A generalized Description Length approach for Sparse and Robust Index Tracking

Davide Ferrari, *University of Melbourne*, davide.ferrari@unimelb.edu.au
Margherita Giuzio, *EBS Universität für Wirtschaft und Recht*, margherita.giuzio@ebs.edu
Sandra Paterlini, *EBS Universität für Wirtschaft und Recht*, sandra.paterlini@ebs.edu

Abstract. We develop a new minimum description length criterion for index tracking, which deals with two main issues affecting portfolio weights: estimation errors and model misspecification. The criterion minimizes the uncertainty related to data distribution and model parameters by means of a generalized q-entropy measure, and performs model selection and estimation in a single step, by assuming a prior distribution on portfolio weights. The new approach results in sparse and robust portfolios in presence of outliers and high correlation, by penalizing observations and parameters that highly diverge from the assumed data model and prior distribution. The Monte Carlo simulations and the empirical study on financial data confirm the properties and the advantages of the proposed approach compared to state-of-art methods.

Keywords. q-entropy, penalized least squares, sparsity, index tracking

1 Introduction

Since Markowitz [1], an optimal portfolio in asset allocation is determined by first considering the risk/return performance of each asset, in terms of mean and variance, and then selecting the portfolio with the best trade-off. Portfolio weights result then to be very sensitive to changes in parameter estimates, especially in presence of model misspecification and high dimensionality of the problem. Thus, estimation bias may heavily affect the optimization process resulting in suboptimal and unsatisfactory performance ([2], [3], [4]). Typically, asset returns are highly correlated with a leptokurtic distribution, which is largely contaminated by outliers [5]. If these statistical regularities are not properly considered, the misspecification of the data model may result in imprecise parameter estimates. To deal with these issues, several methods have been proposed in the financial literature, i.e. robust estimation methods, minimum divergence models and penalized least squares. We formulate a new criterion for portfolio selection that is able to deal with both estimation errors and model misspecification, and develop a general algorithm to obtain robust and sparse portfolios, i.e. with a low number of active positions.

In particular, we propose a description length criterion that codes the uncertainty about the data and the model parameters through a q-entropy, a generalized information measure [6] that accounts for the divergence from the assumed data model and the target prior distribution. It enhances the robustness of the portfolio to model misspecifications by assigning a lower weight to observations and parameter estimates that are not consistent with the assumed models. The whole criterion performs model selection and estimation in a single step and depends on the choice of two tuning parameters,

q and λ. The former manages the trade-off between accuracy and stability of parameter estimates [7], while the latter controls the penalization of portfolio weights.

Section 2 introduces the description length criterion. Section 3 describes the re-weighting algorithm for portfolio selection and the special cases in which data are assumed to follow a Normal or a t-Student distribution, while the prior distribution on the parameters is a Laplace function. Section 4 presents the simulation study comparing the performance of our method to the main state-of-art benchmark. Section 5 illustrates the behaviour of our portfolio selection method in an index tracking framework with real-world financial data. Section 6 concludes.

2 Description Length Criterion

Let a financial portfolio return be defined as $Y = \boldsymbol{\beta}^T \boldsymbol{X}$, where $\boldsymbol{X}$ is a p-dimensional random vector of asset returns with unknown multivariate distribution and $\boldsymbol{\beta}$ is the vector of asset weights. Given observations $x_i, i = 1, \ldots, n$, let μ and σ^2 be the portfolio expected return and variance. Then, the true probability density function of the standardized portfolio return $g(z)$, can be modelled through the function f, which may be for example the standard Normal or the t-Student distribution. Given a mean target value $\mu = \mu^*$, we can then compute portfolio weights $\widehat{\boldsymbol{\beta}}_{q,\lambda}$, by minimizing the following description length criterion:

$$\widehat{D}_{q,\lambda}(\boldsymbol{\beta}, \sigma) = -\sum_{i=1}^{n} L_q \left\{ f\left(\frac{\mathbf{x}_i^T \boldsymbol{\beta} - \mu^*}{\sigma} \right) \right\} - \sum_{j=1}^{p} L_q \left\{ \pi(\beta_j; \lambda) \right\}, \tag{1}$$

for fixed tuning constants $\lambda \geq 0$ and $q \leq 1$. In (1), $L_q(\cdot)$ is the generalized q-logarithm

$$L_q(u) = \begin{cases} (u^{1-q} - 1)/(1-q), & q \neq 1, \\ \log(u), & q = 1, \end{cases} \tag{2}$$

and $\pi(\beta_j; \lambda)$ is a symmetric distribution for β_j with zero mean and variance depending on λ. In the general framework, no restrictions are placed on the vector of portfolio weights $\boldsymbol{\beta}$. We notice that when $q \to 1$, criterion (1) is equal to maximum a posteriori (MAP) estimation of $\boldsymbol{\beta}$, where $\pi(\beta_j; \lambda)$ represent a prior probability density function on β_j. The penalty function $\pi(\beta_j; \lambda)$ controls the model selection and sparsity by shrinking to zero the weights of the assets that do not contribute to obtain a mean target value μ^*. From now on, $\pi()$ is assumed to be a Laplace function and then $L_q(\pi)$ results in a non-convex function. In (1), the first term represents the information provided by the data $\mathbf{x}_i$ given a model, while the second term encodes the information about the model itself, given by the prior distributions $\pi(\beta_j; \lambda)$. Minimizing this criterion results in the most efficient description of the data, including the description of the model itself [8]. Differentiating function (1) with respect to parameters $(\boldsymbol{\beta}, \sigma)^T$, we get the following estimating equations:

$$\mathbf{0} = \nabla \hat{D}_{q,\lambda}(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^{n} w_q(\mathbf{x}_i, \boldsymbol{\beta}, \sigma) \nabla \log f(\sigma^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} - \mu^*)) + \sum_{j=1}^{p} v_q(\boldsymbol{\beta}_j, \lambda) \nabla \log \pi(\beta_j; \lambda), \tag{3}$$

where

$$w_q(\mathbf{x}_i, \boldsymbol{\beta}, \sigma) = f(\sigma^{-1}(\mathbf{x}_i^T \boldsymbol{\beta} - \mu^*))^{1-q}, \qquad v_q(\boldsymbol{\beta}_j, \lambda) = \pi(\beta_j; \lambda)^{1-q} \tag{4}$$

are the vectors of weights applied to the observations and the parameters, respectively. The weights w_q downweight observations $\mathbf{x}_i$ that diverge from the assumed data model f, while v_q downweights the $|\hat{\beta}_j|$ that diverge from the assumed prior distribution π. For example, when $q < 1$, the linear combinations $\mathbf{x}_i^T \boldsymbol{\beta}$ that are far away from the target mean μ^* are assigned a small w_q. If $q \to 1$, $f(z)$ is the normal density function and $\pi(\beta; \lambda)$ is the Laplace function, we recover the popular Lasso method [9], in which $w_i = v_j = 1$. However, as shown by [10], since the weights in Lasso do not affect the optimization process, we may obtain unstable and inaccurate results in presence of large coefficients. Our approach proposes a remedy to such problem.

3 Re-weighting algorithms

The following section describes the weighting algorithm we introduce to estimate optimal portfolios in the general case in which data are assumed to follow a generic ditribution f, and then focus on the specific cases in which f is a Normal or a t-Student distribution. The aim of the optimization process is to obtain the parameter estimates $\widehat{\boldsymbol{\beta}}_{q,\lambda}$ by minimizing criterion (1). Since the L_q terms are typically non-convex in $\boldsymbol{\beta}$, we divide the whole process in several convex optimization steps. In particular, if we fix q, the vectors of weights w_q and v_q become $w_i, i = 1, \ldots, n$ and $v_j, j = 1, \ldots, p$, and the criterion results in a penalized likelihood problem that we can solve with an iteratively re-weighted scheme: given the weights w_i and v_j, we estimate $\widehat{\boldsymbol{\beta}}_{q,\lambda}$ by solving equation (3) and then update the weights using the new parameter estimates. We call this process a doubly re-weighted (2RE) algorithm as the re-weighting is applied to both data and penalty scores.

Algorithm 3.1.
Given the tuning constants $q \leq 1$, $\lambda \geq 0$, and a target portfolio return μ^*, the algorithm consists of the following steps:

Step 0 *At Iteration $s = 0$, compute the parameter estimates $\widehat{\boldsymbol{\beta}}^{(s)}$ and $\widehat{\sigma}^{(s)}$.*

Step 1 *Set $s = s + 1$, and update the vector of weights as*

$$\widehat{w}_i^{(s)} = f((\mathbf{x}_i \widehat{\boldsymbol{\beta}}^{(s-1)} - \mu^*)/\widehat{\sigma}^{(s-1)})^{1-q}, \qquad \widehat{v}_j^{(s)} = \pi(\widehat{\beta}_j^{(s-1)}; \lambda)^{1-q}. \tag{5}$$

Step 2 *Compute the parameter estimates $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\sigma}$ by minimizing*

$$\sum_{i=1}^{n} \widehat{w}_i \log f((\mathbf{x}_i^T \boldsymbol{\beta} - \mu^*)/\sigma) + \sum_{j=1}^{p} \widehat{v}_j \log \pi(\beta_j; \lambda). \tag{6}$$

Step 3 *Update $\widehat{\boldsymbol{\beta}}^{(s)}$ and $\widehat{\sigma}^{(s)}$ by solving $f(\mathbf{x}_i^T \widetilde{\boldsymbol{\beta}} - \mu^*)/\widetilde{\sigma})^q$ for $\boldsymbol{\beta}$ and σ.*

Step 4 *Repeat Steps 1 and 2 until a stopping criterion is satisfied.*

In Step 3, a re-scaling operation re-centers the estimates to correct the bias arising from the weights $w_q(\mathbf{x}_i, \boldsymbol{\beta}, \sigma)$, as suggested by [7].

The parameter λ controls the penalty term on the $\boldsymbol{\beta}$ coefficiants and regulates the sparsity of the portfolio. The literature suggests to choose such tuning parameters by information criteria like the AIC and BIC. As [11], given a certain level of q, we select the optimal values of λ by minimizing the robust Bayesian Information Criterion defined as below, where $k \leq p$ is the number of active positions:

$$\text{BIC}_q = -2 \sum_{i=1}^{n} L_q \left\{ f\left(\frac{\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_{q,\lambda} - \mu^*}{\hat{\sigma}_{q,\lambda}} \right) \right\} + \log(n)k. \tag{7}$$

Normal portfolios

If we assume that data follow a p-variate normal distribution and $\pi(\beta_j; \lambda)$ is a Laplace function, then $Y \sim N(\mu, \sigma^2)$. In this case, the 2RE algorithm can be adapted as follows.

Algorithm 3.2.
Given $q \leq 1$, $\lambda \geq 0$, and a target return μ^*:

Step 0 *At Iteration $s = 0$, initialize $w_i^{(s)}$, $v_j^{(s)}$ and $\sigma^{(s)}$.*

Step 1 *Set $s = s + 1$, and obtain $\widehat{\boldsymbol{\beta}}^{(s)}$ by solving*

$$\widehat{\boldsymbol{\beta}}^{(s)} = \underset{\boldsymbol{\beta}}{argmin} \left\{ \sum_{i=1}^{n} \widehat{w}_i^{(s-1)} \frac{1}{2} \left(\frac{\mu^* - \mathbf{x}_i^T \boldsymbol{\beta}}{\widehat{\sigma}^{(s-1)}} \right)^2 + \lambda \sum_{j=1}^{p} \widehat{v}_j^{(s-1)} |\beta_j| \right\}, \tag{8}$$

Step 2 *Update the vectors of weights as*

$$\widehat{w}_i^{(s-1)} = \left[\frac{1}{\sqrt{2\pi\widehat{\sigma}^{2(s-1)}}}\exp\left\{-\frac{\left(\mu^* - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}^{(s-1)}\right)^2}{2\widehat{\sigma}^{2(s-1)}}\right\}\right]^{1-q}, \widehat{v}_j^{(s-1)} = \left[\frac{\lambda}{2}\exp\left\{-\lambda|\widehat{\beta}_j^{(s-1)}|\right\}\right]^{1-q}. \quad (9)$$

Step 3 *When the portfolio variance is a fixed target σ^{*2}, we set $\widehat{\sigma}^{2(s)} = \sigma^{*2}$, for all $s \geq 0$; otherwise*

$$\widehat{\sigma^{2}}^{(s)} = \frac{\sum_{i=1}^n \widehat{w}_i^{(s-1)}\left(\mu^* - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}^{(s-1)}\right)^2}{q\sum_{i=1}^n \widehat{w}_i^{(s-1)}}. \quad (10)$$

Step 4 *Repeat Steps 1 to 3 until a stopping criterion is satisfied.*

The optimization function in (8) is a weighted L_1-penalized least squares problem that we solve by applying the gradient projection algorithm developed by [12]. Other algorithms, like coordinate wise and quadratic optimization ([9]), could be used to efficiently estimate $\widehat{\boldsymbol{\beta}}^{(s)}$. However, as the gradient projection is faster and updates parameters and solutions by using the optimal values of the previous iteration as warm-start points ([13]), we rely on it for solving the penalized least squares problem.

t-porfolios

If we assume the portfolio Y to be a non standardized t-Student distribution with mean μ, variance σ and number of degrees of freedom $\nu > 1$ (i.e. $Y \sim f_\nu(\mu, \sigma)$), then Step 2 of the 2RE algorithm computes $\{\widehat{\boldsymbol{\beta}}^{(s)}, \widehat{\sigma}^{(s)}\}$ as

$$\underset{\boldsymbol{\beta},\sigma}{\text{argmin}}\left\{-\left(\frac{\nu+1}{2}\right)\sum_{i=1}^n \widehat{w}_i^{(s-1)}\log\left\{1 + \frac{(\mathbf{x}_i\boldsymbol{\beta}^T - \mu^*)^2}{\nu\sigma^2}\right\} + \lambda\sum_{j=1}^p \widehat{v}_j^{(s-1)}|\beta_j|\right\}, \quad (11)$$

where $\sigma > 0$. While the penalty weights $\widehat{v}_j$ are updated as in (9), the data weights $\widehat{w}_i$ are obtained as

$$\widehat{w}_i^{(s-1)} = \left[f_\nu\left(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}^{(s-1)}; \mu, \widehat{\sigma}^{(s-1)}\right)\right]^{1-q}, \quad i = 1, \ldots, n. \quad (12)$$

When data are assumed to follow the nonstandardized t-Student distribution and $\lambda \to 0$, equation (11) results in biased estimates for $\boldsymbol{\beta}$ and σ. Thus, according to Proposition 1 in [7], we solve this issue by adjusting the degrees of freedom parameter: we use $\nu_q = q\nu + (q-1)$ instead of ν. Also, the optimization function (11) represents a non-convex problem, which results in imprecise estimates if solved directly. Therefore, by writing a t-Student observation as a scale mixture of normals $Y_i \sim N(\mu, \sigma^2 Z_i^{-1})$, where Z_i follows a Gamma distribution $Z_i \sim \text{Ga}(\nu/2, \nu/2)$, we derive an EM algorithm, which efficiently estimates the optimal solutions as follows.

Algorithm 3.3.

For any $s > 0$, we set the initial weights $\widehat{z}_i = 1/n$, $i = 1, \ldots, n$ and estimate $\widehat{\boldsymbol{\beta}}^{(s)}$ and $\widehat{\sigma}^{(s)}$ through the expectation-maximization steps:

M-Step *Estimate $\boldsymbol{\beta}$ and σ as*

$$\boldsymbol{\beta}' = \underset{\boldsymbol{\beta}}{\textit{argmin}}\left\{\sum_{i=1}^n \widehat{w}_i^{(s-1)}\widehat{z}_i^{(s-1)}\frac{1}{2}\left(\frac{\mathbf{x}_i^T\boldsymbol{\beta} - \mu}{\widehat{\sigma}^{(s-1)}}\right)^2 + \lambda\sum_{j=1}^p \widehat{v}_j^{(s-1)}|\beta_j|\right\}, \quad (13)$$

$$\sigma'^2 = \frac{\sum_{i=1}^n \widehat{w}_i^{(s-1)}\widehat{z}_i^{(s-1)}\left(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}} - \mu\right)^2}{\sum_{i=1}^n \widehat{w}_i^{(s-1)}\widehat{z}_i^{(s-1)}} \times \frac{\nu}{(\nu+1)q - 1}. \quad (14)$$

E-Step *Update the mixing constants $\widehat{z}_i$, such that*

$$\widehat{z}_i = \frac{(\nu_q + 1)\sigma'^2}{\nu_q\sigma'^2 + \widehat{w}_i^{(s-1)}(\mathbf{x}_i^T\boldsymbol{\beta}' - \mu)^2}, \quad i = 1, \ldots, n, \quad (15)$$

4 Simulation study

In the following simulation study, we evaluate and compare the behaviour of the 2RE algorithm, for both normal (GDL_N) and t-Student portfolios (GDL_t), with respect to the Lasso penalization model. In particular, we want to test the robustness of the proposed methods in presence of outliers and correlated assets $\boldsymbol{X}$. We simulate data from a multivariate t-Student distribution with ν degrees of freedom: $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where $\mu_j = 1$, if $j \leq k$, and $\mu_j = 0$, if $j > k$, and the covariance matrix has diagonal elements $\Sigma_{jj} = 1$, $j = 1, \ldots, p$, and off-diagonal elements $\Sigma_{jk} = \rho$, $0 \leq \rho < 1$ $j \neq k$. We construct four settings by considering four different levels of correlation between assets, $\rho = 0.2, 0.4, 0.6, 0.8$. For each setting, we generate $B = 50$ samples with $n = 500$, $p = 50$, $k = 10$.

We evaluate the average performance of the B portfolios in terms of sparsity, model selection performance and risk/return characteristics with respect to a specifc target $\mu^* = k$. In particular, we compute (i) the number of active positions as $\widehat{k} = \sum_{j=1}^{p} I(|\widehat{\beta}_j| > \tau)$, where $\tau = 0.005$ is a threshold value, below which the estimated weights are set equal to zero; (ii) the F-measure to assess whether the portolios select the "correct" assets, which in our model are the ones in the first k positions; (iii) the Monte Carlo mean squared error to compare the risk/return performance to the specified target:

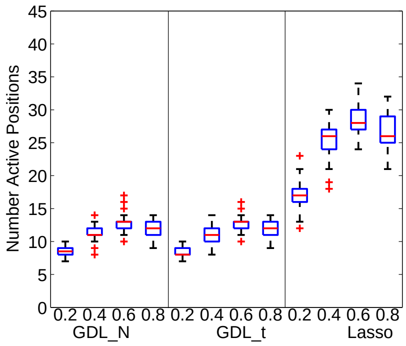
$$\widehat{MSE} = \frac{1}{B} \sum_{b=1}^{B} \left(\frac{\boldsymbol{\mu}^T \widehat{\boldsymbol{\beta}}_b - \mu^\star}{\sqrt{\widehat{\boldsymbol{\beta}}_b^T \boldsymbol{\Sigma} \widehat{\boldsymbol{\beta}}_b}} \right)^2, \qquad F\text{-}measure = 2 \frac{|supp(\boldsymbol{\beta}^*)| \cap |supp(\widehat{\boldsymbol{\beta}})|}{|supp(\boldsymbol{\beta}^*)| + |supp(\widehat{\boldsymbol{\beta}})|}, \tag{16}$$

where, given a vector $\boldsymbol{\beta}$, the support is equal to $\text{supp}(\boldsymbol{\beta}) = \{j : |\beta_j| \geq \tau\}$, and $\boldsymbol{\beta}^*$ represents the vector of weights whose first k positions are equal to 1.

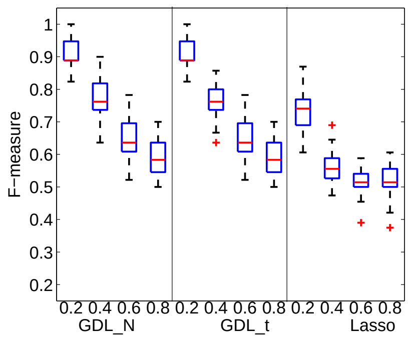
For each setting, we set $q = 0.9$ and select from a grid of values the λ associated to the model with the lowest BIC. We then compare the average portfolio performances with the ones obtained using Lasso. As specified in Section 3, we handle the optimization problem by using the DC-programming as proposed by [13]. As the EM algorithm is very sensitive to the initialization of β, we initialize the Lasso and the GDL_N algorithms with the OLS β estimates, while the GDL_t approach uses instead the optimal estimates obtained by the GDL_N. Finally, the initial vectors of weights w_i and v_j are set equal to $w_i = 1/n$ and $v_j = 1/p$.

Figure 1 shows from left to right the boxplots of the average number of active positions $\hat{k}$ estimated by the GDL methods and Lasso (a), and the relative F-measure (b) and MSE (c) obtained in 50 simulations for different values of correlation $\rho = 0.2, 0.4, 0.6, 0.8$ on the x-axis. We can compare the performance of the three methods in terms of sparsity and selection ability, and analyse their robustness in presence of correlated data.

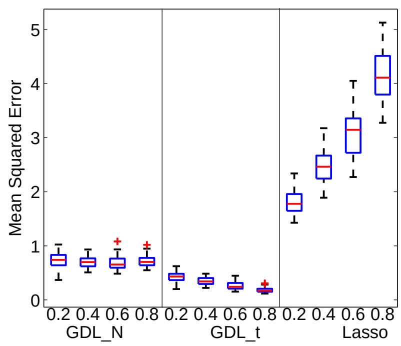
First of all, we notice that the GDL criteria estimate much sparser portfolios than Lasso for each value of ρ. The number of active positions is very close to the optimal value of 10 and it is not influenced by the level of correlation between assets (Panel (a)). The stability of the GDL criteria represents a clear advantage when comparing with Lasso, whose performance becomes worse when ρ increases: on average it selects approximately 17 assets when $\rho = 0.2$ and 27 assets when $\rho = 0.8$, against the 8 and 11 assets selected by the GDL_t with ρ equal to 0.2 and 0.8, respectively. In terms of F-measure, the GDL approaches obtain better performance than Lasso as closer to 1, showing very good model selection properties. However, for all the methods, the average value of F-measure highly depends on the level of ρ (Panel (b)): when data exhibit low correlation, Lasso obtains a value of 0.74 while GDL_N and GDL_t are closer to the maximum of 1, that represents the case in which we select the correct vector of assets $\boldsymbol{\beta}^*$; when data are highly correlated, Lasso presents a value of 0.52, while the GDL methods obtain approximately 0.6. The GDL_N and GDL_t algorithms show similar results in terms of sparsity and F-measure since they both select the same active positions and their estimated weights differ only in magnitude. Finally, we analyse the overall performance of the three methods with respect to the return target μ^* by comparing their MSE. Though the two GDL criteria slightly differ in their results, they both outperform the Lasso, whose performance get much worse when data show high correlation (i.e. with $\rho = 0.8$ the MSE is twice the value obtained with $\rho = 0.2$). As expected, given that the true model is a t-Student one, the GDL_t obtains the lowest MSE in all settings, indicating very good performance. However, this advantage might also result from the initialization of the vector of β as the optimal solution of the GDL_N algorithm.



(a)



(b)



(c)

Figure 1: Average number of estimated active positions $\hat{k}$, F-measure and Mean Squared Error for different levels of correlation ρ in 50 simulations, using GDL for Normal and t-Student, and Lasso methods.

Further simulations considering different set-ups support the main reported findings. Results are available upon request. This study points out the main advantages of the proposed approach with respect to a well-known benchmark: (i) the sparsity of the selected portfolios obtained by penalizing and weighting the vector of asset weights β; (ii) the high robustness of the estimates in presence of correlation between assets, which is ensured by weighting the observations according to their divergence from an assumed distribution.

5 Sparse and Robust Index Tracking

In this section, we test our approach in an index tracking framework, where we try to reproduce the performance obtained by a certain index by selecting a vector of active weights only for some of its components, in order to limit transaction and managing costs. The optimization problem can be described as a regression problem, where the dependent variable $\boldsymbol{y}$ represents the vector of index returns and $\boldsymbol{X}$ is the return matrix of its components.

Using a penalized technique may help to obtain good out-of-sample performance with respect to the index by optimally selecting a small number of components. In order to evaluate the behaviour of the proposed GDL criteria, we focus on three financial indexes by using $n = 1401$ daily return observations of the Fama & French 100, the S&P 200 and the S&P 500, with different number of constituents p, equal to 100, 200 and 500, respectively. For each index, we compare the performance of three strategies: the GDL for Normal and t-Student portfolios, and the Lasso.

We estimate the optimal portfolios using a rolling window sample of 250 observations, and compute the excess return of the first out-of-sample observation with respect to the index. For the GDL criteria, we set $q = 0.9$ and select the λ in each window as described in Section 3. First, we evaluate the risk/return performance of the optimal portfolios through the Information Ratio (IR), which is computed dividing the excess return by the tracking error volatility (TEV). Then, we check sparsity by means of the number of estimated active positions $\hat{k}$ and finally, we test the tracking ability computing the correlation with respect to the index.

Table 1 shows the out-of-sample statistics of each tracking strategy. In terms of IR (Column 4), the GDL_N has the best performance for F&F 100 and S&P 500, while the Lasso outperforms the other strategies in the second dataset, S&P 200. However, the GDL criteria always obtain a lower out-of-sample TEV (Column 3), which is a characteristic already underlined in the simulation study, where the GDL showed smaller MSE than Lasso. This result is even more important if we consider that the GDL strategies select very sparse solutions for each dataset (Column 5). While the Lasso always uses approximately 66 positions, the GDL strategies select 35% of the available assets for the first index, less than 25% for the second index and less than 10% for the third index. In terms of

Strategy	ER (%)	TEV (%)	IR	$\bar{k}$	TO	Cor
PANEL A: F&F 100						
GDL N	0.338	0.624	0.542	37.749	0.068	0.999
GDL t	0.170	0.492	0.346	32.241	0.066	0.999
Lasso	1.030	2.117	0.486	65.939	0.017	0.990
PANEL B: S&P 200						
GDL N	0.319	4.500	0.071	36.950	0.399	0.963
GDL t	-2.421	4.897	-0.494	28.431	0.520	0.933
Lasso	4.760	7.267	0.655	66.532	0.037	0.950
PANEL C: S&P 500						
GDL N	2.906	6.966	0.417	44.564	0.605	0.932
GDL t	1.192	9.018	0.132	27.770	0.811	0.872
Lasso	2.986	10.315	0.289	66.407	0.053	0.926

Table 1: Out-of-sample statistics of each tracking portfolio: strategy (column 1), annualized excess return ER (column 2), tracking error volatility TEV (column 3), Information Ratio IR (column 4), average number of active components $\bar{k}$ (column 5), turnover TO (column 6), correlation w.r.t. index Cor (column 7).

tracking ability, the GDL_N portfolios achieve values of Cor near 1 and outperform the Lasso by closer tracking the indexes, especially in small dataset, where the TEV is lower.

6 Conclusion

In this paper we propose a generalized description length criterion to obtain sparse and robust portfolios in presence of estimation errors and model misspecification. By relying on a q-entropy measure, the approach minimizes the uncertainty about the distributions of data and model parameters by assigning a lower weight to observations and parameters that diverge from the assumed models. After deriving the general estimation algorithm, we specify two interesting cases, in which data are assumed to follow a Normal or a t-Student distribution, and develop the corresponding algorithms, GDL_N and GDL_t. The simulation study supports the theoretical properties of the GDL criterion and shows that it achieves better performance in terms of sparsity, stability and robustness of the estimates with respect to the well-known Lasso benchmark, especially when data exhibit high correlation. The empirical results presented for the index tracking framework show that the GDL criterion is able to obtain good out-of-sample estimates and reproduce the performance of an index by using only a small number of its components in order to limit transaction and managing costs.

Bibliography

[1] H. Markowitz, Portfolio selection, The Journal of Finance 7 (1952) 77–91.

[2] M. J. Best, R. R. Grauer, On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results, The Review of Financial Studies 4(2) (1991) 315342.

[3] R. Jagannathan, T. Ma, Risk reduction in large portfolios: Why imposing the wrong constraints helps, Journal of Finance LVIII (2003) 1651–1683.

[4] V. De Miguel, F. J. Nogales, Portfolio selection with robust estimation, Operations Research 57 (2009) 560–577.

[5] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance 1 (2001) 223–236.

[6] J. Havrda, F. Charvát, Quantification method of classification processes: Concept of structural entropy, Kibernetika 3 (1967) 30–35.

[7] D. Ferrari, D. La Vecchia, On robust estimation via pseudo-additive information, Biometrika 99 (1) (2012) 238–244.

[8] P. Grünwald, The Minimum Description Length Principle (Adaptive Computation and Machine Learning), The MIT Press, 2007.

[9] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society 58 (1996) 267–288.

[10] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of American Statistical Association 96 (2001) 1348–1360.

[11] E. Ronchetti, Robustness aspects of model choice, Statistica Sinica 7 (1997) 327–338.

[12] M. Figueiredo, R. Nowak, S. Wright, 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 (2007) 586–598.

[13] G. Gasso, A. Rakotomamonjy, S. Canu, Recovering sparse signals with a certain family of non-convex penalties and dc programming, IEEE Transactions on Signal Processing 57, 12 (2009) 4686–4698.