Penalized Least Squares for Optimal Sparse Portfolio Selection

Bjoern Fastrich, *University of Giessen*, Bjoern.Fastrich@wirtschaft.uni-giessen.de
Sandra Paterlini, *EBS Universität für Wirtschaft und Recht*, Sandra.Paterlini@ebs.edu
Peter Winker, *University of Giessen*, Peter.Winker@wirtschaft.uni-giessen.de

Abstract. Markowitz portfolios often result in an unsatisfying out-of-sample performance, due to the presence of estimation errors in inputs parameters, and in extreme and unstable asset weights, especially when the number of securities is large. Recently, it has been shown that imposing a penalty on the 1-norm of the asset weights vector not only regularizes the problem, thereby improving the out-of-sample performance, but also allows to automatically select a subset of assets to invest in. Here, we propose a new, simple type of penalty that explicitly considers financial information and consider several alternative non-convex penalties, that allow to improve on the 1-norm penalization approach. Empirical results on U.S.-stock market data support the validity of the proposed penalized least squares methods in selecting portfolios with superior out-of-sample performance with respect to several state-of-art benchmarks.

Keywords. Penalized Least Squares, Regularization, LASSO, Non-convex penalties, Minimum Variance Portfolios

1 Introduction

The Markowitz mean-variance portfolio model [1] is the cornerstone of modern portfolio theory. Given a set of assets with expected return vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, Markowitz's model aims to find the optimal asset weight vector that minimizes the portfolio variance, subject to the constraint that the portfolio exhibits a desired portfolio return. Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, some estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ must be obtained from a finite sample of data to compute the optimal asset allocation vector. As financial literature has largely shown, using sample estimates can hardly provide reliable out-of-sample asset allocations in practical implementations [2],[3],[4],[5],[6]. [7], [8], [2], and [9] already provided strong empirical evidence that estimates of the expected portfolio return and variance are very unreliable. Here, we focus on the minimum-variance portfolio (MVP), which relies solely on the covariance structure and neglects the estimation of expected returns altogether [10],[11],[12],[13],[14],[15],[16]. Somewhat surprisingly, MVPs are usually found to perform better out-of-sample than portfolios that consider asset means [17, 11, 6], because the (co)variances can be estimated more accurately than the means. A superior performance also prevails when performance measures consider both portfolio means and variances. Nevertheless, MVPs still suffer considerably from estimation errors [10],[11],[12].

One stream of research has recently focused on shrinking asset allocation weights by using penalized least squares methods. Among the first contributors, [18] and [19] use ℓ_1-penalization to obtain stable and sparse (i.e. with few active weights) portfolios, which is an adaptation of the Least Absolute Shrinkage and Selection Operator (LASSO) by [20]. The LASSO relies on imposing a constraint on the ℓ_1-norm the regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^K$, where $\ell_1 = |\beta_1| + ... + |\beta_K|$. Recently, [14] provide both theoretical and empirical evidence supporting the use of ℓ_1-penalization to identify sparse and stable portfolios by limiting the gross exposure, showing that this causes no accumulation of estimation errors, the result of which is an outperformance compared to standard Markowitz portfolios. Further examples of penalised methods applied in the Markowitz framework are [21, 22, 23], and [15].

Despite the appeal of using ℓ_1-penalization in portfolio optimization to estimate (numerically stable) asset weights and select the portfolio constituents in a single step by solving a convex optimization problem, [24] show that the ℓ_1-penalty, as a linear function of absolute coefficients, tends to produce biased estimates for large (absolute) coefficients. As a remedy, they suggest using penalties that are singular at the origin, just like the ℓ_1-penalty, in order to promote sparsity, but non-convex, in order to countervail bias. Ideally, a good penalty function should result in an estimator with three properties: unbiasedness, sparsity, and continuity. Then, new non-convex penalties such as the so-called Smoothly Clipped Absolute Deviation (SCAD) , the Zhang-penalty, the Log-penalty and the ℓ_q-penalties with $0 < q < 1$ were introduced (e.g. see [25] for a comparison). The seemingly nice properties of non-convex penalties come at the cost of posing a difficult optimization challenge, which, however, can nowadays be solved quite efficiently by using a dual-convex appraoch, as suggested by [25]. An alternative to non-convex approaches, which can still retain the oracle property, has been suggested by [26]. His approach is now known as the adaptive LASSO and has proven to be able to prevent bias while preserving convexity of the optimization problem, and thus clearly alleviates the optimization challenge as compared to the non-convex approaches.

This work contributes to the literature on portfolio regularization by proposing a new, simple type of convex penalty, which is inspired by the adaptive LASSO and explicitly considers financial information to optimally determine the portfolio composition. Moreover, we are the first to apply non-convex penalties in the Markowitz framework to identify sparse and stable portfolios with desiderable out-of-sample properties, when dealing with a large number of assets.

2 Penalized Approaches for Minimum Variance Portfolios

Given a set of K assets and a penalty function $\rho(\cdot)$, the regularized minimum-variance problem can be stated as:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w} \in \mathbb{R}^K}{\operatorname{argmin}} \left\{ \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} + \lambda \sum_{i=1}^{K} \rho(w_i) \right\} \tag{1}$$

$$\text{subject to} \qquad \mathbf{1}_K'\boldsymbol{w} = 1 \ , \tag{2}$$

where $\boldsymbol{w}^*$ is the optimal (and potentially sparse) $(K \times 1)$-vector of asset weights, $\mathbf{1}_K$ is a $(K \times 1)$-vector of ones and λ is the regularization parameter that controls the intensity of the penalty and thereby the sparsity of the optimal portfolio. The optimization problem (1) can be re-written as a penalized least square problem.

Assuming we estimate $\mathbf{\Sigma}$ by $\widehat{\mathbf{\Sigma}}$ and we set $\lambda = 0$, the solution to problem (1)-(2) is the MVP, where the optimized portfolio weights vector $\boldsymbol{w}^*$ is (over)fitted to the correlation structure in $\widehat{\mathbf{\Sigma}}$, thereby assuming absence of estimation error and unlimited trust in the precision of the estimate $\widehat{\mathbf{\Sigma}}$, which is obviously very naive. On the contrary, whenever $\lambda > 0$, the penalty term $\sum_{i=1}^{K} \rho(w_i)$ will allow to control for the estimation error by selecting only few active weights. The larger λ, the smaller the number of active weights and the total amount of shorting. The optimal solution $\boldsymbol{w}^*$ is thus determined by a trade-off between the estimated portfolio risk and the corresponding penalty term, whose magnitude is controlled by λ.

In this work, we focus on penalty functions $\rho(\cdot)$ that are singular at the origin and thus allow a shrinkage of the components in $\boldsymbol{w}$ to exactly zero. Hence, the corresponding approaches not only stabilize the problem to improve the out-of-sample performance, but simultaneously also conduct the asset selection step. Table 1 reports the definition of the six penalties functions we consider.

The Least Absolute Shrinkage and Selection Operator (LASSO) has already received considerable attention in the portfolio optimization context and therefore we choose it as a benchmark to test the validity of the newly proposed approaches. Due to the budget constraint, the minimum value that $||\boldsymbol{w}||_1$ can be shrunk to is one. This is possible only when the portfolio weights are shrunk towards zero until they are all non-negative, identifying the so-called no-shortsale portfolio. Increasing values of λ cause the construction of portfolios with less shorting, or more precisely, with a shrunken ℓ_1-norm of the portfolio weight vector. This prevents the estimation errors contained in $\widehat{\mathbf{\Sigma}}$ from entering unhindered in the portfolio weight vector. Note that while the intensity of shrinkage is controlled by the value of λ, the decision as to which assets to shrink and to which relative extent is determined by the estimated correlation structure.

The weighted Lasso approach, henceforth $w8Las$, was proposed in its statistical formulation by [26] to countervail the difficulties of the LASSO that are related to potentially biased estimates of large true coefficients [24]. The idea is to replace the equal penalty that is applied to all coefficients (here portfolio weights) with a penalization-scheme that can vary among the K portfolio weights. This can be achieved by introducing a weight ω_i for each of the absolute portfolio weights $|w_i|$. In general, the intuition is to over- or underweight some assets in comparison to the LASSO in order to improve performance. Specifically, this intuition depends on the method used to determine the ω_i, for which no "blueprint" exists in a portfolio optimization context. We suggest determining the (individual) regularization weights λ_i by considering specific financial time series properties that are ignored when many, e.g. $T = 250$, historical observations are used to estimate one (constant) covariance matrix. In particular, we focus on comparing short-term and log-term estimates of the volatilities to extract some signals, such that if the short term volatility is below the long-term volatility estimate, a smaller penalty λ_i is applied and, consequently, a larger portfolio weight in comparison to the LASSO. Due to space limitations, we refer to [27] for a detailed description of the implementation of the $w8Las$ penalty.

While LASSO and $w8Las$ are convex penalties, as Figure 1 shows, the remaining four penalties (i.e. SCAD, Zhang, Log and ℓ_q with $0 < q < 1$) are non-convex and allow to deal with the potentially biased LASSO estimates of large absolute coefficients. The economic intuition behind the non-convex penalties is as follows: if the true correlation of assets is high, shorting can reduce the risk, since it accounts for true similarities of the assets instead of being the result

Table 1: Penalties

penalty		$\lambda\rho(w_i)$	domains
LASSO	=	$\lambda\|w_i\|$	all
$w8Las$	=	$\lambda\omega_i\|w_i\|$	all
SCAD	=	$\begin{cases} \lambda\|w_i\| \\ \frac{-\|w_i\|^2+2a\lambda\|w_i\|-\lambda^2}{2(a-1)} \\ \frac{(a+1)\lambda^2}{2} \end{cases}$	$\|w_i\| \leq \lambda$ $\lambda < \|w_i\| \leq a\lambda$ $a\lambda < \|w_i\|$
Zhang	=	$\begin{cases} \lambda\|w_i\| \\ \lambda\eta \end{cases}$	$\|w_i\| < \eta$ $\eta \leq \|w_i\|$
L_q	=	$\lambda\|w_i\|^q,\ 0<q<1$	all
Log	=	$\lambda\ln(\|w_i\|+\phi)$ $-\lambda\ln(\phi)$	all

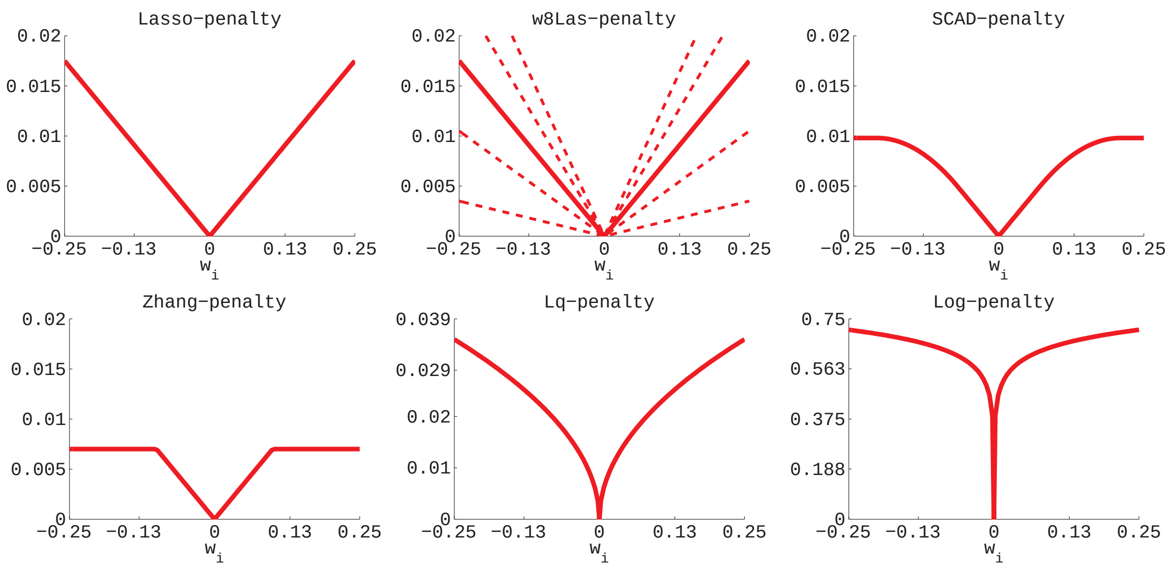

Figure 1: The six (non-)convex penalty functions under consideration in this work.

Table 2: U.S. stock market datasets for the period 23.08.02 to 27.03.08

	dataset	source	obs	K	$\bar{r}$	$\widehat{\sigma}$	$\hat{S}$	$\hat{K}$
S&P200:	largest firms (w.r.t. ME)	Datastream	1401	200	6.57	14.79	0.0487	5.32
S&P500:	largest firms (w.r.t. ME)	Datastream	1401	500	6.57	14.77	0.0410	5.13
S&P1036:	largest firms (w.r.t. ME)	Datastream	1401	1036	6.39	14.88	0.0380	4.99

Table 2 reports the datasets under consideration, the source of the data, the number of assets (K), and the number of observations (obs) in each dataset. For the S&P datasets, value weighted indices are computed whose return distributions are characterized by the mean p.a. $\bar{r}$, the standard deviation p.a. ($\widehat{\sigma}$), the skewness ($\widehat{S}$), and the kurtosis ($\widehat{K}$) given in the last four columns. The S&P indices are market value weighted. The weighting schemes are updated daily and applied the following day.

of overfitting. Analogously, large portfolio weights tend to be appropriate if the true correlations are small. Now, if a correlation structure is "strong enough" to grow absolute portfolio weights – against the counteracting penalty – large enough, it is considered reliable and should therefore enter the portfolio to a greater extend. The main differences between them, as pointed out by Figure 1 is on the intensity on penalizing the different asset weights. The ℓ_q- and the Log-penalty provide a particularly strong incentive to avoid small and presumably dispensable positions in favor of selecting a small subset of presumably indispensable assets. This tendency to construct very sparse and less diversified portfolios coincides with the suggestion of [28] to use the ℓ_q-norm as a diversity measure for portfolios.

3 Empirical Analysis

Data and Experimental Set-Up

We consider daily observations of five different datasets shown in Table 2 that represent the U.S. stock market at different levels of aggregation. Datasets are characterized by a different number of constituents, which include the 200, 500, and 1036 largest individual firms (with respect to the market value on March 27, 2008) of the S&P 1500, which we label as *large* datasets. We refer to [27] for results also on the 48 industry portfolios and the 98 firm portfolios provided by Kenneth French, which could be considered as *small* dataset.

We backtest the out-of-sample performance of the proposed methods with a moving time window procedure, where $\tau = 250$ in-sample observations (corresponding to one year of market data) are used to form a portfolio. The optimized portfolio allocations are then kept unchanged for the subsequent 21 trading days (corresponding to one month of market data) and the out-of-sample returns are recorded. After holding the portfolios unchanged for one month, the time window is moved forward, so that the formerly out-of-sample days become part of the in-sample window and the oldest observations drop out. The updated in-sample window is then used to form a new portfolio, according to which the funds are reallocated. The $T = 1401$ observations allow for the construction of $\Gamma = 54$ portfolios with the corresponding out-of-sample returns.

Table 3 shows the different measures we use to evaluate the out-of-sample performance and the composition of the portfolios, where $F_r^{-1}(p)$ is the value of the inverse cumulated empirical distribution function of the daily out-of-sample returns at point p.

Table 3: Portfolio evaluation measures

Measures based on the out-of-sample portfolio returns		
Portfolio variance (s^2)	*Sharpe ratio (SR)*	*95% Value-at-Risk (VaR)*
$\frac{1}{T-\tau-1}\sum_{t=\tau+1}^{T}(r_t-\bar{r})^2$	$\frac{\bar{r}}{\sqrt{s^2}}$	$\lvert F_r^{-1}(0.05)\rvert$
Measures based on the portfolio composition		
No. active positions (No. act.)	*Shorting (Short)*	*Turnover (TO)*
$\frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma}\lvert\{i \mid w_{i,\gamma}\neq 0\ \forall\ i\}\rvert$	$\frac{1}{\Gamma}\sum_{j=\{i \mid w_{i,\gamma}<0\ \forall\ i\}} -w_{j,\gamma}$	$\frac{1}{\Gamma-1}\sum_{\gamma=2}^{\Gamma}\sum_{i=1}^{K}\left\lvert w_{i,\gamma}-w_{i,\gamma-1}\right\rvert$

For comparative evaluations, we also implement the following standard benchmarks: (i) the shortsale-unconstrained MVP, denoted MVPssu, the shortsale-constrained MVP, denoted MVPssc, the market value weighted portfolio, denoted mvw, and the equally weighted portfolio, denoted 1oK.

To determine the optimal minimum variance portfolio, we choose to focus on three types of frequently used covariance matrix estimators: (i) the sample estimator, (ii) a three-factor model estimator [10] and (iii) the Ledoit-Wolf estimator [12]. However, we report in the following results related to the three-factor model and refer the reader to [27] for a complete empirical analysis.

Determining the Regularization Parameter

Prior to optimizing problem formulation (1)-(2) for any of the six penalization approaches, a value of the regularization parameter λ must be chosen. Since the optimal values λ^* for the various penalties are unknown, we try for each approach a set of 30 ascending values starting from zero. The largest element in each set is chosen such that the resulting portfolios exhibit only few active positions and a high out-of-sample portfolio variance. In this manner, it is most likely that the intervals spanned by zero and the largest regularization parameters cover λ^*.

Each of the 30 regularization parameters corresponds to one specific (optimized) portfolio, which demands a decision about in which one to eventually invest. This difficult decision is the reason we split the empirical experiments into two setups: (i) we keep track of *all* 30 portfolios that correspond to the *entire spectrum* of 30 regularization parameters over all periods; (ii) we invest in only *one* portfolio by applying ten-fold cross-validation to choose a suited value of λ prior to the investment decision in each period. While procedure (ii) is more realistic from an investment perspective,[1] procedure (i) provides valuable insights into the potential benefit of regularization and how different values of λ affect the portfolio performance. However, due to space limitations, we refer the reader to [27] for results related to the entire spectrum of regularization parameters and we focus in the next section on results related to the cross-validation procedure.

[1]The cross-validation procedure is as follows: 21 observations are randomly picked from the in-sample data, portfolios are optimized on the remaining 229 observations for all 30 regularization parameters, and the portfolio variance is computed using the 21 picked observations. This is done ten times and the λ is chosen that corresponds to smallest average portfolio variance.

Table 4: Three-factor model covariance matrix (cross-validation experiment)

	MVPssu	MVPssc	mvw	1oK	Lasso	w8Las	Log	ℓ_q	Zhang	SCAD
					Panel A: S&P 200 individual firms					
$s^2 \cdot 10^5$	3.007	3.162	6.023	6.524	2.843	2.808	3.017	3.009	2.777	2.942
$VaR \cdot 10^2$	0.885	0.898	1.312	1.348	0.828	0.824	0.893	0.916	0.843	0.881
SR	0.054	0.062	0.018	0.050	0.049	0.050	0.054	0.048	0.049	0.054
No. act.	200.0	54.9	200.0	200.0	82.6	91.1	66.1	65.6	93.9	64.8
Short	0.75	0.00	0.00	0.00	0.26	0.29	0.38	0.38	0.32	0.39
TO	0.57	0.52	0.04	0.00	0.59	0.68	0.96	0.98	0.73	0.90
					Panel B: S&P 500 individual firms					
$s^2 \cdot 10^5$	2.883	3.796	6.081	6.799	2.529	2.495	2.617	2.601	2.538	2.643
$VaR \cdot 10^2$	0.923	1.071	1.335	1.385	0.834	0.835	0.794	0.814	0.847	0.842
SR	0.031	0.042	0.018	0.045	0.043	0.043	0.043	0.049	0.042	0.036
No. act.	500.0	278.6	500.0	500.0	131.9	147.6	102.8	108.1	151.6	101.0
Short	0.83	0.00	0.00	0.00	0.20	0.24	0.33	0.35	0.24	0.33
TO	0.61	0.22	0.04	0.00	0.69	0.75	1.11	1.04	0.80	1.09
					Panel C: S&P 1036 individual firms					
$s^2 \cdot 10^5$	2.649	4.593	6.254	9.001	2.382	2.379	2.343	2.356	2.485	2.369
$VaR \cdot 10^2$	0.833	1.166	1.352	1.566	0.802	0.792	0.775	0.789	0.819	0.754
SR	0.031	0.031	0.016	0.028	0.054	0.050	0.041	0.045	0.050	0.044
No. act.	1036.0	572.4	1036.0	1036.0	276.7	308.3	179.6	153.8	298.7	161.3
Short	0.84	0.00	0.00	0.00	0.26	0.30	0.33	0.31	0.28	0.31
TO	0.65	0.22	0.04	0.00	0.84	0.89	1.30	1.13	0.87	1.26

Table 4 shows results of the four benchmarks and the six regularization approaches for the three large datasets and the three-factor model covariance matrix.

Empirical Results

Table 4 shows that the cross-validation approach works well for the considered large datasets. The out-of-sample variances of the penalized approaches are always lower than the constraned minimum variance approach (MVPssc) and the equally weighted (mvw) and often also than the unconstrained minimum variance portfolio (MVPssu). This shows that the possibility of having a stronger shrinkage in some periods but not in others is beneficial. The only exception is for the S&P 200 dataset in Panel A, where the Log- and the ℓ_q-regularized portfolios exhibit even higher risks than the MVPssu. However, this fits the picture that the non-convex approaches perform the better the larger the number of constituents compared to the number of observations, which corresponds to a window size of 250. The $w8Las$ reaches the smallest variance for both S&P200 and S&P500, while the Log-penalty outperforms for S&P1036. In terms of Sharpe Ratio, the equally weighted portfolio is a tough benchmark, especially for S&P500, where only the ℓ_q-penalty allows to reach a slightly larger value by using just an average subset of 108.1 active components. Lasso, $w8Las$ and Zhang penalty reach the largest Sharpe Ratios values for S&P1036, while still investing in an average number of assets much larger than the Log, ℓ_q and SCAD penalties. Clearly, as the non-convex penalties lead often to sparser solutions than other methods, they end up paying a price in terms of turnover rates and identify optimal portfolios with larger shorting amounts, while the extreme risks, as captured by VaR and ES, are still often smaller than the MVPssu, MVPssc and Mvw portfolios.

4 Conclusions

Introducing a penalty in the Markowitz minimum variance framework can allow to determine optimal portfolios that better control for estimation error and have superior out-of-sample performances than the unconstrained approach and the equally weighted benchmark. In particular, we propose a new type of a (convex) penalty whose construction allows for easy processing of all kinds of signals to optimized portfolios, may they be gained from (time series) econometrics, fundamental or technical analysis, or expert knowledge. Moreover, we consider four non-convex penalty functions that have not yet been examined in a portfolio optimization context. It turned out that these approaches perform very well when dealing with very large datasets, where they not only outperformed standard benchmarks but also the (convex) "state-of-the-art" LASSO approach. The success of these approaches stems from their ability to maintain relevant assets in the portfolio with large absolute weights, while only the weights of the remaining assets are shrunk. This allows for a better exploitation of the higher potential to diversify portfolio risk in larger datasets. Further research aims to further develop the underlying signal extraction that could be operationalized in the $w8Las$ approach and investigate alternative cross-validation criteria, which likely will allow for a further improvement of the results.

Bibliography

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

[2] J. Jobson, R. Korkie, Estimation for Markowitz efficient portfolios, Journal of the American Statistical Association 75 (371) (1980) 544–554.

[3] M. Best, J. 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) 315–342.

[4] M. Broadie, Computing efficient frontiers using estimated parameters, Annals of Operations Research 45 (1) (1993) 2158.

[5] M. Britten-Jones, The sampling error in estimates of mean-variance efficient portfolio weights, Annals of Operations Research 54 (2) (1999) 655–671.

[6] V. DeMiguel, J. Garlappi, R. Uppal, Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy?, Review of Financial Studies 22 (5) (2009) 1915–1953.

[7] G. Frankfurter, H. Phillips, J. Seagle, Portfolio selection: The effects of uncertain means, variances, and covariances, Journal of Financial and Quantitive Analysis 6 (5) (1971) 1251–1262.

[8] J. Dickinson, The reliability of estimation procedures in portfolio analysis, Journal of Financial and Quantitive Analyis 9 (3) (1974) 447–462.

[9] P. Frost, J. Savarino, For better performance: Constrain portfolio weights, Journal of Portfolio Management 15 (1) (1988) 29–34.

[10] L. Chan, J. Karceski, J. Lakonishok, On portfolio optpimization: Forecasting covariances and choosing the risk model, The Review of Financial Studies 12 (5) (1999) 937–974.

[11] R. Jagannathan, T. Ma, Risk reduction in large portfolios: Why imposing the wrong constraints helps, The Journal of Finance 58(4) (2003) 1651–1683.

[12] O. Ledoit, M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, Journal of Empirical Finance 10 (5) (2003) 603–621.

[13] V. DeMiguel, F. Nogales, Portfolio selection with robust estimation, Operations Research 57 (3) (2009) 560–577.

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

[15] M. Fernandes, G. Rocha, T. Souza, Regularized minimum-variance portfolios using asset group information, Available from http://webspace.qmul.ac.uk/tsouza/index_arquivos/Page497.htm (2012) 1–28.

[16] P. Behr, A. Guettler, F. Truebenbach, Using industry momentum to improve portfolio performance, Journal of Banking and Finance 36 (5) (2012) 1414–1423.

[17] P. Jorion, Bayes-Stein estimation for portfolio analysis, Journal of Financial and Quantitative Analysis 21 (3) (1986) 279–292.

[18] J. Brodie, I. Daubechies, C. DeMol, D. Giannone, D. Loris, Sparse and stable Markowitz portfolios, Proceedings of the National Academy of Science USA 106 (30) (2009) 1226712272.

[19] V. DeMiguel, L. Garlappi, J. Nogales, R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, Management Science 55 (5) (2009) 798–812.

[20] R. Tibshirani, Regression shrinkage and selection via the Lasso, Royal Statistical Society 58 (1) (1996) 267–288.

[21] Y.-M. Yen, A note on sparse minimum variance portfolios and coordinate-wise descent algorithms, Available from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1604093 (2010) 1–27.

[22] M. Carrasco, N. Noumon, Optimal portfolio selection using regularization, Working Paper University of Montreal; available from http://www.unc.edu/maguilar/metrics/carrasco.pdf.

[23] Y.-M. Yen, T.-J. Yen, Solving norm constrained portfolio optimizations via coordinate-wise descent algorithms, Available from http://personal.lse.ac.uk/yen/sp_090111.pdf (2011) 1–41.

[24] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association 96 (456) (2001) 1348–1360.

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

[26] H. Zou, The adaptive lasso and its oracle properties, Journal of the American Statistical Association 101 (476) (2006) 1418–1429.

[27] B. Fastrich, S. Paterlini, P. Winker, Constructing optimal sparse portfolios using regularization methods, Working paper; available from `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2169062`.

[28] R. Fernholz, R. Garvy, J. Hannon, Diversity weighted indexing, Journal of Portfolio Management 24 (2) (1998) 74–82.