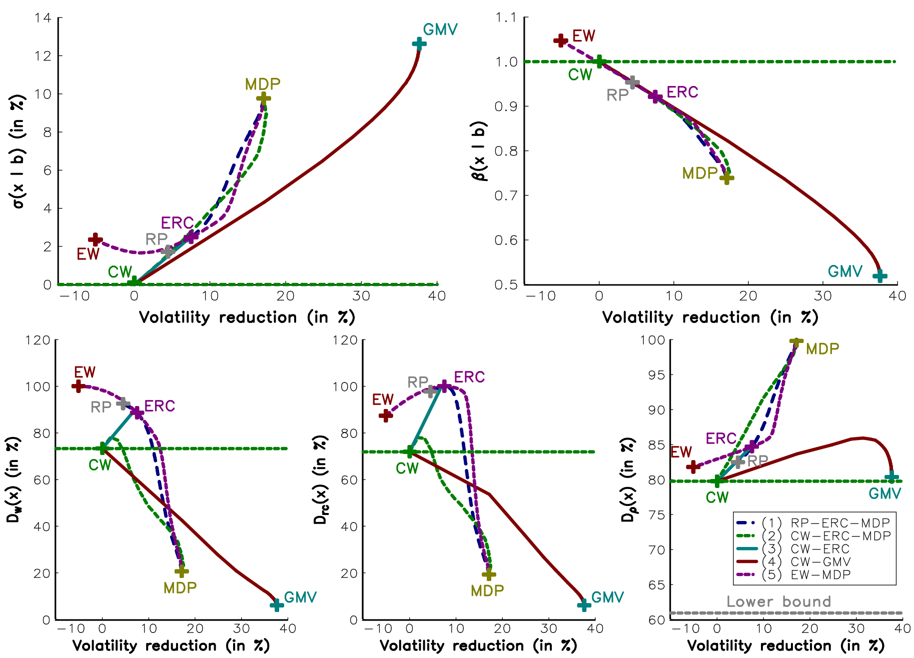
[6]It is equivalent to impose that $\lambda_{\mathrm{gmv}} - \lambda_{\mathrm{h}} + \lambda_{\mathrm{erc}} + \lambda_{\mathrm{mdp}} = 1$ and $\lambda_{\mathrm{erc}} = \lambda_{\mathrm{h}} + \lambda$.

With Problem (14), we can explore new risk-based portfolios by mixing different constraints. We consider the following set of parameters:

Set	Achievable Portfolios	λ	γ	δ	κ
(1)	RP-ERC-MDP	$\in \mathbb{R}_+$	1	0	0
(2)	CW-ERC-MDP	0	1	0	$\in [0,1]$
(3)	CW-ERC	$\in \mathbb{R}_+$	1	0	1
(4)	CW-GMV	0	0	$1-\kappa$	$\in [0,1]$
(5)	EW-MDP	$\in \mathbb{R}_+$	1	$1-e^{-\lambda}$	0

For each set, we indicate the achievable portfolios. For instance, if $\kappa = 1$ (and $\delta = 0$), we obtain the CW portfolio. Depending on the values of λ and γ, we can then build risk-based portfolios between CW and another smart beta portfolio. For instance, if $\lambda \in \mathbb{R}_+$ and $\gamma = 1$, we obtain solutions between the CW portfolio and the ERC portfolio. In Figure 5, we have reported the paths of the different parameter sets.

Figure 5: Trade-off relationships of Problem (14)



3.3 Diversification profile of risk-based portfolios

Radar charts of the different objectives are reported in Figure 6. Each hexagonal chart (represented by dashed lines) corresponds to an improvement of the measure by 15%. In order to compare the different profiles, we use a benchmark profile which has the GMV volatility reduction, a zero tracking error, a beta equal to one, the diversification ratio of the MDP, the weight and risk diversifications of the EW and ERC portfolios. The GMV portfolio focuses on minimizing the volatility but presents a poor diversification in terms of weights and risk contributions. It is also the portfolio with the highest beta and tracking error risk. The EW portfolio performs well to maximize the beta and minimize the tracking

Figure 6: Diversification profile of smart beta portfolios

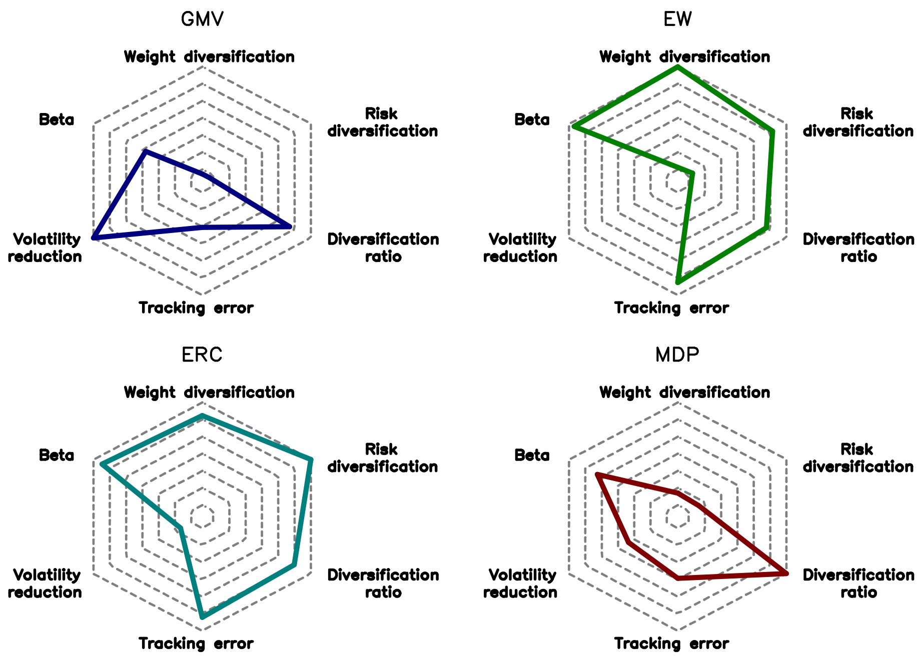
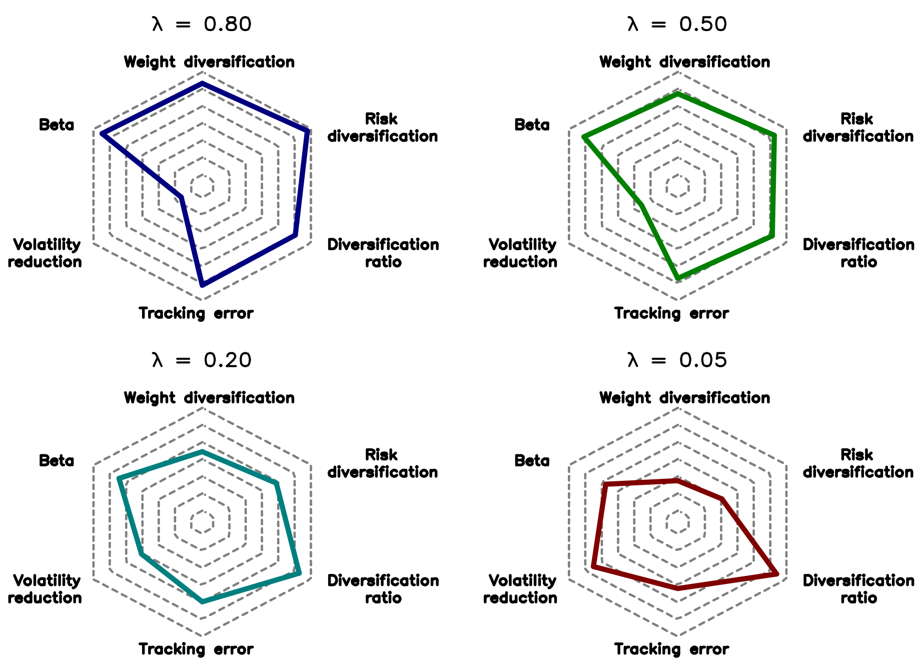


Figure 7: Diversification profile and weight diversification

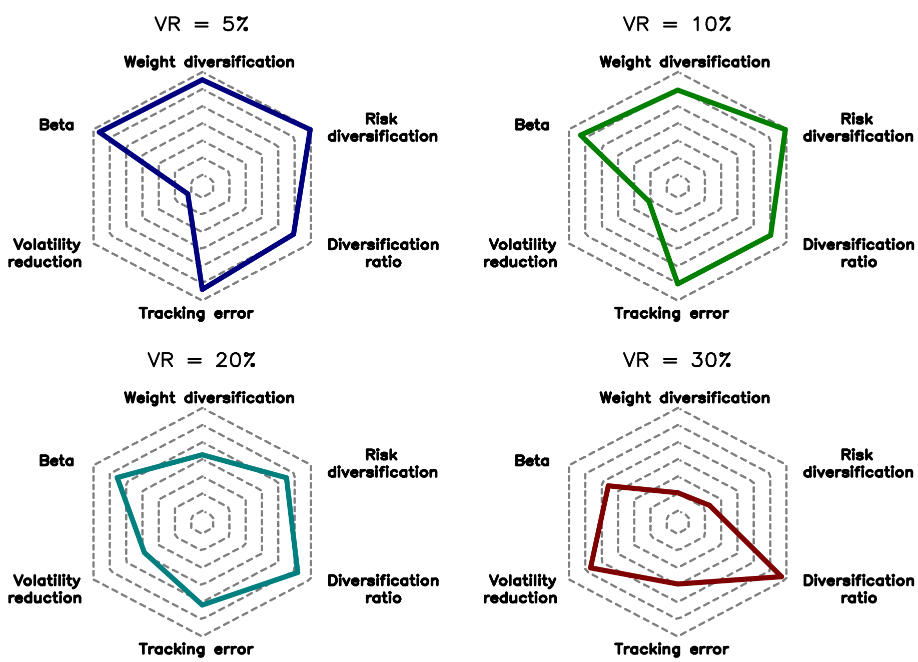


error while diversifying the weights. But this is done with no volatility reduction. The ERC portfolio has a similar profile but pays more attention to volatility reduction. Finally, the MDP profile is similar to the GMV profile, but has a lower beta and tracking error risk.

Figure 7 illustrates that the role of the parameter λ. κ is equal to zero and we fix the other parameters in a balanced manner: $\gamma = \delta = 0.5$. We observe that the volatility reduction is done at the expense of the diversification. Moreover, the weight diversification decreases more quickly than the risk diversification. Indeed, the volatility of the ERC portfolio is always lower than the volatility of the EW portfolio. This means that the impact of the volatility reduction on the diversification is weaker for the ERC.

In Figure 8, we have reported the diversification profile when we specifically target a volatility reduction (5%, 10%, 20% and 30%). In this example, we confirm that the weight diversification decreases more quickly than the risk diversification. The diversification ratio is the less impacted measure by the change in the volatility reduction.

Figure 8: Diversification profile and volatility reduction



4 Understanding the behavior of smart beta portfolios

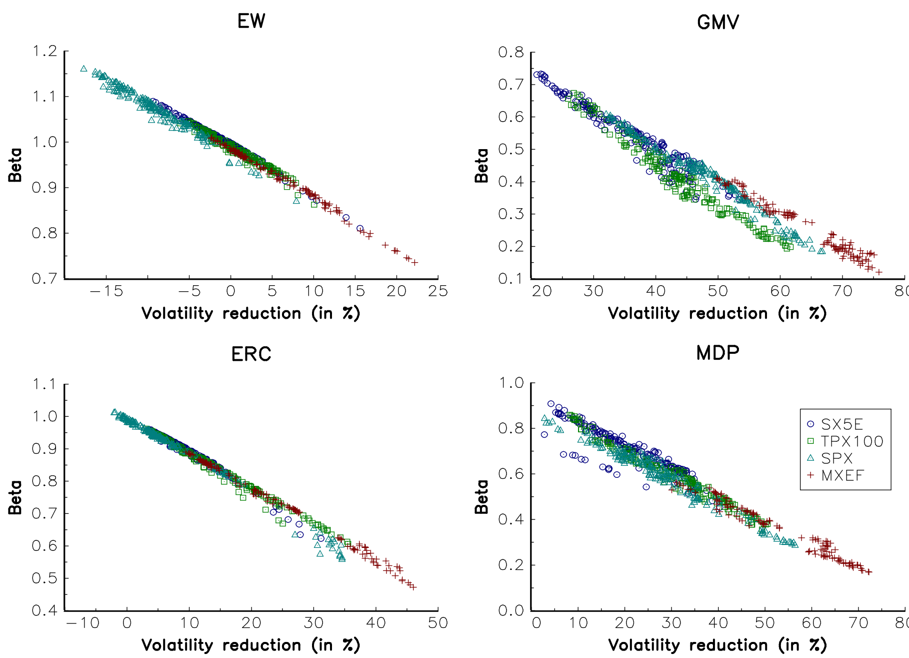
We consider here real-life applications with four different stock universes: the Eurostoxx 50 index (SX5E), the Topix 100 index (TPX100), the S&P 500 index (SPX) and the MSCI EM index (MXEF). We have chosen these stock indices, because they correspond to different regions and different sizes of the universe. For each universe, we compute smart beta portfolios by using the one-year empirical covariance matrix of stock returns. The allocation is rebalanced at a monthly frequency. We conduct backtests from January 2001 to December

2014[7]. Empirical results confirm that there are some trade-off principles. In particular, we obtain a first rule of smart beta indexing:

Rule 1 *There is no free lunch in smart beta. In particular, it is not possible to target high volatility reduction, to be highly diversified and to take low beta risk.*

In Figure 9, we have reported the relationship between the volatility reduction and the beta for the four universes and the four smart beta portfolios (EW, GMV, ERC and MDP). Each point corresponds to a rebalancing date. By reducing the volatility, the smart beta portfolios increase the beta risk. We observe similar results for the other risk measures: tracking error, weight diversification, risk diversification and diversification ratio.

Figure 9: Relationship between the volatility reduction and the beta



4.1 Volatility reduction

Rule 2 *The smart beta portfolios have a time-varying objective of volatility reduction and tracking error.*

This rule shows that the behavior of traditional smart beta portfolios (EW, GMV, ERC, MDP) is not homogeneous across time in terms of volatility reduction and tracking error. We have reported the boxplots in Figures 10 and 11. The bottom and top of the box indicate the first and third quartiles of the statistics, the line inside the box corresponds to the median whereas the ends of whiskers are the minimum and the maximum. We notice that the volatility reduction depends on the underlying index. However, we do not observe a strong relationship with the size of the universe, except for the GMV portfolio. For instance, the EW portfolio has a higher volatility, on average, than the CW portfolio in the case of

[7]For the MSCI EM index, the starting date is February 2005.

Figure 10: Boxplot of the volatility reduction (in %)

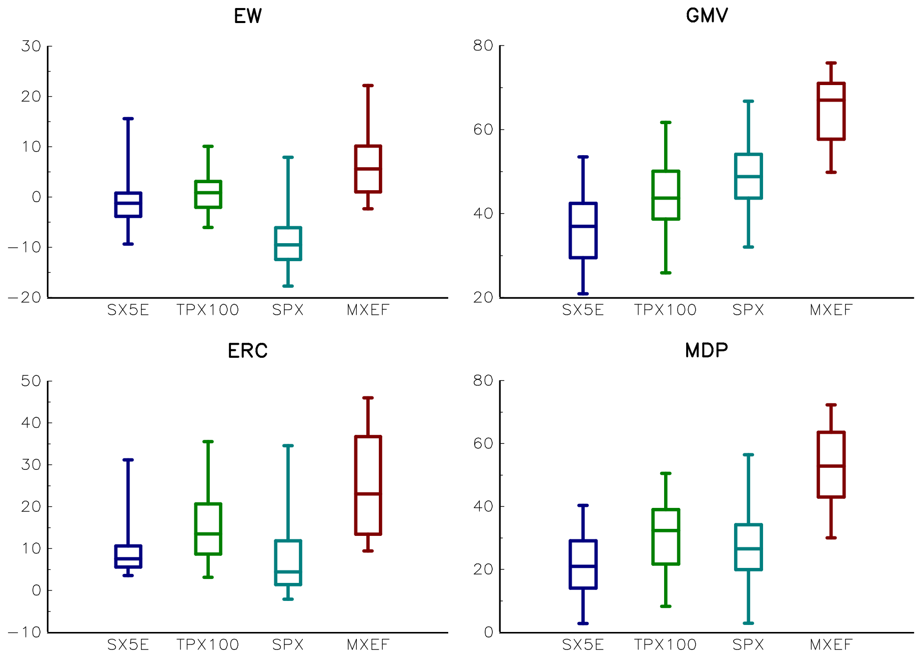
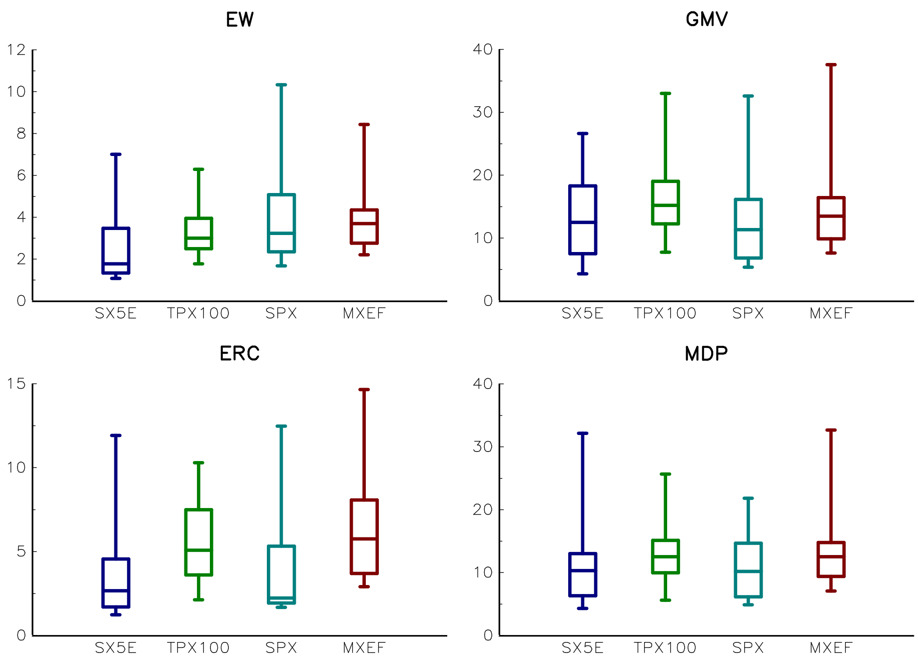


Figure 11: Boxplot of the tracking error (in %)



the S&P 500 index, but the volatility reduction is maximal for the MSCI EM index. If we consider the tracking error, the behavior is even more complex with respect to the underlying index. For instance, the Topix 100 universe presents the highest tracking error in the case of ERC and GMV portfolios.

A statistical analysis shows that the level of the volatility reduction and the tracking error, as well as their variations, cannot be explained by the level or the variation of the volatility of the CW index. We can therefore obtain all the possible existing configurations:

$\sigma(x_{\mathrm{cw}})$	High	Low
$\mathcal{VR}(x \mid x_{\mathrm{cw}})$ or $\sigma(x \mid x_{\mathrm{cw}})$	High/Low	High/Low
$\Delta_t \sigma(x_{\mathrm{cw}})$	$+$	$-$
$\Delta_t \mathcal{VR}(x \mid x_{\mathrm{cw}})$ or $\Delta_t \sigma(x \mid x_{\mathrm{cw}})$	$+/-$	$+/-$

Even if there is no obvious relationship between the volatility of the CW portfolio and the volatility or the tracking error of the smart beta portfolios, there is some similar behavior between the smart beta portfolios themselves. For instance, we report the correlation between the variations during two consecutive rebalancing dates of the volatility reduction in Table 2. We compute the statistic $\rho_{i,j}^{\Delta_t \mathcal{VR}} = \rho\left(\Delta_t \mathcal{VR}\left(x^{(i)} \mid x_{\mathrm{cw}}\right), \Delta_t \mathcal{VR}\left(x^{(j)} \mid x_{\mathrm{cw}}\right)\right)$ for all pairs (i, j) of smart beta portfolios. We notice that the cross-correlations are high except for the EW portfolio in the case of the S&P 500 universe. Results for the tracking error correlation $\rho_{i,j}^{\Delta_t \mathrm{TE}} = \rho\left(\Delta_t \sigma\left(x^{(i)} \mid x_{\mathrm{cw}}\right), \Delta_t \sigma\left(x^{(j)} \mid x_{\mathrm{cw}}\right)\right)$ are also reported in Table 2. Like the volatility reduction, the tracking error cross-correlations are high especially for the pairs (GMV,MDP) and (ERC,MDP).

Table 2: Empirical correlations $\rho_{i,j}^{\Delta_t \mathcal{VR}}$ and $\rho_{i,j}^{\Delta_t \mathrm{TE}}$ (in %)

	Volatility reduction				Tracking error			
(i,j)	SX5E	TPX100	SPX	MXEF	SX5E	TPX100	SPX	MXEF
(EW,GMV)	16.7	44.1	−0.5	37.1	25.0	44.4	44.7	76.3
(EW,ERC)	66.4	70.2	33.1	79.8	26.6	33.1	11.5	76.0
(EW,MDP)	26.9	44.3	5.6	37.9	38.1	37.6	42.0	75.5
(GMV,ERC)	69.2	77.7	49.2	64.5	62.8	41.1	48.1	87.4
(GMV,MDP)	74.7	79.9	45.6	80.3	76.3	90.7	75.5	98.2
(ERC,MDP)	71.9	79.2	66.0	64.8	82.4	53.0	63.0	90.2

Table 3: Empirical correlations $\rho_{i,j}^{\Delta_t \mathcal{D}_{\mathrm{rc}}}$ and $\rho_{i,j}^{\Delta_t \mathcal{D}_{\rho}}$ (in %)

	Risk diversification				Diversification ratio			
(i,j)	SX5E	TPX100	SPX	MXEF	SX5E	TPX100	SPX	MXEF
(EW,GMV)	2.1	−35.4	−2.4	−20.8	73.7	77.1	54.9	87.5
(EW,ERC)	−6.7	8.7	−8.2	−15.5	93.2	93.6	84.8	96.1
(EW,MDP)	0.5	−38.1	−22.5	−29.2	79.2	85.8	81.0	91.1
(GMV,ERC)	34.5	−14.2	−17.9	−7.4	75.5	85.4	65.7	92.4
(GMV,MDP)	25.6	12.9	14.1	42.6	75.3	86.2	74.0	96.2
(ERC,MDP)	23.0	−3.2	15.9	23.5	92.8	92.9	89.7	96.5

If we consider the other risk statistics, we obtain similar results for the beta[8] and the diversification ratio, but not for the weight and risk diversifications. For these two last statistics, the average correlation is close to zero. We report the results for $\rho_{i,j}^{\Delta_t \mathcal{D}_{\mathrm{rc}}}$ and $\rho_{i,j}^{\Delta_t \mathcal{D}_{\rho}}$ in Table 3. We notice that the cross-correlation $\mathcal{D}_\rho$ is extremely high[9].

4.2 Normalizing the smart beta portfolios

In Table 4, we report the average correlation between the three smart beta portfolios (GMV, ERC and MDP) for the different statistics[10]. We notice that the average correlation between returns is less than 90%, implying that the behavior of these smart beta portfolios may be very different in some specific periods.

Table 4: Average correlation between GMV, ERC and MDP portfolios (in %)

Index	$\mathcal{VR}$	TE	β	$\mathcal{D}_w$	$\mathcal{D}_{rc}$	$\mathcal{D}_\rho$	R_t
SX5E	67.4	81.9	73.0	39.2	26.5	95.8	89.6
TPX100	88.2	81.1	87.6	28.9	27.7	93.6	92.8
SX5E	79.9	80.2	82.9	21.3	32.6	97.3	83.2
MXEF	89.3	93.2	93.1	2.4	34.5	97.8	88.5
Average	81.2	84.1	84.1	23.0	30.3	96.1	88.5

We can ask if these differences come from the implied intrinsic constraint of each model, or from the level of volatility reduction targeted by each model. This is why we investigate the behavior of smart beta portfolios when we normalize them by targeting the same level of volatility reduction. Therefore, we calibrate the set of parameters $(\lambda_{\mathrm{gmv}}, \lambda_{\mathrm{h}}, \lambda_{\mathrm{erc}}, \lambda_{\mathrm{mdp}}, \lambda_{\mathrm{te}})$ such that:

$$\mathcal{VR}\left(x^\star\left(\lambda_{\mathrm{gmv}}, \lambda_{\mathrm{h}}, \lambda_{\mathrm{erc}}, \lambda_{\mathrm{mdp}}, \lambda_{\mathrm{te}}\right) \mid x_{\mathrm{cw}}\right) = \eta^\star$$

where $\eta^\star$ is the targeted volatility reduction. For each smart beta portfolio, the calibration is done with one parameter (it is underlined) whereas the other parameters are fixed:

GMV $\lambda_{\mathrm{gmv}} = 1$, $\underline{\lambda_{\mathrm{h}} \in [0, +\infty)}$, $\lambda_{\mathrm{erc}} = 0$, $\lambda_{\mathrm{mdp}} = 0$ and $\lambda_{\mathrm{te}} = 0$;

ERC $\lambda_{\mathrm{gmv}} = -\infty$, $\lambda_{\mathrm{h}} = 0$, $\underline{\lambda_{\mathrm{erc}} \in (0, +\infty)}$, $\lambda_{\mathrm{mdp}} = 0$ and $\lambda_{\mathrm{te}} = 0$;

MDP $\lambda_{\mathrm{gmv}} = 0$, $\lambda_{\mathrm{h}} = 0$, $\lambda_{\mathrm{erc}} = 1$, $\underline{\lambda_{\mathrm{mdp}} \in (-\infty, +\infty)}$ and $\lambda_{\mathrm{te}} = 0$;

For instance, the calibration is done using the Herfindahl parameter λ_{h} in the case of the GMV portfolio. Results are reported in Table 5. We notice that the average correlation between the three smart beta methods has highly increased. This is particularly true for the one-year performance, for which the average correlation is close to 100%. We conclude that the differences between the smart beta methods (GMV, ERC and MDP) are mainly explained by the different level of targeted volatility reduction, which is a consequence of their intrinsic constraints. These results are also valid when we target a level of tracking error. Therefore, we obtain a third rule of smart beta indexing:

Rule 3 *When we impose an objective of volatility reduction or tracking error, the smart beta portfolios becomes comparable.*

[8] We have $\rho_{i,j}^{\Delta_t \beta} \simeq \rho_{i,j}^{\Delta_t \mathcal{VR}}$.

[9] It is equal to 84% on average.

[10] For the risk statistics ($\mathcal{VR}$, TE, β, $\mathcal{D}_w$, $\mathcal{D}_{rc}$ and $\mathcal{D}_\rho$), we consider the monthly series. The correlation between returns R_t is computed using the daily series of the one-year performance.

Table 5: Average correlation between GMV, ERC and MDP portfolios (in %)

Index	$\eta^\star$	$\mathcal{VR}$	TE	β	$\mathcal{D}_w$	$\mathcal{D}_{rc}$	$\mathcal{D}_\rho$	R_t
SX5E	5%	100.0	99.2	100.0	99.3	99.5	99.8	100.0
	10%	100.0	92.1	99.5	86.7	71.6	98.9	99.8
	15%	100.0	91.5	97.4	88.6	76.4	97.2	99.2
TPX100	5%	100.0	99.8	100.0	99.7	99.8	99.9	100.0
	10%	100.0	88.3	98.9	89.1	65.0	98.2	100.0
	15%	100.0	91.5	97.6	92.7	78.4	97.5	99.9
SPX	5%	100.0	96.8	99.8	86.4	63.6	98.2	99.8
	10%	100.0	86.9	97.1	88.4	69.7	93.4	99.0
	15%	100.0	85.6	90.8	88.9	77.6	88.4	97.6
MXEF	5%	100.0	100.0	100.0	99.9	100.0	100.0	100.0
	10%	100.0	100.0	100.0	98.2	99.5	99.8	100.0
	15%	100.0	99.9	100.0	96.1	95.0	99.5	100.0
Average		100.0	94.3	98.4	92.8	83.0	97.6	99.6

4.3 Performance of the smart beta portfolios

Rule 4 *The performance of smart beta portfolios depends on the market risk premium. When this is high, it is better to consider an objective of low volatility reduction (or tracking error volatility). Conversely, it is preferable to target a high volatility reduction when the market risk premium is weak or negative.*

This rule is very logical and easy to understand. Indeed, when the performance of stocks is high, it is better to invest in a more diversified portfolio than the CW portfolio, but with a limited tracking error in order to fully benefit from the bull market. Conversely, in a bear market, a concentrated portfolio of low volatility stocks will do a better job. In Figures 12, 13 and 14, we have reported the relationship[11] between the volatility reduction (in %) of smart beta portfolios and their excess return (in %) measured as the difference between the annualized return and the risk-free rate. The excess return for the CW index corresponds to the horizontal dashed line. Results for the entire study period (January 2001 – December 2014) are given in Figure 12. We notice that the rule is satisfied except for the MSCI EM index. If we consider the financial crisis (July 2007 – February 2009), we observe a positive relationship between volatility reduction and excess return (Figure 13). For this period, it is therefore better to target a high volatility reduction. The opposite is true if we consider the recent recovery period (March 2009 – December 2013).

4.4 Dynamic smart beta strategies

The previous rule can be used to build dynamic smart beta strategies. The idea is to fix the level of volatility reduction with respect to market conditions. If the risk sentiment is high, we would like to have an aggressive portfolio or to target a high level of volatility reduction. If the risk sentiment is low, it is better to have a more diversified portfolio with low tracking error with respect to the CW index.

[11] Because of the third rule, we know that the calibration method to target a given volatility reduction has little impact, particularly on the performance. This is why we only report the results with the GMV approach when the calibration is done by estimating λ_{h}. We obtain the same results if we consider other calibrations schemes.

Figure 12: Relationship between volatility reduction and excess return (2001-2014)

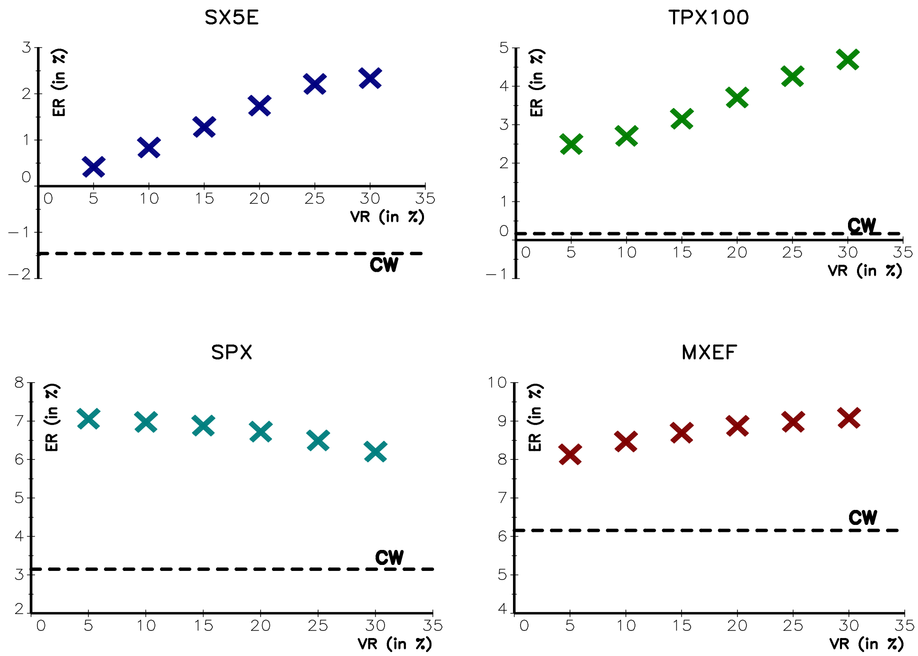


Figure 13: Relationship between volatility reduction and excess return (Jul. 2007–Feb. 2009)

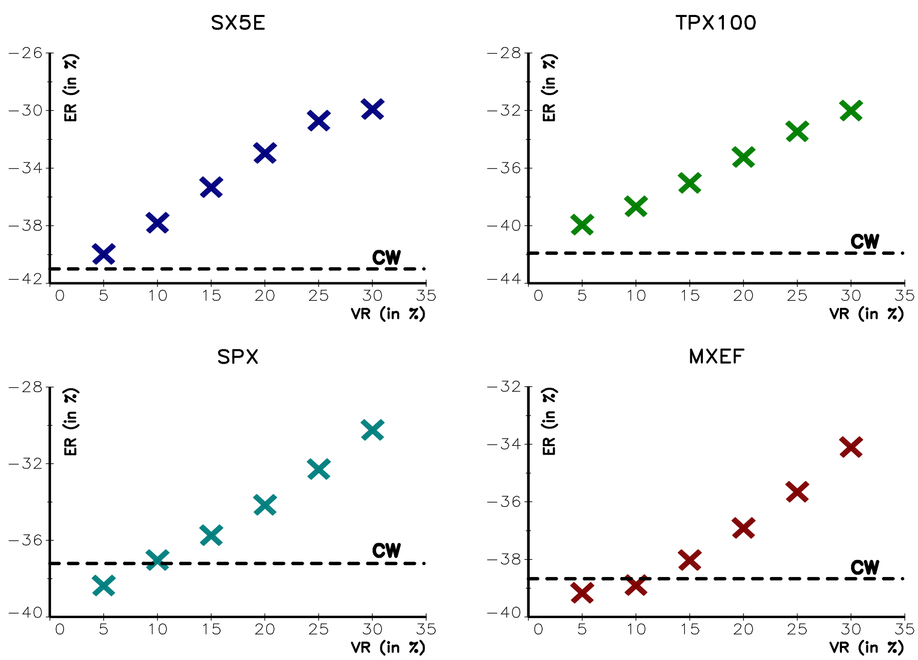
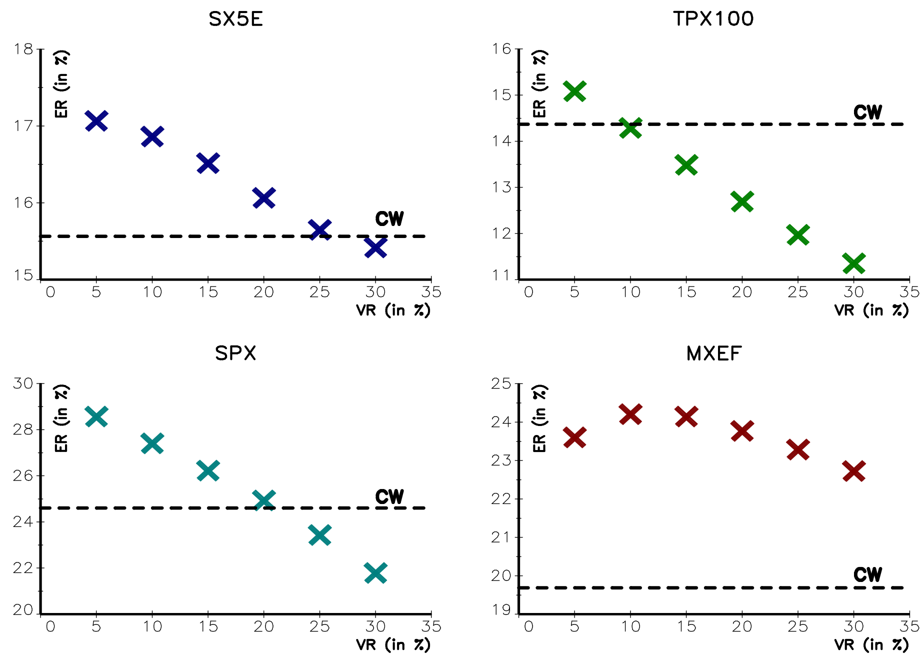


Figure 14: Relationship between volatility reduction and excess return (Mar. 2009–Dec. 2013)



We consider the optimization problem (14) with $\lambda \in [0,1]$, $\gamma = 1$ and $\delta = 1$. In this case, we obtain smart beta portfolios between the GMV portfolio ($\lambda = 0$) and the ERC portfolio ($\lambda = 1$). At each date t, we estimate the market sentiment by computing the cross-sectional volatility σ_t^{csv} of stocks which belong to the CW index[12]. We then consider the following rule to fix λ:

$$\lambda = 1 - \phi \frac{\sigma_t^{\text{csv}} - \sigma_t^-}{\sigma_t^+ - \sigma_t^-}$$

where σ_t^- and σ_t^+ are the minimum and maximum values of σ_t^{csv} observed for the window period $[t-h;t]$ and ϕ is a scalar between 0 and 1. We consider two strategies. $\text{D}_{\#1}$ corresponds to the case $\phi = 1$ and $\lambda \in [0,1]$. For the second strategy $\text{D}_{\#2}$, ϕ is equal to 0.85 meaning that $\lambda \in [0.15,1]$. Backtests[13] for the study period 2001-2014 are reported in Table 6. For each strategy, we calculate the annualized return $\mu(x)$, the annual volatility $\sigma(x)$, the corresponding Sharpe ratio $\text{SR}(x)$, the maximum drawdown $\mathcal{DD}(x)$ and the turnover $\tau(x)$ of the allocation. We notice that the dynamic smart beta strategies $\text{D}_{\#1}$ and $\text{D}_{\#2}$ improve the performance of the GMV and ERC portfolios for three indices (Eurostoxx 50, S&P 500 and MSCI EM). In particular, the second strategy $\text{D}_{\#2}$ is a considerable improvement on the ERC strategy with a higher return, a lower volatility, a reduced drawdown and a limited turnover[14]. Even this application is a toy model as it gives some indications about the benefit of dynamically managing the volatility reduction with respect to the market sentiment.

[12] In order to reduce the noise, we also apply an exponentially weighted moving average with a smoothing coefficient of 0.98 to the cross-sectional volatility.

[13] The lag window h is equal to one year.

[14] On average, its turnover is twice the turnover of the ERC portfolio, but half that of the GMV portfolio.

Table 6: Comparing GMV, ERC and dynamic smart beta strategies (2001-2014)

	CW	GMV	ERC	$D_{\#1}$	$D_{\#2}$	CW	GMV	ERC	$D_{\#1}$	$D_{\#2}$
			SX5E					TPX100		
$\mu(x)$	0.6	3.8	3.4	5.1	4.7	0.4	6.3	3.3	3.6	3.2
$\sigma(x)$	24.5	19.1	23.1	21.3	22.4	24.4	16.3	21.3	18.9	19.8
$\mathrm{SR}(x)$	−0.1	0.1	0.1	0.1	0.1	0.0	0.4	0.1	0.2	0.1
$\mathcal{DD}(x)$	−59.6	−52.4	−54.4	−50.7	−51.5	−62.8	−49.4	−57.4	−51.1	−54.2
$\tau(x)$	0.2	3.4	0.8	3.0	1.9	0.3	3.8	1.0	2.9	1.8
			SPX					MXEF		
$\mu(x)$	5.0	8.3	9.9	11.5	10.5	8.0	12.0	10.8	14.3	12.6
$\sigma(x)$	20.1	12.2	19.2	16.2	18.2	21.6	9.4	16.3	13.0	14.3
$\mathrm{SR}(x)$	0.2	0.5	0.4	0.6	0.5	0.3	1.1	0.6	1.0	0.8
$\mathcal{DD}(x)$	−55.3	−33.3	−55.9	−44.7	−52.5	−65.1	−29.9	−53.8	−34.9	−44.9
$\tau(x)$	0.1	5.9	1.0	3.5	1.6	0.5	5.6	1.6	4.2	2.8

5 Conclusion

Smart beta indexing is becoming increasingly popular with institutional investors and pension funds. It is perceived as a method of reducing risk and increasing performance with respect to capitalization-weighted indexing. However, there are many ways to build a smart beta portfolio. One interesting property is that these different alternative-weighted portfolios belong to the same optimization problem family. They are minimum variance portfolios and differ because of the implied constraint they consider. In this article, we develop a unified analytical framework based on the CCD algorithm in order to show the trade-off between the volatility reduction and the risks of such alternative-weighted solutions. Using this approach, we can illustrate and understand the behaviorial differences of smart beta portfolios. We can also develop new smart beta strategies by explicitly targeting a level of volatility reduction or by dynamically linking this level to the market sentiment.

References

[1] Cazalet Z., Grison P. and Roncalli T. (2014), The Smart Beta Indexing Puzzle, *Journal of Index Investing*, 5(1), pp. 97-119.

[2] Choueifaty Y. and Coignard Y. (2008), Toward Maximum Diversification, *Journal of Portfolio Management*, 35(1), pp. 40-51.

[3] Friedman J., Hastie T. and Tibshirani R. (2010), Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of Statistical Software*, 33(1), pp. 1-22.

[4] Griveau-Billion T., Richard J-C. and Roncalli T. (2013), A Fast Algorithm for Computing High-Dimensional Risk Parity Portfolios, *SSRN*, `www.ssrn.com/abstract=2325255`.

[5] Jensen M.C. (1968), The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance*, 23(2), pp. 389-416.

[6] Jurczenko E., Michel T. and Teiletche J. (2013), Generalized Risk-Based Investing, *SSRN*, `www.ssrn.com/abstract=2205979`.

[7] Maillard S., Roncalli T. and Teiletche J. (2010), The Properties of Equally Weighted Risk Contributions Portfolios, *Journal of Portfolio Management*, 36(4), pp. 60-70.

[8] Roncalli T. (2013), *Introduction to Risk Parity and Budgeting*, Chapman & Hall/CRC Financial Mathematics Series.

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

[10] Tseng P. (2001), Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, *Journal of Optimization Theory and Applications*, 109(3), pp. 475-494.

A Appendix

A.1 Managing the tracking error volatility

Let x_{cw} be the capitalization-weighted portfolio. The tracking error variance of the portfolio x is:

$$\begin{aligned}\sigma^2\left(x \mid x_{\text{cw}}\right) &= \left(x - x_{\text{cw}}\right)^\top \Sigma \left(x - x_{\text{cw}}\right) \\ &= x^\top \Sigma x - 2x^\top \Sigma x_{\text{cw}} + x_{\text{cw}}^\top \Sigma x_{\text{cw}}\end{aligned}$$

Because $x_{\text{cw}}^\top \Sigma x_{\text{cw}}$ is constant, the optimization problem becomes

$$x^\star\left(c_1, c_2\right) = \arg\min \frac{1}{2} x^\top \Sigma x - x^\top \Sigma x_{\text{cw}} \tag{15}$$

We recognize a Markowitz optimization problem where the expected returns μ are equal to Σx_{cw}. We notice that these expected returns are exactly the implied expected returns in the Black-Litterman model[15]. Following Roncalli (2013), we can transform the optimization problem (15) into a μ-problem:

$$\begin{aligned}x^\star\left(c\right) &= \arg\min \frac{1}{2} x^\top \Sigma x \\ &\text{u.c.} \quad \left\{\begin{array}{l} x \in \mathcal{C} \\ \sum_{i=1}^n x_i \left(\Sigma x_{\text{cw}}\right)_i \geq c \\ \mathbf{1}^\top x = 1 \\ x \geq \mathbf{0} \end{array}\right.\end{aligned} \tag{16}$$

with $c \in \left[0, c^+\right]$ with $c^+ = x_{\text{cw}}^\top \Sigma x_{\text{cw}}$. This problem is precisely the formulation (1) used in this paper by adding the constraint $\sum_{i=1}^n x_i \left(\Sigma x_{\text{cw}}\right)_i \geq c$. The limit cases are $x^\star\left(0\right) = x_{\text{gmv}}$ and $x^\star\left(c^+\right) = x_{\text{cw}}$.

We can use this framework to introduce the tracking error constraint in the different optimization problems considered in this study. For instance, we can mix this constraint with the ERC constraint. In this case, we will obtain optimized portfolios with a trade-off between the tracking error volatility and the diversification in terms of risk contributions.

[15] See Roncalli (2013) on page 23.

A.2 Solving the general optimization problem using the CCD algorithm

We consider the following optimization problem:

$$\begin{aligned}
x^{\star}\left(\lambda_{\mathrm{gmv}},\lambda_{\mathrm{h}},\lambda_{\mathrm{erc}},\lambda_{\mathrm{mdp}},\lambda_{\mathrm{te}}\right) &= \arg\min \frac{1}{2} x^{\top}\Sigma x - \\
&\quad \lambda_{\mathrm{gmv}}\left(\sum_{i=1}^{n} x_i\right) + \lambda_{\mathrm{h}}\left(\sum_{i=1}^{n} x_i^2\right) - \\
&\quad \lambda_{\mathrm{erc}}\left(\sum_{i=1}^{n} \ln x_i\right) - \lambda_{\mathrm{mdp}}\left(\sum_{i=1}^{n} x_i\sigma_i\right) - \\
&\quad \lambda_{\mathrm{te}}\left(\sum_{i=1}^{n} x_i\left(\Sigma x_{\mathrm{cw}}\right)_i\right) \\
\text{u.c.} &\quad x \geq \mathbf{0}
\end{aligned} \tag{17}$$

This formulation encompasses the different optimization problems presented in this article. We notice that Problem (17) is of the form:

$$\begin{aligned}
x^{\star}(\lambda) &= \arg\min \frac{1}{2} x^{\top}\Sigma x + \lambda\mathcal{P}(x) \\
\text{u.c.} &\quad x \geq \mathbf{0}
\end{aligned}$$

where $\mathcal{P}(x)$ is a penalty function combining different norms. This penalized optimization is frequent in machine learning and is generally solved using the cyclical coordinate descent algorithm.

A.2.1 CCD algorithm

The main idea behind the cyclical coordinate descent (CCD) algorithm is to minimize a function $f(x_1,\ldots,x_n)$ by minimizing only one direction at each step, whereas classical descent algorithms consider all the directions at the same time. In this case, we find the value of x_i which minimizes the objective function by considering the values taken by x_j for $j \neq i$ as fixed. The procedure is repeated for each direction until the global minimum is reached. This method uses the same principles as Gauss-Seidel or Jacobi algorithms for solving linear systems. The main objective is then to find the update rule.

Convergence of coordinate descent methods requires that $f(x)$ is strictly convex and differentiable. However, Tseng (2001) has extended the convergence properties to a non-differentiable class of functions:

$$f(x_1,...,x_n) = f_0(x_1,...,x_n) + \sum_{k=1}^{m} f_k(x_1,...,x_n)$$

where f_0 is strictly convex and differentiable and the functions f_k are non-differentiable.

Some properties make this algorithm very attractive. First of all, it is very simple to understand and implement. Moreover, the method is efficient for solving large-scale problems. That is why it is used in machine learning theory for computing constrained regressions or supporting vector machine problems (Friedman *et al.*, 2010). A further advantage is that the method does not need stepsize descent tuning as opposed to gradient based methods.

A.2.2 Application to the smart beta problem

In Problem (17), $f_0(x) = \frac{1}{2}x^\top \Sigma x$ is strictly convex and the functions f_k are non-differentiable, meaning that we can apply the CCD algorithm. Let $\mathcal{L}(x)$ be the Lagrange function (17). We have:

$$\frac{\partial\,\mathcal{L}(x)}{\partial\,x_i} = (\Sigma x)_i - \lambda_{\mathrm{gmv}} + 2\lambda_{\mathrm{h}} x_i - \frac{\lambda_{\mathrm{erc}}}{x_i} - \lambda_{\mathrm{mdp}}\sigma_i - \lambda_{\mathrm{te}}(\Sigma x_{\mathrm{cw}})_i$$

Let us assume that $\lambda_{\mathrm{erc}} > 0$. At the optimum, we have $\partial_{x_i}\,\mathcal{L}(x) = 0$ or:

$$x_i(\Sigma x)_i - \lambda_{\mathrm{gmv}} x_i + 2\lambda_{\mathrm{h}} x_i^2 - \lambda_{\mathrm{erc}} - \lambda_{\mathrm{mdp}} x_i \sigma_i - \lambda_{\mathrm{te}} x_i (\Sigma x_{\mathrm{cw}})_i = 0 \tag{18}$$

It follows that:

$$x_i^2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right) + x_i\left(\sigma_i \sum_{j\neq i} x_j \rho_{i,j}\sigma_j - \lambda_{\mathrm{gmv}} - \lambda_{\mathrm{mdp}}\sigma_i - \lambda_{\mathrm{te}}(\Sigma x_{\mathrm{cw}})_i\right) - \lambda_{\mathrm{erc}} = 0$$

We notice that the polynomial function is convex because we have $\sigma_i^2 + 2\lambda_{\mathrm{h}} > 0$. Since the product of the roots is negative[16], we always have two solutions with opposite signs. We deduce that the solution is the positive root of the second degree equation:

$$\begin{aligned} x_i^\star &= \frac{\lambda_{\mathrm{gmv}} + \lambda_{\mathrm{mdp}}\sigma_i + \lambda_{\mathrm{te}}(\Sigma x_{\mathrm{cw}})_i - \sigma_i \sum_{j\neq i} x_j\rho_{i,j}\sigma_j}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \\ &+ \frac{\sqrt{\left(\sigma_i \sum_{j\neq i} x_j\rho_{i,j}\sigma_j - \lambda_{\mathrm{gmv}} - \lambda_{\mathrm{mdp}}\sigma_i - \lambda_{\mathrm{te}}(\Sigma x_{\mathrm{cw}})_i\right)^2 + 4\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)\lambda_{\mathrm{erc}}}}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \end{aligned} \tag{19}$$

If the values of $(x_1, \ldots, x_n)$ are strictly positive, it follows that $x_i^\star$ is strictly positive. The positivity of the solution is then achieved after each iteration if the starting values are positive. The coordinate-wise descent algorithm consists in iterating Equation (19) until convergence and normalizing the solution at the final step.

Remark 6 *When the correlation of the assets is equal to zero, we obtain a closed-form expression:*

$$\begin{aligned} x_i^\star &= \frac{\lambda_{\mathrm{gmv}} + \lambda_{\mathrm{mdp}}\sigma_i + \lambda_{\mathrm{te}} x_{\mathrm{cw},i}\sigma_i^2}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \\ &+ \frac{\sqrt{\left(\lambda_{\mathrm{gmv}} + \lambda_{\mathrm{mdp}}\sigma_i + \lambda_{\mathrm{te}} x_{\mathrm{cw},i}\sigma_i^2\right)^2 + 4\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)\lambda_{\mathrm{erc}}}}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \end{aligned}$$

Remark 7 *We can deduce the risk contributions of risk-based portfolios from Equation (18):*

$$\mathcal{RC}_i \propto \lambda_{\mathrm{gmv}} x_i - \lambda_{\mathrm{h}} x_i^2 + \lambda_{\mathrm{erc}} + \lambda_{\mathrm{mdp}}\sigma_i x_i + \lambda_{\mathrm{te}} x_i (\Sigma x_{\mathrm{cw}})_i$$

We retrieve the different well-known results. For instance, the risk contributions are equal for the ERC portfolio, correspond to the weights for the GMV portfolio and are proportional to $x_i\sigma_i$ for the MDP, etc.

A.2.3 Special cases

In this section, we derive limit cases of Problem (17) by using the CCD formulation of the solution.

[16] We have $-\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)\lambda_{\mathrm{erc}} < 0$.

EW portfolio If we assume that $\lambda_{\mathrm{mdp}} = \lambda_{\mathrm{te}} = \lambda_{\mathrm{h}} = 0$ and $\lambda_{\mathrm{erc}} > 0$, the solution is reduced to:

$$\begin{aligned} x_i^\star &= -\frac{\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \lambda_{\mathrm{gmv}}}{2\sigma_i^2} \\ &+ \frac{\sqrt{\left(\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \lambda_{\mathrm{gmv}}\right)^2 + 4\sigma_i^2 \lambda_{\mathrm{erc}}}}{2\sigma_i^2} \end{aligned}$$

We have $\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \lambda_{\mathrm{gmv}} \approx |\lambda_{\mathrm{gmv}}|$ when $\lambda_{\mathrm{gmv}} = -\infty$. Using a first-order Taylor expansion in the neighborhood of zero, we obtain:

$$\begin{aligned} \lim_{\lambda_{\mathrm{gmv}} \to -\infty} x_i^\star &= \lim_{\lambda_{\mathrm{gmv}} \to -\infty} \frac{-|\lambda_{\mathrm{gmv}}| + |\lambda_{\mathrm{gmv}}| \sqrt{1 + \frac{4\sigma_i^2 \lambda_{\mathrm{erc}}}{|\lambda_{\mathrm{gmv}}|^2}}}{2\sigma_i^2} \\ &\approx \lim_{\lambda_{\mathrm{gmv}} \to -\infty} \frac{-|\lambda_{\mathrm{gmv}}| + |\lambda_{\mathrm{gmv}}| \left(1 + \frac{2\sigma_i^2 \lambda_{\mathrm{erc}}}{|\lambda_{\mathrm{gmv}}|^2}\right)}{2\sigma_i^2} \\ &= \frac{\lambda_{\mathrm{erc}}}{|\lambda_{\mathrm{gmv}}|} \end{aligned}$$

This means that all the weights are constant and equal. We finally obtain the equally-weighted portfolio:

$$\lim_{\lambda_{\mathrm{gmv}} \to -\infty} x^\star = x_{\mathrm{ew}} = \frac{1}{n}$$

There is another way to find the EW portfolio thanks to the Herfindahl index. If we assume that $\lambda_{\mathrm{mdp}} = \lambda_{\mathrm{te}} = \lambda_{\mathrm{gmv}} = 0$ and $\lambda_{\mathrm{erc}} > 0$, the solution is reduced to:

$$\begin{aligned} x_i^\star &= -\frac{\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \\ &+ \frac{\sqrt{\left(\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j\right)^2 + 4\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)\lambda_{\mathrm{erc}}}}{2\left(\sigma_i^2 + 2\lambda_{\mathrm{h}}\right)} \end{aligned}$$

When $\lambda_{\mathrm{h}} \gg +\infty$ we obtain:

$$x_i^\star \approx \frac{\sqrt{\lambda_{erc}}}{\sqrt{2\lambda_{\mathrm{h}}}}$$

Again all the weights are constant and we obtain the equally-weighted portfolio.

RP portfolio If we assume that $\lambda_{\mathrm{gmv}} = \lambda_{\mathrm{te}} = \lambda_{\mathrm{h}} = 0$ and $\lambda_{\mathrm{erc}} > 0$, we have $\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \lambda_{\mathrm{mdp}} \sigma_i \approx |\lambda_{\mathrm{mdp}}| \sigma_i$ when $\lambda_{\mathrm{mdp}} = -\infty$ and:

$$\lim_{\lambda_{\mathrm{mdp}} \to -\infty} x_i^\star = \frac{\lambda_{\mathrm{erc}}}{|\lambda_{\mathrm{mdp}}| \sigma_i}$$

This means that the weight is inversely proportional to the asset volatility. We then obtain the risk parity portfolio:

$$\lim_{\lambda_{\mathrm{mdp}} \to -\infty} x^\star = x_{\mathrm{rp}} = \frac{\sigma^{-1}}{\mathbf{1}^\top \sigma^{-1}}$$

BP portfolio If we assume that $\lambda_{\text{gmv}} = \lambda_{\text{mdp}} = \lambda_{\text{h}} = 0$ and $\lambda_{\text{erc}} > 0$, we have[17] $\sigma_i \sum_{j \neq i} x_j \rho_{i,j} \sigma_j - \lambda_{\text{te}} (\Sigma x_{\text{cw}})_i \approx |\lambda_{\text{te}}| (\Sigma x_{\text{cw}})_i$ when $\lambda_{\text{te}} = -\infty$ and:

$$\lim_{\lambda_{\text{te}} \to -\infty} x_i^\star = \frac{\lambda_{\text{erc}}}{|\lambda_{\text{te}}| (\Sigma x_{\text{cw}})_i}$$

The beta $\beta_i (x_{\text{cw}})$ of the asset i with respect to the CW portfolio x_{cw} is:

$$\beta_i (x_{\text{cw}}) = \frac{(\Sigma x_{\text{cw}})_i}{x_{\text{cw}}^\top \Sigma x_{\text{cw}}}$$

We deduce that the weight is inversely proportional to the asset beta. We finally obtain the 'beta parity' portfolio:

$$\lim_{\lambda_{\text{te}} \to -\infty} x^\star = x_{\text{bp}} = \frac{\beta^{-1} (x_{\text{cw}})}{\sum_{j=1}^n \beta_j^{-1} (x_{\text{cw}})}$$

Summary Finally, the different limit cases are reported in Table 7 where $\lambda \geq 0$ is an arbitrary constant.

Table 7: Limits of the smart beta portfolio $x^\star (\lambda_{\text{gmv}}, \lambda_{\text{h}}, \lambda_{\text{erc}}, \lambda_{\text{mdp}}, \lambda_{\text{te}})$

Parameters	GMV	EW		ERC	MDP	RP	BP	CW
λ_{gmv}	$+\infty$	$-\infty$	0	0	0	0	0	0
λ_{h}	0	0	$+\infty$	0	0	0	0	0
λ_{erc}	λ	$+\infty$	λ	λ	λ	$+\infty$	$+\infty$	λ
λ_{mdp}	0	0	0	0	$+\infty$	$-\infty$	0	0
λ_{te}	0	0	0	0	0	0	$-\infty$	$+\infty$

[17] We must also have $(\Sigma x_{\text{cw}})_i \geq 0$.