

Doubly Regularized Portfolio with Risk Minimization

Weiwei Shen[†] and Jun Wang[‡] and Shiqian Ma^{*}

[†]Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

[‡]Business Analytics and Mathematical Sciences, IBM T. J. Watson Research, Yorktown Heights, NY, USA

^{*}Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong
ws2215@columbia.edu, wangjun@us.ibm.com, sqma@se.cuhk.edu.hk

Abstract

Due to recent empirical success, machine learning algorithms have drawn sufficient attention and are becoming important analysis tools in financial industry. In particular, as the core engine of many financial services such as private wealth and pension fund management, portfolio management calls for the application of those novel algorithms. Most of portfolio allocation strategies do not account for costs from market frictions such as transaction costs and capital gain taxes, as the complexity of sensible cost models often causes the induced problem intractable. In this paper, we propose a doubly regularized portfolio that provides a modest but effective solution to the above difficulty. Specifically, as all kinds of trading costs primarily root in large transaction volumes, to reduce volumes we synergistically combine two penalty terms with classic risk minimization models to ensure: (1) only a small set of assets are selected to invest in each period; (2) portfolios in consecutive trading periods are similar. To assess the new portfolio, we apply standard evaluation criteria and conduct extensive experiments on well-known benchmarks and market datasets. Compared with various state-of-the-art portfolios, the proposed portfolio demonstrates a superior performance of having both higher risk-adjusted returns and dramatically decreased transaction volumes.

Introduction

Academic research in portfolio management and allocation has had a remarkable impact on many aspects of the financial services industry, from mutual fund management to asset pricing and insurance to corporate risk management. Most applications build upon the pathbreaking work of Markowitz (Markowitz 1952), which provides optimal rules of allocating wealth across assets by estimating means and covariances of asset returns in a single *static* period. However, due to the inaccurate estimation of parameters this mean-variance framework is notorious for producing extreme portfolio weights that fluctuate substantially over time and perform poorly under out-of-sample settings (Michaud 1989).

Typically, the mean-variance portfolios aim at maximizing expected returns for a given level of risk tolerance or minimizing the investment risk while achieving a target return. The trade-off between these two critical factors, i.e. *risk* and *return*, plays a central role in the Markowitz portfolio theory. Generally, portfolio allocation with an explicit measure of *risk-adjusted returns* rather than gross growth has been highlighted in finance. One of the common standard criteria for measuring the *risk-adjusted return* is called the Sharpe ratio (Sharpe 1966).

Meanwhile, the rapid growth of globalized markets urges the implementation of advanced data analysis tools in finance. Recently, machine learning algorithms have been identified as important technical analysis tools and have called intensive attention. Different from the perspectives of finance community, machine learning researchers rely more on the real time data stream to design optimal portfolio strategies (Blum and Kalai 1999; Borodin, El-Yaniv, and Gogan 2004; Agarwal et al. 2006; Györfi, Lugosi, and Udina 2006; Li and Hoi 2012; Li et al. 2012). For example, one recent paper (Li and Hoi 2012) estimates on-line moving average reversion from sequential market data and then maximizes expected returns to achieve optimal portfolio selection rules. This approach and many other papers mainly depend upon the effectiveness of the growth optimal portfolio (GOP) strategy (Thorp 1971). However, the principal drawbacks of GOP (Samuelson 1969), such as high return volatility along the investment horizon and extremely long time realization, inevitably appear in those data-driven GOP strategies. Moreover, due to the complexity of cost models characterizing market frictions such as transaction costs, market impacts and capital gain taxes (Cvitanic 2001), many papers either ignore costs or only subtract ad hoc costs afterward. Therefore, huge rebalancing volumes, substantial changes of asset allocation and extreme high asset *turnover* rates can often be observed across all the investment periods. Admittedly, incorporating a full-fledged cost model is difficult. It should involve the factors related to financial rules, policy and specific trading activities, such as commission rules for different asset classes and investors, taxation policy for different institutions and investment periods, and implicit costs from the real trading implementation given a strategy.

In order to overcome those limitations, in this paper, we propose a doubly regularized portfolio allocation strategy to bridge the gap between return chasing and cost reduction. Specifically, we take advantage of the classic mean-variance portfolio framework to control volatility risk and impose two regularization terms on the portfolio structure to reduce trading volumes thus implicitly mitigating the costs from market frictions. The first regularization term is motivated from *structured sparsity* (Huang, Zhang, and Metaxas 2009), where we aim to determine portfolio weights that are allocated across assets sparsely. In other words, only a small number of assets are considered to form a low-risk portfolio. The second regularization term underlines a *consistent allocation*, where we ensure portfolio weights between consecutive investment periods are similar. Those two regularization terms help to control trading volumes such that costs from market frictions can be implicitly reduced. Besides, they represent typical concerns of individual investors in private wealth management. To validate the proposed strategy, we utilize a battery of standard finance metrics, including *Sharpe ratios*, *cumulative wealth*, *turnovers* and *volatility* to measure the performance from various perspectives. Our extensive empirical studies and comparisons over several well-known financial benchmarks and real-world market data clearly illustrate the superiority of the new strategy.

Mean-Variance Portfolio

We briefly review one of the most representative portfolio frameworks, i.e., *Mean-Variance Portfolio* (MVP), which is a cornerstone of Markowitz’s modern portfolio theory (Markowitz 1952). The fundamental assumption of MVP is that investors are risk averse, which means that they tend to choose the portfolio with a lower risk if profits are the same. Accordingly, the only way to achieve a higher profit is to take more risk. Since MVP characterizes the risk by the variance of asset returns and models the profits by the expected asset returns, investors will optimize their investment by selecting mean-variance efficient portfolios. In particular, assume at time t the covariance matrix of asset returns is Σ_t and the expected net return is denoted by a column vector \mathbf{r}_t . The MVP weight ω_t is obtained by solving the following optimization problem:

$$\begin{aligned} \omega_t^* &= \arg \min_{\omega_t} \omega_t^\top \Sigma_t \omega_t - \nu \mathbf{r}_t^\top \omega_t \quad (1) \\ \text{s.t.} \quad &\omega_t^\top \mathbf{1} = 1, \end{aligned}$$

where ω_t indicates the proportion of the invested wealth across all assets, and $\nu > 0$ is a risk tolerance factor. The above cost function that has two components, the variance of portfolio returns $\omega_t^\top \Sigma_t \omega_t$ and the expected return $\mathbf{r}_t^\top \omega_t$, captures the aforementioned risk-return tradeoff.

Since the ingenious work by Markowitz, many variants of MVP have been developed (Brandt 2010). Among them, without considering estimating means, minimum-variance portfolios often perform stronger than mean-variance portfolios for out-of-sample tests (Jagannathan and Ma 2003). That is because it is more difficult to accurately estimate means than covariances of asset returns (Merton 1980) and

Table 1: Summary of key notations

Symbol	Definition
N	Number of assets
T	Number of investing periods
t	Index of investment periods
ω_t	Portfolio weight vector
ω_{t-}	Re-normalized portfolio before rebalancing
$\omega_{(j)_t}$	Portfolio weight on the j^{th} asset at time t
$S_{(j)_t}$	Price of the j^{th} asset at time t
$R_{(j)_t}$	Gross return of the j^{th} asset at time t
$\nabla \omega_{(j)_t}$	Trade of the j^{th} asset at time t
$\hat{\Sigma}_t$	Estimated covariance matrix at time t

errors in estimates of means have a larger impact on portfolio weights than those in covariances. In addition, portfolio performances could be improved when a shrinkage estimator or a one factor model is applied for covariance matrix estimation (Ledoit and Wolf 2003). The recent papers (Brodie et al. 2009; DeMiguel et al. 2009; Fan, Zhang, and Yu 2012) show superior portfolio performances when various types of norm regularities are combined in the MVP framework.

Methodology

We first briefly introduce the notations and financial terms used in this paper. Investment time is modeled as discrete and indexed as $t = 0, \dots, T$, with $t = 0$ the initial period and $t = T$ the terminal period. The investment over a set of assets in the t^{th} time period is denoted by a column vector representing portfolio weights $\omega_t = \{\omega_{(j)_t}\}_{j=1}^N$, where N is the number of assets. Since $\omega_{(j)_t}$ specifies the percentage of invested wealth over the j^{th} asset, the sum of all the allocated portfolio weights always equals one, i.e. $\omega_t^\top \mathbf{1} = \sum_{j=1}^N \omega_{(j)_t} = 1$. In addition, the value $\omega_{(j)_t} > 0$ indicates that investors take a *long* position of the j^{th} asset, and $\omega_{(j)_t} < 0$ means that they take a *short sale* position. A short sale position means investors borrow an asset to sell and then invest its liquidation on other assets. A negative sign of the short sale weight indicates investors will suffer a loss if the price of this asset starts to mount. We also denote $\mathbf{R}_t = \{R_{(j)_t}\}_{j=1}^N$ as the asset gross returns from $t - 1$ to t . The gross return $R_{(j)_t}$ for the j^{th} asset is computed as $R_{(j)_t} = S_{(j)_t}/S_{(j)_{t-1}}$, where $S_{(j)_t}$ and $S_{(j)_{t-1}}$ are the prices of the j^{th} asset at time t and $t - 1$, respectively. Accordingly, the portfolio net return μ_t from time t to $t + 1$ can be easily calculated as $\mu_t = \omega_t^\top \mathbf{R}_{t+1} - 1$. We will use the term “return” to represent “gross return” in the sequel. Table 1 summarizes the key notations in the paper.

Formulation

We consider investors who operate portfolio management in a market containing N assets with the return \mathbf{R}_t at time t . According to the change of the portfolio weights between consecutive rebalancing periods, the trades at time t are

computed by:

$$\nabla \omega_t = \omega_t - \omega_{t-} = [\nabla \omega_{(1)t} \cdots \nabla \omega_{(N)t}]^\top, \quad (2)$$

where $\nabla \omega_{(j)t} = \omega_{(j)t} - \omega_{(j)t-}$ indicates the trades of rebalancing the j^{th} asset at the t^{th} time period; ω_{t-} denotes the re-normalized portfolio weight before the rebalancing at time t . Specifically, investors buy more asset j if $\nabla \omega_{(j)t} > 0$, and sell asset j if $\nabla \omega_{(j)t} < 0$. After a trading period, the portfolio weights have changed due to the asset price fluctuation. Thus, we need to calculate the re-normalized portfolio weight ω_{t-} as

$$\omega_{t-} = \frac{\omega_{t-1} \circ \mathbf{R}_t}{\omega_{t-1}^\top \mathbf{R}_t}, \quad (3)$$

where the symbol \circ denotes the Hadamard product and $\omega_{t-1}^\top \mathbf{R}_t$ represents the portfolio return at time t .

By the principle of risk minimization, we derive the doubly regularized portfolio, where the objective is to choose a limited set of assets to invest and control the changes of the asset positions. The changes of the positions during each rebalancing period are directly related to transaction costs, market impacts and taxes, while the number of selected assets is related to management costs. In particular, the regularization term on the position changes is denoted by the ℓ_2 -norm as

$$\|\omega_t - \omega_{t-}\|_2^2 = \sum_{j=1}^N [\omega_{(j)t} - \omega_{(j)t-}]^2. \quad (4)$$

The sparse selection of assets is realized through imposing the ℓ_1 regularization term as

$$\|\omega_t\|_1 = \sum_{j=1}^N |\omega_{(j)t}|. \quad (5)$$

Then the *doubly regularized* portfolio allocation is formed as

$$\begin{aligned} \min_{\omega_t} \omega_t^\top \hat{\Sigma}_t \omega_t + \lambda_1 \|\omega_t\|_1 + \lambda_2 \|\omega_t - \omega_{t-}\|_2^2 \quad (6) \\ \text{s.t.} \quad \omega_t^\top \mathbf{1} = 1. \end{aligned}$$

The core component in the above objective function is the minimization of the portfolio variance $\omega_t^\top \hat{\Sigma}_t \omega_t$, where $\hat{\Sigma}_t \in \mathbb{R}^{N \times N}$ is the estimated covariance matrix of asset returns from historical data at time t . Besides the unit constraint of the sum of the portfolio weights, we impose two additional regularization terms weighted by two coefficients λ_1 and λ_2 . The first regularization with an ℓ_1 -norm enforces *structured sparsity* to portfolio weights (Huang, Zhang, and Metaxas 2009) so that ω_t will only have a few non-zero elements, which indicates only a small set of assets are selected. In addition, the sparsity regularization term prohibits any extreme holding of *long* or *short-sale* positions of a particular asset. The second regularization with an ℓ_2 -norm $\|\omega_t - \omega_{t-}\|_2^2$ imposes the consistency of portfolio weights before and after each rebalance. Hence, the ℓ_2 regularization term implicitly reduces *turnover* rates and trading costs by moderately constraining the changes of asset positions. In summary, with

Algorithm 1 Doubly Regularized Portfolio (DRP)

Input: Asset returns data points $\{\mathbf{R}_{-\tau+1}, \dots, \mathbf{R}_T\}$,

Initialization:

Investment period $t = 0$;

Initial portfolio weight vector ω_0 ;

for $t = 1 \rightarrow T - 1$ **do**

Estimate the covariance matrix of asset returns:

$$\hat{\Sigma}_t \leftarrow \{\mathbf{R}_{-\tau+t+1}, \dots, \mathbf{R}_t\}$$

Renormalize the portfolio weight vector ω_{t-} using Equation (3);

Compute the optimal portfolio weight vector ω_t by solving Equation (6);

Perform rebalancing $\omega_t - \omega_{t-}$;

end for

Output:

The portfolio weight vectors $\{\omega_t\}_{t=0}^{T-1}$ and the portfolio returns $\{\mu_t\}_{t=0}^{T-1}$.

the core component of minimum-variance and the two structure regularizations, we can derive the *doubly regularized* portfolio (DRP) with risk minimization.

Note that the above doubly regularized quadratic loss in Equation (6) can be linked to existing learning frameworks in both machine learning and statistics. For example, if $\lambda_1 + \lambda_2 = 1$, the doubly regularized form becomes the so called elastic-net penalty, a convex combination of the lasso and ridge penalty (Zou and Hastie 2005). A similar regularization framework has also been combined with the hinge loss to derive the doubly regularized support vector machine (Wang, Zhu, and Zou 2006).

Optimization and Analysis

In the above formulation of DRP, we need to estimate the covariance matrix of asset returns $\hat{\Sigma}_t$. Instead of using a sample covariance matrix, we apply a factor model (Fan, Fan, and Lv 2008), which has been shown effective for estimating high-dimensional covariance matrices. Besides, different from the focus of growth optimal portfolio research (Li and Hoi 2012), we do not rely on any prediction of portfolio returns in the procedure of designing portfolio allocation rules. It is well-known that the prediction of future returns is extremely challenging (Merton 1980; Rapach and Zhou 2012). In addition, we notice that a proposed gross-exposure constrained (GEC) MVP also achieves a sparse selection of assets (Fan, Zhang, and Yu 2012). However, since the consistency of consecutive portfolios has not been considered, there is no guarantee that the GEC portfolio can reduce asset trading volumes and the corresponding costs from market frictions. Hence, the uniqueness of the proposed DRP portfolio lies in the advantages of minimizing investment risks and meanwhile reducing trading costs. We discuss the optimization strategy of the DRP portfolio below.

Without considering the ℓ_1 -norm regularization term in the objective function, the optimization problem defined in Equation (6) is a standard convex quadratic problem (QP). Sparsity-inducing norms have often been adapted in vari-

Table 2: Summary of the tested datasets

Dataset	Frequency	Time Period	# of data	N
FF25	Monthly	07/01/1963 - 12/31/2004	498	25
FF48	Monthly	07/01/1963 - 12/31/2004	498	48
FF100	Monthly	07/01/1963 - 12/31/2004	498	100
ETF	Weekly	01/01/2008 - 10/30/2012	252	139
INDEX	Weekly	02/15/2008 - 10/30/2012	246	26
EQUITY	Weekly	01/01/2008 - 10/30/2012	252	181

ous machine learning problems if the given a priori knowledge supports such assumptions (Bach et al. 2011), similar to our formulation for the sparse portfolio allocation. Without considering the constraint $\omega_t^\top \mathbf{1} = 1$, problem (6) can be viewed as an ℓ_1 -norm regularized least squares problem, which is known as Lasso (Tibshirani 1996) in statistics and a sparse recovery problem in compressed sensing (Candès and Tao 2006). Many efficient algorithms have been proposed for solving lasso or ℓ_1 sparse recovery problems. To name just a few, these methods include FISTA (Beck and Teboulle 2009), FPC (Hale, Yin, and Zhang 2008), TFOCS (Becker, Candès, and Grant 2011), etc. When both the ℓ_1 -norm and the constraint $\omega_t^\top \mathbf{1} = 1$ are present, the problem is much harder than QP and Lasso. Although it is possible to implement an efficient first-order algorithm based on the projected proximal gradient method (Boyd and Vandenberghe 2004) or the alternating direction method of multipliers (Boyd et al. 2011), we use a commercial software toolbox TOMLAB/SNOPT developed by the Stanford Systems Optimization Laboratory (SOL) for simplicity.

Sequential Portfolio Allocation

In order to perform the process of sequential portfolio allocation, we first need to choose a time window τ over which to perform the estimation of $\hat{\Sigma}_t$. To avoid overfitting, we split the entire dataset into two parts. The first τ data points, denoted as $\{\mathbf{R}_{-\tau+1}, \dots, \mathbf{R}_0\}$, are used as the training data to estimate the initial covariance matrix. The sequential portfolio allocation starts from $t = 0$ and lasts a total number of T time periods. We apply the “rolling-horizon” procedure (DeMiguel et al. 2009) to sequentially perform portfolio allocation and evaluate the performance. Specifically, we shift the estimation window along the time axis by including the data point of the next period and dropping the data point of the earliest period. This “rolling-horizon” procedure is repeated until the end of the dataset. At the end of this process, we will have T portfolio vectors derived for each strategy, denoted as $\{\omega_t\}_{t=0}^{T-1}$. The portfolio allocation at the last point $t = T$ is excluded due to the lack of future data for performance evaluation. For calculating the initial portfolio ω_0 , we can either choose a simple heuristics such as an equally-weighted portfolio, or derive the optimal solution of the problem defined in Equation (6) without the consistency regularization. The algorithm chart 1 summarizes the sequential procedure of the proposed DRP strategy.

Table 3: Portfolio Sharpe ratios (%) with p -values

Method	FF25	FF48	FF100	ETF	INDEX	EQUITY
EW	26.58 (1.00)	24.30 (1.00)	26.97 (1.00)	6.01 (1.00)	5.98 (1.00)	9.03 (1.00)
VW	28.53 (0.01)	23.37 (0.22)	29.76 (0.00)	5.85 (0.22)	5.74 (0.60)	8.95 (0.86)
OLMAR	21.50 (0.01)	24.48 (0.93)	23.60 (0.22)	7.61 (0.44)	15.80 (0.06)	12.15 (0.27)
FMMV	39.71 (0.02)	27.34 (0.60)	45.94 (0.00)	11.51 (0.16)	22.54 (0.14)	13.86 (0.07)
GEC	39.85 (0.01)	30.34 (0.13)	46.44 (0.00)	17.87 (0.12)	22.69 (0.15)	13.84 (0.07)
DRP	41.67 (0.00)	30.37 (0.13)	48.27 (0.00)	19.64 (0.15)	23.33 (0.12)	13.88 (0.05)

Experiment

In this section, we will report the results of our empirical studies using several real-world benchmarks.

Data

In the experiments, two types of datasets are chosen for performance validation and comparison. The first type of datasets is from the well-known academic benchmarks called Fama and French datasets (Fama and French 1992). Briefly speaking, based on the data sampled from the U.S. stock market, the benchmarks were formed for different financial segments. For example, FF48 contains monthly returns of 48 portfolios representing different industrial sectors, and FF100 includes 100 portfolios formed on the basis of size and book-to-market ratio. Fama and French datasets have been recognized as standard testbeds and heavily adopted in finance research because of its extensive coverage to asset classes and very long historical data series. The second type of datasets represents three popular financial asset classes, exchange-traded funds (ETF), major world stock market indices (INDEX), and individual equities sampled from the large-cap segment of the Russell 200 index (EQUITY). The data of the prices are all crawled from Yahoo! Finance in a weekly base from 2008 to 2012. They represent the commonly chosen investment opportunity sets by investors. For example, ETFs that have the advantage over conventional mutual funds of low costs, tax efficiency, and stock-like features have been highlighted in both the industry and academia. The 26 major stock market indices cover all the mature world wide financial markets. The selected stocks pool of the large-cap segment of the U.S. stock market measure the performance of the 200 largest U.S. companies and cover 63% of total market capitalization.¹

Table 2 summarizes those benchmark datasets used in our experiments. Note that these two types of datasets provide different perspectives for performance evaluation. The Fama and French datasets essentially emphasize the long-term performance of the proposed strategy, as results of such long historical datasets introduce limited selection bias and

¹After removing those with incomplete historical data, for which the assets did not exist at the beginning of the chosen trading periods, finally we have 181 stocks in the EQUITY dataset.

Table 4: Cumulative wealth

Method	FF25	FF48	FF100	ETF	INDEX	EQUITY
EW	98.06	54.77	123.92	1.20	1.12	1.30
VW	137.05	48.06	198.32	1.19	1.22	1.29
OLMAR	33.56	42.34	57.74	1.21	1.24	1.34
FMMV	246.35	46.78	523.47	1.05	1.30	1.38
GEC	225.37	53.64	496.01	1.22	1.30	1.38
DRP	266.32	54.74	646.91	1.27	1.31	1.38

are difficult to be manipulated. The experiments on the real world short-term datasets underline the robustness of the proposed portfolio with respect to the high volatility environment after the recent financial crisis.

Experimental Settings

We set the length of the estimation window as $\tau = 120$ (DeMiguel et al. 2009), which means that the previous 120 data points are used to make the current decisions of portfolio allocation. For the parameters in DRP such as λ_1 and λ_2 , cross validation is applied to determine the optimal values. For the comparison study, we consider the following methods: a) equally-weighted portfolio (EW); b) value-weighted portfolio (VW); c) factor model based minimum-variance portfolio (FMMV) (Fan, Fan, and Lv 2008); d) gross-exposure constrained (GEC) portfolio (Fan, Zhang, and Yu 2012); e) on-line moving average reversion (OLMAR) based portfolio (Li and Hoi 2012); f) the proposed doubly regularized portfolio (DRP). The first four portfolios typify those in finance. For example, the EW portfolio is a naive approach yet has been empirically shown to mostly outperform 14 models across seven empirical datasets (DeMiguel, Garlappi, and Uppal 2009). VW mimics a market portfolio. The FMMV portfolio adopts a factor model to estimate high-dimensional covariance matrices and has shown a large performance improvement over MVP. The GEC portfolio shows better results than MVP by simply imposing a sparsity constraint. The OLMAR portfolio that is based on GOP represents a more data-driven approach developed by machine learning researchers and has been shown robust and outperforms 12 different portfolio strategies across five datasets. For all the compared approaches, we follow the recommended parameter settings in the corresponding studies.

Performance Metrics

We compare the out-of-sample performance of the portfolios using four standard criteria in finance (Brandt 2010): (i) *Sharpe ratios*; (ii) *cumulative wealth*; (iii) *turnovers*; (iv) *volatility*. These four evaluation metrics represent different focuses on measuring portfolio performance. The *Sharpe ratio* (SR) measures the reward-to-risk ratio of a portfolio strategy, which is computed as the portfolio return normalized by its standard deviation:

$$SR = \frac{\hat{\mu}}{\hat{\sigma}}, \quad \hat{\mu} = \frac{1}{T} \sum_{t=0}^{T-1} \mu_t, \quad \hat{\sigma} = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} (\mu_t - \hat{\mu})^2}. \quad (7)$$

Table 5: Portfolio turnovers (%)

Method	FF25	FF48	FF100	ETF	INDEX	EQUITY
OLMAR	36.88	26.22	32.23	17.22	14.96	15.55
FMMV	43.11	27.39	55.23	99.72	7.60	10.99
GEC	36.80	6.05	46.49	13.23	7.56	10.77
DRP	12.75	4.23	20.42	4.78	5.01	9.62

where $\hat{\mu}$ is the mean of portfolio net returns and $\hat{\sigma}$ represents the standard deviation of portfolio net returns.

The *cumulative wealth* of a portfolio measures the total profit yield from the portfolio strategy across investment periods without considering any risks and costs. By starting with one dollar, the *cumulative wealth* (CW) is computed by

$$CW = \prod_{t=0}^{T-1} \omega_t^\top \mathbf{R}_{t+1}. \quad (8)$$

The portfolio *turnover* indicates the frequency of trading activities and the volumes of rebalancing. It has been recognized as an important performance metric for portfolio management, since a high turnover inevitably generates high explicit and implicit trading costs thus degenerating the real return. The portfolio *turnover* rate is calculated as an average absolute value of the rebalancing trades across all the assets and over all the trading periods:

$$\text{Turnover} = \frac{1}{T-1} \sum_{t=1}^{T-1} \|\omega_t - \omega_t^-\|_1, \quad (9)$$

where $T-1$ indicates the total number of the rebalancing periods and ω_t^- is the re-normalized portfolio weight vector before rebalance.

Finally, the *volatility* represents a quantitative measurement of investment risk, which is computed as the standard deviation of returns. To achieve a fair comparison of the portfolios with the different rebalancing periods, we compute the annualized volatility by $\sqrt{M}\hat{\sigma}$, where M is the frequency of the rebalance each year, where $\hat{\sigma}$ is computed as (7). For the monthly and weekly rebalancing frequency, $M = 12$ and $M = 52$, respectively.

To measure the statistical significance of the difference between the volatility and the Sharpe ratio for two comparing portfolios, we report the p -values under the corresponding results in Tables 3 and 6. To compute the p -values for the case of no i.i.d. returns, we adopt the studentized circular block bootstrapping methodology (Ledoit and Wolf 2011). We test the statistical significance by using the EW portfolio as the benchmark, 1000 bootstrap resamples, a 95% significance level, and a block size equal to 5.

Results

Tables 3, 4, 5 and 6 summarize portfolio performance evaluated by the *Sharpe ratios*, *cumulative wealth*, *turnovers* and *volatility* for all the tested benchmarks, respectively. Among the comparisons of various methods, the values in bold denote the winners' performance. The proposed DRP strategy achieves the best performance in most of the cases.

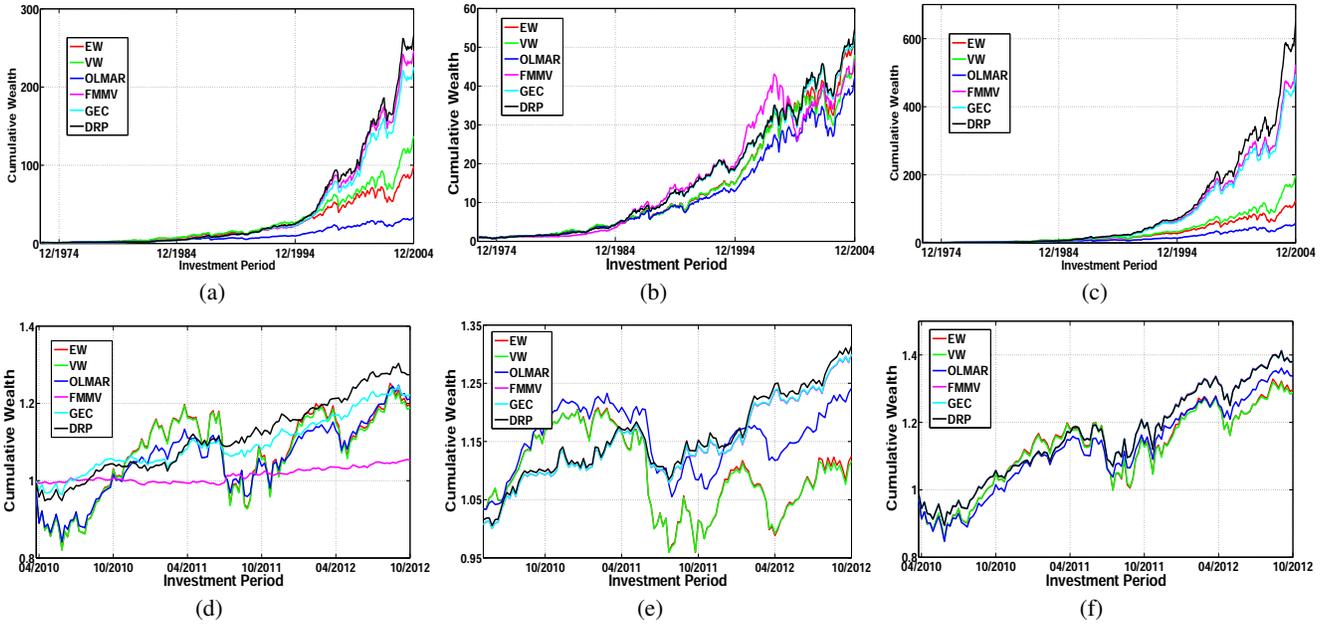


Figure 1: The curves of *cumulative wealth* across the investment periods for different portfolios on a) FF25, b) FF48, c) FF100, d) ETF, e) INDEX, and f) EQUITY datasets.

In Table 3, where we report the annualized Sharpe ratios in percentage, DRP generates consistent higher Sharpe ratios than the rest, which indicates that it achieves higher risk-adjusted returns. The GEC portfolio archives approximately as high Sharpe ratios, as its strategy embodies a similar idea of risk minimization and sparse portfolio selection. Even for the portfolio returns measured by cumulative wealth, DRP is among of the top performers in most of the cases, as shown in Table 4. In addition, Table 5 illustrates that DRP has dramatically lower turnover rates, reduced by 10.7% ~ 66.2% compared with GEC. The reduction of the turnover rates is attributed to the unique consistency regularization in the DRP formulation. Accordingly, DRP not only improves risk-adjusted returns but also ensures much less trading costs. VW has a zero turnover rate by definition and EW incurs the equal portfolio allocation thus having negligible trading volumes. Hence we exclude these two heuristic portfolios and only report the results from the other portfolios in Table 5. Finally, Table 6 unsurprisingly illustrates that the DRP strategy has the lowest volatility risk in most of the cases. Although FMMV is designed to create a minimal variance portfolio, without having constraints or regularities on the portfolio structure its out-of-sample variance cannot be ensured minimal.

Furthermore, in order to compare the trend and dynamics of the return growth, Figure 1 illustrates the curves of the cumulative wealth over the investment periods for the different datasets. Apparently, DRP outperforms the others with the visible margins on four of the tested datasets in most of the time periods, as shown in Figures 1(a), 1(c), 1(d) and 1(f). Those figures show DRP grows more steadily together with a reduced volatility across most of the investment periods.

Table 6: Portfolio volatilities (%) with p -values

Method	FF25	FF48	FF100	ETF	INDEX	EQUITY
EW	17.66 (1.00)	16.97 (1.00)	18.33 (1.00)	21.98 (1.00)	13.32 (1.00)	18.34 (1.00)
VW	17.32 (0.23)	16.76 (0.16)	18.10 (0.05)	21.79 (0.02)	13.04 (0.00)	18.00 (0.00)
OLMAR	16.97 (0.32)	15.49 (0.01)	17.66 (0.14)	16.12 (0.00)	7.21 (0.00)	14.42 (0.00)
FMMV	13.42 (0.00)	13.86 (0.00)	13.42 (0.00)	2.51 (0.00)	6.80 (0.00)	13.62 (0.00)
GEC	13.42 (0.00)	13.05 (0.00)	13.1 (0.00)	6.29 (0.00)	6.79 (0.00)	13.61 (0.00)
DRP	12.96 (0.00)	12.82 (0.00)	12.77 (0.00)	6.92 (0.00)	6.85 (0.00)	13.61 (0.00)

Conclusion

We have presented a novel doubly regularized portfolio strategy through leveraging structure regularization, i.e. *sparsity* and *consistency*, to the conventional minimum-variance portfolio framework. In particular, the *sparsity* regularization enforces to choose a small set of assets with a minimum risk and the *consistency* regularization produces similar asset positions in consecutive investment periods. Extensive experiments and comparisons on both long-term and short-term financial benchmarks and real-world market datasets are conducted. The experimental results show that, by virtue of the synergistic combination of the classic risk minimization framework and the novel structure regularization, the proposed portfolio constantly achieves higher risk-adjusted returns with much lower implied costs measured by *turnover* rates. Our future work includes appropriately incorporating market friction models into the studied framework, such as capital gain taxes and market price impacts.

References

- Agarwal, A.; Hazan, E.; Kale, S.; and Schapire, R. E. 2006. Algorithms for portfolio management based on the Newton method. In *Proceedings of the 23th international conference on machine learning*, 9–16.
- Bach, F.; Jenatton, R.; Mairal, J.; and Obozinski, G. 2011. Convex optimization with sparsity-inducing norms. *Optimization for Machine Learning* 19–53.
- Beck, A., and Teboulle, M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal Imaging Sciences* 2(1):183–202.
- Becker, S.; Candès, E. J.; and Grant, M. 2011. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation* 3(3):165–218.
- Blum, A., and Kalai, A. 1999. Universal portfolios with and without transaction costs. *Machine Learning* 35(3):193–205.
- Borodin, A.; El-Yaniv, R.; and Gogan, V. 2004. Can we learn to beat the best stock? *Journal of Artificial Intelligence* 21:579–594.
- Boyd, S., and Vandenberghe, L. 2004. *Convex optimization*. Cambridge University Press.
- Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* 3(1):1–122.
- Brandt, M. W. 2010. Portfolio choice problems. In Ait-Sahalia, Y., and Hansen, L. P., eds., *Handbooks of Financial Econometrics*. Elsevier. 269–336.
- Brodie, J.; Daubechies, I.; Mol, C. D.; Giannoned, D.; and Loris, I. 2009. Sparse and stable Markowitz portfolios. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 106, 12267–12272.
- Candès, E. J., and Tao, T. 2006. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory* 52(1):5406–5425.
- Cvitanic, J. 2001. Theory of portfolio optimization in markets with frictions. In Jouini, E.; Cvitanic, J.; and Musiela, M., eds., *Handbooks in Mathematical Finance*. Cambridge University Press.
- DeMiguel, V.; Garlappi, L.; Nogales, F. J.; and Uppal, R. 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55:798–812.
- DeMiguel, V.; Garlappi, L.; and Uppal, R. 2009. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *The Review of Financial Study* 22:1915–1953.
- Fama, E. F., and French, K. R. 1992. The cross-section of expected stock returns. *Journal of Finance* 47(2):427–465.
- Fan, J.; Fan, Y.; and Lv, J. 2008. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147:186–197.
- Fan, J.; Zhang, J.; and Yu, K. 2012. Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *Journal of American Statistical Association* 412–428.
- Györfi, L.; Lugosi, G.; and Udina, F. 2006. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance* 16:337–357.
- Hale, E. T.; Yin, W.; and Zhang, Y. 2008. Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization* 19(3):1107–1130.
- Huang, J.; Zhang, T.; and Metaxas, D. 2009. Learning with structured sparsity. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 417–424.
- Jagannathan, R., and Ma, T. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58:1651–1684.
- Ledoit, O., and Wolf, M. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10:603–621.
- Ledoit, O., and Wolf, M. 2011. Robust performance hypothesis testing with the variance. *Wilmott Magazine* 86–89.
- Li, B., and Hoi, S. C. 2012. On-line portfolio selection with moving average reversion. In *Proceedings of the 29th international conference on machine learning*.
- Li, B.; Hoi, P.; Hoi, S. C.; and Gopalkrishnan, V. 2012. PAMR: passive aggressive mean reversion strategy for portfolio selection. *Machine Learning* 87:221–258.
- Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7:77–91.
- Merton, R. C. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8:323–361.
- Michaud, R. O. 1989. The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal* 31–42.
- Rapach, D. E., and Zhou, G. 2012. Forecasting stock returns. In Elliott, G., and Timmermann, A., eds., *Handbook of Forecasting II*. North-Holland.
- Samuelson, P. A. 1969. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics* 51:239–246.
- Sharpe, W. F. 1966. Mutual fund performance. *Journal of Business* 39:119–138.
- Thorp, E. O. 1971. Portfolio choice and the Kelly criterion. *Business and Economics Statistics Section of Proceedings of the American Statistical Association* 215–224.
- Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1):267–288.
- Wang, L.; Zhu, J.; and Zou, H. 2006. The doubly regularized support vector machine. *Statistica Sinica* 16(2):589–615.
- Zou, H., and Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(2):301–320.