A Toolkit for Exploiting Contemporaneous Stock Correlations *

Kazuhiro Hiraki † Chuanping Sun ‡

Abstract

Contemporaneous correlations are important for portfolio optimization problems. We propose a newly developed machine learning tool, the OWL shrinkage method, which explicitly exploits stocks' contemporaneous correlations by assigning similar positions to correlated stocks (the grouping property). We find strong evidence that OWL-based portfolio strategies outperform other benchmark strategies in the literature when stocks exhibit strong correlations. In particular, the OWL shrinkage method bridges the gap between the naive (but well performing) 1/N portfolio strategy (DeMiguel et al., 2009b) and the portfolio optimization framework: our OWL-based portfolio strategies yield very similar portfolio weights to (yet not the same as) the 1/N portfolio strategy, but outperform the 1/N portfolio strategy in terms of both the Sharpe ratio and turnovers. We also show that the superior performance in Sharpe ratio against the 1/N portfolio is significant.

JEL classification: G11, C61
Keywords: Portfolio Optimization, LASSO, Machine Learning, 1/N Portfolio Strategy, Stock Correlation, Norm Constraints, Model Confidence Set

*We are grateful to Liudas Giraitis, Emmanuel Guerre, Ian Marsh and two anonymous referees for their advice, comments and suggestions. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Bank of Japan.

†Institute for Monetary and Economic Studies, Bank of Japan, 2-1-1, Nihonbashi-Hongokucho, Chuo-ku, Tokyo, 103-0021, Japan, kazuhiro.hiraki@boj.or.jp.

‡Corresponding author. Faculty of Finance, Bayes Business School (formerly Cass), City, University of London, 106 Bunhill Row, London EC1Y 8TZ, U.K. Email: chuanping.sun@city.ac.uk.

1 Introduction

Despite the theoretical elegance of the mean variance efficient (MVE) portfolio theory, put forward by Markowitz (1952), its empirical performance is often criticized due to the difficulties in precisely estimating two important ingredients in its formula: the expected asset returns and the variance-covariance matrix. Since then, continuous attempts and contributions have been made to improve the empirical performance of the MVE portfolio. Michaud (1989), Jagannathan and Ma (2003) and DeMiguel et al. (2009b) argue that the estimation error in the expected returns is so large that nothing much is lost in ignoring the mean altogether when no further information about the population mean. Ledoit and Wolf (2003, 2017), DeMiguel et al. (2009a) and Jagannathan and Ma (2003) find that constraining portfolio weight in the optimization framework and/or shrinking the variance-covariance matrix improve portfolio performance. Although many efforts have been made to improve the performance of optimized portfolio strategies, DeMiguel et al. (2009b) demonstrate that the simple 1/N (equally weighted) portfolio strategy outperforms 14 other optimized portfolio strategies, making this non-optimized naive diversification strategy a competitive benchmark in comparing portfolio selection strategies.

This paper builds upon and extends the norm constrained optimization framework by DeMiguel et al. (2009a). However, our method differs in an important way: we introduce a newly developed machine learning tool, the Ordered-Weighted-LASSO (OWL) shrinkage method, which admits two important properties. First, it achieves *sparsity*, that is, it shrinks unimportant stocks' positions to zero. This means that the OWL shrinkage method encompasses the LASSO shrinkage method considered in DeMiguel et al. (2009a). Second, the OWL shrinkage method admits the *grouping property* which distinguishes itself from the LASSO shrinkage method: it identifies correlated stocks and assigns similar portfolio weights to them. Note that although the MVE optimization framework takes into account the stock correlations through the variance-covariance matrix, large estimation errors in the sample analog of the variance-covariance matrix erode all gains from optimization if no further regularization considered. Therefore, the OWL shrinkage method exploits stock correlations from a different channel compared to the standard MVE optimization framework. Empirically, we find that exploiting contemporaneous stock correlations substantially improves the performance of the MVE portfolio. This point becomes clear when we consider transaction costs in measuring the performance of portfolio strategies. The grouping property achieves low turnovers while the LASSO shrinkage method may result in high turnovers due to unstable stock selections from highly correlated stocks.

To better understand the grouping property, we theoretically show that three elements determine the degree of grouping (Theorems 3.1 and 3.2). 1) The correlation between stocks. Highly correlated stocks encourage stock grouping. 2) There is a hyper parameter in the design of the OWL shrinkage component that determines the grouping intensity. While we determine the value of this hyper parameter by cross-validation, investors can conveniently determine an ideal level of the grouping intensity depending on their prior information. It is worth noting that the OWL shrinkage component collapses to the LASSO method when this hyper parameter is set to zero.[1] 3) For the MVE portfolio construction, differences in the mean returns also play a role. Holding everything else the same, stocks with similar mean returns are likely to be assigned with similar positions for the MVE portfolio.

This paper also relates to DeMiguel et al. (2014), which implement a VAR(1) model to exploit time-series correlations between stocks and find that lagged stock returns can be utilized to improve portfolio performance. In contrast, our paper focuses on utilizing *contemporaneous* stock correlations to improve portfolio performance. This paper also contribute to the algorithmic method for solving portfolio optimization problems with multiple constraints. We devise an efficient algorithm to solve the OWL regularized optimization problem with multiple constraints on portfolio weights.[2] For example, investors can incorporate upper and/or lower bound constraints for each individual stock. This bound constraint can be obtained based on their prior beliefs or other existing portfolio strategies. Therefore, our optimization framework can be used to improve any existing portfolio strategy, where the existing portfolio strategy provides information on the additional bound constraints.

Empirically, we test candidate portfolio strategies in five asset classes. First, we consider

[1]See Lemma 1 for details.

[2]A tailored ADMM algorithm will be introduced which achieves fast convergence. More details see Section 3.3.

the Fama-French 25 (FF25) portfolios because of their popularity as test assets in the finance literature. It is worth noting that, because they are sorted portfolios, they are less prone to large variations in returns compared to individual stocks. Second, we consider the S&P 500 (SP500) stocks with *daily* return frequency and we rebalance our hedge portfolio either weekly or monthly. Third, we also consider the S&P 100 (SP100) stocks with *monthly* return frequency and we rebalance our hedge portfolio monthly.[3] S&P 500 and S&P 100 stocks are usually the largest stocks in the market. To test candidate portfolio strategies on small and medium stocks, we follow Jagannathan and Ma (2003) and DeMiguel et al. (2009b) and adopt the randomly selected stocks approach: in April each year, we randomly select 500 stocks with daily return series from the CRSP dataset (and 100 stocks for monthly return series) which have no missing data in the past three years (10 years for monthly return series) and in the next one year. The randomly selected 500 stocks with *daily* returns and 100 stocks with *monthly* returns consist of our fourth and fifth test assets classes.

We adopt an out-of-sample procedure to compare candidate portfolio strategies. At each point of time, we use a rolling window to estimate each stock's weight (portfolio's weight for the FF25 case) to invest for the next period, then we roll the training sample forward until the next rebalancing point. In the end, we obtain a sequence of out-of-sample returns and portfolio weights, from which we can compute the out-of-sample Sharpe ratio and turnovers.[4] Notably, transaction cost, which is a monotonically increasing function of turnovers, is an important consideration for investors. Therefore, we also consider a transaction cost adjusted Sharpe ratio (TCadjSR), which will be one of our main comparison criteria. On the other hand, Sharpe ratio, formulated as the ratio between portfolio return and portfolio risk, is often dominated by the portfolio risk component. Therefore, we also introduce the model confidence set (MCS) method of Hansen et al. (2011) to answer the following question: which candidate strategies offer *statistically* the best out-of-sample returns, while portfolio risk determines the confidence band?

Our empirical findings complement some stands of existing literature and shed light on new perspectives of portfolio selection strategies. First and foremost, DeMiguel et al.

[3] We require an invertible sample covariance matrix as an input in our optimization problem, thus we need the time-series dimension larger than the cross-sectional dimension to obtain a non-singular covariance matrix.

[4] Turnover is the change in portfolio weights right before and after rebalancing.

(2009b) show that the naive 1/N strategy outperforms 14 optimization-based strategies. Our method bridges the gap between the naive diversification strategy and a well-defined optimization framework. We show that the OWL-based portfolio strategies yield very similar positions to (but not the same as) the 1/N strategy. However, they outperform the 1/N strategy in terms of both the Sharpe ratio and turnovers. It is a remarkable discovery, as in the existing literature, it is difficult to find a portfolio strategy that can outperform the 1/N strategy by *both* the Sharpe ratio *and* the turnover criteria, see DeMiguel et al. (2009b) for example. Second, OWL-based strategies perform better in large stocks than small stocks and in monthly returns than daily returns. Generally, the OWL-based portfolio strategies score higher than other portfolio strategies when using Fama-French 25 portfolios and the SP100 stocks with monthly returns. A stylized fact is that large stocks with monthly returns exhibit higher cross-sectional correlation than smaller stocks and daily returns. However, this superior performance against other portfolio strategies is less obvious while using randomly selected 500 stocks from the CRSP dataset. Nonetheless, the superior performance of the OWL-based strategies against the 1/N strategy is prevalent across all asset classes. This finding offers a guidance of effectiveness and suitability of OWL-based strategies. Third, the MVE portfolio performs poorly due to excessive estimation errors in expected returns and the variance-covariance matrix. However, the OWL embedded MVE (MVE-OWL) strategies together with weight constraints produce sizable Sharpe ratio and low turnovers. The MCS test confirms that the MVE-OWL strategies yield *significantly* larger out-of-sample returns than all other portfolio strategies using FF25 as test assets. This finding challenges some common stances in the existing literature. Because of the excessive estimation error in the expected stock returns, it is better off to ignore the expected return component altogether in the MVE framework, see Jagannathan and Ma (2003). Therefore, the majority of the empirical portfolio selection literature is focusing on optimizing the minimum-variance portfolio. Our finding suggests that it is still beneficial to consider the MVE portfolio while using the OWL shrinkage method and additional weight constraints.

The rest of this paper is organized as follows. Section 2 reviews related literature. Section 3.1 outlines some popular portfolio optimization strategies in the related literature.

They are also used as benchmarks in our empirical analysis. Sections 3.2 - 3.3 introduce the OWL shrinkage method and discuss its statistical properties before devising an ADMM algorithm to solve the OWL optimization problem with multiple constraints. Section 4 applies OWL-based portfolio strategies using five different asset classes and compares them with other benchmarks.

2 Literature review

This paper naturally builds on a strand of literature devoted to exploring the portfolio optimization theory. Since the groundbreaking work of Markowitz (1952), modern portfolio theory has evolved rapidly. However, Markowitz's portfolio theory has long been criticized for working poorly empirically, because one needs to obtain the *ex-ante* returns and variance-covariances matrix of stock returns, which are difficult to be estimated with precision. Michaud (1989) looks into the "Markowitz optimization enigma" and finds that the mean variance optimization is in fact "error maximization". DeMiguel et al. (2009b) study the simple equal weighted strategy and find it outperforms 14 other optimization based strategies. They argue that estimation error in the expected asset return and the variance-covariance matrix erodes any gains from optimization. Kan and Zhou (2007) show that using the sample analogs of the expected returns and the variance-covariance matrix can lead to very poor out-of-sample performance due to parameter uncertainty. They find that holding the tangent portfolio and the risk free asset is no longer optimal, though holding some other risky portfolios will help reduce the portfolio risk caused by parameter uncertainty. Ledoit and Wolf (2003) propose a shrinkage-based estimation method for the variance-covariance matrix. They suggest that shrinking the sample covariance matrix linearly towards a target matrix (for example the identity matrix) will improve the out-of-sample performance of the minimum variance portfolio. Ledoit and Wolf (2017) propose a non-linear version of shrinkage estimator for the covariance matrix which shows better performance than the linear version. Jagannathan and Ma (2003) suggest no-short-sale constraint on all stocks and find significant gains in out-of-sample Sharpe ratio for the minimum variance portfolio. They argue that such constraint helps reduce the upward biased estimation errors in the variance-covariance matrix. DeMiguel et al. (2014) implement a VAR(1) model to exploit time-series correlations between stocks and demonstrate substantial gains in portfolio performance. DeMiguel et al. (2020) consider a portfolio optimization problem by selecting a large number of firm characteristics while embedding the transaction cost in the object function.

This paper is also related to a new and fast growing field which uses machine learning techniques for portfolio optimization problems. DeMiguel et al. (2009a) propose norm constraints on portfolio weights for the minimum variance portfolio. In particular, they consider separately the LASSO and Ridge penalties on portfolio weights and find significant improvement on the out-of-sample Sharpe ratio of the minimum variance portfolio. Our paper paper is closely related to DeMiguel et al. (2009a) in the sense that the OWL shrinkage method is an extension of the LASSO shrinkage method (i.e. both 1-norm penalized optimization problems) employed in DeMiguel et al. (2009a). However, our OWL method admits the grouping property which is absent from the LASSO shrinkage method. Empirical analysis shows that the grouping property dominates our OWL-based portfolio strategies, whereas the shrinkage effect is minimal in comparison[5]. Therefore, our OWL-optimized portfolio weights would differ substantially from the LASSO shrinkage method as in DeMiguel et al. (2009a). Ao et al. (2018) combine the unconstrained regression with LASSO penalty and achieve superior portfolio performance. Inspired by the adaptive LASSO in Zou (2006), Fastrich et al. (2015) incorporate the financial information into the adaptive weights to determine the portfolio composition. Figueiredo and Nowak (2016) study the ordered and weighted LASSO estimator and show that it has appealing property of clustering correlated variables by assigning them with similar coefficients.

This paper is closely related to DeMiguel et al. (2014) and DeMiguel et al. (2009a). However, our portfolio optimization method differs in several ways. First, we endeavor to exploit the *contemporaneous* correlation between stocks instead of time-series correlations considered in DeMiguel et al. (2014). Therefore, our strategies are more relevant to "stock-picking" investors. Second, our OWL shrinkage method not only encompasses the LASSO norm constraint in DeMiguel et al. (2009a), it also exploits contemporaneous stock correlations. Third, we devise an optimization framework that enables investors to incorporate

[5]Note that this result is entirely data-driven. The OWL shrinkage method is not always dominated by the grouping property and it depends on the data structure.

their prior beliefs into the optimization problem (i.e., investors can set upper/lower bounds on portfolio weights for each individual asset).

3 Methodology

We first consider a simple case of the OWL shrinkage method in the portfolio optimization problem, where no constraints are imposed on the portfolio weights. Then we move on to impose additional constraints on portfolio weights for the optimization framework.

3.1 Setup

Consider N assets in the investment universe. Denote by R_t the returns of N assets in the excess of risk-free rate at time t. Denote by μ $(N \times 1)$ and Σ $(N \times N)$ the population mean and population variance-covariance matrix of N asset returns, while $\hat{\mu}$ and $\hat{\Sigma}$ are their sample estimates. An investor, according to Markowitz (1952)'s classical portfolio theory, aims to maximize the risk-adjusted portfolio returns, or equivalently:

$$\begin{aligned} &\min_{w} \quad \left(\frac{\gamma}{2} w'\Sigma w - \mu' w\right) \\ &s.t. \quad w'e = 1 \end{aligned} \tag{1}$$

where γ is a scalar that represents the investor's absolute risk aversion, w is the $N \times 1$ weighting vector of N assets, also referred to as positions, and e is a column vector of ones. The closed-form solution of the above optimization problem is $w = \frac{1}{\gamma}\Sigma^{-1}\mu$. However, Σ and μ are unobservable. Typically, we use the sample analogs $\hat{\mu}$ and $\hat{\Sigma}$ in the above equation, which gives

$$w_{MVE} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{2}$$

Michaud (1989), DeMiguel et al. (2009b) and Jagannathan and Ma (2003) have pointed out that the sample analogs of μ and Σ are subject to large estimation errors. In particular, the estimation of the expected asset returns (μ) proves to be extra challenging. In fact, the estimation error (of μ) is so large that it offsets all gains from optimization (Jagannathan and Ma, 2003). So in practice, focusing on the minimum variance (minVar) portfolio proves to have better out-of-sample results than the mean-variance efficient (MVE) portfolio. In this regard, the optimal weights for the minVar portfolio are obtained by optimizing (1) while setting $\mu = 0$, that is

$$w_{minVar} = \frac{\hat{\Sigma}^{-1}e}{e'\hat{\Sigma}^{-1}e}, \tag{3}$$

where we use the sample analog for Σ. For the estimation error in the sample covariance matrix, shrinkage estimator proves to be a useful remedy. For instance, Ledoit and Wolf (2003) propose to shrink the sample covariance matrix towards a target matrix. They suggest the following estimator

$$\hat{\Sigma}_{LW03} = \delta\hat{\Sigma} + (1-\delta)\hat{\Sigma}_{target}, \tag{4}$$

where $\delta \in (0,1)$ is a shrinkage intensity parameter and $\hat{\Sigma}_{target}$ is a target estimator, which can be, for example, the identity matrix.

DeMiguel et al. (2009a) show that imposing norm constraints on portfolio weights to shrink them towards zeros substantially improves the out-of-sample Sharpe ratio of the hedged portfolios.[6] They suggest the following norm shrinkage methods

$$\min_{w} \quad \left(\frac{\gamma}{2}w'\Sigma w - \mu' w + \lambda ||w||_1\right), \tag{5}$$

or

$$\min_{w} \quad \left(\frac{\gamma}{2}w'\Sigma w - \mu' w + \lambda ||w||_2^2\right), \tag{6}$$

where

$$||w||_1 = \sum_{i=1}^{N} |w_i|, \qquad ||w||_2^2 = \sum_{i=1}^{N} w_i^2,$$

and λ is a shrinkage intensity parameter. Shrinkage method in (5) is broadly known as LASSO shrinkage, which produces sparse estimator for w, while (6) is referred to as Ridge shrinkage, which shrinks all elements in w towards zero.

[6] We set $\mu = 0$ for the minimum variance portfolio. Otherwise, we are optimizing the mean-variance efficient portfolio.

Jagannathan and Ma (2003) find that imposing no-short-sale constraints on portfolio weights helps to improve out-of-sample Sharpe ratio. They propose the following constraint in addition to (1):

$$w_i \geq 0, \quad \text{for all} \ \ i \in \{1, ..., N\}. \tag{7}$$

They argue that imposing this constraint leads to substantial reduction of extreme negative positions of stocks which are caused by upward biased estimation of variances.

In addition, DeMiguel et al. (2014) reveal that stock correlations matter for portfolio construction. They propose a vector-autoregressive (VAR) model to capture stocks' serial correlations and find that VAR-based portfolios outperform traditional unconditional portfolios. To fix ideas, let us assume that the vector of asset return R_t follows a VAR(1) process,

$$R_{t+1} = a + BR_t + \epsilon_{t+1}, \tag{8}$$

where a is a $N \times 1$ vector of intercepts, B is a $N \times N$ matrix of parameters, and ϵ_t is the $i.i.d.$ error term. Equation (8) is a reduced model, which suggests that tomorrow's expected stock returns depend linearly on today's return. The linear dependence is characterized by the coefficient matrix B, which describes the lagged cross-sectional and serial dependence. On the other hand, contemporaneous correlations between stocks are left unexplained.

This paper builds on and extends DeMiguel et al. (2009a) and DeMiguel et al. (2014). We introduce a newly developed machine learning tool, the ordered and weighted LASSO (OWL), which (1) encompasses the LASSO shrinkage method in DeMiguel et al. (2009a); (2) exploits *contemporaneous* correlations between stocks, drawing a distinctive line between our portfolio optimization approach and that in DeMiguel et al. (2014); (3) enables adopting bespoke constraints on portfolio weights if investors have prior beliefs.[7] We devise efficient algorithms to solve the OWL optimization problem with/without additional constraints on portfolio weights.

[7] Prior beliefs could come from investors' exclusive information, or an existing trading strategy. In that respect, it can be used as an improvement/refinement of any existing strategies.

3.2 The OWL shrinkage method

We follow the idea of DeMiguel et al. (2009a) to add a penalty term in the object function $f(w)$, which measures the loss given portfolio weight w. The optimization problem can be written as

$$\hat{w} = \arg\min_{w} \; f(w) + \Omega_\omega(w), \qquad \Omega_\omega(w) = \omega'|w|_\downarrow, \tag{9}$$

where ω is a pre-specified weighting vector which will be specified in (11) below. w is a vector of stock weights (positions) and $|w|_\downarrow$ is the vector that stores the absolute value of stock weights, decreasingly ordered by its magnitude. Both ω and $|w|_\downarrow$ take values in a monotone non-negative cone κ, which is defined as $\kappa := \{x \in R^N : x_1 \geq x_2 \geq ... \geq x_N \geq 0\}$. $f(w)$ can be any continuously differentiable function of w. However, in this paper we focus on the mean-variance efficient portfolio (or the minimum variance portfolio if we set $\mu = 0$), which corresponds to

$$f(w) = \frac{\gamma}{2} w'\Sigma w - \mu' w, \tag{10}$$

where γ is agent's risk aversion level. We further specify ω to have a linear weighting structure

$$\omega_i = \lambda_1 + (N - i)\lambda_2, \tag{11}$$

where λ_1 and λ_2 are two hyper parameters which pin down ω, and the values of λ_1 and λ_2 are determined through cross validation.[8] In order to solve the optimization problem in (9) - (11), we use the proximal descent algorithm, more details about this algorithm can be found in Sun (2019).

Next, we discuss some econometric properties of the OWL shrinkage method.

Lemma 1. *Suppose that the pre-specified weighting vector ω of the OWL shrinkage method is defined in (11). If λ_2 is set to be zero, then the OWL shrinkage method is equivalent to*

[8] In particular, we set a grid value of λ_1 and λ_2. Then, at each point on the grid, we split the sample into 5 folds, using 4 folds to evaluate the model and obtain the estimated parameters. Then we use the other 1 fold as out-of-sample to evaluate the Mean Square Forecast Errors (MSE). We rotate each fold as the out-of-sample fold, and compute the average MSE. We repeat these procedures on each grid, and compare the average MSE for each point on the grid. The one receiving the smallest average MSE will determine the hyper parameter values.

the LASSO shrinkage as in (5), or equivalently

$$\lambda\|w\|_1 = \Omega_\omega(w).$$

Proof: see Appendix B.1.2.

Lemma 1 shows that the OWL shrinkage method encompasses the LASSO shrinkage method used by DeMiguel et al. (2009a). Furthermore, once we adopt a linear weighting scheme for ω as in (11), the OWL shrinkage method is linked to the OSCAR regularization introduced in Bondell and Reich (2008), which has appealing properties of clustering correlated features. The OSCAR regularization unit is defined as

$$\Omega_{OSCAR}(w) = \lambda_1||w||_1 + \lambda_2\sum_{i<j}\max\{|w_i|, |w_j|\}, \tag{12}$$

which is a combination of the LASSO regularization (ℓ_1 norm) unit and a pair-wise ℓ_∞ norm unit.

Lemma 2. *Suppose that the pre-specified weighting vector ω of the OWL shrinkage method is defined in (11). Then the OWL shrinkage method is equivalent to the OSCAR shrinkage method as in (12), or equivalently*

$$\Omega_{OSCAR}(w) = \Omega_\omega(w).$$

Proof: see Appendix B.1.3.

Lemma 2 shows that by adopting a linear decreasing weighting scheme for ω as in (11), the OWL shrinkage method is equivalent to the OSCAR regularization, which has property of clustering correlated variables. However, the OWL shrinkage is a more general method than the OSCAR regularization. For instance, by adopting a non-linear (for instance, the inverse of the normal cumulative distribution function) weighting scheme for ω, the OWL shrinkage method is equivalent to the SLOPE estimator proposed by Bogdan et al. (2015), which is widely used in multiple testing. In the scope of this paper, we restrict the weighting vector ω as defined in (11) because of the clustering property offered by the

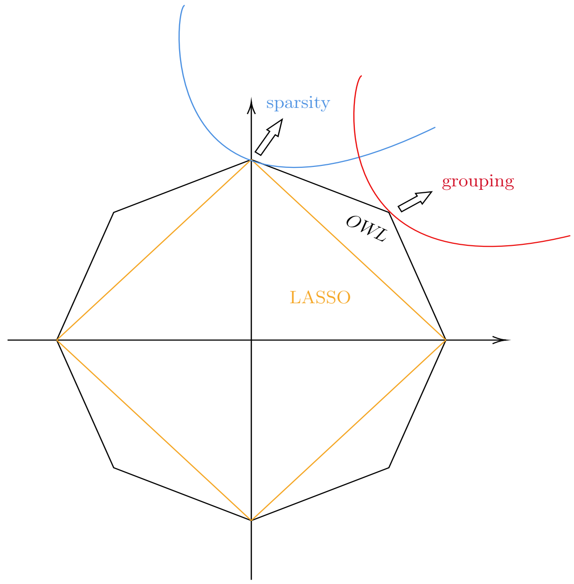


Figure 1. Geometric interpretation of the atomic norm of LASSO and OWL regularization

OSCAR regularization and our objective of exploiting the contemporaneous correlations between stocks. To gain some impression of how the OWL shrinkage method achieves sparse selection and correlation identification simultaneously, we first look at the geometric interpretation of the atomic norm of $\Omega_\omega(w)$ in Figure 1.

Figure 1 depicts the atomic norm of OWL and LASSO regularization in a two-dimensional space. We can see that the atomic norm of LASSO has all vertices on axes, which encourages sparse selection of variables. On the other hand, the atomic norm of the OWL regularization is octagonally shaped, having vertices on both axes and the ± 45 degree lines. The former (vertices on axes) encourages sparse selection and the latter (vertices on the ± 45 degree lines) encourages variable grouping.[9]

The geometric interpretation offers a ballpark explanation of how the OWL shrinkage achieves both sparse selection and correlation identification (grouping) simultaneously. Next, we formally investigate some econometric properties for the OWL shrinkage method. There is a rich literature in finance focusing on the sparse selection property offered by LASSO type of estimators, see DeMiguel et al. (2020, 2009a), Fastrich et al. (2015) and

[9]Sparse selection means the vertices on axes will assign one variable zero coefficient and another non-zero (in this 2-dimensional space), thus performing sparse selection. The variable assigned with zero coefficient is shrunk off. Variable grouping means variables exhibiting high correlations will be assigned with the same or similar coefficients. The vertices on the ± 45 degree lines will dictate the tangent point with the contour from the un-regularized solutions, which give the same or similar coefficients to both variables.

Chinco et al. (2019) for example. The OWL shrinkage method encompasses and shares similar sparse-selection-properties of the LASSO estimator, and see Sun (2019) for a formal investigation of the asymptotic property of the OWL estimator. For this reason we focus on investigating the grouping property in this paper. Theorems 3.1 and 3.2 below state the factors and condition that affect the grouping property.

Theorem 3.1. *Let $\Sigma_{i.}$ and $\Sigma_{j.}$ denote the i^{th} and j^{th} columns of the variance-covariance matrix and λ_2 be the parameter defined as in (11). Suppose the loss function is defined as in (10) while setting $\mu = 0$ (i.e. minimum variance portfolio). If*

$$\|\Sigma_{i.} - \Sigma_{j.}\|_2 < \lambda_2,$$

then $\hat{w}_i = \hat{w}_j$, where $\hat{w}_i$ and $\hat{w}_j$ are obtained by optimizing (9).
Proof: see Appendix B.1.

Theorem 3.1 shows that in the minimum variance portfolio optimization problem, if two assets are highly correlated, i.e. $\|\Sigma_{i.} - \Sigma_{j.}\|_2$ is small, then they will receive the same positions $\hat{w}_i = \hat{w}_j$. We regard this as the grouping property.[10] The tuning parameter λ_2 plays an active role in influencing the grouping property: a larger λ_2 makes the condition in Theorem 3.1 more likely to be satisfied and, therefore, contributes positively to portfolio weights grouping.

Theorem 3.2. *Let $\Sigma_{i.}$ and $\Sigma_{j.}$ be defined as in Theorem 3.1. Denote by μ_i, μ_j the expected returns of the i^{th} and j^{th} asset. Let γ represent investor's risk aversion level. Suppose the loss function is defined as in (10) (i.e. mean-variance efficient portfolio). If*

$$\gamma\|\Sigma_{i.} - \Sigma_{j.}\|_2 + |\mu_i - \mu_j| < \lambda_2,$$

then $\hat{w}_i = \hat{w}_j$, where $\hat{w}_i$ and $\hat{w}_j$ are obtained by optimizing (9).
Proof: see Appendix B.1.1

[10] Note that the grouping property also includes the case in which assets are assigned with similar weights. For the ease of mathematical proof, Theorem 3.1 and 3.2 give the condition of exact equality for two assets' weights and, they shed light on factors that influence the grouping property. From Figure 1 we can see that the grouping property entails assigning similar weights to assets, if conditions in Theorem 3.1 and 3.2 are nearly satisfied.

Theorem 3.2 extends Theorem 3.1 into a mean-variance efficient portfolio optimization problem where investors care about both risks and expected returns. We find that, compared to the minimum variance portfolio, both the difference in the expected returns $|\mu_i - \mu_j|$ and correlation with other assets $\|\Sigma_{i.} - \Sigma_{j.}\|_2$ influence the grouping property: if two assets have similar expected returns and are similarly correlated with other assets (i.e. $|\mu_i - \mu_j|$ and $\|\Sigma_{i.} - \Sigma_{j.}\|_2$ are small), then they are likely to be grouped together (i.e. $\hat{w}_i \approx \hat{w}_j$). The risk aversion parameter γ can be viewed as a scaling parameter adjusting weights between the risk component and the expected return component. Also, large λ_2 encourages grouping.

It is worth stressing that we derive the grouping property using the population values of the variance-covariance matrix Σ and the expected returns μ, which are unobservable and difficult to estimate with precision. However, the proof of Theorem 3.1 and 3.2 does not depend on the asymptotic properties of Σ or μ. In other words, we arrive at those results only using properties of the OWL regularization, and those results are also applicable to $\hat{\Sigma}$ and $\hat{\mu}$, which are sample analogs of Σ and μ. It is well known that large estimation errors in those sample analogs erode any gains in optimization. In the next subsection, we set out to constrain portfolio weights while using these sample analogs $\hat{\mu}$ and $\hat{\Sigma}$ to mitigate estimation errors.

3.3 The OWL optimization problem and the ADMM algorithm

Let us consider a more common problem, where investors have some prior information on stocks. For instance, an investor may hold positive opinions on some specific stocks while negative on others, thus she may want to impose some bounds constraints on the weight of those stocks. To generalize those constraints, we impose the following inequality

$$\mathbf{lb} \preceq w \preceq \mathbf{ub}, \tag{13}$$

where $\mathbf{lb}$ ($\mathbf{ub}$) is a lower (upper) bound for the vector of portfolio weights w. For any $x, y \in R^N$, $x \preceq y$ implies $x_i \leq y_i$, for all $i \in \{1, ..., N\}$. However, the optimization problem of (9) with constraint (13) is challenging to solve with gradient descent algorithm, which is commonly used in machine learning algorithms. Hence, we introduce a newly developed ADMM (Alternating Direction Method of Multiplier, Boyd et al. (2010)) algorithm to solve this constrained optimization problem. The outline of the algorithm is the following: equation (9) consists of two components, one is $f(w)$ which is differentiable with respect to w, another is $\Omega_\omega(w)$ which is not differentiable with respect to w. In order to make computation easier and tangible, we introduce a new variable v, and replace it with w in the undifferentiable component $\Omega_\omega(w)$, so that we can optimize each component separately. In addition, we impose an extra constraint that these two random variables are equal $w = v$. For this reason, this algorithm is named "alternating direction method of multiplier". Therefore, the constrained OWL optimization problem can be written as

$$\min_{w,v \in R^N} \left[\frac{\gamma}{2} w'\hat{\Sigma}w - \hat{\mu}'w + \Omega_\omega(v)\right], \tag{14}$$

$$s.t. \quad w = v, \tag{15}$$

$$w'e = 1, \tag{16}$$

$$w \succeq \mathbf{lb}, \tag{17}$$

$$w \preceq \mathbf{ub}, \tag{18}$$

where $\Omega_\omega(v) = \omega'|v|_\downarrow$ defined similarly as in (9), $\hat{\Sigma}$ and $\hat{\mu}$ are sample analogs of Σ and μ, and e is a column vector of ones. For the technical details of the ADMM algorithm, see Appendix A.

4 Empirical Analysis

In this section, we apply the OWL shrinkage method with or without additional constraints on portfolio weights and compare them with other portfolio strategies in the literature. We consider five different asset classes with daily and/or monthly returns. The variety in characteristics of these asset classes summarizes the pros and cons of each strategy which we will discuss in details later.

In Section 4.1, we first introduce the data (five asset classes) and all candidate strategies. Section 4.2 explains the empirical method we employ to compare strategies and Section

4.3 outlines the comparison criteria we use for ranking strategies. Section 4.4 argues the performance of the candidate strategies mainly by their Sharpe ratio criterion. To test the significance of difference in Sharpe ratios, we employ a bootstrap based Sharpe ratio test by Ledoit and Wolf (2008). In Section 4.5, we compare the performance of candidate strategies based on their transaction costs adjusted returns because the Sharpe ratio criterion tends to be dominated by the variance component. To test which strategies statistically produce higher returns, we use the model confidence set (MCS) test by Hansen et al. (2011).

4.1 Data

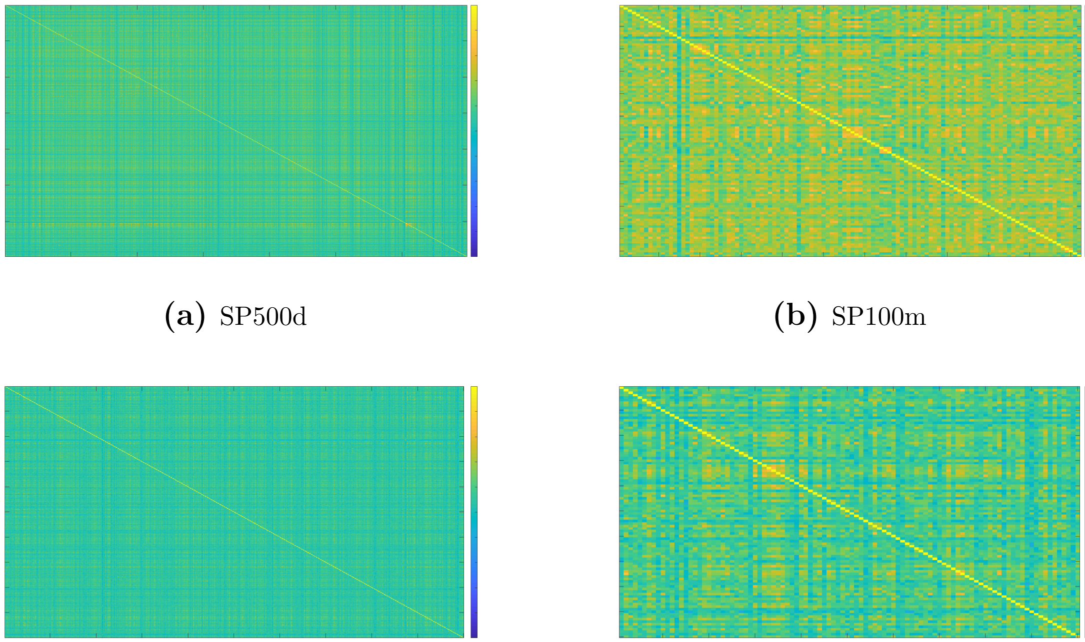
We first consider the Fama French 25 portfolios (FF25) from July 1927 to December 2017.[11] FF25 is obtained by sorting stocks into five by five tranches according to their size and book-to-market ratio. The return of each tranche is the average returns of a large number of stocks allocated to the tranche sharing similar characteristics (size and book-to-market ratio in this case). Since returns of these tranches are averaged returns of many stocks, they are less prone to large variations caused by idiosyncratic shocks.

We then consider the S&P 500 stocks with *daily* returns (SP500d), from 1st January 1978 to 31st December 2017 and the S&P 100 stocks with *monthly* returns (SP100m) from January 1978 to December 2017. Stock return data are obtained from the CRSP stock return files from Wharton Research Data Services.

While S&P 500 and S&P 100 stocks are typically the largest stocks in the market, to investigate stocks with medium or small sizes, we follow DeMiguel et al. (2009a) and Jagannathan and Ma (2003) to consider the randomly selected 500 stocks from the CRSP *daily* return file (CRSP500d). We also conduct a similar procedure to randomly select 100 stocks from the CRSP *monthly* return file (CRSP100m).

Figure 2 shows the correlation coefficient matrices of SP500d, SP100m, CRSP500d and CRSP100m stocks, respectively. We observe that SP500d returns exhibit higher correlations than the randomly selected CRSP500d returns. Similarly, we observe that the SP100m exhibit higher correlations than CRSP100m returns. These patterns may reflect the fact that large stocks are less prone to idiosyncratic noises and more affected by market-wide

[11] Data is downloaded from Kenneth French's website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html



(a) SP500d **(b)** SP100m **(c)** CRSP500d **(d)** CRSP100m

Figure 2. Correlation coefficient matrices of four asset classes

Note: yellow and deep blue indicate high correlation, while green indicating low correlation.

common factors than small stocks, so large stocks exhibit higher correlations than small stocks. Meanwhile, we also observe that, by comparing the left panels and right panels in Figure 2, monthly returns shows higher correlations than daily returns. This can be characterized as the Epps effect (Epps, 1979): the sample correlation tends to be biased towards zero when the sampling frequency progressively shrinks.

Table 1. Candidate strategies

Abbreviation	Strategies	Source
EW (1/N)	equal weighted	DeMiguel et al. (2009b)
minVar	minimum variance portfolio	N/A
minVar-JM	minVar with no-short-sale constraint	Jagannathan and Ma (2003)
minVar-LW	minVar with Ledoit-Wolf shrinkage	Ledoit and Wolf (2003)
minVa-OWL	OWL shrinkage on minVar	New Proposal
minVar-OWL-Pos	OWL shrinkage with no-short-sale constraint	New Proposal
minVar-OWL-bounds	OWL shrinkage with bounds constraints	New Proposal
minVar-LW-OWL	OWL with LW shrinkage on Cov matrix	New Proposal
minVar-hard-OWL	OWL with hard-thresholding for Cov matrix	New Proposal
MVE-OWL-Pos	OWL shrinkage with no-short-sale constraint on MVE	New Proposal
MVE-OWL-bounds	OWL shrinkage with bounds constraints on MVE	New Proposal

Next, Table 1 lists all considered candidate strategies. First, we consider the equal weighted (EW, also known as 1/N) strategy which has attracted great attention after DeMiguel et al. (2009b) showing that this non-optimized naive diversification strategy achieves superior out-of-sample performance against other optimized ones. We also consider the no-short-sale constraint on the minimum variance portfolio by Jagannathan and Ma (2003) and the linear shrinkage method by Ledoit and Wolf (2003) in our candidate strategies. In the newly proposed OWL shrinkage methods, we focus on the "minVar-OWL" method which implements the OWL shrinkage method on the minimum variance portfolio without additional constraints. The rest are some enhanced OWL strategies. For instance, "Pos" indicates that we further impose a no-short-sale constraint on stock weights. "Bounds" indicates that we impose upper and lower bounds for each stock. In this case, since we do not hold any additional information about each stock in our exercise, we blindly impose a bound constraint between -5% and 30% for all stocks. "Hard" indicates a hard-thresholding method for estimating the covariance matrix as in Bickel and Levina (2008) and Dendramis et al. (2019).

Next, we set out to conduct out-of-sample based empirical methods to implement each strategy and compare their performances using various criteria.

4.2 Empirical methods

For the FF25 asset class, since returns are sorted portfolio returns, we have balanced panel data, which is convenient for our analysis. We choose a rolling window size, say five years (60 months). At time t, we use the recent 60 months (from $t-59$ to t) data to estimate the model with each strategy and obtain the weighting vector for the next period. At the beginning of $t+1$ we invest in each 25 portfolios according to the weighting vector we obtained at time t. Then, at the end of $t+1$, returns will be realized, so we can compute the returns for the hedge portfolio. Next, we roll the window one month forward (from $t-58$ to $t+1$) to estimate the weight for next month's investment.

For SP500d, we first find all stocks that have been in SP500 index at least once between January 1978 and December 2017, total 1439 stocks. Then we implement a rolling window scheme, with rolling window size equal to 750 working days, approximately 3 years. In each rolling window, we remove stocks having missing data, which typically leaves 500 to 700 stocks in the investment universe in each rolling window. We then perform various portfolio selection strategies, and get weights for stocks which constitute next period's investment amount. We consider two rebalance frequencies, weekly or monthly. When the rebalance period is met, we compute the portfolio's return and turnover. Then we move forward to the next rolling window and repeat these steps until the end of out-of-sample period. We follow a similar procedure for the SP100m dataset except we rebalance only monthly and use a rolling window size of 10 years (120 months).

For CRSP500d, we follow DeMiguel et al. (2009a)'s procedure. In April each year, we randomly choose 500 stocks that have no missing data for the past 10 years as well as the following one year. Then in each rolling window, with window size equal to 750 working days, we estimate weights using various strategies. We also consider rebalancing portfolios weekly or monthly. At each rebalance point, we compute out-of-sample portfolio returns and turnovers. At the end of the out-of-sample period we can compute out-of-sample returns, standard deviation, Sharpe ratio and turnovers. We follow a similar procedure for CRSP100m, except the rolling window size is 10 years (120 months), and rebalance monthly only.

4.3 Out-of-sample comparison

To compare our OWL-based strategies with other existing ones in the literature, we consider the following criteria: (1) the out-of-sample Sharpe ratio (SR), (2) portfolio turnovers (transaction cost), (3) transaction cost adjusted Sharpe ratios (TCadjSR), (4) transaction cost adjusted out-of-sample returns (TCadjR). We follow the methodology of DeMiguel et al. (2009a) to construct the first two criteria. We add the third criterion because the first two criteria look at Sharpe ratio and turnover separately, which leads to (on many occasions) contradictory preferences: one strategy that delivers higher SR usually entails higher turnover (transaction cost), and vice verse. TCadjSR allows one to look into and compare strategies in a complete fashion. Nonetheless, Sharpe ratio comparison is usually dominated by its variance component and, by definition, the minVar portfolio typically delivers much lower portfolio risk than the MVE portfolio. Hence, in addition to the above mentioned criteria, we want to find that which portfolio strategies yield the best (transaction cost adjusted) out-of-sample returns. To answer that question, we employ the model confidence set (MCS) of Hansen et al. (2011). MCS compares TCadjR and puts all strategies that *significantly* produce the highest returns in a set while portfolio risk determines the confidence band. More details about the MCS method are included in Section 4.5. We argue that looking at the criteria three and four together gives a more completed profile of portfolio performance.

To fix ideas, let $r_t = (r_{1,t}, r_{2,t}, \cdots, r_{N,t})'$ denote the vector of N asset returns in excess of the risk-free rate $r_{f,t}$ at time t and w_t denote the vector storing portfolio weights of N assets at time t. For the monthly dataset, we choose a rolling window size $\tau = 120$ months, where $\tau \ll T$ and T is the total number of time-series observations. We set the rebalancing frequency as monthly ($q = 1$ month). For the daily data set, we choose a rolling window size of three calendar years (about 756 time-series observations) and rebalance either weekly ($q = 5$ days) or monthly ($q = 21$ days). At time t, we estimate portfolio weights w_t using data from $t-\tau+1$ to t for each strategy. w_t will be the investment amount at the beginning of time $t+1$ and we hold this position until the next rebalancing point $t+q$. At time $t+q$, before rebalancing we need to compute the *weight before rebalance*. The weight of each asset changes between the beginning and the end of time t due to the price fluctuation. We re-calculate the weight at the end of time t using new stock prices and call it the "weight before rebalance" (w_{t^+}). Then at the beginning of time $t+1$ we invest according to the new weight (w_{t+1}) obtained at the end of time t. The difference between them ($|w_{t+1} - w_{t^+}|$) is the *turnover*. More specifically, we compute the summation of absolute value of this difference at each point of time, then take the average across time. Then, we consider the following comparison criteria. For strategy i, the standard deviation of the out-of-sample return is

$$\hat{\sigma}^i = \sqrt{\frac{1}{|\Upsilon|} \sum_{t \in \Upsilon} \left(w_t^{i\prime} r_{t+q} - \hat{\mu}^i \right)^2}, \tag{19}$$

where $\hat{\mu}^i = \frac{1}{|\Upsilon|} \sum_{t \in \Upsilon} w_t^{i\prime} r_{t+q}$ is the mean of OOS returns and $t \in \Upsilon := \{x \mid \{x = \tau, \tau + q, \tau + 2q, \cdots\} \cap \{x \leq T - q\}\}$. $|\Upsilon|$ denotes the cardinality of the set Υ which represents the number of times rebalancing portfolio. The Sharpe ratio of portfolio strategy i is

$$SR^i = \frac{\hat{\mu}^i}{\hat{\sigma}^i}, \tag{20}$$

and the turnover of portfolio strategy i is defined as

$$TO^i = \frac{1}{|\Upsilon|} \sum_{t \in \Upsilon} \sum_{j=1}^{N} \left(|w^i_{j,t+q} - w^i_{j,t^+}| \right), \tag{21}$$

where w_{t^+} indicates the weight before rebalancing due to price fluctuation between two consecutive rebalancing points, and w^i_{j,t^+} is the j^{th} element in $w^i_{t^+}$ for portfolio strategy i, where $j \in \{1, 2, \cdots, N\}$. Hence, $|w^i_{j,t+q} - w^i_{j,t^+}|$ measures the change in portfolio weight for asset j at rebalancing point $t+q$, and TO measures the average weight change for all $t \in \Upsilon$. Therefore, the transaction cost adjusted Sharpe ratio for portfolio strategy i is defined as

$$TCadjSR^i = \frac{\hat{\mu}^i - TC^i}{\hat{\sigma}^i}, \tag{22}$$

where

$$TC^i = TO^i * |\Upsilon| * cost\_per\_transaction \tag{23}$$

denotes the transaction cost. Recall that $|\Upsilon|$ is the cardinality of the set Υ, measuring the total number of times investors rebalance their portfolios. Rebalancing frequency q directly affects $|\Upsilon|$: higher frequency of rebalancing (i.e. smaller q) will result in larger $|\Upsilon|$, thus greater transaction cost. $cost\_per\_transaction$ can be interpreted as per US dollar transaction cost to trade stocks. In line with DeMiguel et al. (2013), we set $cost\_per\_transaction =$ 50 basis points. Finally, the transaction cost adjusted returns for portfolio strategy i are defined as

$$TCadjR^i_t = w^{i\prime}_t r_{t+q} - TC^i. \tag{24}$$

Note that $t \in \Upsilon$ is a subscript indicating each rebalancing point. For Sharpe ratio comparison, we further utilize a Sharpe ratio test devised by Ledoit and Wolf (2008) to reveal whether Sharpe ratios are statistically different between portfolio strategies by pair-wise comparison. Specifically, we implement a circular block bootstrap method, which is robust to correlated returns.[12] Let μ^i (μ^j) denote the mean (excess) return and σ^i (σ^j) denote the standard deviation for strategy i (j). The null hypothesis is

$$H_0: \quad \frac{\mu^i}{\sigma^i} - \frac{\mu^j}{\sigma^j} = 0. \tag{25}$$

4.4 Empirical results

In this subsection, we compare the criteria described above for all strategies listed in Table 1 and discuss their implications. We also provide robustness check for the OWL-ADMM algorithm and we illustrate detailed weight distributions for each asset class using various portfolio strategies in Appendix D.

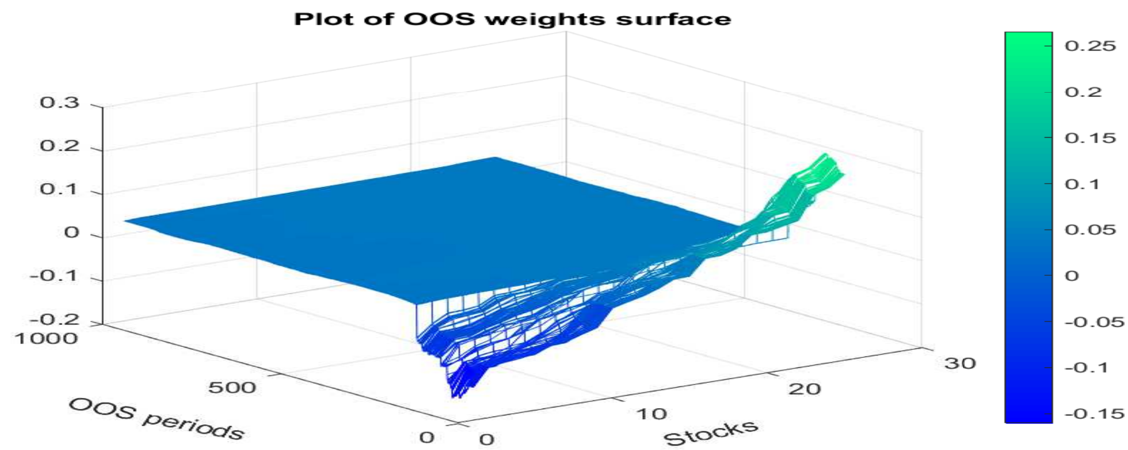
4.4.1 Fama-French 25 portfolios

Figure 3 displays the weight distributions of the FF25 portfolios in three-dimensional graphs, using three OWL-based portfolio selection strategies.[13] It is worth noting that the full sample period for FF25 portfolios is from July 1926 to December 2017. Since we are using a rolling window scheme with a 120-month rolling window size to evaluate portfolio performance, the beginning of the out-of-sample is July 1936, where the portfolio positions are evaluated using the first rolling window from July 1926 to June 1936. Considering the "great depression" happened between the end of 1920s and the early 1930s, we argue that this is the main contributor to the volatile portfolio positions in the first 150 months, or thereabouts, of the out-of-sample period. After that, portfolio positions are much closer to the equal-weighted positions due to the grouping property as discussed before. It is worth noting that each panel in this figure has different color scales.[14] Note that Figure 3 panel (a) implements the OWL shrinkage method on the minimum variance portfolio. Panel (b) implements the OWL shrinkage method as well as the no-short-sale constraint,

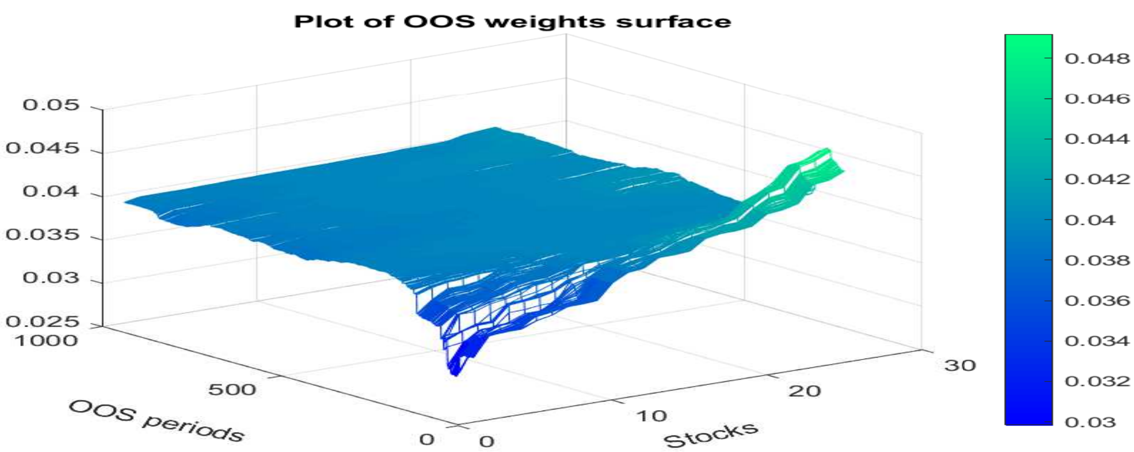
[12] We download the code from https://www.econ.uzh.ch/en/people/faculty/wolf/publications and we compute the two-sides p-values with 1000 (B=1000) bootstrap random draws and block size sets to be 5 (b=5).

[13] Note that the distribution of the 25 portfolios at each point of time during the out-of-sample period is reordered in an ascending order to enable clear three-dimensional display.

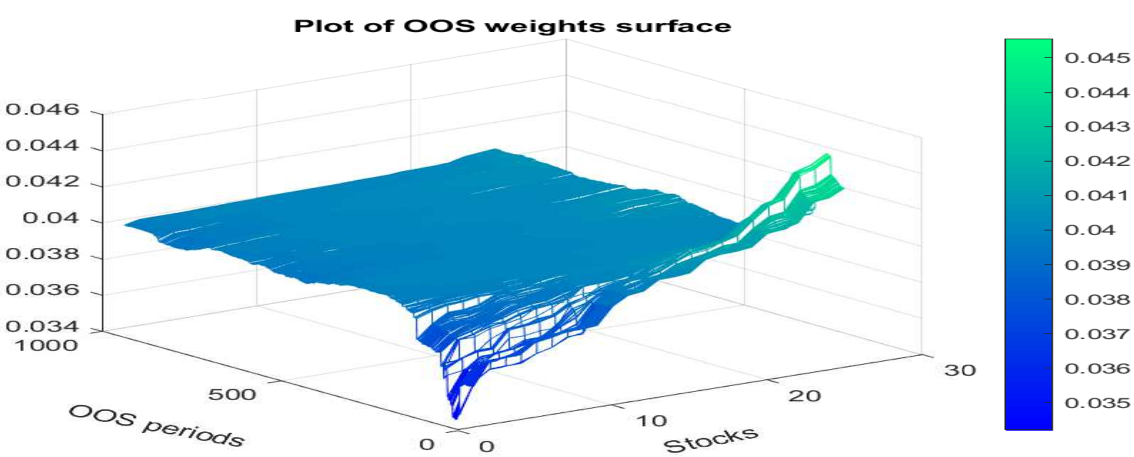
[14] For example, in Figure 3 panel (a), the scale is from -0.15 to 0.25, whereas in panel (b) the scale is from 0.03 to 0.048. Therefore, although panel (a) appears to have flatter distribution of portfolio weights than does panel (b), it is panel (b) actually having much more similar distribution to the equal-weighted portfolio positions.



(a) minVar-OWL



(b) minVar-OWL-Pos



(c) minVar-OWL-bounds

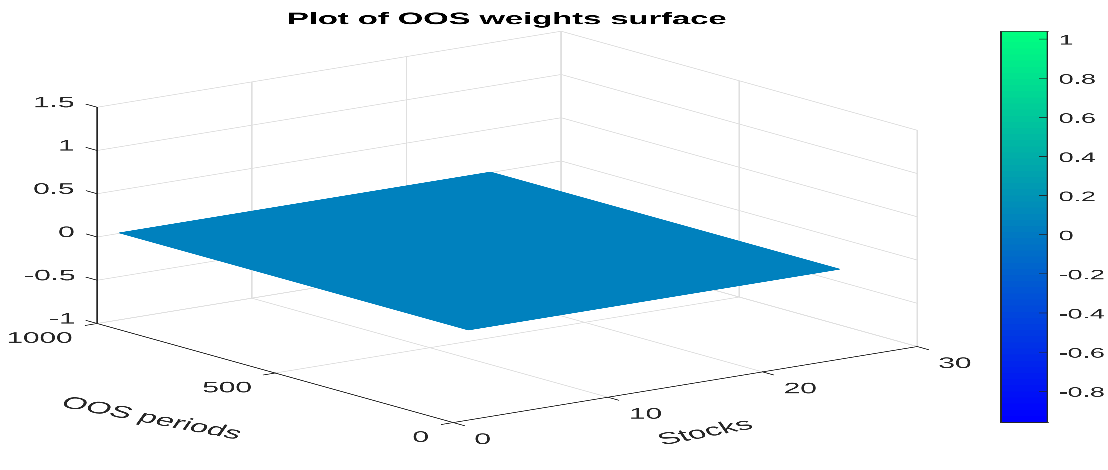
Figure 3. Weight distribution of various strategies using FF25

while panel (c) implements the OWL shrinkage method as well as a bound constraint (i.e., each asset's position is constrained between -5% and 30%). We find that the constrained and unconstrained OWL shrinkage methods display similar patterns for the weight distribution across the out-of-sample period and the FF25 portfolios. The near-equal-weighted portfolio positions are influenced by the grouping property of the OWL shrinkage component. Furthermore, once we consider the constrained OWL shrinkage method, we obtain a similar pattern to the unconstrained version but in a controlled manner, for example, all asset positions are non-negative (with respect to the no-short-sale constraint) or restricted between a specified range (with respect to the bound constraint).

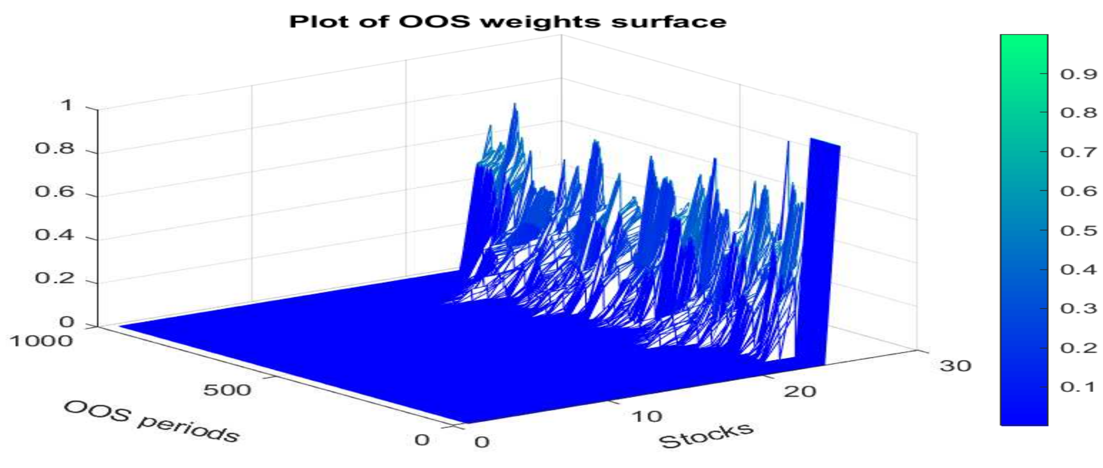
Figure 4 is similar to Figure 3 except that it considers three alternative portfolio selection strategies: panel (a) displays the weight distribution for the equal-weighted portfolio strategy, while panel (b) and (c) display the weight distribution for the no-short-sale constrained portfolio strategy by Jagannathan and Ma (2003) (minVar-JM) and the shrinkage method by Ledoit and Wolf (2003) (minVar-LW), respectively. We find largely dispersed portfolio distributions in the last two strategies. minVar-JM strategy exhibits portfolio weights between zero and 90%, where the majority of the 25 portfolios take the zero positions. On the other hand, the minVar-LW strategy exhibits portfolio weights between -40% and 80%. This figure also explains why the latter two portfolio strategies occur high transaction cost due to excessive position changes at rebalancing.

Because of limited display space, we provide more detailed insights in Appendix D.3 regarding each portfolio strategy's performance on sparse selection, out-of-sample returns and turnovers.

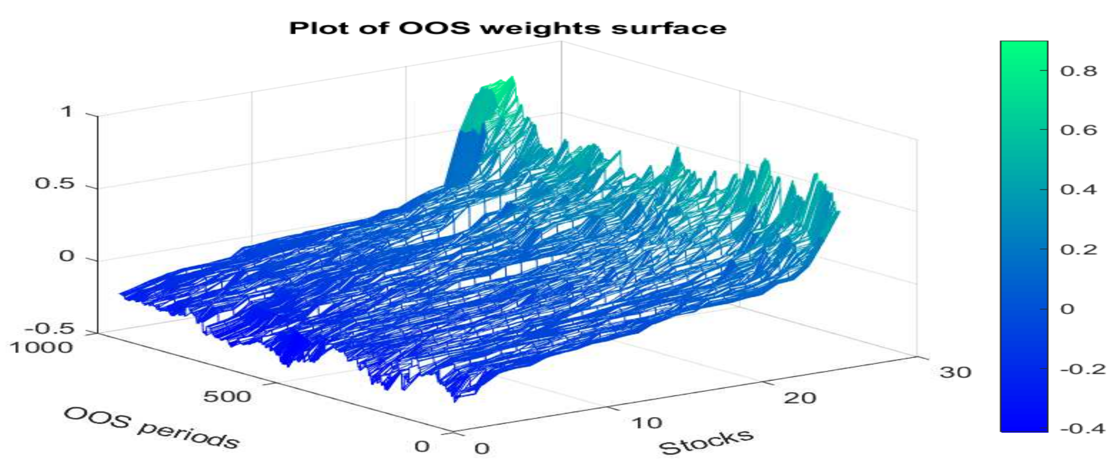
Next, we move on to evaluate the performance scores and compare each candidate strategies. Table 2 reports the out-of-sample performance scores using various criteria including the Sharpe ratio, standard deviation, turnover, (annualized) mean returns, transaction cost and transaction cost adjusted Sharpe ratios for each trading strategy. We find that the plain minVar strategy achieves highest OOS Sharpe ratio and lowest standard deviation. This may be because the FF25 portfolios are less prone to idiosyncratic noises and hence less prone to estimation errors in the sample covariance matrix compared to other asset classes. Moreover, the FF25 portfolios have relatively small cross-sectional dimension but



(a) EW



(b) minVar-JM



(c) minVar-LW

Figure 4. Weight distribution of various strategies using FF25

Table 2. OOS scores using FF25

	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	0.7271	0.0549	0.0172	0.1384	0.0010	0.7216
minVar	0.9785	0.0401	0.7558	0.1358	0.0453	0.6518
minVar-JM	0.8020	0.0425	0.0678	0.1182	0.0041	0.7744
minVar-LW	0.8613	0.0425	0.2952	0.1268	0.0177	0.7410
minVar-OWL	0.7727	0.0484	0.0278	0.1295	0.0017	0.7627
minVar-OWL-Pos	0.7323	0.0544	0.0172	0.1379	0.0010	0.7268
minVar-OWL-bounds	0.7299	0.0549	0.0178	0.1389	0.0011	0.7243
minVar-hard-OWL	0.0141	0.4128	9.8011	0.0202	0.5881	-0.3971
minVar-LW-OWL	0.7739	0.0483	0.0295	0.1295	0.0018	0.7633
MVE-OWL-Pos	0.7344	0.0547	0.0147	0.1391	0.0009	0.7298
MVE-OWL-bounds	0.7311	0.0551	0.0163	0.1395	0.0010	0.7260

Note: this table reports performance scores for various strategies using Fam-French 25 portfolios. The transaction cost is calibrated to be 50 base points for trading 1 US dollar.

have large time-series dimension. This helps to obtain a relatively precise estimate of the covariance matrix, which is crucial in our optimization problems. Meanwhile, the minVar-JM and minVar-LW also do well in achieving high Sharpe ratios. On the other hand, the minVar strategy, although it produces high Sharpe ratio, suffers from high turnovers. The EW strategy produces the smallest turnovers, closely followed by some OWL related strategies. Note that we calibrate the cost of trading stocks worth one US dollar to be 50 basis points, and this can be viewed as a scaling parameter to tilting weights between Sharpe ratio and the transaction cost; a higher value on this parameter will favor strategies with low transaction cost. By looking at the transaction cost adjusted Sharpe ratio, we find that the plain minVar strategy is outperformed by many other competitors. Notably, the mean-variance efficient (MVE) portfolio typically performs poorly, but once regularized by OWL and further imposing (no-short-sale or bounds) constraints, the MVE portfolio achieves sizeable Sharpe ratio and low transaction cost. It is worth stressing that some of those transaction cost adjusted Sharpe ratios are very similar. To see the significance in their performance, we run a bootstrap based test outlined in Section 4.3.

Table 3 reports the p-values of the pair-wise comparison of Sharpe ratios between any two strategies using the Fama-French 25 portfolios. First of all, we find that the minVar-OWL strategy is not statistically different from the Equal weighted, minVar-JM, and minVar-LW strategies which exhibit high Sharpe ratios in Table 2. Similarly, this insignificance also appears after comparing minVar-JM, minVar-LW, and equal weighted

Table 3. Pairwise Sharpe ratio test using FF25

		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0190	N/A									
minVar-JM	3	0.1089	0.0470	N/A								
minVar-LW	4	0.2537	0.1029	0.5135	N/A							
minVar-OWL	5	0.2298	0.0470	0.5554	0.4306	N/A						
minVar-OWL-Pos	6	0.0150	0.0220	0.1399	0.2957	0.2498	N/A					
minVar-OWL-bounds	7	0.0160	0.0190	0.1239	0.2488	0.2358	0.0140	N/A				
minVar-hard-OWL	8	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	N/A			
minVar-LW-OWL	9	0.2398	0.0490	0.5704	0.4605	0.8641	0.2787	0.2627	0.0010	N/A		
MVE-OWL-Pos	10	0.0010	0.0300	0.1518	0.2567	0.2907	0.2368	0.0050	0.0010	0.3227	N/A	
MVE-OWL-bounds	11	0.0010	0.0230	0.1129	0.2587	0.2687	0.4346	0.2777	0.0010	0.2767	0.0010	N/A

Note: this table reports the p-values of the pairwise Sharpe ratio tests in Ledoit and Wolf (2008) using the Fama-French 25 portfolios. If p-value is great than 5%, then we do not reject the hypothesis that these two strategies yield the same (TC adjusted) Sharpe ratio.

strategies, indicating these strategies are not significantly different in producing Sharpe ratios.

4.4.2 SP500 daily returns

Table 4 reports performance scores for various strategies using the S&P 500 daily returns with weekly or monthly rebalancing frequency.[15] The transaction cost is calibrated to be 50 basis points for trading one US dollar of stocks. First of all, we find when using individual stocks as test assets, particularly if using daily return series, estimation errors in the sample covariance matrix become more evident: by looking into SR and $\hat{\sigma}$, the plain minVar strategy becomes inferior to many competitors, resulting from elevated estimation error in the sample covariance matrix. Then, by looking into turnovers, we find that the equal weighted strategy and some OWL related strategies (particularly for the MVE-OWL-bounds strategy) produce the smallest turnovers. It is worth stressing that we find that the "minVar-OWL" outperforms the equal weighted strategy in both Sharpe ratio and turnovers, with either weekly or monthly rebalancing frequency, which is a remarkable finding as it is difficult to find a strategy that outperforms the equal weighted strategy in both Sharpe ratio and turnovers. Next, we compare Sharpe ratios between strategies pairwisely and test for significance.

[15] Due to limited display space, we only include the out-of-sample comparison scores from this subsection onward. All details regarding the estimation results are available upon request.

Table 4. OOS scores using SP500

Panel A: SP500 daily returns with weekly rebalancing						
	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	1.0046	0.0244	0.0314	0.1771	0.0082	0.9584
minVar	0.5338	0.0540	12.9184	0.2079	3.3588	-8.0889
minVar-JM	1.5831	0.0143	0.0849	0.1629	0.0221	1.3684
minVar-LW	1.3914	0.0130	0.6057	0.1306	0.1575	-0.2866
minVar-OWL	1.0568	0.0227	0.0306	0.1728	0.0080	1.0082
minVar-OWL-Pos	1.0216	0.0238	0.0310	0.1756	0.0080	0.9748
minVar-OWL-bounds	1.0128	0.0240	0.0319	0.1751	0.0083	0.9649
minVar-hard-OWL	1.0656	0.0223	0.0309	0.1717	0.0080	1.0158
minVar-LW-OWL	1.0513	0.0227	0.0305	0.1722	0.0079	1.0030
MVE-OWL-Pos	0.9811	0.0257	0.0561	0.1815	0.0146	0.9023
MVE-OWL-bounds	0.9913	0.0244	0.0266	0.1742	0.0069	0.9520
Panel B: SP500 daily returns with monthly rebalancing						
	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	1.0399	0.0484	0.0684	0.1745	0.0041	1.0154
minVar	0.6107	0.0831	22.3297	0.1758	1.3398	-4.0433
minVar-JM	1.5013	0.0318	0.2021	0.1654	0.0121	1.3912
minVar-LW	1.3267	0.0288	1.3261	0.1326	0.0796	0.5306
minVar-OWL	1.0850	0.0455	0.0673	0.1711	0.0040	1.0594
minVar-OWL-Pos	1.0552	0.0474	0.0676	0.1732	0.0041	1.0305
minVar-OWL-bounds	1.0472	0.0476	0.0681	0.1726	0.0041	1.0223
minVar-hard-OWL	1.0888	0.0450	0.0684	0.1699	0.0041	1.0624
minVar-LW-OWL	1.0798	0.0456	0.0669	0.1705	0.0040	1.0544
MVE-OWL-Pos	1.0694	0.0494	0.1138	0.1829	0.0068	1.0294
MVE-OWL-bounds	1.0568	0.0473	0.0560	0.1733	0.0034	1.0363

Note: this table reports performance scores for various strategies using the Standard & Poor 500 stocks daily returns with weekly or monthly rebalancing frequency. The transaction cost is calibrated to be 50 base points for trading 1 US dollar.

Table 5 reports the p-values of the Sharpe ratio test outlined in Section 4.3 using SP500 stocks with daily returns. The pairwise comparison suggests that the minVar-OWL strategy is statistically outperforming the equal weighted strategy with weekly rebalancing frequency, while insignificant for the monthly rebalancing frequency. Meanwhile, we find that for the SP500d returns, the best performing strategies in terms of Sharpe ratios are minVar-JM and minVar-LW and their superior performance against other strategies is significant suggested by the Sharpe ratio test. Next, we set out to test the SP100 stocks with monthly returns, which exhibit higher correlations between stock returns compared to SP500d stocks.

Table 5. Sharpe ratio test using SP500d

Panel A: SP500 daily returns with weekly rebalancing												
		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0330	N/A									
minVar-JM	3	0.0010	0.0010	N/A								
minVar-LW	4	0.0709	0.0010	0.2308	N/A							
minVar-OWL	5	0.0150	0.0250	0.0010	0.1059	N/A						
minVar-OWL-Pos	6	0.0030	0.0320	0.0010	0.0899	0.0160	N/A					
minVar-OWL-bounds	7	0.0100	0.0230	0.0020	0.0729	0.0110	0.0030	N/A				
minVar-hard-OWL	8	0.0110	0.0140	0.0010	0.1079	0.0989	0.0150	0.0140	N/A			
minVar-LW-OWL	9	0.0160	0.0180	0.0030	0.0929	0.0010	0.0340	0.0230	0.0290	N/A		
MVE-OWL-Pos	10	0.6174	0.0500	0.0010	0.0679	0.1678	0.3956	0.5095	0.1049	0.1848	N/A	
MVE-OWL-bounds	11	0.5884	0.0370	0.0010	0.0559	0.0340	0.2118	0.3556	0.0240	0.0440	0.7193	N/A
Panel B: SP500 daily reurns with monthly rebalancing												
		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.1059	N/A									
minVar-JM	3	0.0080	0.0070	N/A								
minVar-LW	4	0.2757	0.0030	0.2977	N/A							
minVar-OWL	5	0.0899	0.0939	0.0110	0.3387	N/A						
minVar-OWL-Pos	6	0.0829	0.0989	0.0090	0.2977	0.1119	N/A					
minVar-OWL-bounds	7	0.0999	0.0909	0.0090	0.2817	0.1009	0.0669	N/A				
minVar-hard-OWL	8	0.1149	0.0929	0.0160	0.3107	0.4835	0.1439	0.1059	N/A			
minVar-LW-OWL	9	0.1189	0.0839	0.0060	0.3347	0.0020	0.1578	0.1149	0.1159	N/A		
MVE-OWL-Pos	10	0.7033	0.0919	0.0180	0.3427	0.8861	0.8771	0.7612	0.8332	0.8881	N/A	
MVE-OWL-bounds	11	0.6484	0.1089	0.0080	0.3057	0.5604	0.9700	0.7972	0.5235	0.6374	0.7722	N/A

Note: this table reports p-values of pair-wise Sharpe ratio test according to Ledoit and Wolf (2008) using SP500d returns. The rebalancing frequency in Panel A is weekly, and in Panel B is monthly.

4.4.3 SP100 monthly returns

Table 6 reports performance scores using SP100 stocks with monthly returns, and we rebalance the portfolio monthly. First of all, we find that minVar-OWL and some other OWL related strategies consistently outperform the equal weighted strategy in terms of the Sharpe ratio and turnovers. The MVE-OWL-bounds strategy yields the smallest turnover while the turnover of the minVar-JM strategy doubles that of OWL related strategies. Second, the raw Sharpe ratio (i.e. not adjusted by transaction cost) of the minVar-JM strategy tops the ranking, and it is closely followed by OWL related strategies. However, after being adjusted by transaction cost, the minVar-OWL and minVar-hard-OWL strategies top the ranking. Third, by comparing Table 6 and Table 4, we find that the performance of OWL related strategies has improved, and we reckon that is because SP100 stocks with monthly returns exhibit higher correlation between stocks which is a desirable property for the OWL shrinkage method to work well. Next, we apply the Sharpe ratio test in Section

Table 6. OOS scores using SP100m

	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	0.9233	0.0460	0.0571	0.1472	0.0034	0.9018
minVar	0.0624	0.0851	4.4201	0.0184	0.2652	-0.8371
minVar-JM	1.0065	0.0345	0.1140	0.1203	0.0068	0.9493
minVar-LW	0.8537	0.0359	0.2974	0.1061	0.0178	0.7102
minVar-OWL	0.9811	0.0420	0.0555	0.1426	0.0033	0.9582
minVar-OWL-Pos	0.9418	0.0444	0.0566	0.1448	0.0034	0.9197
minVar-OWL-bounds	0.9257	0.0461	0.0572	0.1477	0.0034	0.9042
minVar-hard-OWL	0.9862	0.0415	0.0565	0.1419	0.0034	0.9626
minVar-LW-OWL	0.9633	0.0425	0.0546	0.1417	0.0033	0.9411
MVE-OWL-Pos	0.9302	0.0451	0.0581	0.1452	0.0035	0.9079
MVE-OWL-bounds	0.9293	0.0458	0.0516	0.1475	0.0031	0.9098

Note: this table reports performance scores for various strategies using the Standard & Poor 100 stocks with monthly returns and rebalanced monthly. The transaction cost is calibrated to be 50 base points for trading 1 US dollar.

4.3 to check the significance between strategies.

Table 7. Sharpe ratio test using SP100m

		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0020	N/A									
minVar-JM	3	0.5195	0.0010	N/A								
minVar-LW	4	0.7982	0.0010	0.3526	N/A							
minVar-OWL	5	0.0030	0.0010	0.8202	0.6074	N/A						
minVar-OWL-Pos	6	0.0050	0.0020	0.6084	0.7123	0.0010	N/A					
minVar-OWL-bounds	7	0.0040	0.0030	0.5105	0.7822	0.0040	0.0060	N/A				
minVar-hard-OWL	8	0.0040	0.0010	0.8551	0.5844	0.2577	0.0090	0.0060	N/A			
minVar-LW-OWL	9	0.0320	0.0020	0.7233	0.6454	0.0010	0.1019	0.0280	0.0020	N/A		
MVE-OWL-Pos	10	0.8911	0.0030	0.5634	0.7802	0.2997	0.8062	0.9141	0.2687	0.0030	N/A	
MVE-OWL-bounds	11	0.4575	0.0050	0.5205	0.7862	0.0110	0.1998	0.6204	0.0170	0.1099	0.9740	N/A

Note: this table reports the p-values of the Sharpe ratio test according to Ledoit and Wolf (2008) using the SP100 monthly returns.

Table 7 reveals that the performance between the minVar-OWL strategy and the minVar-JM, minVar-LW strategies is not significantly different. However, it has statistically higher Sharpe ratios than that of the equal weighted strategy.

It is worth stressing that our main target is to draw attention to the comparison between the minVar-OWL strategy and the equal weighted strategy. They receive very similar weight distributions, but we show that the minVar-OWL strategy outperforms the equal weighted strategy in both Sharpe ratio and turnover. In appendix D.3, we show (3-dimensional) graphs that illustrate the weight distribution for some strategies and find that the minVar-OWL strategy has a very similar distribution to the equal weighted strategy. This near-equal-weighted weight distribution is caused by the grouping property as discussed in Section 3.2. We further find the superior performance in Sharpe ratio against the equal weighted strategy is indeed statistically significant after applying a bootstrap based test outlined in Section 4.3. Similar exercises are applied and tested on the CRSP500d stock returns and CRSP100m stock returns. We put those empirical results in Appendix D.2.

So far, we have focused our comparison criteria on Sharpe ratios and turnovers (transaction cost). We find that the minVar-JM strategy delivers impressive Sharpe ratios, although in the SP100m and FF25 asset classes its Sharpe ratio is not significantly different from the minVar-OWL strategy after running a Sharpe ratio test. On the other hand, the minVar-OWL strategy and other OWL related strategies consistently yield the smallest turnovers.

In addition, we stress that we developed a flexible algorithm that can incorporate bespoke weight constraints on individual stocks in the optimization problem. However, in our empirical analysis, we applied (blindly) a -5% to 30% bound for all stocks, since we do not hold any further information on individual stocks. Thus, the bound-constrained OWL strategies can potentially do better if more information about individual stocks becomes available.

Although the Sharpe ratio incorporates both the mean portfolio returns and portfolio risk in its formula, it is often dominated by the portfolio risk component when portfolio returns are small. Alternatively, we use the model confidence set (MCS) method to compare strategies: it includes all the best performing strategies in a set where the average portfolio returns are the highest, while using the portfolio risk to control the confidence band of this set.

4.5 Model confidence set for comparing transaction cost adjusted returns

Hansen et al. (2011) propose the model confidence set (MCS) to compare loss sequences of candidate models and put the best candidates in a "confidence set". MCS avoids comparing models pairwisely, which often leads to inconclusive decisions. Instead, MCS enables us to compare multiple models while returning a set that includes all (single or multiple) best performing models. In our application, we want to compare out-of-sample portfolio returns. We want to answer the following question: which strategies produce the highest returns while taking account of transaction cost, where portfolio risk determines the confidence band for including the best candidates in a set?

To fix ideas, let M^0 denote a set of finite candidate models (i.e. M^0 collects all candidate models) and M be the active model confidence set with size m.[16] Denote by $L_{i,t}$ the loss function of model i at time t.[17] Then,

$$d_{ij,t} = L_{i,t} - L_{j,t}, \qquad \forall i, j \in M^0, \tag{26}$$

is the loss difference function between model i and j at time t. Then, we denote

$$\begin{aligned} \mu_{ij} &= \mathrm{E}(d_{ij,t}), \\ \bar{d}_{ij} &= n^{-1} \sum_{t=1}^{n} d_{ij,t}, \end{aligned} \tag{27}$$

where μ_{ij} is the expected value of the loss difference between model i and j, and $\bar{d}_{ij}$ is the sample analogy of μ_{ij}. Denote

$$\bar{d}_{i\cdot} \equiv m^{-1} \sum_{j \in M} \bar{d}_{ij}, \tag{28}$$

where m is the cardinality of set M, and M is the active set which collects models that need to be tested. $\bar{d}_{i\cdot}$ is the average loss difference sequence of model i with all models left in the active set M. Then, the model confidence set is defined as

$$M^* \equiv \{i \in M^0 : \mu_{ij} \leq 0 \quad \forall j \in M^0\}. \tag{29}$$

A detailed testing procedure of MCS is included in Appendix C.

Table 8 reports the p-values of the MCS test. It compares transaction cost adjusted

[16]We set $M = M^0$ at the beginning, then run a series of tests to remove inferior models from the active set M. In the end, what are left in the active set M will be the final model confidence set.

[17]The MCS compares models using a loss function of each model. Since returns are "gains" rather than "losses", we use the negative values of returns to measure each strategy's "loss".

Table 8. MCS test for transaction cost adjusted returns

	FF25	SP500d, w	SP500d, m	SP100m	CRSP500d, w	CRSP500d, m	CRSP100m
EW	0.0020	0.4210	0.4650	0.0040	0.0000	0.8620	0.4170
minVar	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
minVar-JM	0.0000	0.0050	0.0010	0.0040	0.0000	0.0000	0.4070
minVar-hard-OWL	0.0000	0.0130	0.0080	0.0040	0.0000	0.0070	1.0000
minVar-LW	0.0000	0.0000	0.0000	0.0040	0.8270	1.0000	0.4070
minVar-LW-OWL	0.0000	1.0000	1.0000	0.8930	0.0000	0.0040	0.3430
minVar-OWL	0.0000	0.0130	0.0080	0.0040	0.0000	0.0000	0.4070
minVar-OWL-bounds	0.0000	0.0050	0.0010	0.0040	0.0000	0.0000	0.4170
minVar-OWL-Pos	0.0000	0.4210	0.4650	0.0040	1.0000	0.8860	0.4170
MVE-OWL-bounds	1.0000	0.0000	0.5520	1.0000	0.0000	0.0000	0.0000
MVE-OWL-Pos	0.0840	0.0050	0.0000	0.0040	0.0000	0.0000	0.4070

Note: this table reports the p-values of the MCS test. It compares transaction cost adjusted returns using various strategies and using different asset classes. We consider both weekly ('w') and monthly ('m') rebalancing frequencies for daily returns. If p-value is greater than 5%, then the corresponding strategy will be included in the MCS.

returns of various strategies within each asset class. We consider both weekly and monthly rebalancing frequencies for daily returns. If p-value is greater than 5%, then the corresponding strategy will be included in the MCS.

First of all, we notice that the equal weighted strategy has been included in the MCS four times, confirming that the naive 1/N strategy performs well in terms of producing sizeable returns. On the other hand, we find that the minVar-OWL-Pos strategy (OWL regularized minimum variance portfolio with no-short-sale constraint) has been included in the MCS fives times, which makes it the only strategy that has been included in the MCS more often than the equal weighted strategy.

We also notice that the MVE-OWL-bound and MVE-OWL-Pos strategies performs particularly well with the Fama-French portfolios. The OWL shrinkage method with further constraints on portfolio weights helps to utilize the optimization gains from the mean variance efficient portfolio. We reckon this is because sorted portfolio returns are less prone to idiosyncratic noises and thus the sample estimate of expected asset return and asset covariances are less biased compared to individual stock returns. Also, for the FF25 asset class, large T and small N (i.e. large in time-series dimension and small in cross sectional dimension compared to other asset classes) help to improve the precision of the sample estimate of the covariance matrix.

Meanwhile, the minVar-JM strategy, which performs well when using Sharpe ratio as comparison criterion, performs poorly if we use MCS to compare portfolio returns: the minVar-JM strategy has been included in MCS only once, indicating that the minVar-JM strategy produces significantly lower returns than other strategies using various test assets. We find that the MCS for CRSP100m asset class (which consists of 100 randomly selected (usually small) stocks from the CRSP dataset) includes many (9 out of 11) candidate strategies, which is caused by large variations in out-of-sample returns for each strategy using this asset class.

5 Conclusion

In this paper, we introduce the OWL shrinkage method for efficient portfolio construction problems. The OWL shrinkage method encompasses the LASSO shrinkage setup and exploits contemporaneous correlations between stocks, thereby extending the LASSO shrinkage method in DeMiguel et al. (2009a) and the VAR(1) model in DeMiguel et al. (2014). We develop an efficient algorithm that incorporates the OWL shrinkage method together with bespoke constraints on individual stocks if prior information is available. We apply our OWL portfolio strategies on five asset classes and find that the OWL shrinkage method outperforms other benchmarks when stocks exhibit high correlations. DeMiguel et al. (2009b) compare the naive 1/N portfolio strategy with the other 14 optimization-based strategies, finding superb out-of-sample performance in the naive 1/N portfolio. In this paper, we bridge the gap between the naive 1/N portfolio strategy and an optimization based method: our OWL optimization problem yields similar portfolio weights to the 1/N portfolio strategy due to the grouping property, yet our OWL-based portfolio strategies outperform the 1/N strategy in terms of Sharpe ratios and turnovers. A bootstrap based Sharpe ratio test by Ledoit and Wolf (2008) also confirms that this difference in Sharpe ratio is significant.

Appendix

A ADMM algorithm to solve the constrained OWL optimization problem

Boyd et al. (2010) proposed a general optimization algorithm which utilizes the augmented Lagrangian function and can decompose a complex optimization problem into two parts which share different characteristics in computational complexity. This algorithm optimizes these two parts separately and in an orderly fashion, hence gains the name of "alternating directions".

A.1 Augmented Lagrangian

First, define the augmented Lagrangian of the optimization problem (14) - (18) as

$$\begin{aligned}\ell_\rho(w, v, \alpha, \beta, \theta, \xi) = \;& \frac{\gamma}{2} w'\hat{\Sigma}w - \hat{\mu}'w + \Omega_\omega(v) + \alpha'(w - v) + \beta(w'e - 1) \\ & + \theta'(lb - w) + \xi'(w - ub) + \frac{\rho}{2}(||w - v||_2^2 + (w'e - 1)^2 \\ & + ||lb - w||_2^2 + ||w - ub||_2^2),\end{aligned} \tag{A.1}$$

where α, β, θ and ξ are Lagrangian multipliers, ρ is a parameter to control penalty and e is a column vector of ones. The ADMM algorithm consists of these updates for each step:

$$w^{k+1} = \arg\min_{w} \quad \ell_\rho(w, v^k, \alpha^k, \beta^k, \theta^k, \xi^k), \tag{A.2}$$

$$w_i^{k+1} = lb_i; \quad if \quad w_i^{k+1} < lb_i \quad \forall \quad i = 1, 2, ..., N, \tag{A.3}$$

$$w_i^{k+1} = ub_i; \quad if \quad w_i^{k+1} > ub_i \quad \forall \quad i = 1, 2, ..., N, \tag{A.4}$$

$$v^{k+1} = \arg\min_{v} \quad \ell_\rho(w^{k+1}, v, \alpha^k, \beta^k, \theta^k, \xi^k), \tag{A.5}$$

$$\alpha^{k+1} = \alpha^k + \rho(w^{k+1} - v^{k+1}), \tag{A.6}$$

$$\beta^{k+1} = \beta^k + \rho(e'w^{k+1} - 1), \tag{A.7}$$

$$\theta^{k+1} = \theta^k + \rho(lb - w^{k+1}), \tag{A.8}$$

$$\theta_i^{k+1} = 0; \quad if \quad lb_i - w_i^{k+1} < 0 \quad \forall \quad i = 1, 2, ..., N, \tag{A.9}$$

$$\xi^{k+1} = \xi^k + \rho(w^{k+1} - ub), \tag{A.10}$$

$$\xi_i^{k+1} = 0; \quad if \quad w_i^{k+1} - ub_i < 0 \quad \forall \quad i = 1, 2, ..., N, \tag{A.11}$$

where k is a superscript indicating the step number. We refer to equations (A.3) and (A.4) as *primal feasibility* conditions, and equations (A.9) and (A.11) as *complementary slackness* conditions. Moreover, (A.2) can be simplified as

$$\begin{aligned}
w^{k+1} &= \arg\min_{w} \quad \ell_\rho(w, v^k, \alpha^k, \beta^k, \theta^k, \xi^k) \\
&= \arg\min_{w} \quad \Big[\frac{\gamma}{2} w'\hat{\Sigma} w - \hat{\mu}' w + \alpha^{k\prime}(w - v^k) + \beta^k(e'w - 1) + \theta^{k\prime}(lb - w) + \xi^{k\prime}(w - ub) \\
&\qquad + \frac{\rho}{2}(||w - v^k||_2^2 + (w'e - 1)^2 + ||lb - w||_2^2 + ||w - ub||_2^2)\Big] \\
&= \arg\min_{w} \quad \Big[\frac{\gamma}{2} w'\hat{\Sigma} w - (\hat{\mu} - \alpha^k - \beta^k e + \theta^k - \xi^k)' w \\
&\qquad + \frac{\rho}{2}(||w - v^k||_2^2 + (w'e - 1)^2 + ||lb - w||_2^2 + ||w - ub||_2^2))\Big] \\
&= \arg\min_{w} \quad \Big[\frac{1}{2} w'(\gamma\hat{\Sigma} + \rho(3I + ee'))w - (\hat{\mu} - \alpha^k - \beta^k e + \theta^k - \xi^k \\
&\qquad + \rho(v^k + e - lb - ub))' w\Big] \\
&= \ (\gamma\hat{\Sigma} + \rho(3I + ee'))^{-1}(\hat{\mu} - \alpha^k - \beta^k e + \theta^k - \xi^k + \rho(v^k + e - lb - ub)).
\end{aligned}$$

Meanwhile, equation (A.5) can be simplified as

$$\begin{aligned} v^{k+1} &= \arg\min_{v} \left[\frac{\rho}{2} ||v - w^{k+1} - \frac{1}{\rho}\alpha^k||_2^2 + \Omega_\omega(v) \right] \\ &= \text{prox}_\Omega(w^{k+1} + \frac{1}{\rho}\alpha^k), \end{aligned}$$

where $\text{prox}_\Omega(.)$ is a proximal function for the OWL shrinkage method. Discussion of how to find a minimizer of the proximal function $\text{prox}_\Omega(.)$ can be found in Sun (2019).

A.1.1 Optimality conditions

Suppose w^* and v^* are optimizers of the optimization problem (14) - (18). Then, the optimality conditions of (A.1) consist of the *primal feasibility* and the *dual feasibility* conditions. The *primal feasibility* concerns the following conditions

$$w^* - v^* = \mathbf{0}, \tag{A.12}$$

$$w^{*\prime} e - 1 = 0, \tag{A.13}$$

$$w^* \succeq \mathbf{lb}, \tag{A.14}$$

$$w^* \preceq \mathbf{ub}. \tag{A.15}$$

Equations (A.2) and (A.5) command the *dual feasibility* condition, which requires

$$\triangledown f(w^*) + \alpha^* + \beta^* e - \theta^* + \xi^* = 0, \tag{A.16}$$

$$\triangledown \Omega(v^*) - \alpha^* = 0, \tag{A.17}$$

where $f(w) = \frac{\gamma}{2} w' \hat{\Sigma} w - \hat{\mu}' w$. By equation (A.5), v^{k+1} minimizes the function $\ell_\rho(w^{k+1}, v, \alpha^k, \beta^k, \theta^k, \xi^k)$ w.r.t v, so we have

$$0 = \triangledown \ell_\rho(v^{k+1}) = \triangledown \Omega(v^{k+1}) - \alpha^k - \rho(w^{k+1} - v^{k+1}) = \triangledown \Omega(v^{k+1}) - \alpha^{k+1},$$

which makes (A.17) hold automatically. Similarly, by (A.2), w^{k+1} minimizes the function $\ell_\rho(w, v^k, \alpha^k, \beta^k, \theta^k, \xi^k)$ w.r.t w, so we obtain

$$
\begin{aligned}
0 &= \bigtriangledown f(w^{k+1}) + \alpha^k + \beta^k e - \theta^k + \xi^k + \rho(w^{k+1} - v^k) \\
&\quad + \rho(w^{k+1\prime} e - 1)e - \rho(lb - w^{k+1}) + \rho(w^{k+1} - ub) \\
&= \bigtriangledown f(w^{k+1}) + \alpha^{k+1} + \beta^{k+1} e - \theta^{k+1} + \xi^{k+1} + \rho(v^{k+1} - v^k).
\end{aligned}
$$

Rearranging the above equation gives

$$\bigtriangledown f(w^{k+1}) + \alpha^{k+1} + \beta^{k+1} e - \theta^{k+1} + \xi^{k+1} = -\rho(v^{k+1} - v^k) := s^{k+1},$$

where we denote $s^{k+1} := -\rho(v^{k+1} - v^k)$ as the *dual residual* at step $k+1$, because s^{k+1} is the deviation from a dual feasibility condition in (A.16). Similarly, the *primal residual* at step k w.r.t the primal feasibility conditions in (A.12) and (A.13) is defined as

$$||r^k||_2 = \sqrt{||w^k - v^k||^2 + (w^{k\prime} e - 1)^2}.$$

A.1.2 Stopping criterion and the penalty parameter ρ

The stopping criterion for k suggested by Boyd et al. (2010) is such that k satisfies

$$
\begin{aligned}
||r^k||_2 &\leq \epsilon^{pri} \quad \text{and} \quad ||s^k||_2 \leq \epsilon^{dual}, \\
\epsilon^{pri} &= \sqrt{N}\epsilon^{abs} + \epsilon^{rel} \max\{||w^k||_2, ||v^k||_2\}, \\
\epsilon^{dual} &= \sqrt{T}\epsilon^{abs} + \epsilon^{rel} ||\alpha^k + \beta^k e||_2,
\end{aligned}
$$

where ϵ^{rel} and ϵ^{abs} are calibrated to be 0.001.

Boyd et al. (2010) also argues that allowing ρ to change along steps makes computation more efficient and suggests the following scheme for the values of ρ:

$$
\rho^{k+1} = \begin{cases} \tau \rho^k & \text{if } ||r^k||_2 > \eta ||s^k||_2 \ , \\ \rho^k / \tau & \text{if } ||s^k||_2 > \eta ||r^k||_2 \ , \\ \rho^k & \text{otherwise} \ , \end{cases}
$$

where $\eta, \tau > 1$ are two tuning parameters, which are calibrated such that $\eta = 10$, $\tau = 2$ in our exercise.

B Technical proofs

B.1 Proof of Theorem 3.1

Proof. The proof of Theorem 3.1 relies on the Pigou-Dalton transfer principle and the directional derivative lemma at the minimum of a convex function. It follows using a similar argument as in Figueiredo and Nowak (2016), except that we are dealing with different loss functions.

Lemma 3 (Pigou-Dalton transfer principle). *Let be given vector $x \in R^p_+$, and its two components x_i, x_j are such that $x_i > x_j$. Let $\epsilon \in (0, (x_i - x_j)/2)$, $z_i = x_i - \epsilon$, $z_j = x_j + \epsilon$, and $z_k = x_k$, for $k \neq i, j$. Set $\Omega_\omega(x) = \omega' x$, where $\omega \in R^p_+$, and $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_p$. Then it holds*

$$\Omega_\omega(x) - \Omega_\omega(z) \geq \Delta_\omega \epsilon, \qquad \Delta_\omega := \min_{i=1,\cdots,p-1} (\omega_{i+1} - \omega_i).$$

Lemma 4 (Directional derivative). *The directional derivative of function $f : R^K \to R$ at $x \in dom(f)$, in the direction $\xi \in R^K$ is given by*

$$f'(x, \xi) = \lim_{\alpha \to 0^+} [f(x + \alpha\xi) - f(x)]/\alpha, \quad \alpha > 0.$$

If f is a convex function, then $x^* \in \arg\min(f)$ if and only if $f'(x^*, \xi) \geq 0$ for any direction $\xi \in R^K$.

Denote the objective function as $Q(w) = \frac{1}{2} w' \Sigma w + \Omega_\omega(w)$. By definition, if $\hat{w}$ is the minimizer, then $Q(\hat{w}) \leq Q(w)$ for all w. Thus by Lemma 4, for any $\xi \in R^N$,

$$Q'(\hat{w}, \xi) \geq 0. \tag{B.18}$$

Recall that $\Sigma_{i.}$ and $\Sigma_{j.}$ denote the i^{th} and j^{th} columns of the $N \times N$ variance-covariance matrix Σ. Suppose

$$\|\Sigma_{i.} - \Sigma_{j.}\|_2 < \lambda_2, \tag{B.19}$$

and assume $\hat{w}_i \neq \hat{w}_j$. We will show contradiction between assumption $\hat{w}_i \neq \hat{w}_j$ and (B.18). Without loss of generality, assume $\hat{w}_i > \hat{w}_j$, $i < j$. First we define a special directional vector $\xi = (\xi_1, \xi_2, \cdots, \xi_N)'$. Set $\xi_i = 1, \xi_j = -1$ and $\xi_k = 0$ for all $k \neq i, j$. The directional derivative of Q at $\hat{w}$ with such ξ is

$$Q'(\hat{w}, \xi) = \lim_{\alpha \to 0^+} \left(QL_\alpha(\hat{w}, \xi) + RP_\alpha(\hat{w}, \xi) \right), \tag{B.20}$$

where

$$\begin{aligned} QL_\alpha(\hat{w}, \xi) &= \frac{(\hat{w} + \alpha\xi)'\Sigma(\hat{w} + \alpha\xi) - \hat{w}\Sigma\hat{w}}{2\alpha} \\ &= \frac{\alpha\hat{w}'\Sigma\xi + \alpha\xi'\Sigma\hat{w} + \alpha^2\xi'\Sigma\xi}{2\alpha}, \end{aligned} \tag{B.21}$$

and

$$RP_\alpha(\hat{w}, \xi) = \frac{\Omega_\omega(\hat{w} + \alpha\xi) - \Omega_\omega(\hat{w})}{\alpha}. \tag{B.22}$$

Note that $\Sigma' = \Sigma$ and $\hat{w}'\Sigma\xi$ is a scaler, so we have $\hat{w}'\Sigma\xi = \xi'\Sigma\hat{w}$. Then it follows

$$\lim_{\alpha \to 0^+} QL_\alpha(\hat{w}, \xi) = \hat{w}'\Sigma\xi = \text{trace}(\hat{w}'\Sigma\xi) = \text{trace}(\xi\hat{w}'\Sigma). \tag{B.23}$$

Observe that $\xi\hat{w}'$ is a $N \times N$ matrix with i^{th} row as $\hat{w}'$, j^{th} row as $-\hat{w}'$ and the remaining rows are filled with zeros. Then we have

$$\lim_{\alpha \to 0^+} QL_\alpha(\hat{w}, \xi) = \text{trace}(\xi\hat{w}'\Sigma) = \hat{w}'(\Sigma_{i.} - \Sigma_{j.}), \tag{B.24}$$

where $\Sigma_{i.}$ and $\Sigma_{j.}$ are the i^{th} and j^{th} columns of Σ.

Applying the Pigou-Dalton transfer principle on $RP_\alpha(\hat{w}, \xi)$ with $\epsilon = \alpha$, we obtain

$$-RP_\alpha(\hat{w}, \xi)\alpha = \Omega_\omega(\hat{w}) - \Omega_\omega(\hat{w} + \alpha\xi) \geq \Delta_\omega\alpha. \tag{B.25}$$

So for any α and ξ,

$$RP_\alpha(\hat{w}, \xi) \leq -\frac{\Delta_\omega\alpha}{\alpha} = -\Delta_\omega.$$

By the definition of ω in (11), $\Delta_\omega = \lambda_2$. Therefore, applying the above bound in (B.20), we obtain

$$\begin{aligned} Q'(\hat{w},\xi) &\leq \hat{w}'(\Sigma_{i.}-\Sigma_{j.})-\Delta_\omega \\ &= \hat{w}'(\Sigma_{i.}-\Sigma_{j.})-\lambda_2. \end{aligned} \tag{B.26}$$

Using Cauchy-Schwarz inequality, we have

$$\hat{w}'(\Sigma_{i.}-\Sigma_{j.}) \leq \|\hat{w}\|_2\|\Sigma_{i.}-\Sigma_{j.}\|_2 \leq \|\hat{w}\|_1\|\Sigma_{i.}-\Sigma_{j.}\|_2 = \|\Sigma_{i.}-\Sigma_{j.}\|_2,$$

so (B.26) becomes

$$Q'(\hat{w},\xi) \leq \|\Sigma_{i.}-\Sigma_{j.}\|_2-\lambda_2. \tag{B.27}$$

Then, (B.27) together with (B.19) implies

$$Q'(\hat{w},\xi) < 0,$$

which violates (B.18). Hence, there is a contradiction between $\hat{w}_i \neq \hat{w}_j$ and (B.19). So we must have

$$\hat{w}_i = \hat{w}_j,$$

which completes the proof. □

B.1.1 Proof of Theorem 3.2

Proof. The proof of Theorem 3.2 is similar to Theorem 3.1, except we have a different objective function, that is

$$Q(w) = \frac{\gamma}{2}w'\Sigma w - \mu' w + \Omega_\omega(w),$$

where γ is investors' risk aversion level and μ is the vector of expected returns. Following similar procedures as in the proof of Theorem 3.1 and by Lemma 4, we have that for any $\xi \in R^N$,

$$Q'(\hat{w},\xi) \geq 0. \tag{B.28}$$

Suppose

$$\gamma\|\Sigma_{i.} - \Sigma_{j.}\|_2 + |\mu_i - \mu_j| < \lambda_2 \tag{B.29}$$

and assume $\hat{w}_i \neq \hat{w}_j$. Without loss of generality, assume $\hat{w}_i > \hat{w}_j$, $i < j$. Define a special directional vector $\xi = (\xi_1, \xi_2, \cdots, \xi_N)'$. Set $\xi_i = 1, \xi_j = -1$ and $\xi_k = 0$ for all $k \neq i, j$. The directional derivative of Q at $\hat{w}$ with such ξ is

$$Q'(\hat{w}, \xi) = \lim_{\alpha \to 0^+} \left(QL_\alpha(\hat{w}, \xi) + RP_\alpha(\hat{w}, \xi) \right), \tag{B.30}$$

where

$$\begin{aligned} QL_\alpha(\hat{w}, \xi) &= \frac{\frac{\gamma}{2}[(\hat{w} + \alpha\xi)'\Sigma(\hat{w} + \alpha\xi) - \hat{w}\Sigma\hat{w}] - \mu'(\hat{w} + \alpha\xi) + \mu'\hat{w}}{\alpha} \\ &= \frac{\frac{\gamma}{2}(\alpha\hat{w}'\Sigma\xi + \alpha\xi'\Sigma\hat{w} + \alpha^2\xi'\Sigma\xi) - \alpha\mu'\xi}{\alpha}, \end{aligned} \tag{B.31}$$

and

$$RP_\alpha(\hat{w}, \xi) = \frac{\Omega_\omega(\hat{w} + \alpha\xi) - \Omega_\omega(\hat{w})}{\alpha}. \tag{B.32}$$

Note that $\Sigma' = \Sigma$ and $\hat{w}'\Sigma\xi$ is a scalar, so we have $\hat{w}'\Sigma\xi = \xi'\Sigma\hat{w}$. Then it follows

$$\lim_{\alpha \to 0^+} QL_\alpha(\hat{w}, \xi) = \gamma\hat{w}'\Sigma\xi - \mu'\xi, \tag{B.33}$$

where $\gamma\hat{w}'\Sigma\xi = \text{trace}(\gamma\hat{w}'\Sigma\xi) = \gamma\ \text{trace}(\xi\hat{w}'\Sigma)$. Observe that $\xi\hat{w}'$ is a $N \times N$ matrix with i^{th} row as $\hat{w}$, j^{th} row as $-\hat{w}'$ and the remaining rows are filled with zeros. Then we have

$$\lim_{\alpha \to 0^+} QL_\alpha(\hat{w}, \xi) = \gamma\ \text{trace}(\xi\hat{w}'\Sigma) - \mu'\xi = \gamma\hat{w}'(\Sigma_{i.} - \Sigma_{j.}) - \mu'\xi, \tag{B.34}$$

where $\Sigma_{i.}$ and $\Sigma_{j.}$ are the i^{th} and j^{th} columns of Σ.

Similarly to the procedures we used to handle $RP_\alpha(\hat{w}, \xi)$ in the proof of Theorem 3.1, we can obtain

$$RP_\alpha(\hat{w}, \xi) \leq -\lambda_2.$$

Therefore,

$$Q'(\hat{w}, \xi) \leq \gamma\hat{w}'(\Sigma_{i.} - \Sigma_{j.}) - \mu'\xi - \lambda_2. \tag{B.35}$$

Using Cauchy-Schwarz inequality, we have

$$\hat{w}'(\Sigma_{i.} - \Sigma_{j.}) \leq \|\hat{w}\|_2 \|\Sigma_{i.} - \Sigma_{j.}\|_2 \leq \|\hat{w}\|_1 \|\Sigma_{i.} - \Sigma_{j.}\|_2 = \|\Sigma_{i.} - \Sigma_{j.}\|_2,$$

Observe that $\mu'\xi = \mu_i - \mu_j \geq -|\mu_i - \mu_j|$, so (B.35) becomes

$$Q'(\hat{w}, \xi) \leq \gamma \|\Sigma_{i.} - \Sigma_{j.}\|_2 + |\mu_i - \mu_j| - \lambda_2. \tag{B.36}$$

Then (B.36) together with (B.29) implies

$$Q'(\hat{w}, \xi) < 0,$$

which violates (B.28). Hence, there is a contradiction between $\hat{w}_i \neq \hat{w}_j$ and (B.29). So we must have

$$\hat{w}_i = \hat{w}_j,$$

which completes the proof. □

B.1.2 Proof of Lemma 1

Proof. If $\lambda_2 = 0$, then $\omega = (\lambda_1, \lambda_1, ..., \lambda_1)' \in R^N$. So we have

$$\Omega_\omega(w) = \omega'|w|_\downarrow = \lambda_1 e'|w|_\downarrow = \lambda_1 \|w\|_1,$$

where e is a column vector of ones. If we set $\lambda_1 = \lambda$, then $\Omega_\omega(w) = \lambda\|w\|_1$ which completes the proof. □

B.1.3 Proof of Lemma 2

Proof. Note that $|w|_\downarrow = (|w|_{[1]}, |w|_{[2]}, \cdots, |w|_{[N]})'$ reorders the elements in vector $|w| = (|w_1|, |w_2|, \cdots, |w_N|)'$ decreasingly according to the absolute value of each element. Denote by $|w_j|$ and $|w|_{[j]}$ the j^{th} element of $|w|$ and $|w|_\downarrow$, respectively. So we have $|w|_{[1]} > |w|_{[2]} >$

$\cdots > |w|_{[N]}$. Then by the definition of the OSCAR penalty term, we obtain

$$\begin{aligned}\Omega_{OSCAR}(w) &= \lambda_1 ||w||_1 + \lambda_2 \sum_{1\leq i<j\leq N} \max\{|w_i|, |w_j|\} \\ &= \sum_{i=1}^{N} \left[\lambda_1 + \lambda_2 (N-i)\right] |w|_{[i]} \\ &= \omega' |w|_{\downarrow} = \Omega_{\omega}(w),\end{aligned}$$

which completes the proof. □

C MCS testing procedure

The MCS (model confidence set, Hansen et al. (2011)) testing procedure consists of the following steps:

1. Initialize the active model confidence set $M \leftarrow M^0$, where M^0 contains all candidate models.

2. Compute the t-statistics for any pairwise loss difference sequences and the average t-statistics for each model:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_{ij})}} \quad \text{and} \quad t_{i\cdot} = \frac{\bar{d}_{i\cdot}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_{i\cdot})}}, \quad \text{for all } i, j \in M. \tag{C.37}$$

3. Find the model with the largest t-statistic

$$T_{max,M} = \max_{i\in M} \quad t_{i\cdot} \tag{C.38}$$

   and test whether $T_{max,M}$ is significantly different from zero.

4. If $T_{max,M}$ is statistically different (greater) than zero, eliminate this model from the model confidence set, and go back to step 1 while removing this model from set M. Repeat this procedure until $T_{max,M}$ is not significantly different from zero. What remains in M will be the model confidence set.

D Robustness check

D.1 Convergence of OWL-ADMM algorithm

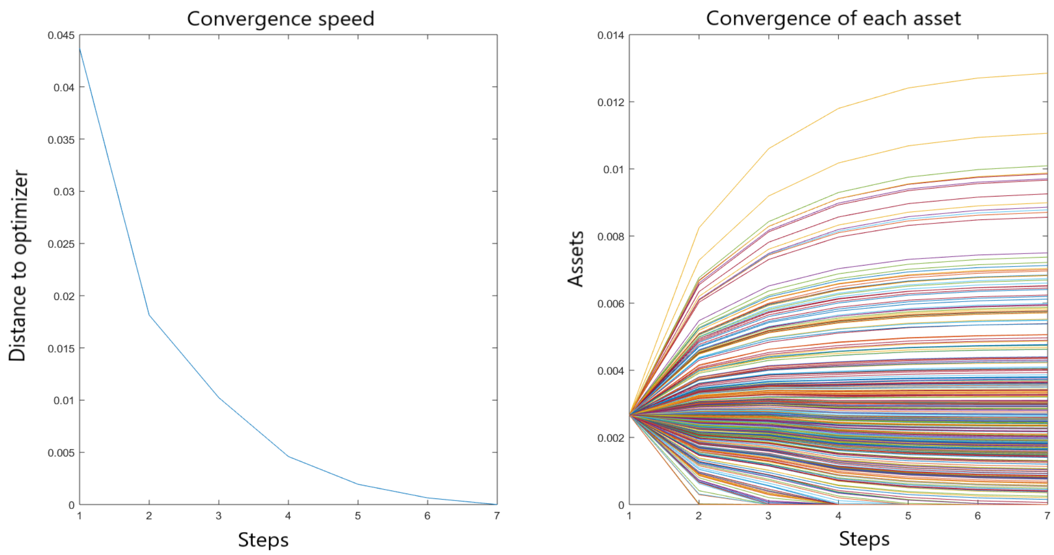


Figure 5. Convergence check for ADMM algorithm using SP500 stocks

Figure 5 shows the convergence diagram for the OWL-ADMM algorithm used to solve the optimization problem in (14) - (18) using SP500 daily returns. Left panel shows the distance (i.e. the ℓ_2 norm of two vectors) of the estimated portfolio weights at each step to the final optimizer. Compared to other algorithms, such as gradient descent, ADMM offers a much faster convergence speed. The right panel shows the individual stock's weights at each step until convergence. Each colored line represents one stock, and note that we initialize the portfolio weights as equal weighted at the beginning. We find that the ADMM algorithm is fast to find the optimizer, typically requiring less than 10 steps.

D.2 Empirical application using randomly selected stocks from CRSP dataset

Panels A and B in Table 9 report the OOS performance scores using 500 randomly selected stocks with daily returns from the CRSP dataset and Panels A and B in Table 10 report the p-values of the Sharpe ratio test by comparing portfolio strategies pair-wisely. We find

Table 9. OOS score using CRSP500d and CRSP100m

Panel A: CRSP500 daily returns with weekly rebalancing						
	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	1.0491	0.0244	0.0000	0.1844	0.0000	1.0491
minVar	1.4801	0.0125	0.6168	0.1337	0.1604	-0.2954
minVar-JM	1.8328	0.0114	0.0699	0.1512	0.0182	1.6124
minVar-LW	1.5108	0.0108	0.2054	0.1182	0.0534	0.8280
minVar-OWL	1.1281	0.0229	0.0027	0.1860	0.0007	1.1239
minVar-OWL-Pos	1.0687	0.0238	0.0018	0.1838	0.0005	1.0661
minVar-OWL-bounds	1.0588	0.0239	0.0107	0.1822	0.0028	1.0425
minVar-hard-OWL	1.1203	0.0228	0.0026	0.1843	0.0007	1.1161
minVar-LW-OWL	1.1049	0.0230	0.0023	0.1830	0.0006	1.1013
MVE-OWL-Pos	0.9150	0.0256	0.0644	0.1690	0.0167	0.8244
MVE-OWL-bounds	0.9797	0.0241	0.0280	0.1706	0.0073	0.9379
Panel B: CRSP500 daily return with monthly rebalancing						
	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	0.8525	0.0547	0.0000	0.1616	0.0000	0.8525
minVar	1.3888	0.0295	1.2631	0.1417	0.0758	0.6463
minVar-JM	1.4227	0.0302	0.1576	0.1488	0.0095	1.3322
minVar-LW	1.4136	0.0270	0.4481	0.1323	0.0269	1.1262
minVar-OWL	0.8887	0.0528	0.0071	0.1627	0.0004	0.8864
minVar-OWL-Pos	0.8664	0.0538	0.0026	0.1615	0.0002	0.8655
minVar-OWL-bounds	0.8591	0.0537	0.0108	0.1599	0.0006	0.8556
minVar-hard-OWL	0.8857	0.0527	0.0070	0.1616	0.0004	0.8834
minVar-LW-OWL	0.8778	0.0528	0.0061	0.1605	0.0004	0.8757
MVE-OWL-Pos	0.7724	0.0586	0.1240	0.1569	0.0074	0.7358
MVE-OWL-bounds	0.8136	0.0549	0.0529	0.1548	0.0032	0.7969
Panel C: CRSP100 monthly return with monthly rebalancing						
	SR	$\hat{\sigma}$	TO	$\hat{\mu}$(annualized)	TC	$TCadjSR$
EW	0.8533	0.0505	0.0794	0.1492	0.0048	0.8261
minVar	0.4439	0.0640	1.6118	0.0985	0.0967	0.0079
minVar-JM	1.2616	0.0334	0.1102	0.1458	0.0066	1.2044
minVar-LW	1.0283	0.0346	0.1623	0.1232	0.0097	0.9470
minVar-OWL	0.8954	0.0435	0.0801	0.1348	0.0048	0.8634
minVar-OWL-Pos	0.8819	0.0480	0.0769	0.1467	0.0046	0.8542
minVar-OWL-bounds	0.8698	0.0490	0.0778	0.1476	0.0047	0.8423
minVar-hard-OWL	0.9094	0.0429	0.0861	0.1350	0.0052	0.8746
minVar-LW-OWL	0.8808	0.0444	0.0793	0.1354	0.0048	0.8499
MVE-OWL-Pos	0.7168	0.0542	0.0814	0.1346	0.0049	0.6908
MVE-OWL-bounds	0.7720	0.0513	0.0661	0.1371	0.0040	0.7497

Note: this table reports performance scores for various strategies using randomly selected 500 stocks (Panel A and B, daily returns) and 100 stocks (Panel C, monthly returns) from CRSP and rebalanced weekly or monthly. The transaction cost is calibrated to be 50 base points for trading 1 US dollar.

Table 10. Sharpe ratio test using CRSP500d and CRSP100m

Panel A: CRSP 500 daily returns, weekly rebalancing

		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0240	N/A									
minVar-JM	3	0.0010	0.0370	N/A								
minVar-LW	4	0.0400	0.8192	0.0559	N/A							
minVar-OWL	5	0.0150	0.0699	0.0010	0.0859	N/A						
minVar-OWL-Pos	6	0.0010	0.0310	0.0010	0.0519	0.0300	N/A					
minVar-OWL-bounds	7	0.0010	0.0280	0.0010	0.0400	0.0160	0.0010	N/A				
minVar-hard-OWL	8	0.0240	0.0450	0.0010	0.0959	0.0010	0.0619	0.0230	N/A			
minVar-LW-OWL	9	0.0509	0.0559	0.0010	0.0719	0.0010	0.1868	0.1049	0.0010	N/A		
MVE-OWL-Pos	10	0.0020	0.0070	0.0010	0.0120	0.0010	0.0010	0.0030	0.0020	0.0020	N/A	
MVE-OWL-bounds	11	0.0030	0.0080	0.0010	0.0250	0.0010	0.0010	0.0010	0.0010	0.0010	0.0090	N/A

Panel B: CRSP 500 daily returns, monthly rebalancing

		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0160	N/A									
minVar-JM	3	0.0140	0.7782	N/A								
minVar-LW	4	0.0549	0.8352	0.9590	N/A							
minVar-OWL	5	0.2488	0.0320	0.0120	0.0819	N/A						
minVar-OWL-Pos	6	0.0100	0.0210	0.0120	0.0679	0.4266	N/A					
minVar-OWL-bounds	7	0.0210	0.0180	0.0090	0.0679	0.2957	0.0040	N/A				
minVar-hard-OWL	8	0.2488	0.0310	0.0070	0.0679	0.0330	0.4745	0.3467	N/A			
minVar-LW-OWL	9	0.3906	0.0220	0.0140	0.0689	0.0030	0.6623	0.4745	0.0050	N/A		
MVE-OWL-Pos	10	0.1259	0.0110	0.0050	0.0390	0.0779	0.0929	0.0959	0.1069	0.1099	N/A	
MVE-OWL-bounds	11	0.1309	0.0210	0.0040	0.0629	0.0849	0.0719	0.0999	0.1009	0.1259	0.1459	N/A

Panel C: CRSP 100 monthly returns, monthly rebalancing

		1	2	3	4	5	6	7	8	9	10	11
EW	1	N/A										
minVar	2	0.0689	N/A									
minVar-JM	3	0.0030	0.0010	N/A								
minVar-LW	4	0.4226	0.0060	0.0889	N/A							
minVar-OWL	5	0.4915	0.0519	0.0090	0.5345	N/A						
minVar-OWL-Pos	6	0.0260	0.0719	0.0030	0.4955	0.8012	N/A					
minVar-OWL-bounds	7	0.0140	0.0829	0.0070	0.4675	0.7023	0.0210	N/A				
minVar-hard-OWL	8	0.4226	0.0420	0.0110	0.5534	0.2567	0.6683	0.5834	N/A			
minVar-LW-OWL	9	0.6583	0.0569	0.0030	0.4655	0.4266	0.9910	0.8611	0.1429	N/A		
MVE-OWL-Pos	10	0.0060	0.2098	0.0010	0.1768	0.0440	0.0070	0.0030	0.0350	0.0500	N/A	
MVE-OWL-bounds	11	0.0160	0.1648	0.0020	0.2498	0.0859	0.0050	0.0080	0.0599	0.1189	0.0110	N/A

Note: this table reports the p-values of Sharpe ratio test according to Ledoit and Wolf (2008) using randomly selected 500 stocks (with daily returns) and 100 stocks (with monthly returns) from CRSP data-set.

that the OWL related strategies consistently offer smaller turnovers (transaction costs) than minVar-JM and minVar-LW strategies, and the minVar-OWL strategy consistently and significantly outperforms the equal weighted strategy in terms of Sharpe ratio. Panel C in Table 9 and 10 report the OOS scores and p-values of Sharpe ratio tests using 100 randomly selected stocks with monthly returns from the CRSP data-set. These results confirm the previous findings that OWL related strategies consistently and *significantly* outperform equal weighted strategy.

D.3 Plot of out-of-sample portfolio performance

We provide extra details on each portfolio strategy, including the sparsity and turnovers of a portfolio selection strategy.

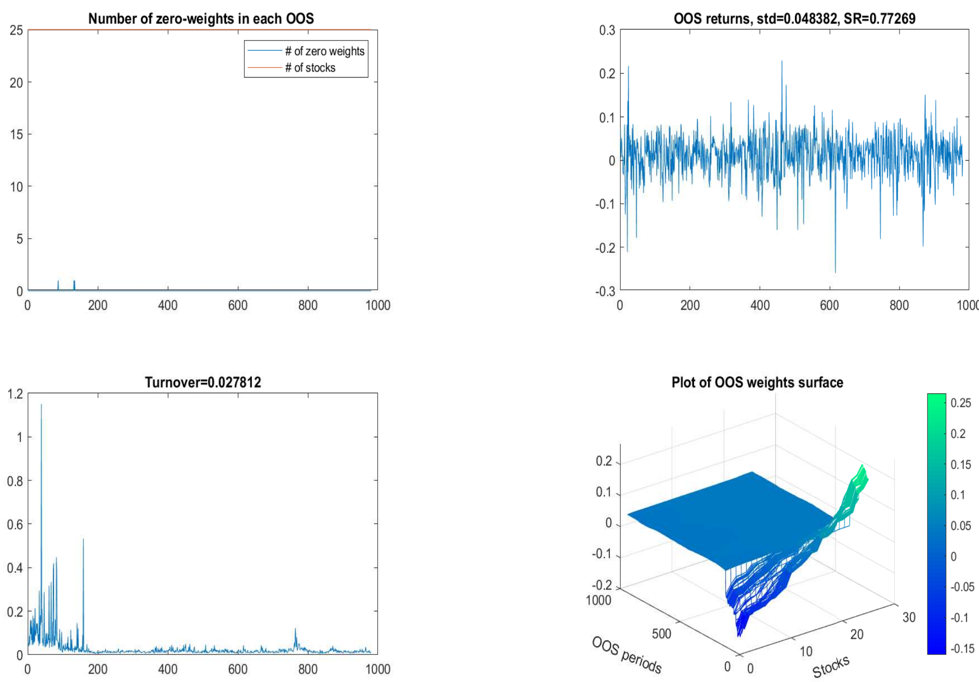


Figure 6. Sparsity, OOS returns, turnovers and weight distribution of the minVar-OWL strategy using FF25

Figures 6 to 9 provide out-of-sample portfolio performance using four portfolio selection strategies.[18] In each figure, the top-left panel displays, at each point in the OOS period,

[18]Restricted to the limited display space here, we only present 4 strategies using the FF25 portfolios as

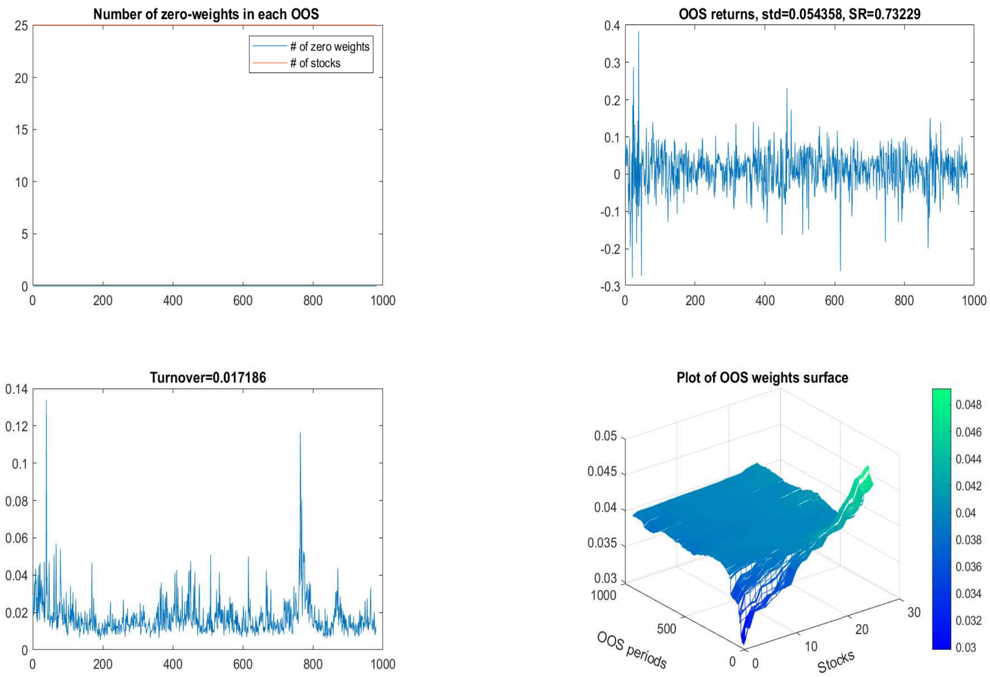


Figure 7. Sparsity, OOS returns, turnovers and weight distribution of the minVar-OWL strategy with no-short-sale constraint using FF25

the total number of assets (red) and in which, how many of them receives zero position (blue). Top-right panel displays the portfolio returns for each OOS period. Bottom-left panel displays the weight change at each rebalancing point (in the case of FF25 portfolios, we rebalance at each month during the OOS period). Bottom-right panel displays the portfolio weight distribution in a three-dimensional graph.

It is worth noting that sub-figures have different scales when comparing different figures. OWL-based portfolio strategies have much smaller turnovers than the minVar-JM and minVar-LW portfolio strategies. The unconstrained minVar-OWL portfolio strategy has relatively high turnovers in the first 160 months (1936-1949) before becoming less volatile (from the bottom-right panel, we can see the weight distribution is close to the equal-weighted portfolio strategy after 160 months). However, its magnitude of portfolio change still dwarfed by the minVar-JM and minVar-LW portfolio strategies. If further impose the

test assets. More graphs regarding out-of-sample performance using different portfolio strategies on various test assets are available upon request.

no-short-sale constraint on the OWL portfolio strategy, we can find substantial decline in turnovers, leaving it as one of the most attractive strategies of yielding minimal transaction cost.

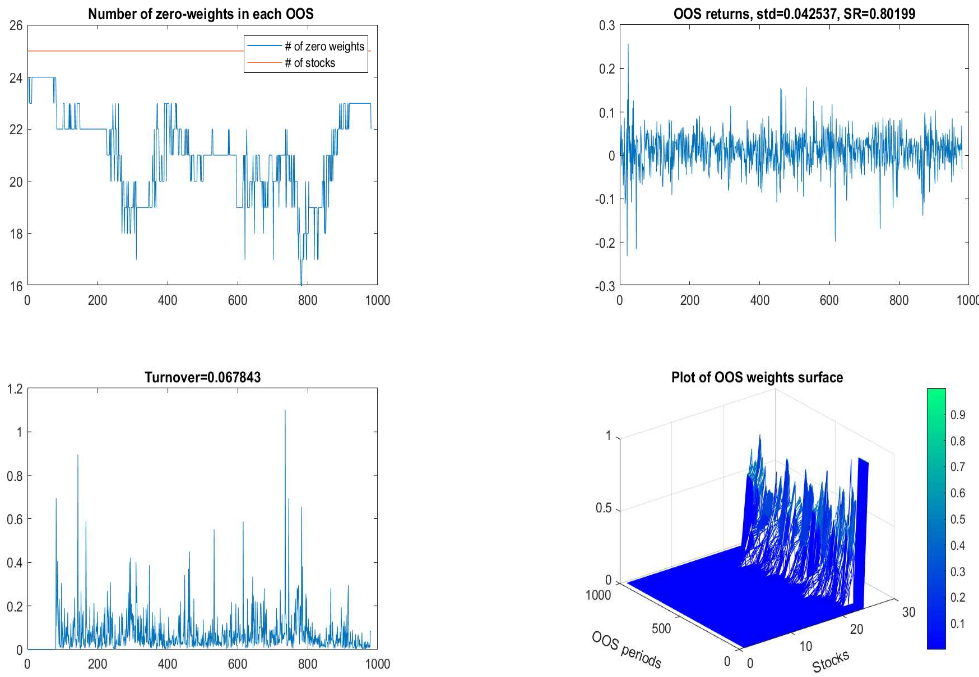


Figure 8. Sparsity, OOS returns, turnovers and weight distribution of the minVar-JM strategy using FF25

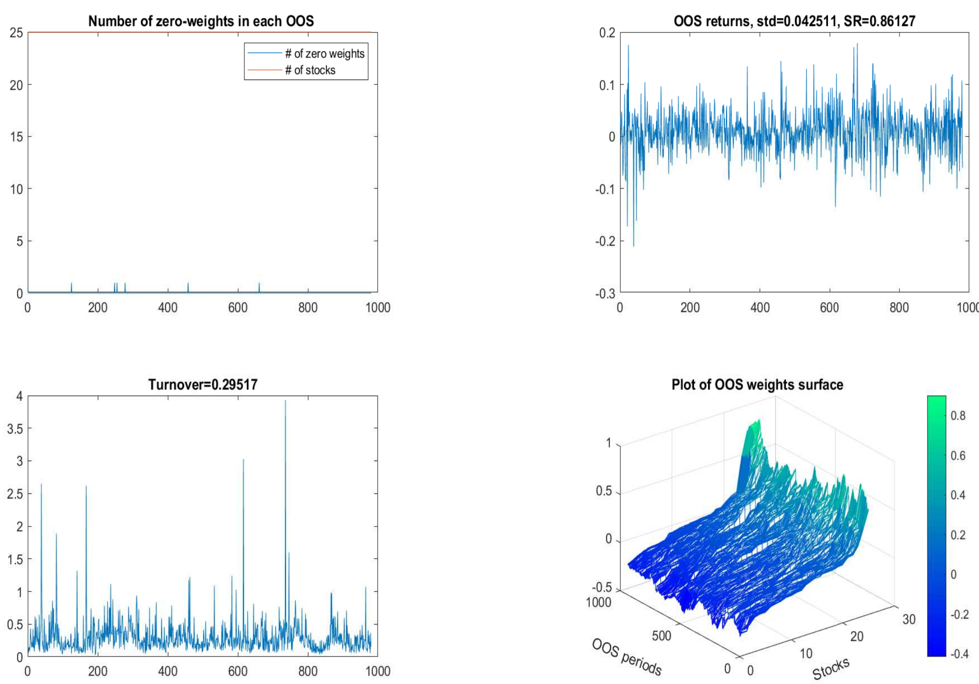


Figure 9. Sparsity, OOS returns, turnovers and weight distribution of the minVar-LW strategy using FF25

References

Ao, M., Y. Li, and X. Zheng (2019): "Approaching Mean-Variance Efficiency for Large Portfolios," *The Review of Financial Studies*, 32(7), 2890–2919.

Bickel, P. J. and E. Levina (2008): "Covariance Regularization by Thresholding," *The Annals of statistics*, 36, 2577–2604.

Bogdan, M., E. van den Berg, C. Sabatti, W. Su, and E. J. Candès (2015): "SLOPE - Adaptive Variable Selection via Convex Optimization," *Annals of Applied Statistics*, 9, 1103–1140.

Bondell, H. D. and B. J. Reich (2008): "Simultaneous Regression Shrinkage, Variable Selection, and Supervised Clustering of Predictors with OSCAR," *Biometrics*, 64, 115–123.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein (2010): "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, 3, 1–122.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019): "Sparse Signals in the Cross-Section of Returns," *Journal of Finance*, 74, 449–492.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009a): "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms," *Management Science*, 55, 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009b): "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *The Review of Financial Studies*, 22, 1915–1953.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020): "A Transaction-Cost Perspective on the Multitude of Firm Characteristics," *The Review of Financial Studies*, 33, 2180–2222.

DeMiguel, V., F. J. Nogales, and R. Uppal (2014): "Stock Return Serial Dependence and Out-of-Sample Portfolio Performance," *The Review of Financial Studies*, 27, 1031–1073.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov (2013): "Improving Portfolio Selection Using Option-Implied Volatility and Skewness," *The Journal of Financial and Quantitative Analysis*, 48, 1813–1845.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2021): "Estimation of Time-Varying Covariance Matrices for Large Datasets," *Econometric Theory*, 0, 1–35.

Epps, T. W. (1979): "Comovements in Stock Prices in the Very Short Run," *Journal of the American Statistical Association*, 74, 291–298.

Fastrich, B., S. Paterlini, and P. Winker (2015): "Constructing Optimal Sparse Portfolios Using Regularization Methods," *Computational Management Science*, 12, 417–434.

Figueiredo, M. A. T. and R. D. Nowak (2016): "Ordered Weighted L1 Regularized Regression with Strongly Correlated Covariates: Theoretical Aspects," *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 41, 930–938.

HANSEN, P. R., A. LUNDE, and J. M. NASON (2011): "The Model Confidence Set," *Econometrica*, 79, 453–497.

Jagannathan, R. and T. Ma (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraint Helps," *Journal of Finance*, 58, 1651–1684.

Kan, R. and G. Zhou (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *The Journal of Financial and Quantitative Analysis*, 42, 621–656.

Ledoit, O. and M. Wolf (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

——— (2008): "Robust Performance Hypothesis Testing with the Sharpe Ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2017): "Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks," *The Review of Financial Studies*, 30, 4349–4388.

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

Michaud, R. O. (1989): "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal*, 45, 31–42.

Sun, C. (2019): "Dissecting the Factor Zoo : A Correlation-Robust Approach," *Queen Mary University of London Working Paper.*

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.