w o r k i n g
p a p e r

16 28

Undiversifying during Crises: Is It a Good Idea?

Margherita Giuzio and Sandra Paterlini

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 16-28 December 2016

Undiversifying during Crises: Is It a Good Idea?

Margherita Giuzio and Sandra Paterlini

High levels of correlation among financial assets, as well as extreme losses, are typical during crisis periods. In such situations, quantitative asset allocation models are often not robust enough to deal with estimation errors and lead to identifying underperforming investment strategies. It is an open question if in such periods, it would be better to hold diversified portfolios, such as the equally weighted, rather than investing in few selected assets. In this paper, we show that alternative strategies developed by constraining the level of diversification of the portfolio, by means of a regularization constraint on the sparse l_q-norm of portfolio weights, can better deal with the trade-off between risk diversification and estimation error. In fact, the proposed approach automatically selects portfolios with a small number of active weights and low risk exposure. Insights on the diversification relationships between the classical minimum variance portfolio, risk budgeting strategies, and diversification-constrained portfolios are also provided. Finally, we show empirically that the diversification-constrained-based l_q-strategy outperforms state-of-art methods during crises, with remarkable out-of-sample performance in risk minimization.

Keywords: minimum variance portfolio, sparsity, diversification, regularization methods.

JEL classification: G11, C58.

Suggested citation: Giuzio, Margherita, and Sandra Paterlini, "Undiversifying during Crises: Is It a Good Idea?" Federal Reserve Bank of Cleveland, Working Paper no. 16-28.

Margherita Giuzio (margherita.giuzio@ebs.edu) and Sandra Paterlini (sandra.paterlini@ebs.edu) are at EBS Universität für Wirtschaft und Recht.

1 Introduction

Diversification is one of the core principles behind the development of effective asset allocation strategies. The debate between Warren Buffet, who promotes careful selection of few investments, resulting in concentrated portfolios (Buffet, 1979), and the use instead of diversified strategies, which allocate capital among a variety of assets, is still far from its end (Boyle et al., 2012). In fact, the simplest diversification strategy, namely the equally weighted portfolio (EW), which assigns the same weight to each security, is known to be a tough benchmark to beat, given its remarkable out-of-sample performance due to its shrinkage properties (De Miguel et al., 2009b). However, despite the simplicity and the absence of estimation errors, the EW strategy only allows for naive diversification, based on equal asset weights. It completely ignores the risk and the correlation between assets, which instead are the fundamentals of Markowitz's mean-variance theory (Markowitz, 1952). In the mean-variance framework, portfolio components are chosen according to their risk-return profile and their correlation with other securities. Optimal portfolios minimize the expected risk, given a minimum level of expected return, subject to a budget constraint. As proved by Carrasco and Noumon (2012), errors in estimating the expected return and risk significantly affect the composition and the stability of the optimal portfolios, especially when the pool of candidate assets is large and the problem is high dimensional (the number of observations is much lower than the number of assets). Furthermore, it has been shown that after reaching a certain portfolio size, portfolio risk cannot be further decreased by adding new components, as systematic risk cannot be diversified away (Jagannathan and Ma, 2003; De Miguel and Nogales, 2009; Brodie et al., 2009; Fan et al., 2012; Fastrich et al., 2015), suggesting that diversification, interpreted as an increasing number of active positions, is not always beneficial, as also discussed later in Section 3.

As Merton (1980), Chopra and Ziemba (1993) and Jagannathan and Ma (2003) have suggested long ago, estimation errors in the expected returns are much larger than those in the covariance matrix and result in portfolios sensitive to changes in assets means. Hence, as a first step, here we exclude such inputs from the objective function and focus on the well-known minimum variance (MV) and risk-parity frameworks. One important drawback of the MV allocation is that the portfolio is highly concentrated with extreme weights in few low volatile stocks, due to the fact that the only objective is to obtain the portfolio with the lowest risk possible. The risk-parity framework instead is a risk budgeting strategy, which allows for risk diversification in allocating wealth among different asset classes according to some risk targets (Maillard et al., 2010). In particular, the equal risk contribution strategy (ERC) selects portfolio weights inversely proportional to the estimated riskiness and can be shown to be equivalent to the mean-variance problem, in which all pairwise correlations and Sharpe Ratios of the assets are equal. Recently, risk-parity strategies have become increasingly attractive in portfolio allocation, as they provide benefits from reduced turnover and risk diversification (Bruder and Roncalli, 2012). On the other hand, in periods of financial distress and extreme losses, diversification strategies might not be robust to high levels of volatility and correlation among financial assets, leading to underperforming investment strategies. For this reason, in practice, many portfolio managers rather invest in concentrated portfolios, claiming that focusing on few securities yields better risk-returns performance, with lower trading and monitoring costs (Kacperczyk et al., 2005; Brands et al., 2005; Ivkovic et al., 2008).

Portfolio sparsity (i.e. a small number of non-zero weights) may be enhanced by means of the so-called regularization techniques, which allow to prevent data overfitting through a penalty term on the asset weights. One of the most popular methods nowadays is the so-called Lasso or ℓ_1-regularization, which relies on constraining the ℓ_1-norm of the asset weights[1] (Brodie et al., 2009; De Miguel et al., 2009a; Carrasco and Noumon, 2012; Fan et al., 2012). De Miguel and Nogales (2009) showed that portfolios with constraints on the ℓ_1 and ℓ_2-norms of the asset weights can outperform the EW strategy. In particular, the ℓ_1-regularization not only leads to sparse and stable portfolios, but also constraints their gross exposure, i.e. limits the amount of short-selling by shrinking the covariance matrix of asset returns. Indeed, when the penalty is included in the objective function, it can be shown that the approximation error is bounded and does not accumulate, resulting in more robust portfolio estimates (Fan et al., 2012). Still, even before the ℓ_1-method was formally introduced in portfolio selection, Jagannathan and Ma (2003) had already

[1] The ℓ_q-norm of a vector $\boldsymbol{w}$ of n elements is defined here, for $0 < q < \infty$, as $\ell_q = \|\boldsymbol{w}\|_q^q = \sum_{i=1}^n |w_i|^q$, with slight abuse of terminology. In fact, the ℓ_q-norm would be $\|\boldsymbol{w}\|_q = \left(\sum_{i=1}^n |w_i|^q\right)^{1/q}$. Note that for $0 < q < 1$, the q-norm $\|\boldsymbol{w}\|_q$ is a pseudo-norm.

discussed why no short-selling in presence of the budget constraint (i.e. the ℓ_1-norm is equal to 1) would allow to prevent overfitting and reduce the out-of-sample portfolio risk. However, the Lasso portfolios might provide biased estimates for large absolute weights and still might result in many active positions (Gasso et al., 2009). For these reasons, different regularization methods have been applied to portfolio selection recently, such as the elastic net (Zou and Hastie, 2005; Yen and Yen, 2014), which is a weighted sum of the ℓ_1 and ℓ_2-penalties, the ℓ_∞-norm[2] and other non-convex penalties, which are able to overcome this bias and obtain remarkable out-of-sample performance (Fastrich et al., 2014, 2015; Xing et al., 2014).

Alternatively, portfolio sparsity may be imposed by means of a cardinality constraint, which leads to select an optimal portfolio with a given (maximum) number of active positions and it is equivalent to constraining the ℓ_0-norm[2] of the asset weights. Numerous optimization approaches have then been proposed in the literature to deal with such non-convex constraints, even in high dimensions. Still, estimating the optimal number of active weights to use as upper bound for the ℓ_0-norm might be difficult, and other types of penalties might offer a better alternative to naturally deal with the trade-off between diversification and sparsity. Probably, one of the most interesting penalties is the so-called ℓ_q-norm with $0 < q \leq 1$. Recently, Chen et al. (2016) have introduced an optimization approach including the ℓ_q-norm constraint in the mean-variance framework, which allows to trade-off between portfolio sparsity and good out-of-sample performance, measured by the level of Sharpe Ratio. The resulting portfolios obtain comparable performance to cardinality constrained portfolios, even when accounting for transaction costs.

Here, we provide a financial interpretation of the ℓ_q-norm as diversification constraint and discuss its effectiveness when compared to other types of penalties. Moreover, we show that the approximation error bound is tighter for $0 < q < 1$ than for $q \geq 1$. Furthermore, we interpret the ℓ_q-approach in a Bayesian framework, by pointing out explicitly the prior on the distribution of portfolio weights. Also, we analyse the behavior of the ℓ_q-penalty in terms of sparsity, amount of shorting and diversification, and show that it better deals with the trade-off between number of active components, amount of shorting, size of asset weights and portfolio diversification compared

[2] Given a vector $\boldsymbol{w}$ of n elements, $\ell_\infty = \|\boldsymbol{w}\|_\infty = \max(|w_1|, \dots, |w_n|)$ and $\ell_0 = \|\boldsymbol{w}\|_0 = \sum_{i=1}^{n} \mathbb{1}(w_i \neq 0)$.

to other regularization methods, by penalizing both concentrated solutions with extreme short positions and fully diversified allocations. Finally, we observe empirically that this characteristic is especially beneficial in periods of financial distress and bear markets, i.e. in presence of extreme losses and high positive correlation between assets.

The paper is structured as follows. Section 2 introduces the diversification and regularization strategies in the minimum-variance framework and discusses their main characteristics. Section 3 shows the out-of-sample performance of the different strategies on real-data and points out the trade-off between diversification and sparsity in low and highly volatile markets. Section 4 concludes.

2 Methods

2.1 Risk Diversification Strategies

Markowitz (1952) is considered the pioneer of the concept of diversification in asset allocation, i.e. the idea that the overall portfolio risk can be reduced by investing in different assets, based on their volatilites and correlations. Let $\boldsymbol{w} = [w_1, \ldots, w_n]'$ be the $n \times 1$ vector of portfolio weights and $\boldsymbol{\Sigma} = [\sigma_{ij}]$ with $i, j = 1, \ldots, n$, the $n \times n$ covariance matrix of asset returns. Then, the so-called minimum variance portfolio can be computed by solving the following optimization problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} = \sigma_p^2 \tag{1}$$
$$\mathbf{1}'\boldsymbol{w} = 1$$

where $\mathbf{1}$ is a $n \times 1$ vector of ones and σ_p^2 is the portfolio variance. Here, asset returns are assumed to be normal and the dependence can then be fully captured by linear correlation. The minimum variance portfolio is then a quadratic optimization problem, which requires an estimate $\widehat{\boldsymbol{\Sigma}}$ of the covariance matrix $\boldsymbol{\Sigma}$. If no constraint on the weights is imposed, the problem has an analytic solution with no element equal to zero. However, this optimal solution, characterized by extreme weights, highly sensitive to correlation and estimation errors, can hardly be implemented in practice (Michaud, 1989). The EW strategy might then be preferred (De Miguel et al., 2009b), but it still implies to hold positions in a large number of assets, which is often unrealistic for many investors. Recently, Brodie et al. (2009), De Miguel and Nogales (2009) and Fan et al. (2012) pointed out that diversifying risk does not necessarily imply investing in a large number of assets, but identifying the ones that do not move together and are less risky. This may be particularly challenging in a period of financial distress, in presence of extreme losses, which result in an increased volatility and higher level of positive correlation between assets. Therefore, investors may not be able to fully exploit risk diversification when they need it most.

Understanding that the EW strategy allows diversifying the investment in terms of capital, but not necessarily in terms of risk (i.e. assets may contribute differently to the overall portfolio risk), Maillard et al. (2010) introduced the risk-parity portfolio, also called equal risk contribution (ERC). Such portfolio allows for risk diversification by allocating capital so that each asset contributes equally to the overall portfolio risk (see A for more details). In particular, the authors proposed an optimization framework to find risk-parity portfolios that minimizes the variance of the risk contributions between assets, i.e. minimizes deviations from the so-called risk-parity, such that

$$\begin{aligned} \min \; & \sum_{i=1}^{n} \sum_{j=1}^{n} \left(w_i(\Sigma w)_i - w_j(\Sigma w)_j\right)^2 \\ & w_i \geq 0, \qquad i = 1, \ldots, n \\ & \mathbf{1}'\boldsymbol{w} = 1 \end{aligned} \tag{2}$$

where $(\Sigma w)_i = \sum_{i,j=1}^{n} \sigma_{ij} w_j$ represents the marginal contribution of asset i to the overall portfolio risk σ_p^2. One of the problems arising from this formulation is that we need to minimize a non-convex function of the asset weights, which might be time-consuming and present some convergence issues for problems with a large number of assets. Interestingly, Kolm et al. (2014) showed that in the long-only case the risk-parity portfolio can be computed by solving the following optimization problem:

$$
\begin{aligned}
&\min \;\; \boldsymbol{w}'\Sigma\boldsymbol{w} - \log(\boldsymbol{w}) \\
&w_i \geq 0, \qquad i = 1, \ldots, n \\
&\mathbf{1}'\boldsymbol{w} = 1
\end{aligned}
\tag{3}
$$

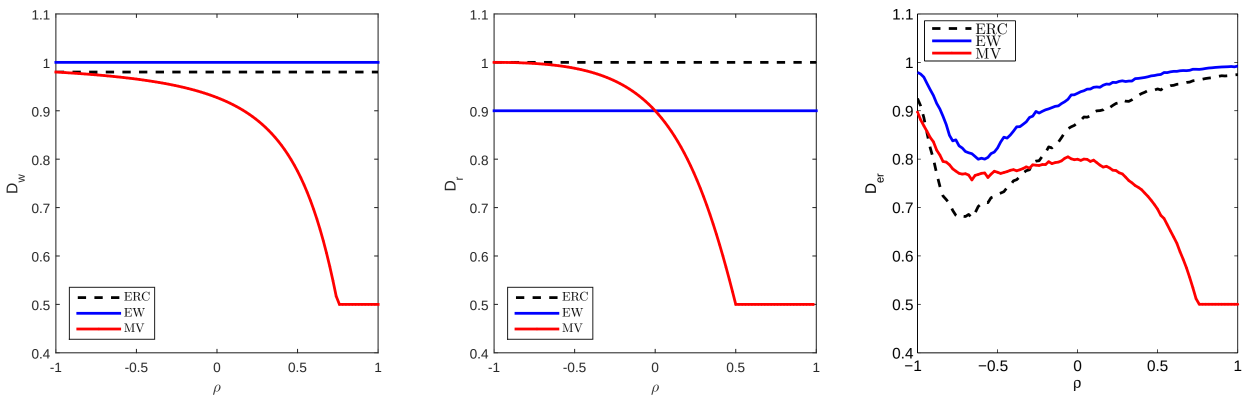
where $\log(\boldsymbol{w})$ represents a logarithmic penalty on the vector of non-negative portfolio weights. Notice that then problem (3) can be interpreted as a penalized minimum variance problem, with penalty equal to $-\log(\boldsymbol{w})$, which does not promote sparsity as pre-multiplies by -1 (see Section 2.2).

2.1.1 Diversification Measures.

The concept of diversification in portfolio allocation might be difficult to capture by a single definition. Hence, different measures have been introduced to quantify it. The number of active positions, the concentration of the weights (i.e. from $1/n$ for EW to 1 for a portfolio totally invested in one asset) and different risk functions play a role in assessing diversification. In order to evaluate the diversification performance of a portfolio composed of n assets, we consider four measures: D_{ℓ_0}, D_w, D_r and D_{er}, defined as

$$
D_{\ell_0} = \frac{\sum_{i=1}^n \mathbb{1}(w_i \neq 0)}{n}, \qquad D_w = \frac{1}{n\sum_{i=1}^n w_i^2}, \qquad D_r = \frac{1}{n\sum_{i=1}^n RC_i^2}, \qquad D_{er} = \frac{1}{n\sum_{i=1}^n CES_i^2},
$$

where w_i is the weight of the i-th asset and RC_i and CES_i are its risk contributions to the overall portfolio risk and Expected Shortfall, respectively (see A). The concentration index D_{ℓ_0} measures the proportion of active weights in a portfolio. Typically, the ℓ_0-norm or number of active positions is reported. Here, we build the D_{ℓ_0} index to be consistent with the other diversification measures by having the maximum value for a portfolio invested in all n assets ($D_{\ell_0} = 1$) and the minimum value for a portfolio totally concentrated in one position ($D_{\ell_0} = 1/n$). The weight diversification measure D_w relies on the popular Herfindal Index $H(\boldsymbol{w}) = \sum_{i=1}^n w_i^2$ to assess the level of weight concentration in a portfolio. D_w takes the value of $1/n$ if the portfolio is totally concentrated in one asset and the value of 1 if the portfolio is equally weighted. Starting from this index, Cazalet et al. (2014) introduced the risk diversification measure D_r, which has value of $1/n$ if the portfolio risk is completely determined by one asset and the value of 1 if portfolio risk is equally spread among assets. From a risk budgeting perspective, it may be useful to understand the composition of a portfolio also in terms of extreme risk. For this reason, we consider the extreme risk diversification measure D_{er}, where risk is quantified by the Expected Shortfall.[3] Similar to the other measures, D_{er} will take a value of $1/n$ if portfolio extreme risk is totally concentrated in one asset and a value of 1 if it is equally spread among assets.



(a) (b) (c)

Figure 1: Diversification indices D_w (panel (a)), D_r (panel (b)) and D_{er} (panel (c)) with respect to correlation coefficients ρ in EW, ERC and MV portfolios when $\sigma_1 = 0.6$, $\sigma_2 = 0.8$. Returns are simulated from a multivariate t-Student distribution with 6 degrees of freedom, for a total of 500 simulations with $T = 2000$ observations. D_{ℓ_0} is constant and equal to 0.5 for each ρ and strategy.

To better understand the differences between these measures, let's consider a small investment universe of only two assets, with volatilities σ_1 and σ_2. Figure 1 shows how the diversification measures change when the correlation between assets, ρ increases. We set $\sigma_1 = 0.6$ and $\sigma_2 = 0.8$. To compute D_{er}, returns are simulated from a multivariate t-Student distribution with six degrees of freedom, for a total of 500 simulations with $T = 2000$ observations. From Figure 1a, we notice that, as expected, the EW portfolio is the most disperse in terms of weight, as it assigns an equal position to both assets. The ERC portfolio reaches analogous results by selecting similar positions for the two assets.[4] On the contrary, the MV strategy returns more concentrated portfolios as ρ

[3]See Bauer and Zanjani (2016) for a discussion about risk exposures based on different risk measures.

[4]In the two-asset case, the ERC strategy selects the assets weight according to their volatilities. As the difference

increases, since it estimates larger absolute weights for the asset with the lowest volatility (σ_1 in our examples).

Figure 1b shows that the ERC strategy returns, as expected, the most disperse portfolio in terms of risk, since the two assets contribute equally to the overall portfolio volatility. Similar results are achieved by the EW portfolio. On the other hand, as ρ increases, the MV strategy concentrates the risk of the portfolio once again in the asset with the lowest volatility. Figure 1c shows how the average D_{er} changes with ρ in 500 simulations. We notice that the EW portfolio is also the most disperse in terms of extreme risk. As the correlation between assets increases, the Expected Shortfall of both EW and ERC portfolios become more diversified. The opposite is true for the MV portfolio. However, despite the good diversification properties, investing in these portfolios implies selecting an active position for each asset. D_{ℓ_0} would then be always equal to $1/n$ for each strategy and correlation level.

2.2 Regularization Strategies

One natural way to select optimal portfolios with few active weights, typically not extreme in size, is to include some penalty functions on the asset weight vector and then solve the following penalized minimum variance problem

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} + \lambda \sum_{i=1}^{n} g(w_i) \tag{4}$$

where $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ is the portfolio variance and λ is the scalar that controls the intensity of the penalty $g(\boldsymbol{w})$. Two interesting penalties are the ℓ_q with $0 < q < 1$ and the log-penalties specified below

$$\ell_q = \|\boldsymbol{w}\|_q^q \ = \sum_{i=1}^{n} |w_i|^q \tag{5}$$

$$\log = \sum_{i=1}^{n} (\log(|w_i| + \phi) - \log(\phi)) \tag{6}$$

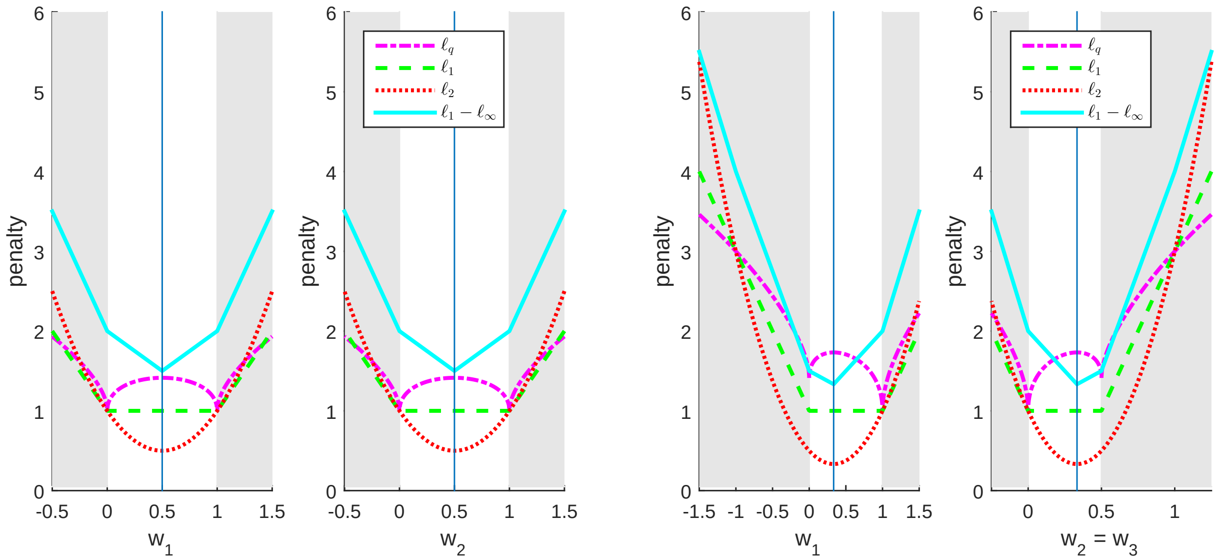
$\sigma_1 - \sigma_2$ is not large in our examples, $w_1 \simeq w_2$.

where $0 < \phi < 1$ is a constant such that the logarithmic function is defined also for $w_i = 0$. Notice that the $\lim_{q \to 0^+} \sum_{i=1}^{n} |w_i|^q = \|\boldsymbol{w}\|_0$ [5]. Moreover, it can be shown that when $q \to 0^+$, the ℓ_q-penalty converges to the log-penalty. Here, we will focus mostly on the ℓ_q-penalty as the results might be extended to the log-penalty as well. The ℓ_q with $0 < q < \infty$ was introduced in the statistical literature as Bridge regression (Frank and Friedman, 1993). Since then, many well-known penalties that belong to the ℓ_q-regularization framework have been studied, such as the well-known Lasso when $q = 1$ (i.e. ℓ_1-penalty). Knight and Fu (2000) and Huang et al. (2008) investigated the asymptotic properties of Bridge estimators in terms of sparsity, normality and consistency. In finance, Brodie et al. (2009), De Miguel et al. (2009a) and Fan et al. (2012) show that portfolios constructed by using ℓ_1 and ℓ_2-penalties achieve higher Sharpe Ratios than their benchmarks, such as EW and MV, and reduce the sensitivity of the weights to estimation errors.

Such results have attracted recently further research on the use of regularization methods in finance. Most studies (e.g. Xing et al., 2014) point out that ℓ_1 and ℓ_2-norms still have shortcomings, such as the fact that the ℓ_2-norm does not encourage sparsity, but typically assigns an active weight to all securities (De Miguel et al., 2009a). Selecting a larger number of assets though does not necessarily yield to lower risk. On the other hand, the Lasso penalty, despite promoting sparsity when $\lambda > 0$, provides often impractical solutions with large absolute positions, due to the bias on large coefficients, and it is ineffective when the no-short-selling constraint is imposed. As pointed out by Gasso et al. (2009) and Fastrich et al. (2014, 2015), non-convex penalties are able to reach much sparser allocations and overcome this bias. Chen et al. (2016) showed recently that the ℓ_q-penalty can achieve lower risk than the ℓ_1 with a smaller amount of shorting. Here, we focus on the ℓ_q-penalty with $0 < q < 1$ as it promotes sparsity by controlling for the trade-off between concentration and diversification.

Figure 2 displays the behavior of the ℓ_q-norm with respect to ℓ_1, ℓ_2 and $\ell_1 - \ell_\infty$-penalty (Xing et al., 2014), for portfolios of two and three assets in presence of the budget constraint. So far, most of the literature on regularized portfolios focuses on plotting the penalties without

[5] Imposing an upper bound on the 0-norm, such that $\|\boldsymbol{w}\|_0 \leq k$, results in the well-studied cardinality constraint, which is known to make the optimization problem NP-hard when n and k are large. Indeed, such constraint would require computing $\binom{n}{k}$ number of estimators, which grows exponentially with k.



(a) (b)

Figure 2: Geometric interpretation of the regularized portfolios with ℓ_q with $q = 0.5$, ℓ_1, ℓ_2 and $\ell_1 - \ell_\infty$-penalties, for 2 (panel (a)) and 3 assets (panel (b)) in presence of the budget constraint (i.e. $\sum_{i=1}^{n} w_i = 1$). The grey area represents solutions with short-selling, while the vertical blue line indicates the EW portfolio.

explicitly incorporating the budget constraint. Instead, we believe that Figure 2 allows to better understand the effect of such penalties in portfolio settings, where asset weights have to sum up to 1 and typically have values between -1 and 1. In the no-short-selling case (white area), we notice that the ℓ_q-penalty with $0 < q < 1$ reaches its maximum value ($\ell_q = n^{1-q}$) for the equally weighted portfolio, corresponding to the vertical blue line, and minimum value ($\ell_q = 1$) for a portfolio totally invested in only one asset, becoming then a natural way to control for weight diversification. The opposite is true for the $\ell_1 - \ell_\infty$- and ℓ_2-penalties, which assign higher values to more concentrated positions, such as $\boldsymbol{w}$ = [-1 0] or $\boldsymbol{w}$ =[0 1], than the EW $\boldsymbol{w}$ =[0.5 0.5] (see Figure 2a). As weights are larger than zero and sum to one, the ℓ_1-penalty is constant and therefore ineffective. When short positions are allowed (grey area), we notice that the ℓ_1, ℓ_2 and $\ell_1 - \ell_\infty$-norms increase approximately linearly with the amount of shorting, while the ℓ_q-norm assigns a lower penalty to large absolute weights.

Finally, we can see that, constraining the norms to be smaller than a value $c > 0$, sparsity is preferred only with ℓ_1 and ℓ_q-norms, while $\ell_1 - \ell_\infty$ and ℓ_2-norms reach their minimum for EW portfolios, as sparser solutions represent only local optima. However, notice that ℓ_1 is ineffective when no-short-selling is imposed. Furthermore, as shown in Figure 2b, the ℓ_q-penalty takes the same value for the EW portfolio and for solutions with small short positions. In such situations, the portfolio that minimizes Eq. (4) will be the one with the lowest variance, i.e. the one which invests in the least volatile assets by setting to zero, with equal strength, both positions with limited shorting and small positive weights. We will see that this is relevant especially in situations characterized by similar level of correlation, as typical during crisis (see Section 3). Sparse solutions are always local minima for the ℓ_q-approach, with at least one of them being global optimum. Then, a strong penalty for large amount of shorting is applied with the same intensity for EW and moderate shorting. Note that ℓ_q is the only approach that strongly penalizes both concentrated portfolios with extreme shorting (i.e. with low D_w) and fully diversified solutions, such as the EW (i.e. with high D_{ℓ_0}). This is an interesting property as, typically, portfolio managers want to avoid EW and large amount of shorting. Moreover, empirical results in Section 3 point out how the ℓ_q-sparsity property is useful to automatically select few assets, which are robust to the high volatilities and correlation settings typical of financial crises.

By trading-off between approximation and estimation errors, i.e. errors in selecting the optimal assets and in estimating the optimal weights, respectively, the ℓ_q-penalty is able to construct sparse portfolios that perform well out-of-sample. When the number of active positions increases, the approximation error monotonically decreases, since more assets are included. However, the estimation error increases because of overfitting. Following Leung and Barron (2006), we can link sparsity and approximation error of the estimates $\boldsymbol{w}$, obtained with ℓ_q-regularization strategies, with $0 < q \leq 2$, and show that the approximation error is more tightly bounded for $0 < q < 1$.

Proposition 1. *In presence of a no-short-selling constraint, i.e.* $0 \leq w_i \leq 1$, *with* $i = 1, \ldots, n$, *the* ℓ_q*-norm is maximum for the equally weighted portfolio and minimum for a portfolio totally invested in one asset, with bounds*

$$1 \leq \|\boldsymbol{w}\|_q^q \leq n^{1-q} \ . \tag{7}$$

Proposition 2. *Let* $\{w_{(j)}\}$ *be the* j*-th element of the vector of weights* $\{\boldsymbol{w}\}$ *sorted descendingly, such that* $|w_{(1)}| \geq |w_{(2)}| \geq \cdots \geq |w_{(n)}|$*, and* k *be the number of non-zero coefficients, i.e.* $k = \sum_{i=1}^{n} \mathbb{1}(w_i \neq 0)$*. Then, we can bound the approximation error as*

$$\sum_{j>k+1} w_{(j)}^2 \leq \frac{\|\boldsymbol{w}\|_q^2}{(k)^{(2-q)/q}} \ . \tag{8}$$

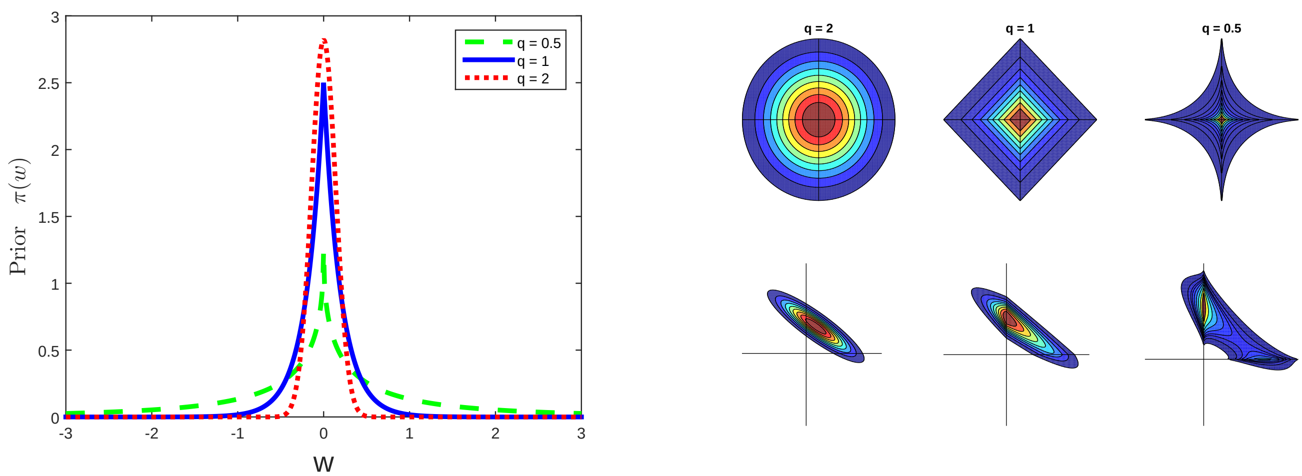
From this inequality, it is clear that the number of active components k plays an important role in bounding the approximation error: the larger k, the lower the bound. Moreover, the decrease in approximation error is much faster for $0 < q < 1$ than for $q = 1$, and when $q \to 0$, the error vanishes. For example, for $q = 1$, we get an upper bound equal to $\|\boldsymbol{w}\|_1^2/k$, while for $q = 0.5$ it is $\|\boldsymbol{w}\|_{0.5}^2/k^3$ (see proof in B).

2.3 Bayesian Interpretation

The ℓ_q-framework has a Bayesian interpretation by considering the maximum a posteriori (MAP) estimation of portfolio weights with prior probability distribution for $\pi(\boldsymbol{w})$ to be an exponential power distribution:

$$\pi(w_i|\mu, a, q) = \frac{q}{2a\Gamma(1/q)} \exp\left[-\left(\frac{|w_i - \mu|}{a}\right)^q\right], \qquad i = 1, \ldots, n \tag{9}$$

where μ is the location parameter, a the scale parameter and q the shape parameter controlling the kurtosis of the distribution (Seeger, 2008; Murphy, 2012). Therefore, if $q = 2$, the resulting prior would be a normal distribution, while if $0 < q < 2$, $\pi(\boldsymbol{w})$ would be leptokurtic, i.e. would present fatter tails than the normal case, as shown in Figure 3a. An interesting example is the Laplace distribution, obtained when $q = 1$, which is particularly important because it results in a convex optimization problem in Eq. (4), with sparse solutions. Non-convex penalties, i.e. ℓ_q-penalties with $0 < q < 1$, correspond to prior distributions with fat tails and assign higher probability to large asset weights, ensuring the selection of relevant coefficients while still inducing strong sparsity. Figure 3b shows the isocurves of the prior (first row) and posterior (second row) distributions, corresponding to the ℓ_q-penalties with $q = 2, 1, 0.5$, for a portfolio of two assets. In the first row, we notice that the normal density function ($q = 2$) distributes the probability homogeneously around the axes, while the other two priors ($q = 1$ and $q = 0.5$) attach more probability to solutions on the axes (i.e. where $w_1 = 0$ or $w_2 = 0$), resulting in sparse portfolios. This behavior is confirmed by the posterior density functions in the second row, obtained by multiplying the likelihood and the prior distribution. In particular, we notice that, while the normal posterior is symmetric, with its mode lying away from both axes, the Laplace posterior is slightly skewed and its mode lies on the vertical axis, corresponding to $w_1 = 0$. Interestingly, the exponential posterior is bimodal when $q = 0.5$ and shows a higher shrinkage towards the axes. In this case, as suggested by Seeger (2008) and Mazumder et al. (2011), the number of posterior modes may increase exponentially with the number of assets for $0 < q < 1$, typically ending up with sparser solutions than $q = 1$ with superior model selection properties while retaining similar prediction accuracy.



(a) (b)

Figure 3: Panel (a): Exponential power distributions for MAP estimation corresponding to ℓ_q-penalties with $q = 0.5, 1, 2$, $\mu = 0$ and $a = 1$. Panel (b) - First row: Isolines of the generalized normal distributions (priors) with different values of q, $\mu = 0$ and $a = 1$. Panel (b) - Second row: Isolines of the posterior distributions, obtained by multiplying the prior and the likelihood, for $q = 0.5, 1, 2$, $\mu = 0$ and $a = 1$. The areas between isocurves are filled according to the distribution's density, i.e. the brighter the color, the higher the density.

3 Empirical Results

In the following empirical analysis, we compare the behavior of the regularization and diversification strategies within the MV framework and evaluate their performance in terms of risk, sparsity and diversification in low and highly volatile markets. The investment goal is to obtain an optimal portfolio with a low level of risk and a good balance between weight and risk diversification. Therefore, the MV optimization problem with regularization and diversification criteria can be described by the penalized objective function in Eq. (4), subject to a budget constraint. In particular, we consider the following penalties: ℓ_2 (Ridge), ℓ_1 (Lasso), logarithmic (Log) and ℓ_q with $q = 0.5$ (ℓ_q).

We test the performance of the different strategies on two dataset including the daily returns of the Standard and Poor's 100 (S&P 100) and 500 (S&P 500) components, from 01.01.2005 to 29.05.2015. We leave out few minor constituents that do not belong to the indices throughout the whole sample period. Therefore, the dataset involves $T = 2734$ observations of $n = 93$ assets for the S&P 100 and $n = 452$ assets for the S&P 500. We divide the sample in two sub-periods of $T_1 = 1304$ observations from 01.01.2005 to 31.12.2009 including the financial crisis of 2007-2008, and $T_2 = 1430$ observations from 01.01.2010 to 29.05.2015 in the post-crisis period. Table 1 shows the average descriptive statistics of S&P 100 and S&P 500 components in the two sub-samples. The first period is characterized by high volatility and extreme losses, quantified by high levels of drawdowns (around 0.6 in both dataset) and large Values-at-Risk (VaR) and Expected Shortfalls (ES), as shown in Figure 4a. We estimate the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ with the shrinkage estimator introduced by Ledoit and Wolf (2004), with rolling windows of 500 in-sample observations. The condition number of $\widehat{\boldsymbol{\Sigma}}$ indicates that our estimate is highly sensitive to small changes in the inputs, in particular during and immediately after the financial crisis (see Figure 4b), implying high level of correlation between assets returns in such periods. Figure 5 displays the correlation matrices between asset returns during (left panels) and after (right panels) the financial crisis. We notice that assets tend to behave more similarly during the crisis, indicating the presence of a global market movement (Kotkatvuori-Örnberg et al., 2013). This characteristic may be crucial for our analysis, because the effect of diversification is reduced when the correlation rises (Choueifaty and Coignard, 2008; You and Daigler, 2010). On top of that, data are far from being normal: the average skewness and kurtosis show that log-returns are leptokurtic, i.e. with higher peaks at the mean and fatter tails than normally distributed data. Indeed, the Jarque-Bera test rejects the null hypothesis that asset returns are normally distributed in both periods at 99% confidence level.

Period	T	n	$\hat{\mu}$	$\hat{\sigma}$	$Skew$	$Kurt$	DD
			S&P 100				
2005 - 2009	1304	93	3.4957e-05	0.0230	-0.0517	14.0324	0.6008
2010 - 2015	1430	93	4.6708e-04	0.0145	-0.1916	8.3050	0.3186
			S&P 500				
2005 - 2009	1304	452	7.5578e-05	0.0259	-0.1564	13.8196	0.6519
2010 - 2015	1430	452	4.7523e-04	0.0166	-0.1541	11.4972	0.3621

Table 1: Descriptive statistics of S&P 100 and S&P 500 components in the two sub-periods: number of observations T, number of constituents n, mean $\hat{\mu}$, standard deviation $\hat{\sigma}$, skewness $Skew$, kurtosis $Kurt$, maximum drawdown DD of the returns.

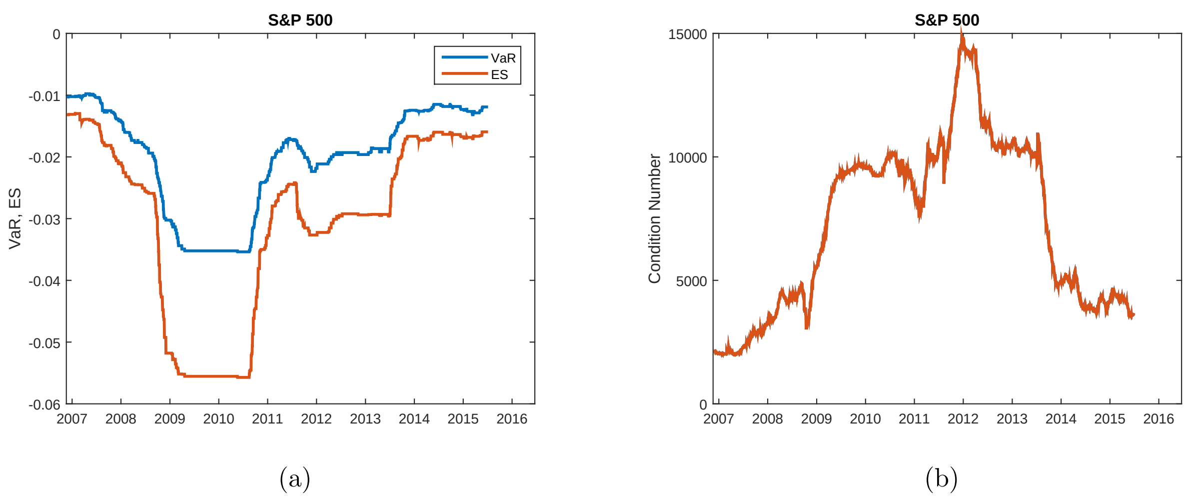


Figure 4: Panel (a): Evolution of Value-at-Risk (VaR) and Expected Shortfall (ES) of S&P 500 index, estimated by using a window of 500 days. Panel (b): Condition Number of the covariance matrix of S&P 500 components, estimated by using a window of 500 days.

We use a rolling-window scheme with window size $WS = 500$ to replicate the situation in which the investor selects his portfolio using the last two years of information and holds it for one day before re-balancing. Therefore, based on 500 in-sample observations, we revise the asset weights daily, moving each time window ahead by one observation and discarding the oldest data point. In total, we end up with $M_1 = 804$ and $M_2 = 930$ rolling windows and out-of-sample observations

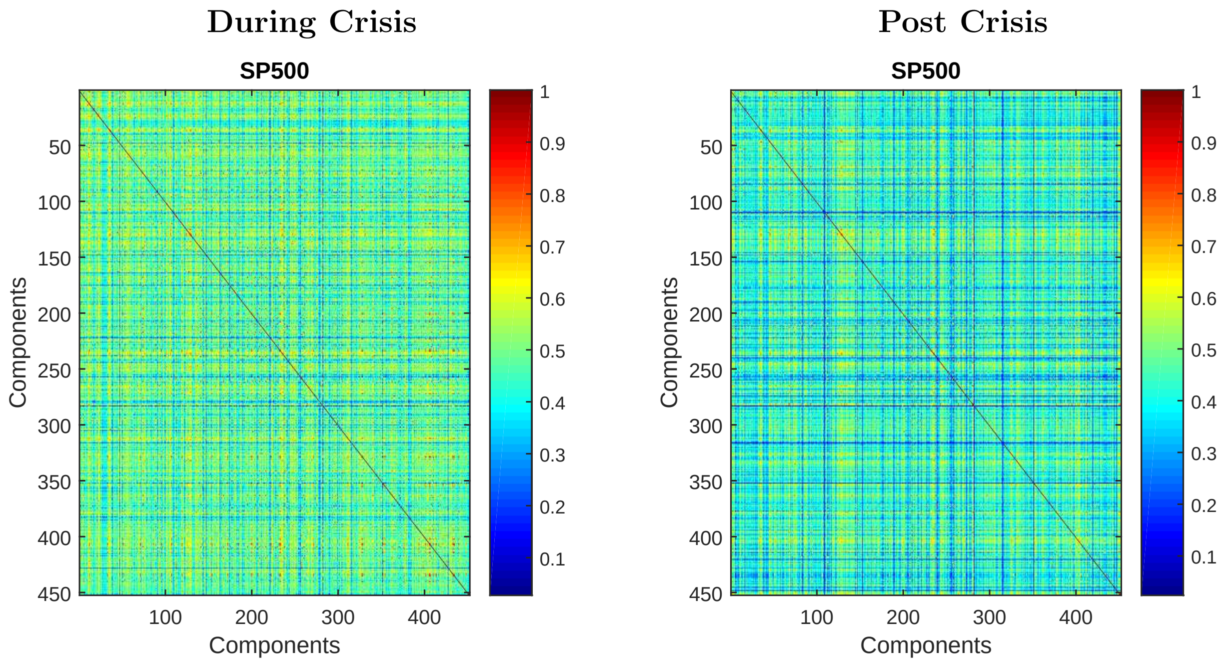


Figure 5: Correlation matrices of S&P 500 components during the crisis (left panel) and post-crisis periods (right panel).

in the first and second period, respectively. We constraint the asset weights to vary between -1 and 1. For each window, we optimize the penalized minimum variance problem with the different penalties by using the gradient projection algorithm developed by Figueiredo et al. (2007), which has been shown to be more efficient than quadratic programming and coordinate-wise optimization (Gasso et al., 2009). We use a vector of 10 λs and select for each strategy the portfolio with the lowest value of the objective function (4) in the in-sample window.

Tables 2 and 3 compare the average in-sample (IS) and out-of-sample (OOS) results obtained by the different strategies, including the equally weighted (EW), the risk-parity (ERC) and the unconstrained minimum variance (MV) portfolios, during the crisis and post-crisis periods, respectively. We report the in- and out-of-sample annual standard deviation, mean return and Sharpe Ratio in Columns 2 to 6 to evaluate the average risk-return performance of the portfolios. In particular, we are interested in comparing the average out-of-sample annual risk (Column 3). Then, we evaluate the performance of the portfolios in terms of transaction and monitoring costs by considering how many assets are selected and how much their weights vary over time. Column

7 shows the portfolio turnover, which represents the average change in portfolio weights between two consecutive windows (i.e. TO $= \sum_{m=2}^{M} |\boldsymbol{w}_m - \boldsymbol{w}_{m-1}|/(M-1)$). This measure can be used as a proxy for the transaction costs that an investor would pay to re-balance a certain strategy every day: the higher the turnover, the larger the transaction costs. Column 8 reports the average number of active positions, $\bar{k}$ (i.e. sparsity of the portfolio) that an investor would need to monitor. Therefore, the lower the turnover and $\bar{k}$, the cheaper the strategy. Furthermore, we analyse the diversification performance of the portfolios by calculating the weight and risk diversification indices, D_{ℓ_0} (Columns 9), D_w (Columns 10), D_r (Columns 11) and D_{er} (Columns 12). Finally, we report for each strategy the annual Value-at-Risk (Column 13), defined as the maximum loss not exceeded in 95% of cases, and the annual Expected Shortfall (Column 14), i.e. the expected return on the portfolio in the worst 5% of cases.

As expected, the in-sample portfolio risk is minimized for the MV strategy and maximized for the EW strategy. Furthermore, we find that ℓ_q, Log and ℓ_1-portfolios reach much lower in-sample risk values than the ERC strategy in all dataset, while the ℓ_2 obtains risk values similar to ERC and EW. As pointed out by Behr et al. (2013), the good in-sample performance of the MV portfolio are not confirmed out-of-sample, where the regularization strategies, and in particular the ℓ_q-portfolios, obtain the lowest risk values. This remarkable performance may be explained by first decomposing the objective function (4) as

$$\min_{\boldsymbol{w}} \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_i w_j \rho_{ij} \sigma_{ij} + \lambda \sum_{i=1}^{n} g(w_i) \ , \tag{10}$$

where the sum of the first two terms is equal to the portfolio variance $\boldsymbol{w}'\Sigma\boldsymbol{w}$, and then looking at the average risk of the assets selected by the ℓ_q-strategy ($w_i \neq 0$, $1 \leq i \leq n$), compared to the risk of the assets not selected by the strategy ($w_i = 0$, $1 \leq i \leq n$). Figure 6 shows the boxplots of the average in-sample variances, drawdowns and correlations within the two groups of securities, during and after the financial crisis for the S&P 500 dataset. We notice that, since assets tend to be highly correlated, especially during the crisis period (i.e. there are very large values of ρ_{ij}), the ℓ_q-strategy selects on average the lowest-volatile assets in order to minimize function (10), discarding the securities that display extreme losses in-sample. Moreover, the ℓ_q-strategy leads to optimally select small amount of shorting rather than increasing the number of constituents, which implies some negative values of the second term in function (10) (i.e. the average amount of shorting of the ℓ_q-portfolio in the S&P 500 dataset is equal to 19% and 43% during the crisis and post-crisis periods, respectively). With respect to the other strategies, the ℓ_q-approach adapts automatically to the different market conditions and better deals with estimation error, achieving also lower OOS Values-at-Risk and Expected Shortfalls than the other strategies and thereby smaller exposure to extreme risk.

In general, during the crisis period (see Table 2), the performance of all strategies suffer from the highly volatile market conditions, resulting in values of portfolio risk that are much higher than the ones obtained in the post-crisis period (see Table 3). We are not able to compare the risk-return performance of the strategies during the crisis because of their negative annual returns; while, in the post-crisis period, the highest Sharpe Ratios are achieved by the ℓ_2 and the ℓ_1-portfolios, respectively for the S&P 100 and S&P 500 dataset. The ERC and EW portfolios also obtain good risk-return performance especially in the first dataset, as shown in Figure 7, which plots the OOS compounded returns of the different portfolios over the two sub-periods.

The EW portfolio has by definition a turnover equal to zero, since we assume the vector of weights to be constant over the windows. Similar performance are achieved by the ℓ_2 and ERC portfolios, which are very stable over time. On the other hand, these three strategies do not promote sparsity and assign a non-zero weight to all available assets, as shown by $\bar{k}$ in Column 8. The Log and the ℓ_q-strategies instead achieve remarkable performance in terms of sparsity by selecting the portfolios with the smallest number of active weights in all dataset. However, they update their investment solutions more frequently or significantly than EW and ERC strategies, thus suffering from high turnover, as reported by TO in Column 7. From Tables 2 and 3, we notice that the EW and ERC portfolios are the most diversified investment solutions in all dataset, as expected. Both strategies invest in all the available assets and distribute the weights and the risk nearly equally among them. The ℓ_2-portfolio reaches similar performance, both in terms of weight and risk diversification. The Log and ℓ_q-strategies show instead an opposite behavior: they are exposed to a few investments with larger weights, resulting in low D_{ℓ_0}, D_w and D_r, but allow more diversification in terms of extreme risk than MV and ℓ_1-strategies. In fact, the MV portfolio, despite investing in all assets, is the most concentrated in terms of weight and extreme risk since it selects extreme positions in very few low volatile assets.

From the empirical results, we confirm the existence of a trade-off between out-of-sample risk, sparsity and diversification of a portfolio. None of the strategies is able to reach all objectives at the same time. EW, ERC and ℓ_2-portfolios achieve similar performance, as shown in Table 4, which reports the correlation coefficients between the OOS returns obtained in the two dataset. These portfolios are well-diversified in terms of weight and risk, they invest in all the available assets and lead to high levels of out-of-sample risk. On the contrary, the MV portfolio can reach satisfactory out-of-sample risk values, but it is neither diversified nor sparse. According to Table 4, ℓ_1, Log and ℓ_q-strategies achieve similar risk performance to MV, but with much sparser portfolios. In fact, they also select on average the lowest volatile assets that do not suffer from extreme losses, but differently from MV, they assign a zero weight to all other securities. However, for this reason, they are not able to diversify their allocation in terms of weight and risk. Among the regularization strategies, the ℓ_q outperforms ℓ_1, by achieving less OOS risk levels with much sparser portfolios. In the S&P 500 dataset, for example, the ℓ_q-portfolio selects on average 93 and 90 assets during the crisis and post-crisis periods, respectively, while the ℓ_1-portfolio selects an average of 365 and 331 assets, which results in much higher transaction and monitoring costs.

We confirm this behavior on simulated data, where we can further test the performance of the ℓ_1 and ℓ_q-portfolios in terms of empirical and actual risk. In particular, we simulate 500 returns of 100 assets from a Fama-French Three-Factor Model with known covariance matrix $\boldsymbol{\Sigma}$, and we construct 50 ℓ_1 and ℓ_q-portfolios with different number of active positions, by increasing the penalization parameter λ. Let $\boldsymbol{w}$ and $\boldsymbol{w}_{opt}$ be the theoretical and empirical optimal allocation vectors, solving the penalized optimization problems $\boldsymbol{w} = \text{argmin}_{\mathbf{1}'\boldsymbol{w}=1} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} + \lambda \sum_{i=1}^{n} g(w_i)$ and $\boldsymbol{w}_{opt} = \text{argmin}_{\mathbf{1}'\boldsymbol{w}=1} \boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{w} + \lambda \sum_{i=1}^{n} g(w_i)$, where $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}$ are the theoretical covariance matrix and its estimate, respectively. Then, we denote the oracle, empirical and actual risks of such portfolios as $R(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$, $R_n(\boldsymbol{w}_{opt}) = \boldsymbol{w}'_{opt}\hat{\boldsymbol{\Sigma}}\boldsymbol{w}_{opt}$ and $R(\boldsymbol{w}_{opt}) = \boldsymbol{w}'_{opt}\boldsymbol{\Sigma}\boldsymbol{w}_{opt}$, respectively (Fan et al., 2012). Figure 8a shows the empirical and actual risks of the 50 ℓ_1 and ℓ_q-portfolios. First of all, we notice that increasing the number of active positions does not necessarily yield lower risk, as also pointed out by Fan et al. (2012). On the contrary, both the empirical and actual risks increase after a certain portfolio size. Furthermore, the risks of ℓ_q-portfolios are lower than the risks of ℓ_1-portfolios with the same number of active positions, except for the most diversified solutions. From Figure 8b, we also notice that the ℓ_q-penalty represents a stronger constraint on short-selling than the ℓ_1-penalty, i.e. for a given number of active positions, the former yields portfolios with less amount of shorting. This behavior is confirmed by looking at the regularization path of the ℓ_1 and ℓ_q-portfolios, in Figures 8c and 8d. Here, we notice that increasing λ, the number of active positions decreases with both penalties and as expected with faster rates for the ℓ_q-approach. Also, the absolute value of the remaining weights increases much more for the ℓ_1 than for the ℓ_q-penalty, resulting in more conservative extreme estimates for the ℓ_q-approach (Gasso et al., 2009).

Interestingly, our analysis suggests that diversifying the portfolio by increasing the number of active positions does not lead to risk minimization out-of-sample, especially during crisis periods. The benefits of diversification in terms of risk reduction rather decrease after reaching a certain portfolio size, particularly when the volatility and correlation between assets increase and the distribution of asset returns is far from being normal (Doganoglu et al., 2007; Desmoulins-Lebeault and Kharoubi-Rakotomalala, 2012; Mainik et al., 2015). Therefore, as shown in Figure 7, regularization techniques that lead to more concentrated portfolio solutions, like the Log and ℓ_q-strategies, can represent a better choice in bear markets.

Method	σ_p (%)		μ_p (%)		SR	TO	$\bar{k}$	D_{ℓ_0}	D_w	D_r	D_{er}	VaR	ES
	IS	OOS	IS	OOS	OOS								
S&P 100, T_1 = 1304 n = 93													
EW	36.04	29.89	-7.72	-4.38	-0.147	-	93.000	1.000	1.000	0.885	0.504	-7.335	-12.181
ERC	30.96	26.34	-4.56	-3.87	-0.147	0.004	93.000	1.000	0.887	1.000	0.552	-6.527	-10.691
MV	12.01	14.81	6.24	-4.36	-0.294	0.102	93.000	1.000	0.062	0.062	0.016	-3.507	-6.273
ℓ_2	35.37	29.62	-7.17	-4.42	-0.149	0.001	93.000	1.000	0.999	0.899	0.515	-7.264	-12.071
ℓ_1	12.03	14.85	6.37	-4.42	-0.298	0.103	84.432	0.908	0.064	0.016	0.019	-3.449	-6.282
Log	12.14	14.90	6.92	-4.75	-0.319	0.128	46.012	0.495	0.063	0.015	0.019	-3.543	-6.347
ℓ_q	12.13	14.85	6.76	-4.32	-0.291	0.103	53.333	0.573	0.067	0.015	0.021	-3.539	-6.336
S&P 500, T_1 = 1304 n = 452													
EW	39.26	32.53	-6.43	-4.75	-0.146	-	452.000	1.000	1.000	0.883	0.496	-7.969	-13.295
ERC	33.80	28.64	-3.44	-5.04	-0.176	0.006	452.000	1.000	0.898	1.000	0.536	-6.984	-11.680
MV	4.19	14.68	1.76	-10.12	-0.689	0.406	452.000	1.000	0.015	0.015	0.002	-3.497	-5.935
ℓ_2	35.97	31.18	-4.38	-5.34	-0.171	0.001	451.759	1.000	0.988	0.930	0.525	-7.532	-12.740
ℓ_1	4.54	14.26	2.51	-9.03	-0.633	0.379	364.966	0.807	0.021	0.003	0.002	-3.440	-5.834
Log	6.12	14.37	4.42	-5.62	-0.391	0.456	85.580	0.189	0.018	0.003	0.002	-3.485	-6.167
ℓ_q	6.26	13.94	3.66	-6.36	-0.456	0.393	92.882	0.205	0.021	0.004	0.003	-3.204	-5.912

Table 2: Average statistics of the different portfolios in the period 01.01.2005 - 31.12.2009 on S&P 100 and S&P 500: IS and OOS annual risk, IS and OOS annual return, OOS annual Sharpe Ratio, turnover (TO), number of active positions ($\bar{k}$), concentration index (D_{ℓ_0}), weight diversification index (D_w), risk diversification index (D_r), extreme risk diversification index (D_{er}), annual Value-at-Risk, annual Expected Shortfall.

Method	σ_p (%)		μ_p (%)		SR	TO	$\bar{k}$	D_{ℓ_0}	D_w	D_r	D_{er}	VaR	ES
	IS	OOS	IS	OOS	OOS								
S&P 100, T_2 = 1430 n = 93													
EW	11.07	11.84	11.02	15.07	1.273	-	93.000	1.000	1.000	0.909	0.206	-3.070	-4.228
ERC	10.56	10.95	10.06	14.12	1.290	0.002	93.000	1.000	0.904	1.000	0.171	-2.744	-3.906
MV	7.47	9.33	-1.55	5.80	0.622	0.124	93.000	1.000	0.041	0.041	0.014	-2.261	-3.268
ℓ_2	10.86	11.00	10.66	14.37	1.306	0.001	92.147	0.999	0.930	0.965	0.176	-2.803	-3.930
ℓ_1	7.54	9.20	-1.20	5.68	0.618	0.112	84.881	0.913	0.047	0.011	0.015	-2.163	-3.218
Log	7.76	9.15	-0.22	6.76	0.739	0.136	39.869	0.429	0.047	0.011	0.017	-2.125	-3.237
ℓ_q	7.87	9.06	0.46	6.58	0.726	0.118	44.782	0.481	0.052	0.011	0.017	-2.147	-3.184
S&P 500, T_2 = 1430 n = 452													
EW	11.34	12.56	10.68	14.17	1.128	-	452.000	1.000	1.000	0.910	0.177	-3.364	-4.512
ERC	10.68	11.49	9.89	13.47	1.172	0.004	452.000	1.000	0.901	1.000	0.158	-3.132	-4.122
MV	2.45	8.79	-2.90	12.35	1.405	0.580	452.000	1.000	0.008	0.008	0.001	-2.139	-2.843
ℓ_2	11.31	12.42	10.65	14.09	1.134	0.001	452.000	1.000	0.999	0.927	0.177	-3.359	-4.461
ℓ_1	3.27	8.01	-1.29	11.69	1.459	0.484	331.344	0.733	0.014	0.002	0.002	-1.973	-2.633
Log	4.85	8.22	1.28	10.50	1.277	0.550	69.509	0.154	0.012	0.002	0.002	-2.013	-2.794
ℓ_q	4.75	7.97	2.50	9.79	1.229	0.397	90.031	0.200	0.015	0.002	0.003	-1.957	-2.736

Table 3: Average statistics of the different portfolios in the period 01.01.2010 - 29.05.2015 on S&P 100 and S&P 500: IS and OOS annual risk, IS and OOS annual return, OOS annual Sharpe Ratio, turnover (TO), number of active positions ($\bar{k}$), concentration index (D_{ℓ_0}), weight diversification index (D_w), risk diversification index (D_r), extreme risk diversification index (D_{er}), annual Value-at-Risk, annual Expected Shortfall.

	EW	ERC	MV	ℓ_1	Log	ℓ_q	ℓ_2
EW	1.000	**0.996**	0.330	0.351	0.321	0.339	**0.999**
ERC	**0.995**	1.000	0.375	0.400	0.372	0.393	**0.998**
MV	0.438	0.498	1.000	**0.972**	**0.846**	**0.852**	0.345
ℓ_1	0.442	0.503	**0.999**	1.000	**0.904**	**0.913**	0.366
Log	0.458	0.519	**0.981**	**0.985**	1.000	**0.997**	0.469
ℓ_q	0.461	0.523	**0.983**	**0.988**	**0.997**	1.000	0.358
ℓ_2	**0.999**	**0.996**	0.449	0.454	0.469	0.473	1.000

Table 4: Correlation coefficients between the OOS returns of the different strategies in the whole period for S&P 100 (below the diagonal) and S&P 500 (above the diagonal).

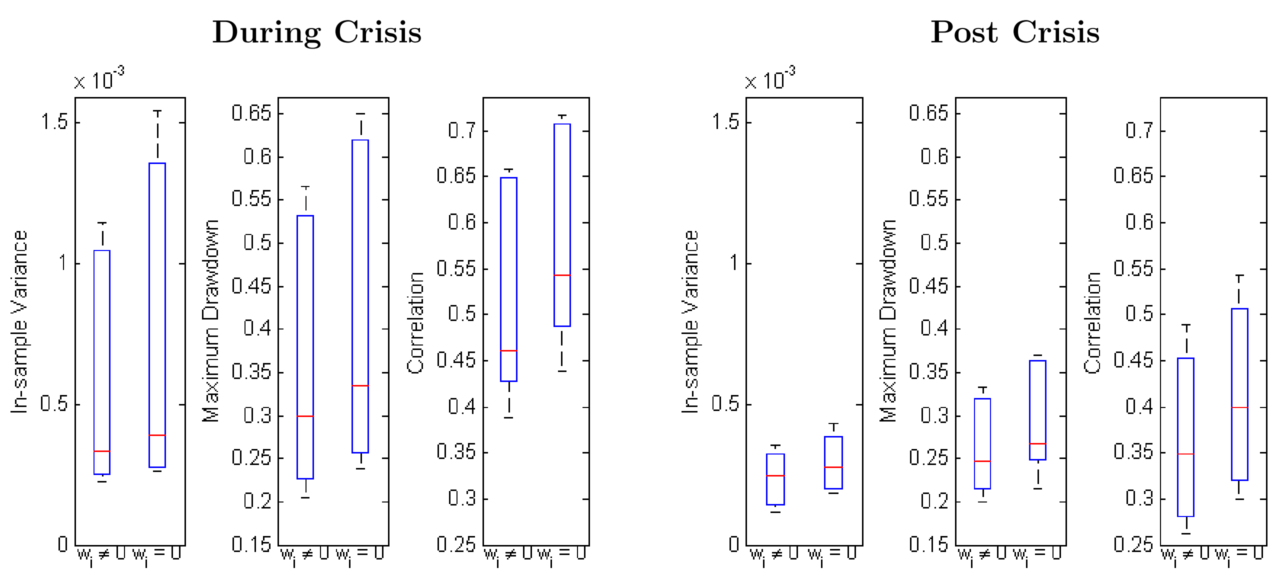
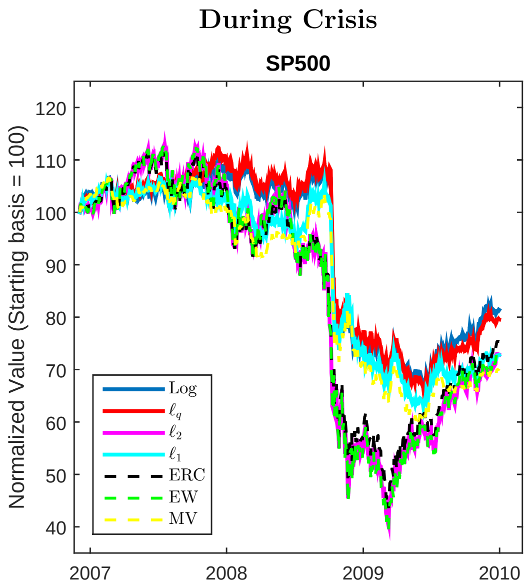


Figure 6: In-sample variance, maximum drawdown and correlation within the S&P 500 components selected ($w_i \neq 0$, $1 \leq i \leq n$) vs discarded ($w_i = 0$, $1 \leq i \leq n$) by ℓ_q during the crisis (left panel) and post-crisis periods (right panel).



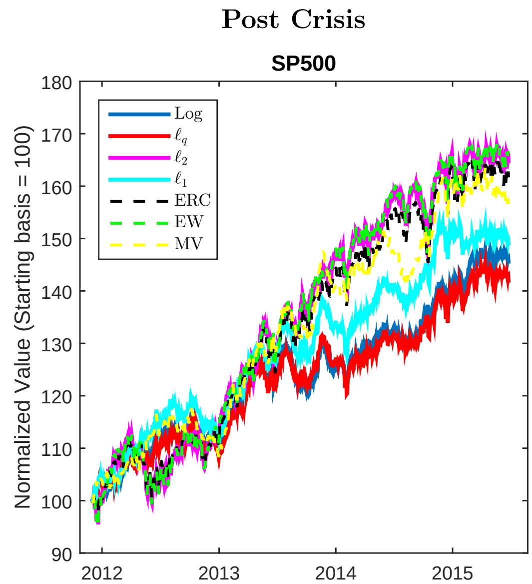


Figure 7: Out-of-sample returns of S&P 500 portfolios in period 01.01.2005 - 31.12.2009 (left panel) and in period 01.01.2010 - 29.05.2015 (right panel).

3.1 Properties

We summarize here the diversification and regularization strategies by comparing their main properties. In particular, Table 5 shows the different behaviors of the portfolios with respect to high levels of correlation, out-of-sample performance, amount of shorting and diversification.

	MV	EW	ERC	ℓ_2	ℓ_1	ℓ_q
Robustness to correlation	-	$\circ$	$\circ$	+	-	+
Constraint on shorting	-	$\circ$	$\circ$	-	+	+
D_{ℓ_0}	+	+	+	+	-	-
D_w	-	+	+	+	-	-
D_r	-	+	+	+	-	-
D_{er}	-	+	+	+	-	-

Table 5: Main properties of portfolio strategies: positive (+), neutral ($\circ$), and negative(-).

Despite investing in all securities, the unconstrained MV strategy leads to portfolios with large number of constituents and extreme positions. Furthermore, the stability of the resulting weights may be strongly affected by high correlations and estimation errors in model parameters. On the contrary, diversification strategies, such as the EW and ERC, select all the available assets to exploit the potential weight and risk diversification effects. However, they fail to provide

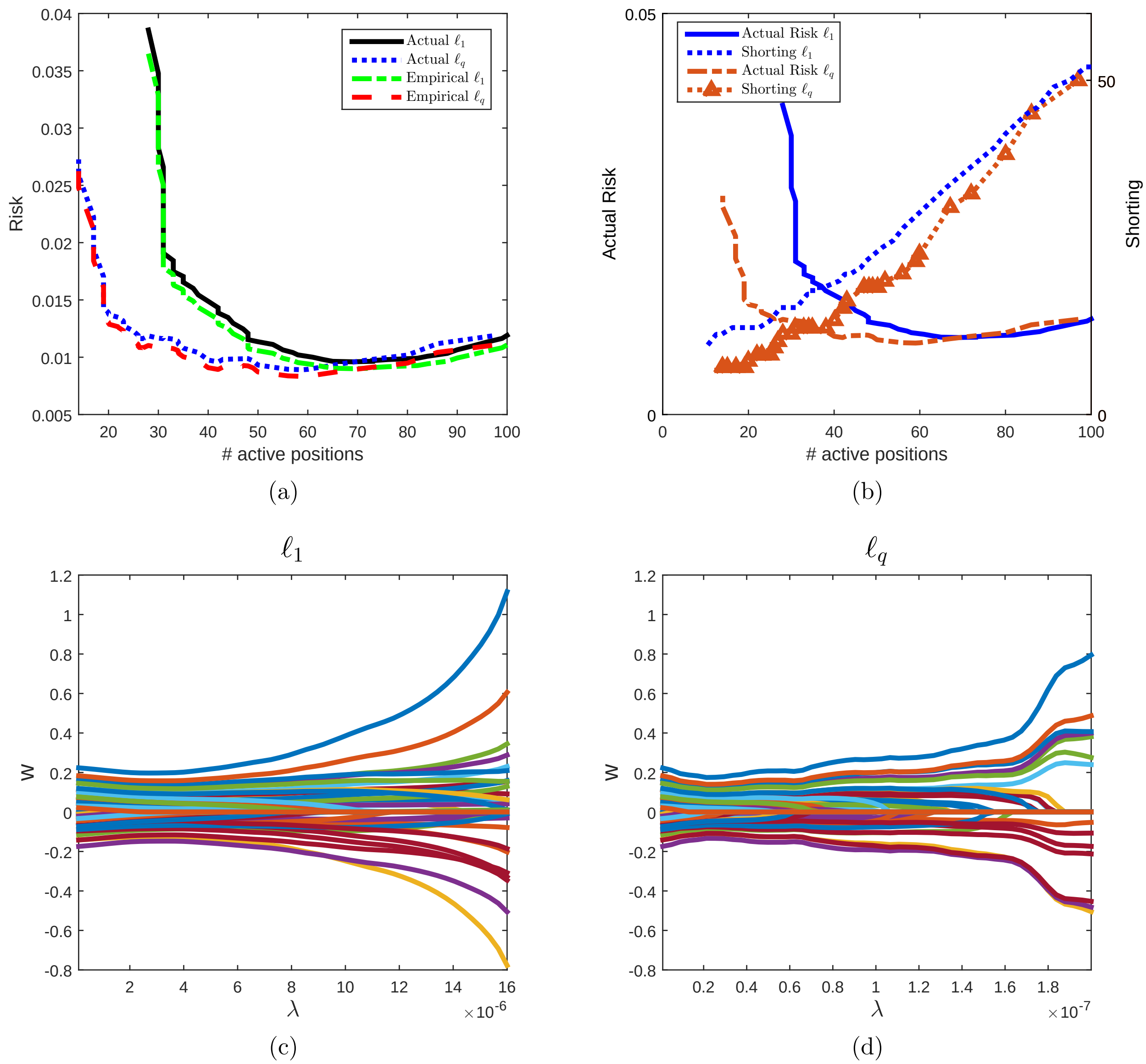


Figure 8: Panel (a): Empirical and actual risks of 50 portfolios obtained by varying the penalization parameter λ in the ℓ_1 and ℓ_q-penalty function. We simulate 500 returns of 100 assets from a Fama-French Three-Factor Model with known covariance matrix $\mathbf{\Sigma}$. The empirical and actual risks are computed by using the sample and the true covariance matrices. Panel (b): Actual risk and amount of shorting of the 50 portfolios. Panels (c) and (d): Regularization path of the 50 ℓ_1 and ℓ_q portfolios, respectively.

sparse solutions, with the ERC portfolio still very sensitive to changes in parameter estimates. Regularization techniques help to mitigate the effect of estimation errors, resulting in better OOS properties, by shrinking the covariance matrix of asset returns. In particular, the ℓ_q with $0 < q < 1$ strategy is able to better deal with estimation errors deriving from extreme observations and multicollinearity, by privileging sparse solutions characterized by low volatilities. These characteristics have been shown to be especially useful in periods of high uncertainty, when assets tend to display high levels of spurious correlation and volatility.

4 Conclusion

In this paper, we investigate the trade-off between diversification and concentration within a risk minimization framework in portfolio selection. Ideally, investors target asset allocations, characterized by few active positions, but still with the right amount of diversification. In the minimum variance optimization problem, we consider the Log and the ℓ_q-norm as penalty terms that can naturally deal with the trade-off between number of active components, size of the asset weights and diversification level. Thus, imposing such constraints, we are able to control the level of diversification of the portfolio. We show that the resulting investment solution has a smaller number of active weights than EW, ERC and MV portfolios, which also results in a lower risk exposure. Moreover, we study the relationship between the diversification constrained optimization methods and risk-parity portfolios and discuss in detail their properties.

In the empirical analysis, we evaluate the out-of-sample performance of the different strategies in terms of risk, sparsity and diversification, and compare the results obtained in low and highly volatile markets. We observe that the Log and the ℓ_q-portfolios achieve lower OOS risk than classical diversification strategies, like EW and ERC, with a much smaller number of active positions. Furthermore, we notice that the ℓ_q-strategy selects on average the lowest-volatile assets and assigns a zero weight to the securities that display extreme losses in-sample. This behavior results in better OOS Values-at-Risk and Expected Shortfalls with respect to other strategies, especially during crisis and bear markets, when the benefits of diversification in terms of risk reduction are decreased by the higher correlation between assets.

Going back to our original question: is it a good idea to un-diversify during crisis? Our results so far suggest "Yes". Still, further research on how to better control the turnover of sparse ℓ_q-portfolios is needed. One possible way to encourage the stability of the ℓ_q-estimates would be to penalize both the portfolio weights and their differences in time, by extending the fused Lasso approach (Tibshirani et al., 2005). This would require to find efficient ways to deal with computational costs, especially when the dimensionality of the problem is high. Furthermore, we plan to extend our analysis by explicitly considering not only risk minimization, but also profitability of the investment strategies. Dealing with such issues is currently high on our agenda.

References

Bauer, D. and Zanjani, G. (2016). The marginal cost of risk, risk measures, and capital allocation. *Management Science*, 62(5):1431 – 1457.

Behr, P., Guettler, A., and Miebs, F. (2013). On portfolio optimization: imposing the right constraint. *Journal of Banking & Finance*, 37(4):1232 – 1242.

Benoit, S., Colletaz, G., Hurlin, C., and Perignon, C. (2013). A theoretical and empirical comparison of systemic risk measures. *HEC Paris Research Paper*, (FIN - 2014 - 1030).

Boyle, P., Garlappi, L., Uppal, R., and Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, 58(2):253 – 272.

Brands, S., Brown, S., and Gallagher, D. (2005). Portfolio concentration and investment manager performance. *International Review of Finance*, 5(3):149 – 174.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., and Loris, I. (2009). Sparse and stable markowitz portfoios. *Proceedings of the National Academy of Science*, 106(30):12267 – 12272.

Bruder, B. and Roncalli, T. (2012). Managing risk exposures using the risk budgeting approach. Working Paper.

Buffet, W. (1979). Chairman's letter.

Carrasco, M. and Noumon, N. (2012). Optimal portfolio selection using regularization. Working Paper University of Montreal.

Cazalet, Z., Grison, P., and Roncalli, T. (2014). The smart beta indexing puzzle. *Journal of Index Investing*, 5(1):97 – 119.

Chen, C., Li, X., Tolman, C., Wang, S., and Ye, Y. (2016). Sparse portfolio selection via quasi-norm regularization. *Management Science*. forthcoming.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6 – 11.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, 34(4):40 – 51.

De Miguel, V., Garlappi, L., Nogales, F., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norm. *Management Science*, 55:798 – 812.

De Miguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22 (5):1915 – 1953.

De Miguel, V. and Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, 57:560 – 577.

Desmoulins-Lebeault, F. and Kharoubi-Rakotomalala, C. (2012). Non-gaussian diversification: When size matters. *Journal of Banking & Finance*, 36(7):1987 – 1996.

Doganoglu, T., Hartz, C., and Mittnik, S. (2007). Portfolio optimization when risk factors are conditionally varying and heavy tailed. *Computational Economics*, 29(3):333 – 354.

Fan, J., Zhang, J., and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592 – 606.

Fastrich, B., Paterlini, S., and Winker, P. (2014). Cardinality versus q-norm constraints for index tracking. *Quantitative Finance*, 14(11):2019 – 2032.

Fastrich, B., Paterlini, S., and Winker, P. (2015). Constructing optimal sparse portfolios using regularization methods. *Computational Management Science*, 12(3):417 – 434.

Figueiredo, M., Nowak, R., and Wright, S. (2007). Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing*, 1(4):586 – 597.

Frank, I. and Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109 – 135.

Gasso, G., Rakotomamonjy, A., and Canu, S. (2009). Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Transactions on Signal Processing*, 57(12):4686 – 4698.

Huang, J., Horowitz, J., and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics*, 30:587 – 613.

Ivkovic, Z., Sialm, C., and Weisbenner, S. (2008). Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis*, 43(3):613 – 656.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651 – 1684.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *Journal of Finance*, 60(4):1983 – 2011.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28(5):1356 – 1378.

Kolm, P. N., Tütüncü, R., and Fabozzi, F. (2014). 60 years following harry markowitz's contribution to portfolio theory and operations research. *European Journal of Operational Research*, 234 (2):343 – 582.

Kotkatvuori-Örnberg, J., Nikkinen, J., and Äijö, J. (2013). Stock market correlations during the financial crisis if 2008 - 2009: evidence from 50 equity markets. *International Review of Financial Analysis*, 28:70 – 78.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(1):365 – 411.

Leung, G. and Barron, A. (2006). Information theory and mixing least-squares regressions. *IEEE transactions on information theory*, 52(8):3396 – 3410.

Maillard, S., Roncalli, T., and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, 36:60 – 70.

Mainik, G., Mitov, G., and Rüschendorf, L. (2015). Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *Journal of Empirical Finance*, 32:115 – 134.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77 – 91.

Mazumder, R., Friedman, J. H., and Hastie, T. (2011). Sparsenet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, 106(495):1125 – 1138.

Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323 – 361.

Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45(1):31 – 45.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* The MIT Press.

Seeger, M. W. (2008). Bayesian inference and optimal design for the sparse linear model. *Journal of Machine Learning Research*, 9:759 – 813.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67(1):91 – 108.

Xing, X., Hub, J., and Yang, Y. (2014). Robust minimum variance portfolio with L-infinity constraints. *Journal of Banking & Finance*, 46:107 – 117.

Yen, Y. and Yen, T. (2014). Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. *Computational Statistics & Data Analysis*, 76:737 – 759.

You, L. and Daigler, R. (2010). Is international diversification really beneficial? *Journal of Banking & Finance*, 34(1):163 – 173.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67(2):301 – 320.

A Risk Decomposition

In risk management, it is important to quantify the contribution of each asset to the overall portfolio risk. One common indicator is given by the sensitivity of portfolio risk to a small change in asset allocation. In this section, we derive this measure for the portfolio standard deviation and Expected Shortfall.

Let $\boldsymbol{w}$ be the $n \times 1$ vector of portfolio weights and $\boldsymbol{\Sigma}$ be the $n \times n$ covariance matrix of n asset returns. Then, the risk of the portfolio, typically measured by the standard deviation of portfolio returns σ_p, can be expressed as follows:

$$\sigma_p = \sqrt{\boldsymbol{w}'\Sigma\boldsymbol{w}}.$$

In order to measure the contribution of each asset to the whole portfolio risk, we can compute the Marginal Risk Contribution of asset i as the partial derivative of σ_p with respect to w_i

$$MRC_i = \frac{\partial \sigma_p}{\partial w_i} = \frac{\sum_{i,j=1}^{n} \sigma_{ij} w_j}{\sigma_p}.$$

MRC_i can be also expressed as a function of $(\boldsymbol{\Sigma w})$, the product of the covariance matrix and the weights vector, as follows:

$$MRC_i = \frac{(\Sigma w)_i}{\sigma_p}$$

where $(\Sigma w)_i = \sum_{i,j=1}^{n} \sigma_{ij} w_j$ represents the i-th component of the column vector $(\boldsymbol{\Sigma w})$. The risk contribution of asset i is then defined as the weighted MRC_i and represents the share of portfolio risk corresponding to the i-asset:

$$RC_i = w_i MRC_i = \sigma_i$$

$$\sum_{i=1}^{n} RC_i = \sum_{i=1}^{n} \sigma_i = \sqrt{\boldsymbol{w}' \Sigma \boldsymbol{w}}$$

The sum of all RC_i is the total portfolio risk, quantified by the standard deviation of the portfolio returns. The relative risk contribution of asset i is defined as

$$RRC_i = \frac{RC_i}{\sigma_p} = \frac{w_i (\Sigma w)_i}{\sigma_p^2} = \frac{w_i (\Sigma w)_i}{\boldsymbol{w}' \Sigma \boldsymbol{w}}$$

By construction, the risk-parity portfolio has a $\mathrm{RC}_i = \sigma_p / n$, which implies an $\mathrm{RRC}_i = 1/n$.

From a risk budgeting perspective, it may be useful to know the composition of a portfolio also in terms of extreme risk. Let's denote with μ_i the return of asset i (with $i = 1, \ldots, n$) and with μ_p the return of the portfolio obtained as the weighted return of its components:

$$\mu_p = \sum_{i=1}^{n} w_i \mu_i \ .$$

Given a constant $0 \leq \alpha \leq 1$, we measure the extreme risk of a portfolio by the Expected Shortfall $ES_{p,\alpha}$, which represents the expected return of the portfolio in the worst $\alpha\%$ of the cases or equivalently the expected return of the portfolio given that μ_p exceeds a threshold C:

$$ES_{p,\alpha} = \mathbb{E}(\mu_p | \mu_p < C) = \mathbb{E}\left(\sum_{i=1}^{n} w_i \mu_i | \mu_p < C \right) \ .$$

To compute the contribution of each asset to the whole portfolio $ES_{p,\alpha}$, we first calculate the

Marginal Expected Shortfall of asset i as the partial derivative of $ES_{p,\alpha}$ with respect to w_i:

$$MES_{i,\alpha} = \frac{\partial ES_{p,\alpha}}{\partial w_i} = \mathbb{E}(\mu_i | \mu_p < C) \ .$$

$MES_{i,\alpha}$ represents the increase in portfolio extreme risk caused by a marginal increase of the weight of asset i. Then, as suggested by Benoit et al. (2013), the extreme risk contribution of each asset $CES_{i,\alpha}$ can be defined as the weighted $MES_{i,\alpha}$ and indicates the share of $ES_{p,\alpha}$ corresponding to the i-asset:

$$\begin{aligned} CES_{i,\alpha} &= w_i MES_{i,\alpha} \\ \sum_{i=1}^{n} CES_{i,\alpha} &= \sum_{i=1}^{n} w_i MES_{i,\alpha} = ES_{p,\alpha} \ . \end{aligned}$$

The sum of all the Contributions to Expected Shortfall $CES_{i,\alpha}$ is the total portfolio Expected Shortfall.

B Risk Approximation

Let's consider the risk minimization problem

$$\begin{aligned} \min_{\boldsymbol{w}} \ & \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \\ & \mathbf{1}'\boldsymbol{w} = 1 \\ & \|\boldsymbol{w}\|_q^q \ \leq c^q \end{aligned} \tag{11}$$

where $0 < q \leq 1$ and $c^q > 0$ is the threshold of the ℓ_q-norm. This optimization could be solved as the following penalized problem (despite convergence to the global optimum is not guaranteed as the ℓ_q-penalty is non-convex).

$$\begin{aligned} \min_{\boldsymbol{w}} \ & \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} + \lambda \|\boldsymbol{w}\|_q^q \\ & \mathbf{1}'\boldsymbol{w} = 1 \end{aligned} \tag{12}$$

with $\lambda > 0$ as a scalar controlling the intensity of the penalty. If $c \to n^{1-q}$, then the solution to problem (11) converges to the EW portfolio, while if $c \to 1$, it converges to the most concentrated portfolio with just one active weight, as $q \to 0^+$.

Proof. Proof of Proposition 1 To prove that the ℓ_q-norm, with $0 < q \leq 1$, is bounded by 1 and n^{1-q}, under the no-short-selling and budget constraints, i.e. $0 \leq w_i \leq 1$, $\sum_{i=1}^n w_1 = 1$, we compute its extreme values, corresponding to the most concentrated (i.e. totally invested) and the EW portfolios. Let's assume weights are sorted from the largest to the smallest such that $w_{(1)} \geq w_{(2)} \geq \dots w_{(n)}$. Then, let $w_{(1)}$ be equal to 1 and therefore $w_{(j)} = 0$, $j = 2, \dots n$. It follows that for the totally invested portfolio

$$\ell_q = \|\boldsymbol{w}\|_q^q = \sum_{i=1}^n |w_i|^q = 1^q = 1 \ .$$

The other limit case is for the EW portfolio, when $w_1 = w_2 = \cdots = w_n = 1/n$. Then,

$$\ell_q = \|\boldsymbol{w}\|_q^q = \sum_{i=1}^n |w_i|^q = \sum_{i=1}^n \left|\frac{1}{n}\right|^q = n^{1-q} \ .$$

As $\ell_1 = \|\boldsymbol{w}\|_1 = \sum_{i=1}^n |w_i| = 1$, $i = 1, \dots, n$, the following relationship between norms holds true:

$$1 \leq \|\boldsymbol{w}\|_1 \leq \|\boldsymbol{w}\|_q^q \leq n^{1-q} \ .$$

□

Proof. Proof of Proposition 2

As the minimum variance problem (4) can be restated as a regression problem (see Section 3.1 Fan et al. (2012) for details), we can use results from regression analysis to derive some bounds for the approximation error of the minimum variance problem. Let's consider the regression problem with T observations and n regressors, where a dictionary of n predictions represents the initial estimate of the unknown true regression function. Using this estimate, we construct a linearly combined estimator that performs best among all linear combinations, i.e. the estimator with the smallest approximation error $\|\boldsymbol{w}\|_2^2 = \sum_{i=1}^n w_i^2$.

We know that the weights $|w_{(j)}|^q$ sum to $\|\boldsymbol{w}\|_q^q$ and are non-increasing. Therefore, the following inequality holds

$$|w_{(j)}|^q \leq \frac{\|\boldsymbol{w}\|_q^q}{j} \ .$$

Furthermore, we can write the approximation error as

$$\sum_{j>k+1} w_{(j)}^2 = \sum_{j>k+1} |w_{(j)}|^{2-q} |w_{(j)}|^q$$

and by using the following inequality in the first sum

$$|w_{(j)}| \leq \frac{\|\boldsymbol{w}\|_q}{(k)^{1/q}}$$

we have then

$$\sum_{j>k+1} w_{(j)}^2 \leq \frac{\|\boldsymbol{w}\|_q^{2-q}}{(k)^{(2-q)/q}} \|\boldsymbol{w}\|_q^q \leq \frac{\|\boldsymbol{w}\|_q^2}{(k)^{(2-q)/q}} \ .$$

□

Let $\boldsymbol{w}$ and $\boldsymbol{w}_{opt}$ be the theoretical and empirical optimal allocation vectors, solving the optimization problems $\boldsymbol{w} = \text{argmin}_{\mathbf{1}'\boldsymbol{w}=1, \|\boldsymbol{w}\|_q^q \leq c^q} \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$ and $\boldsymbol{w}_{opt} = \text{argmin}_{\mathbf{1}'\boldsymbol{w}=1, \|\boldsymbol{w}\|_q^q \leq c^q} \boldsymbol{w}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{w}$, where $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}$ are the theoretical covariance matrix and its estimate, respectively. Then, we define the oracle, empirical and actual risks as Section 3.

Proposition 3. *Let a_n represent the maximum componentwise estimation error, i.e. $a_n = \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty$. Then, under the assumptions in Fan et al. (2012), we have*

$$|R(\boldsymbol{w}) - R_n(\boldsymbol{w}_{opt})| \leq a_n c^2$$
$$|R(\boldsymbol{w}_{opt}) - R_n(\boldsymbol{w}_{opt})| \leq a_n c^2$$
$$|R(\boldsymbol{w}_{opt}) - R(\boldsymbol{w})| \leq 2a_n c^2.$$

These inequalities hold without any condition on the weights and show that the differences between oracle, empirical and actual risks are very small as long as c is not too large and the covariance estimate is precise.

Proof. Proof of Proposition 3 First, let's recall Theorem 1 in Fan et al. (2012), which states the following relationships between oracle, empirical and actual risk of a constrained minimum variance portfolio $\boldsymbol{w}$, with $\|\boldsymbol{w}\|_q \leq c$ (i.e. $\sqrt[q]{\ell_q} \leq c$)

$$
\begin{aligned}
|R_n(\boldsymbol{w}) - R(\boldsymbol{w})| &\leq a_n c^2 \\
|R(\boldsymbol{w}) - R_n(\boldsymbol{w}_{opt})| &\leq a_n c^2 \\
|R(\boldsymbol{w}_{opt}) - R(\boldsymbol{w})| &\leq 2 a_n c^2 \ .
\end{aligned}
$$

From norm inequalities, if $0 < q < p$, we know

$$
\|\boldsymbol{w}\|_p \leq \|\boldsymbol{w}\|_q \leq n^{1/q - 1/p} \|\boldsymbol{w}\|_p \ .
$$

Then, if $p = 1$ and $0 < q \leq 1$

$$
\|\boldsymbol{w}\|_1 \leq \|\boldsymbol{w}\|_q
$$

or, equivalently

$$
\ell_1 \leq \sqrt[q]{\ell_q} \ .
$$

As we solve the optimization problem (11) for $\|\boldsymbol{w}\|_q^q \leq c^q$, then $\|\boldsymbol{w}\|_1 \leq c$. The bounds on the differences between oracle, empirical and actual risks, reported in Theorem 1 in Fan et al. (2012), still hold. □