Portfolio Selection: A Statistical Learning Approach

Yiming Peng
Northwestern University
United States
ypeng@u.northwestern.edu

Vadim Linetsky
Northwestern University
United States
vlinetsky@northwestern.edu

ABSTRACT

We propose a new portfolio optimization framework, partially egalitarian portfolio selection (PEPS). Inspired by the celebrated LASSO regression, we regularize the mean-variance portfolio optimization by adding two regularizing terms that essentially zero out portfolio weights of some of the assets in the portfolio and select and shrink the portfolio weights of the remaining assets towards the equal weights to hedge against parameter estimation risk. We solve our PEPS formulations by applying recent advances in mixed integer optimization that allow us to tackle large-scale portfolio problems. We also build a predictive regression model for expected return using two cross-sectional factors, the short-term reversal factor and the medium-term momentum factor, that are shown to be the more significant predictive factors among the hundreds of factors tested in the empirical finance literature. We then incorporate our predictive regression into PEPS by replacing the historical mean. We test our PEPS formulations against an array of classical portfolio optimization strategies on a number of datasets in the US equity markets. The PEPS portfolios enhanced with the predictive regression estimates of the expected stock returns exhibit the highest out-of-sample Sharpe ratios in all instances.

ACM Reference Format:
Yiming Peng and Vadim Linetsky. 2022. Portfolio Selection: A Statistical Learning Approach. In *3rd ACM International Conference on AI in Finance (ICAIF '22), November 2–4, 2022, New York, NY, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3533271.3561707

1 INTRODUCTION

Portfolio optimization problems involving allocation of investment funds across multiple assets are of fundamental importance in financial management and are faced daily by financial institutions, asset managers, pension plans, university endowments, insurance companies, and individual investors. The pioneering work of [32] on mean-variance portfolio optimization laid the theoretical foundations of applying optimization to the asset allocation problem. Notwithstanding the enormous importance and influence of these classical contributions, the fact remains that, from the applied point of view, portfolio optimization methods are often outperformed by the naive $1/N$ equal weights portfolio out of sample due to the difficulties in estimating assets' expected returns, volatilities and correlations from historical data. In this paper, we aim to improve out-of-sample performance of asset allocation methods by introducing a new portfolio optimization framework inspired by recent advances in machine learning and optimization.

Notations. We use bold symbols for vectors and matrices, as opposed to scalars. For example, $\boldsymbol{w}$ is a vector and w_i is the i-th element of $\boldsymbol{w}$.

2 BACKGROUND

2.1 Mean-variance Portfolio Optimization

The Markowitz mean-variance portfolio optimization model allocating funds among N assets and holding for a single time period can be formulated as the quadratic utility maximization

$$\max_{\boldsymbol{w}} \quad \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} \tag{1}$$

$$\text{s.t.} \quad \boldsymbol{w}'\mathbf{1} = 1 \tag{2}$$

with the assets' vector of expected returns $\boldsymbol{\mu}$ and the covariance matrix Σ assumed known, the investor's coefficient of risk aversion $\gamma > 0$ given, and portfolio weights subject to the budget constraint $\boldsymbol{w}'\mathbf{1} = 1$. The quadratic utility with the risk aversion coefficient γ governs the risk-return trade-off.

2.2 The Equal Weights Puzzle

The equal weights puzzle in portfolio optimization refers to the observation that the naive $1/N$ equal weights portfolio allocation rule beats sophisticated portfolio optimization approaches in many empirical applications. [15] test 14 portfolio optimization approaches across 7 distinct empirical data sets. The baseline approach is taken to be the naive $1/N$ rule that simply allocates equal fraction of the total available funds to each of the N assets without any regard to historical data. The other 13 approaches include the classical Markowitz mean-variance portfolio based on sample moments, as well as 12 descendants of Markowitz with progressively more sophistication in dealing with parameter estimation errors and other issues: Bayesian shrinkage rules ([25], [33]), portfolios with moment restrictions ([31]) and short sale-constrained portfolios ([22]). Yet, the outcome of their study is: "Of the 14 models we evaluate across seven empirical data sets, none is consistently better than the $1/N$ rule in terms of Sharpe ratio, certainty-equivalent return, or turnover, which indicate that, out of sample, the gain from optimal diversification is more than offset by estimation error." The question we are interested in is: can anything reliably beat the naive $1/N$ portfolio rule out of sample?

While the discovery of the equal weights puzzle in portfolio optimization is relatively recent, it has a direct counterpart in the well-known equal weights puzzle in the forecast combinations literature in econometrics and statistics. Suppose that we are interested

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICAIF '22, November 2–4, 2022, New York, NY, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9376-8/22/11…$15.00
https://doi.org/10.1145/3533271.3561707

in predicting some economic quantity y given a set of N competing unbiased forecasts f^i, $i = 1, ..., N$. In the economic forecast combinations literature, these forecasts are either outcomes of competing econometric forecasting models or come from surveys of professional forecasters, in which case the observer does not know the data and the model each forecaster uses to form their forecast. It is also common to combine econometric models and expert forecasts. One possibility is to choose one of the forecasts and zero out all others. However, in reality forecasting models are likely to be misspecified so a single model will unlikely always dominate all the others. Even if a dominant model exists, one is unlikely to identify it. The same applies to trying to select expert forecasters. A more reasonable approach is to combine forecasts to gain from diversification in a portfolio of forecasts. The seminal work of [3] considers a (linearly) combined forecast $C = \boldsymbol{w}'\boldsymbol{f}$ that minimizes the mean squared error $\boldsymbol{w}'\Sigma\boldsymbol{w}$ with the weights vector satisfying $\boldsymbol{w}'\mathbf{1} = 1$. Here Σ is the covariance matrix of the forecast errors. The optimal weights is

$$\boldsymbol{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma\mathbf{1}}. \tag{3}$$

When the weights are estimated from historical time series data of forecasts f_t^i and observed targets y_t, $t = 1, ..., T$, the optimal weight vector is also the coefficient vector of the OLS regression of $\boldsymbol{y}$ onto the forecasts f_i subject to the coefficients adding up to one and no intercept ([19])

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & ||\boldsymbol{y} - \boldsymbol{f}\boldsymbol{w}||_2^2 \\ \text{s.t} \quad & \boldsymbol{w}'\mathbf{1} = 1, \end{aligned} \tag{4}$$

where $\boldsymbol{y}$ is the T-dimensional vector of (y_t) and $\boldsymbol{f}$ is the $T \times N$ matrix of (f_t^i).

Equal weights $w_i = 1/N$ are generally suboptimal, unless Σ satisfies a highly restrictive necessary and sufficient condition for the optimality of equal weights that the vector of ones $\mathbf{1}$ is an eigenvector of Σ ([18]). Yet, a large literature on empirical performance of forecast combinations developed over fifty years since [3] finds that simple equal weights that do not require any knowledge of the covariance matrix frequently outperform optimal weights out of sample. This is the well-known equal weights puzzle in the forecast combinations literature. Surveys and discussions can be found in [9], [36].

2.3 PELASSO

Against this backdrop in the forecast combinations literature, [17] recently propose a forecast combining algorithm inspired by LASSO [37]. First, the equal weights puzzle suggests that equal weights averaging is a natural shrinkage direction, blending data (likelihood) information with prior information. This amounts to a Bayesian approach with the prior centered on equal weights averaging, going back to [16]. The selection issue arises when combining large numbers of forecast, as some forecasts may be largely redundant and not worth including in the combination, going back to the best subset averaging of Elliott (2011). LASSO accomplishes both selection and shrinkage. However, while LASSO selects to zero, it also shrinks to zero, while given the equal weights puzzle, we would like to shrink to equal weights instead. The novel formulation proposed by [17] is the partially egalitarian LASSO (PELASSO) that selects some weights to zero and selects and shrinks the surviving weights to equality ([17] do not impose the additional constraint that the weights add up to one in their formulation)

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}||\boldsymbol{y} - \boldsymbol{f}\boldsymbol{w}||_2^2 + \lambda_1||\boldsymbol{w}||_1 + \lambda_2\left\|\boldsymbol{w} - ||\boldsymbol{w}||_0^{-1}\mathbf{1}\right\|_1. \tag{5}$$

[3] becomes a special case which solve the penalty-free formulation with $\lambda_1 = \lambda_2 = 0$. The penalty term includes the standard ℓ_1 LASSO penalty that selects to zero, as well as the second ℓ_1 penalty that selects and shrinks to equal weights due to subtracting $||\boldsymbol{w}||_0^{-1}\mathbf{1}$, where $||\boldsymbol{w}||_0$ is the ℓ_0 norm (pseudo-norm in fact) counting the number of non-zero elements in $\boldsymbol{w}$. Thus the name *partially egalitarian* – select some of the weights to zero and select and shrink the surviving non-zero weights towards equality.

However, the PELASSO formulation is nonconvex and nonsmooth due to the presence of the ℓ_0 norm that enters in the denominator. For computational tractability [17] instead uses a two-step formulation that first imposes only the standard LASSO penalty ($\lambda_1 > 0$ and $\lambda_2 = 0$), then takes the forecasts that survived with nonzero weights in the first step and solves the egalitarian LASSO (eLASSO) formulation only for those surviving forecasts ($\lambda_1 = 0$ and $\lambda_2 > 0$) to determine the weights with selection and shrinkage towards equality. After the first step the number of nonzero weights $n < N$ is known, and the eLASSO penalty in the second step, which is now a problem of size n, is thus simply $||\boldsymbol{w} - n^{-1}\mathbf{1}||_1$, which is the shifted LASSO. Therefore this two-step approximation can be solved as efficiently as LASSO.

3 PEPS

While the forecast combinations literature and the portfolio optimization literature evolved virtually independently over the years with little overlap and essentially no cross referencing, by connecting quadratic portfolio optimization formulations with regressions we can view the equal weights puzzle in portfolio optimization and the equal weights puzzle in forecast combinations as one and the same. While a number of works exploit the parallels between portfolio optimization formulations and linear regressions, namely [6], [7], [15] and a number of subsequent studies, the parallel between the equal weights puzzle in forecast combinations and portfolio optimization has not been explicitly emphasized in the literature. This view then leads us to applying PELASSO of [17] and its variants to portfolio optimization. We propose a new portfolio optimization framework which we call *partially egalitarian portfolio selection* (PEPS)

PEPS (1).

$$\min_{\boldsymbol{w}} \quad \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu} + \lambda_1||\boldsymbol{w}||_1 + \lambda_2\left\|\boldsymbol{w} - ||\boldsymbol{w}||_0^{-1}\mathbf{1}\right\|_1 \tag{6}$$

$$\text{s.t.} \quad \boldsymbol{w}'\mathbf{1} = 1, \tag{7}$$

$$(w_i \geq 0, \quad i = 1, ..., N). \tag{8}$$

with the PELASSO penalty that selects some of the portfolio weights to zero and selects some of the surviving weights to equal weights and shrinks the others in the direction of equal weights. We note that in this formulation the risk aversion parameter γ can be treated as another hyperparameter, along with hyperparameters λ_i, to be tuned from the data by maximizing out-of-sample Sharpe ratio.

The shortsale constraint (8) is optional. However, when included, it deserves special attention. As noted in [14], (8) is equivalent to a

1-norm constraint

$$||\boldsymbol{w}||_1 \leq 1 \tag{9}$$

therefore making the penalty term $\lambda_1||\boldsymbol{w}||_1$ in (6) redundant. As a result, whenever we impose the shortsale constraint (8), we will implicitly drop $\lambda_1||\boldsymbol{w}||_1$ in the objective function (6).

Some special cases of PEPS include: (1) $\lambda_1 = 0$, and $\lambda_2 \to \infty$ recovers the $1/N$ portfolio if all weights are nonzero; (2) $\lambda_1 = 0, \lambda_2 = 0$ recovers the Markowitz mean-variance portfolio. Thus PEPS can be viewed as a portfolio which interpolates continuously the mean-variance portfolio and the $1/N$ portfolio.

Another special case is when expected return is dropped in (6)

GMV-PEPS.

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} + \lambda_1||\boldsymbol{w}||_1 + \lambda_2\left\|\boldsymbol{w} - ||\boldsymbol{w}||_0^{-1}\mathbf{1}\right\|_1 \tag{10}$$

$$\text{s.t.} \quad \boldsymbol{w}'\mathbf{1} = 1, \tag{11}$$

$$(w_i \geq 0, \quad i = 1, ..., N). \tag{12}$$

which leads to the global minimum variance (GMV) portfolio corresponding to infinite risk aversion, penalized by PELASSO (GMV-PEPS). We observe that the unpenalized GMV portfolio weights are identical to the Bates and Granger weights in the forecast combinations problem, and the GMV-PEPS formulation is exactly the same as in [17].

We note that in portfolio selection applications assets are often grouped in groups based on some attributes (e.g. companies in industry sectors, portfolio managers in management styles, etc.). Such group structures are also highly amenable to penalized formulations and cardinality constrained formulations. In particular we mention group LASSO ([38]) and sparse group LASSO ([35]). Based on these ideas PEPS formulations can be developed with selection and shrinkage to equal averaging at group and/or component levels, depending on the application at hand. For example, a group PEPS model can be formulated by replacing (6) with

Group PEPS.

$$\min_{\boldsymbol{w}} \quad \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu} + \lambda_1\sum_{j=1}^{J}||\boldsymbol{w}^{(j)}||_2 + \lambda_2\left\|\boldsymbol{w} - ||\boldsymbol{w}||_0^{-1}\mathbf{1}\right\|_1 \tag{13}$$

where $\boldsymbol{w} = (\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, ..., \boldsymbol{w}^{(J)})$ is divided into J groups.

4 COMPUTATION

We observe that the PELASSO penalty transforms PEPS into a nonconvex and nonsmooth optimization problem that becomes NP-hard. Following [17], we first develop an approximate, two-step method for PEPS before presenting our one-step solution method:

Step 1 ("Select to zero"):

$$\tilde{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}}\left\{\frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu} + \lambda_1||\boldsymbol{w}||_1\right\}, \tag{14}$$

$$n := ||\tilde{\boldsymbol{w}}||_0, \tag{15}$$

$$\tilde{\boldsymbol{\mu}} = \text{sub-vector of } \boldsymbol{\mu} \text{ corresponding to } \tilde{\boldsymbol{w}}, \tag{16}$$

$$\tilde{\Sigma} = \text{sub-matrix of } \Sigma \text{ corresponding to } \tilde{\boldsymbol{w}}, \tag{17}$$

Step 2 ("Shrink towards equal weights")

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}}\left\{\frac{\gamma}{2}\boldsymbol{w}'\tilde{\Sigma}\boldsymbol{w} - \boldsymbol{w}'\tilde{\boldsymbol{\mu}} + \lambda_2\left\|\boldsymbol{w} - n^{-1}\mathbf{1}\right\|_1\right\}, \tag{18}$$

where each step is subject to the constraints. Since each step solves a convex optimization problem, the two-step method is readily scalable to large portfolio problems. However, it is guaranteed to be sub-optimal.

To develop an exact, one-step solution method for PEPS, we pursue a Mixed Integer Optimization (MIO) approach. Following [5] and [4], we first present a mixed integer quadratically constrained programming (MIQCP) formulation

PEPS (2).

$$\min_{\boldsymbol{w},z,y} \quad \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu} + \lambda_1||\boldsymbol{w}||_1 + \lambda_2\left\|\boldsymbol{w} - y\mathbf{1}\right\|_1 \tag{19}$$

$$\text{s.t.} \quad -\mathcal{M}z_i \leq w_i \leq \mathcal{M}z_i, \quad i = 1, ..., N, \tag{20}$$

$$\sum_{i=1}^{N} yz_i = 1, \tag{21}$$

$$z_i \in \{0, 1\}, \quad i = 1, ..., N, \tag{22}$$

$$\boldsymbol{w}'\mathbf{1} = 1, \tag{23}$$

$$(w_i \geq 0, \quad i = 1, ..., N) \tag{24}$$

where $\mathcal{M}$ is a constant such that if $\hat{\boldsymbol{w}}$ is a minimizer of PEPS (1), then $\mathcal{M} \geq ||\hat{\boldsymbol{w}}||_\infty$. Provided that $\mathcal{M}$ is chosen to be sufficiently large with $\mathcal{M} \geq ||\hat{\boldsymbol{w}}||_\infty$, a minimizer of PEPS (2) will be a minimizer of PEPS (1). Although $\mathcal{M}$ is not known a priori and its choice will affect the strength of the formulation, in practice appropriate values for $\mathcal{M}$ often arise naturally in the context portfolio optimization. For example, the budget constraint implies naturally that $\mathcal{M} = 1$ is a suitable choice.

Moreover we can rewrite the MIQCP in the form which is more easily handled by MIO solvers

PEPS (3).

$$\min_{\boldsymbol{w},z,y,\boldsymbol{u},\boldsymbol{v},s,t} \quad \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\mu} + \lambda_1 s + \lambda_2 t \tag{25}$$

$$\text{s.t.} \quad -\mathcal{M}z_i \leq w_i \leq \mathcal{M}z_i, \quad i = 1, ..., N, \tag{26}$$

$$\sum_{i=1}^{N} yz_i = 1, \tag{27}$$

$$-u_i \leq w_i \leq u_i, \quad i = 1, ..., N \tag{28}$$

$$\sum_{1}^{N} u_i = s, \tag{29}$$

$$-v_i \leq w_i - y \leq v_i, \quad i = 1, ..., N \tag{30}$$

$$\sum_{1}^{N} v_i = t, \tag{31}$$

$$u_i \geq 0, \quad i = 1, ..., N \tag{32}$$

$$v_i \geq 0, \quad i = 1, ..., N \tag{33}$$

$$z_i \in \{0, 1\}, \quad i = 1, ..., N, \tag{34}$$

$$\boldsymbol{w}'\mathbf{1} = 1, \tag{35}$$

$$(w_i \geq 0, \quad i = 1, ..., N) \tag{36}$$

We employ Gurobi ([20]) version 9.0 for the solver, and we emphasize that Gurobi 9.0 features a novel bilinear solver that enables users to solve problems with nonconvex quadratic objectives and constraints, which were previously unavailable.

It is worth noting that $||w||_0 = \sum_1^N z_i$ is essential for the MIQCP formulation to be exact. This requirement, however, is not guaranteed for each optimal solution returned by Gurobi. In reality, Gurobi may deliver a (slightly) suboptimal solution in cases when this is not true. To solve this, we develop a 1-opt local search algorithm (Algorithm 1) that improves the solution returned by Gurobi. In step 3 of Algorithm 1, we examine each of the ties in random order until an improved solution is discovered.

Algorithm 1: 1-Opt Local Search

Result: (Improved) PEPS solution $\mathbf{w}^*$

1 $\mathbf{w}^* \leftarrow$ PEPS solution returned by solver (e.g. Gurobi);
2 $V^* \leftarrow$ PEPS objective value associated with $\mathbf{w}^*$;
3 $\hat{w} \leftarrow$ smallest nonzero element of $\mathbf{w}^*$ in absolute value;
4 $\tilde{\mathbf{w}} \leftarrow \mathbf{w}^*$ except $\hat{w}$ is set to zero;
5 $\tilde{\boldsymbol{\mu}} \leftarrow$ sub-vector of $\boldsymbol{\mu}$ corresponding to $\tilde{\mathbf{w}}$;
6 $\tilde{\Sigma} \leftarrow$ sub-matrix of Σ corresponding to $\tilde{\mathbf{w}}$;
7 $\mathbf{w}^{\mathbf{OPT}} \leftarrow \arg\min_{\boldsymbol{w}} \left\{ \frac{\gamma}{2} \boldsymbol{w}' \tilde{\Sigma} \boldsymbol{w} - \boldsymbol{w}' \tilde{\boldsymbol{\mu}} + \lambda_2 \left\| \boldsymbol{w} - ||\tilde{\mathbf{w}}||_0^{-1} \mathbf{1} \right\|_1 \right\}$,
subject to the original constraints;
8 $V^{OPT} \leftarrow$ PEPS objective value associated with $\mathbf{w}^{\mathbf{OPT}}$;
9 **if** $V^{OPT} < V^*$ **then**
10 $\quad \mathbf{w}^* \leftarrow \mathbf{w}^{\mathbf{OPT}}$, $V^* \leftarrow V^{OPT}$, and go to step 3;
11 **else**
12 $\quad$ **return** $\mathbf{w}^*$;
13 **end**

5 PREDICTIVE EXPECTED RETURN MODEL

The majority of the portfolio optimization literature do not make any attempts to deploy any predictive model for expected return and use the classical historical mean. We study whether having a more sophisticated predictive model for expected returns than the historical mean can result in further performance improvements in PEPS portfolio strategies. We are interested in understanding the capability of PEPS in terms of capturing and harvesting monthly predictability, or the so-called anomalies, in the cross section of stock returns.

The empirical asset pricing literature finds significant cross-sectional predictability in expected stock returns. Many cross-sectional relationships between firm characteristics and future stock returns have been discovered in academic finance research. Size, value, momentum, and investment are all identified by researchers to contain predictive power for future stock returns. A good number of the documented predictors or factors are highly significant and robust, indicating that they are unlikely to be consequences of data mining ([28]).

In this paper, we focus on two of the most prominent cross-sectional factors among the hundreds documented in the empirical asset pricing literature ([10]).

The first factor is the short-term reversal factor, or reversal for short, documented by [23] and [27]. The short-term reversal factor is both one of the strongest and one of the most straightforward phenomena documented in the empirical asset pricing literature ([1]), observing that top performers in a given month often tend to underperform in the subsequent month and vice versa. In this chapter, the (monthly) reversal factor of stock i for month t is measured as the return of the stock during the month t

$$Rev_{i,t} = R_{i,t}, \tag{37}$$

where $Rev_{i,t}$ denotes reversal and $R_{i,t}$ is the return of stock i in month t.

The second factor is the medium-term momentum factor, or momentum for short, documented by [24]. [2], [12] and [21] develop behavioral models in which the momentum phenomenon emerges as a consequence of investors' delayed response and overreaction to information. [11], [29], and [30] provide rational models in which the momentum phenomenon is a manifestation of cross-sectionally persistent expected stock returns. In this chapter, the (monthly) momentum of stock i measured at the end of month t is the return of the stock during the 11-month period from months $t-11$ to $t-1$

$$Mom_{i,t} = \prod_{s=t-11}^{t-1} (1 + R_{i,s}) - 1 \tag{38}$$

where $Mom_{i,t}$ denotes momentum and $R_{i,s}$ is the return of stock s in month t. We refer to the monograph [1] for more details and references on the short-term reversal and mid-term momentum factors. Here we note that these two factors are by far the most significant of all the cross-sectional factors examined in [1], thus motivating our choice of these two factors to include in our predictive regression for the expected return.

We build a predictive regression model for expected return utilizing these two cross-sectional factors

$$r_{i,t+1} = \alpha + \beta_1 Mom_{i,t} + \beta_2 Rev_{i,t} + \epsilon_{i,t+1} \tag{39}$$

where $r_{i,t+1}$ is the excess return of the stock i in month $t+1$. For month t, we estimate the predictive regression from the panel data via pooled OLS. Since empirical asset pricing literature suggests that β_1 should be positive and β_2 should be negative, we apply these constraints to the predictive regression (3.3). [8] argue that predictive regressions can outperform the historical average out-of-sample if one imposes some economically meaningful restrictions on the parameter space. Their insights are to use finance theory to reduce the number of parameters that must be freely estimated from the data and to restrict the estimated equity premium to a reasonable range. They require the estimated slope coefficient β to have the same sign as the theoretically expected slope coefficient estimated over the full sample and is implemented by setting the estimated slope coefficient to zero when its sign in inconsistent with the theory. In our setting, we set $\hat{\beta}_1$ to zero if $\hat{\beta}_1 < 0$. Similarly we set $\hat{\beta}_2$ to zero if $\hat{\beta}_2 > 0$. After applying these constraints to the estimated parameters, we use our model to predict expected return in the next month.

6 EMPIRICAL ANALYSES

Following [15], we conduct an empirical evaluation of the out-of-sample performance of our proposed PEPS framework and compare it to classical portfolio optimization strategies using monthly data from the US equities market.

We obtain monthly stock returns from the Center for Research in Security Prices (CRSP) for all firms listed in the NYSE, AMEX, and NASDAQ. Our sample begins in Januray 1990 and ends in December

2020. We also obtain the Treasury-bill rate from Kenneth French's website to proxy for the risk-free rate from which we calculate individual excess returns.

We construct 5 datasets, or trading universes, which are listed in Table 1. They are the top N most liquid stocks in CRSP for $N = 50, 100, 200, 500, 1000$. For a given value of N and each month t, we rank stocks in decreasing order by their average trading volume (measured in dollars traded) during the previous 12 months, then select the top N stocks in the dataset of the top N most liquid stocks.

In each dataset, we compute the out-of-sample Sharpe ratio of the strategies using a time series of monthly out-of-sample returns generated by each strategy. The out-of-sample Sharpe ratio of a strategy is defined as the sample mean of out-of-sample excess returns (over the risk-free interest rate), $\hat{\mu}$, divided by their sample standard deviation, $\hat{\sigma}$

$$SR = \frac{\hat{\mu}}{\hat{\sigma}}. \tag{40}$$

Our empirical studies are conducted using a rolling-window design where we maintain a rolling window of length 60 months. Using data from the prior 60 months, we estimate the parameters necessary to implement a given portfolio in month t. The parameters estimated are used to construct portfolio weights, which are then utilized to compute the month $t + 1$ return. Besides using the historical mean, we estimate expected return by the predictive regression model (34). For covariance matrix, we apply the [26] covariance matrix to all the CRSP portfolios since regularization is crucial when dealing with large N.

For hyperparameter tuning in PEPS portfolios, we apply time series cross validation with a rolling validation window of length 36 months. The configuration of hyperparameters which generates the highest validation Sharpe ratio is chosen by the time series cross validation procedure. Due to the random noise created by calculating the Sharpe ratio using sample moments, we instead choose hyperparameters via a more robust method based on bootstrap. We first use the bootstrap to generate a 95% confidence interval for the Sharpe ratio in the validation window, then we rank these confidence intervals by their lower limits in descending order. Finally we choose the hyperparamter configuration corresponding to the largest lower limit in these confidence interval. Given that volatility is known to be highly persistent, a time series bootstrap is used, combining the stationary bootstrap of [34] with an average block length of 12 months and the percentile method ([13]). The risk aversion coefficient, γ, is set to 5.

We evaluate 15 portfolio strategies (Table 2) in total. The results are presented in Table 3. We make the following observations.

- The unconstrained mean-variance optimization performs poorly for all data sets, having low and even negative Sharpe ratios. Adding the no-short-sales constraints helps the mean-variance portfolios across the board, but the constrained mean-variance portfolios still underperform the naive $1/N$ portfolios in all instances.
- As expected, the PEPS-1 portfolios outperform the naive $1/N$ portfolios and the mean-variance portfolios, as well as the constrained mean-variance portfolios in all instances.
- The unconstrained global minimum variance (GMV) portfolios do better than unconstrained mean-variance portfolios, but underperform the $1/N$ portfolios in all five instances. The PEPS-1 regularization of the GMV portfolios outperform the original GMV portfolios in all instances.
- Adding the no-short-sales constraints to the GMV portfolios improves their performance, with performance close to the $1/N$ portfolios. The PEPS-1 regularization of the GMV-c portfolios outperforms the original GMV-c portfolios in all instances and, in fact, is the best performing portfolio optimization method among all the methods considered in this table and in all instances. It also beats the $1/N$ portfolios in all instances. The historical mean appears to be unhelpful to portfolio optimization in these empirical tests, while the knowledge of the historical covariance matrix harvested through the use of PEPS-1 combined with short sales constraints adds significant value to portfolio optimization.
- The mean-variance portfolio employing the momentum-reversal predictive regression (model) even without no-short-sales constraint outperforms the $1/N$ portfolio in all five instances. Moreover, it drastically outperforms the mean-variance portfolios (both unconstrained and constrained) based on the historical mean. The predictive regression for the mean significantly contributes to the portfolio performance.
- Adding no-short-sales constraint further improves the mean-variance portfolios with the model mean.
- Unconstrained PEPS-1 portfolios based on the model mean outperform both the PEPS-1 portfolios based on the historical mean considered in the previous chapter, as well as the mean-variance portfolios based on the model mean.
- PEPS-1 portfolios with no-short-sales constraint have the best performance in all five instances.
- Finally, the 2-step PEPS-2 approximations based on the approach of [17] are suboptimal relative to the exact PEPS-1 in all instances.

7 CONCLUSION

In this paper, we develop a new machine learning model, Partially Egalitarian Portfolio Selection, for portfolio optimization. We regularize the mean-variance portfolio optimization by adding two regularizing terms that essentially zero out portfolio weights of some of the assets in the portfolio and select and shrink the portfolio weights of the remaining assets towards the equal weights to hedge against parameter estimation risk. This portfolio optimization framework subsumes the classical mean-variance and the equal weights portfolios as limiting cases. We solve our PEPS formulations by applying recent advances in mixed integer optimization that allows us to tackle large-scale portfolio problems. We apply PEPS to the cross section of expected stock returns by developing a predictive regression model for expected return based on the momentum and reversal factors. We combine the expected return model with the PEPS portfolio optimization methodology and test them on empirical datasets. We demonstrate that the PEPS methodology yields significant out-of-sample portfolio performance gains when combined with a predictive model for the expected return.

Table 1: List of datasets

No.	Dataset	Source	N	Time period	Abbreviation
1	Top 50 most liquid stocks	CRSP	50	01/1990-12/2020	CRSP50
2	Top 100 most liquid stocks	CRSP	100	01/1990-12/2020	CRSP100
3	Top 200 most liquid stocks	CRSP	200	01/1990-12/2020	CRSP200
4	Top 500 most liquid stocks	CRSP	500	01/1990-12/2020	CRSP500
5	Top 1000 most liquid stocks	CRSP	1000	01/1990-12/2020	CRSP1000

Table 2: List of portfolio strategies

No.	Strategy Description	Abbreviation
1	Equal weight portfolio	$1/N$
2	Mean-variance portfolio (historical mean)	MV (hist)
3	Mean-variance portfolio (predictive regression model)	MV (model)
4	Mean-variance portfolio, shortsale-constrained (historical mean)	MV-c (hist)
5	Mean-variance portfolio, shortsale-constrained (predictive regression model)	MV-c (model)
6	Global minimum variance portfolio	GMV
7	Global minimum variance portfolio, shortsale-constrained	GMV-c
8	PEPS, one-step (historical mean)	PEPS-1 (hist)
9	PEPS, one-step (predictive regression model)	PEPS-1 (model)
10	PEPS, two-step (historical mean)	PEPS-2 (hist)
11	PEPS, two-step (predictive regression model)	PEPS-2 (model)
12	PEPS, one-step, shortsale-constrained (historical mean)	PEPS-1-c (hist)
13	PEPS, one-step, shortsale-constrained (predictive regression model)	PEPS-1-c (model)
14	PEPS, two-step, shortsale-constrained (historical mean)	PEPS-2-c (hist)
15	PEPS, two-step, shortsale-constrained (predictive regression model)	PEPS-2-c (model)

Table 3: Sharpe ratios of portfolio strategies

No.	Strategy	CRSP50	CRSP100	CRSP200	CRSP500	CRSP1000
1	$1/N$	0.3452	0.3534	0.4070	0.4841	0.5344
2	MV (hist)	0.0625	-0.0387	0.1796	0.0277	0.0945
3	MV (model)	0.3733	0.3881	0.4631	0.5081	0.5488
4	MV-c (hist)	0.1579	0.3382	0.3709	0.4667	0.4985
5	MV-c (model)	0.4523	0.4693	0.4927	0.5297	0.5801
6	GMV	0.3289	0.3302	0.3842	0.4553	0.4770
7	GMV-c	0.3453	0.3499	0.4072	0.4859	0.5020
8	PEPS-1 (hist)	0.3487	0.3552	0.4079	0.4841	0.5389
9	PEPS-1 (model)	0.4791	0.4983	0.5186	0.5807	0.6159
10	PEPS-2 (hist)	0.3452	0.3512	0.4070	0.4832	0.5193
11	PEPS-2 (model)	0.4635	0.4746	0.4984	0.5344	0.5882
12	PEPS-1-c (hist)	0.2170	0.3190	0.4667	0.5266	0.5577
13	PEPS-1-c (model)	0.4990	0.5126	0.5454	0.6034	0.6391
14	PEPS-2-c (hist)	0.2076	0.2739	0.3624	0.4603	0.5223
15	PEPS-2-c (model)	0.4733	0.4885	0.5090	0.5520	0.6138

ACKNOWLEDGMENTS

This research was supported by the grant from the National Science Foundation CMMI-1916616.

REFERENCES

[1] Turan G Bali, Robert F Engle, and Scott Murray. 2016. *Empirical asset pricing: The cross section of stock returns.* John Wiley & Sons.
[2] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49, 3 (1998), 307–343.
[3] John M Bates and Clive WJ Granger. 1969. The combination of forecasts. *Journal of the Operational Research Society* 20, 4 (1969), 451–468.
[4] Dimitris Bertsimas and Ryan Cory-Wright. 2018. A scalable algorithm for sparse and robust portfolios. *arXiv preprint arXiv:1811.00138* (2018).
[5] Dimitris Bertsimas, Angela King, and Rahul Mazumder. 2016. Best subset selection via a modern optimization lens. *Annals of Statistics* (2016), 813–852.
[6] Mark Britten-Jones. 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* 54, 2 (1999), 655–671.
[7] Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. 2009. Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences* 106, 30 (2009), 12267–12272.
[8] John Y Campbell and Samuel B Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 4 (2008), 1509–1531.
[9] Robert T Clemen. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5, 4 (1989), 559–583.
[10] John H Cochrane. 2011. Presidential address: Discount rates. *Journal of finance* 66, 4 (2011), 1047–1108.
[11] Jennifer Conrad and Gautam Kaul. 1989. Mean reversion in short-horizon expected returns. *Review of Financial Studies* 2, 2 (1989), 225–240.
[12] Kent Daniel, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investor psychology and security market under-and overreactions. *Journal of Finance* 53, 6 (1998), 1839–1885.
[13] Anthony Christopher Davison and David Victor Hinkley. 1997. *Bootstrap methods and their application.* Cambridge university press.
[14] Victor DeMiguel, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal. 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55, 5 (2009), 798–812.
[15] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The review of Financial studies* 22, 5 (2009), 1915–1953.
[16] Francis X Diebold and Peter Pauly. 1990. The use of prior information in forecast combination. *International Journal of Forecasting* 6, 4 (1990), 503–508.
[17] Francis X Diebold and Minchul Shin. 2019. Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting* 35, 4 (2019), 1679–1691.
[18] Graham Elliott. 2011. Averaging and the optimal combination of forecasts. (2011).
[19] Clive WJ Granger and Ramu Ramanathan. 1984. Improved methods of combining forecasts. *Journal of Forecasting* 3, 2 (1984), 197–204.
[20] Incorporate Gurobi Optimization. 2020. Gurobi optimizer reference manual. *URL http://www. gurobi. com* (2020).
[21] Harrison Hong and Jeremy C Stein. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of finance* 54, 6 (1999), 2143–2184.
[22] Ravi Jagannathan and Tongshu Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* 58, 4 (2003), 1651–1683.
[23] Narasimhan Jegadeesh. 1990. Evidence of predictable behavior of security returns. *Journal of finance* 45, 3 (1990), 881–898.
[24] Narasimhan Jegadeesh and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of finance* 48, 1 (1993), 65–91.
[25] Philippe Jorion. 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis* 21, 3 (1986), 279–292.
[26] Olivier Ledoit and Michael Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88 (2004), 365–411.
[27] Bruce N Lehmann. 1990. Fads, martingales, and market efficiency. *Quarterly Journal of Economics* 105, 1 (1990), 1–28.
[28] Jonathan Lewellen et al. 2015. The Cross-section of Expected Stock Returns. *Critical Finance Review* 4, 1 (2015), 1–44.
[29] Andrew W Lo and A Craig MacKinlay. 1990. An econometric analysis of non-synchronous trading. *Journal of Econometrics* 45, 1-2 (1990), 181–211.
[30] Andrew W Lo and A Craig MacKinlay. 1990. When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3, 2 (1990), 175–205.
[31] A Craig MacKinlay and L'uboš Pástor. 2000. Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Review of Financial Studies* (2000), 883–916.
[32] Harry M Markowitz. 1952. Portfolio selection. *Journal of Finance* 7 (1952), 77–91.
[33] L'uboš Pástor and Robert F Stambaugh. 2000. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* 56, 3 (2000), 335–381.
[34] Dimitris N Politis and Joseph P Romano. 1994. The stationary bootstrap. *Journal of the American Statistical association* 89, 428 (1994), 1303–1313.
[35] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2013. A sparse-group lasso. *Journal of Computational and Graphical Statistics* 22, 2 (2013), 231–245.
[36] Jeremy Smith and Kenneth F Wallis. 2009. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* 71, 3 (2009), 331–355.
[37] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288.
[38] Ming Yuan and Yi Lin. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 1 (2006), 49–67.