INTEGRATION OF SUPPORT VECTOR MACHINES AND MEAN-VARIANCE OPTIMIZATION FOR CAPITAL ALLOCATION

Abstract

This work introduces a novel methodology for portfolio optimization that uses support vector machines (SVM) within the context of mean-variance optimization. This work builds on the classical approach to portfolio optimization pioneered by Markowitz (1952) where the representative investor, in addition to optimizing a risk-return trade-off, also requires that the assets eligible for investment satisfy an eligibility criterion modelled by a separating hyper-plane with maximum margin. However, the mean-variance model with SVM-type constraints is a challenging mixed binary quadratic program. To address this, a partial alternating direction method algorithm is proposed that maintains high solution quality and requires less than half as much time as a standard commercial branch and bound solver. The proposed approach, motivated by the added benefits of portfolio constraints and investor behaviours, is shown to be capable of delivering improved out-of-sample risk-adjusted returns on a universe of 286 stocks and provides an insightful interpretation of investor behaviours.

1 Introduction

The problem of capital allocation is top of mind for professionals and researchers alike and rose to enormous popularity due to the work of Markowitz (1952). The fundamental idea is to select a portfolio to optimize a trade-off between risk and expected return. This idea has formed a pillar of modern portfolio theory. However, it has been observed in practice that the mean-variance optimal (MVO) solutions exhibit sensitivity to the estimated input parameters and have high transaction costs. Furthermore, in its standard form, mean-variance optimization takes little consideration of the process through which investment is made as it only relies on estimated expected returns and covariance matrices.

Many approaches have gained popularity in the literature to immunize against estimation error. One approach is to solve robust optimization models of the classic mean-variance optimization problem whereby the parameters belong to an i) uncertainty set, ii) distribution or iii) an uncertain distribution within an ambiguity set; see (Blanchet et al., 2018; Goldfarb and Iyengar, 2003; Lai et al., 2011) for examples of such works, respectively. In some instances, an equivalent approach to robust optimization is to apply regularization terms to the objective function, which can be equivalently modelled as constraints on the portfolio allocation. Constraints on the portfolio allocation have been shown to improve the performance of the solutions obtained via mean-variance optimization and be equivalent to shrinkage on the estimated parameters, which by construction reduces the estimation error present in the estimated parameters (Behr et al., 2013; Jagannathan and Ma, 2002).

With regards to transaction costs, Olivares-Nadal and DeMiguel (2018) show that certain instances of the aforementioned robust and regularized problems are equivalent to a portfolio problem considering transaction costs. Another common approach to mitigating transaction cost is to impose that only a subset of assets is selected for investment, which has been proved to be an NP-hard problem solved via approximate and exact algorithms in many problem domains. Approximate solutions are obtained by the use of l_1 norm penalties in the objective and evolutionary algorithms and exact algorithms as summarized in the review by (Tillmann et al., 2021). One approach that is of particular relevance in this work is the alternating direction method proposed by Costa et al. (2020) for solving cardinality constrained optimization problems. Costa et al. (2020) applies their proposed alternating direction method approach to solve a cardinality constrained portfolio problem and show that it exhibits good solution quality in less time than commercial solvers.

There are quantitative techniques that incorporate the subjective aspects of investment into portfolio optimization. One approach that includes subjective information is the incorporation of investor views pioneered by Black and Litterman (1991), which allowed portfolio managers to merge their views on asset returns into the prior return distribution that is typically estimated based on historical data. Since the seminal work by Black and Litterman (1991), there have been a series of modifications and extensions of the original framework that offer increased relaxations of market assumptions, additional types of views and improved operational capabilities due to reduced computation overhead (Meucci, 2010). In addition to view processing, portfolio managers can also incorporate subjective information by estimating the optimization model's parameters. For example, an investor can devise a model for stock returns that decomposes the returns into factors (Clarke et al., 2005). Using the factor model, the investor can then devise expectations for a set of factors at the investment horizon via some form of subjective analysis and use the estimated expectations within the factor model to compute an expected return for the stock. The particular form of the investor's hypothesis on stock returns defines their investment style. For example, one style of investment, value investing, places a significant emphasis on balance sheet information such as the price to earnings ratio of a company, whereas another style known as momentum-following investing looks to buy securities that have shown recent trends of rising value. Although there is a plethora of different investing styles, Pedersen (2015) provides a detailed exposition of different investing styles used by professional money managers around the globe.

An alternative way to view portfolio management is that an investor is a learner who hopefully has learned what attributes make an investment worthwhile. The most concrete example of this idea is explored by Fan and Palaniswami (2001), where historical price and fundamentals data is used to learn a support vector machine (SVM) to predict assets that will be in the top 25% of performers over the investment period. Fan and Palaniswami (2001) form labels from the historical data by considering the top 25% of assets and labelling those as positive classes and the rest as negative classes. They use the trained SVM on a set of new investments to predict their class based on the available attribute data and then form equal-weighted portfolios in the predicted positive class. Fan and Palaniswami (2001), show that their proposed approach generated returns that are in excess of the market during their period of study. Paiva et al. (2019) builds on Fan and Palaniswami (2001)'s work by using an SVM to decide what stocks to invest in. However, they augment the decision process with a mean-variance optimal allocation among the predicted out-performing stocks as opposed to an equally weighted portfolio. In addition, Paiva et al. (2019) use a return threshold to determine the classes for training the SVM as opposed to the quantile-based class definition used by Fan and Palaniswami (2001). Although it has been empirically shown that the aforementioned SVM-type approaches do deliver promising risk-adjusted returns, they do not handle risk in their estimation as they are explicitly trained on classes defined by return considerations. This explicit definition may lead to portfolio decisions that are too concentrated in particular sectors during cyclical markets. For example, if pharmaceutical stocks are experiencing a bubble-like growth one year, it would be the case in the historical training dataset that tech stocks preferred. In the following year, the decision rules would dictate that one invest in tech-related assets *after* a period of frothy growth. This is a dangerous result!

Contributions

It is well known that investors rely on specific criteria to determine if an investment is eligible for capital allocation. In addition, constraints in the form of robust optimization, cardinality constraints, regularization, and shrinkage have been shown to improve risk-adjusted returns and portfolio turnover. Furthermore, current statistical learning alternatives to MVO that aim at learning investor criteria do not give a complete treatment of risk in deciding which assets are eligible for investment. Given these considerations, it stands to reason that an investor may trade-off between expected risk and returns while requiring that the potential investments satisfy an internal logic. Inspired by the flexibility and demonstrated effectiveness of SVM's in a portfolio management setting, this paper proposes a representative investor whose asset eligibility logic takes the form of a support vector machine. This representative investor simultaneously solves for a mean-variance optimal portfolio, with a cardinality constraint such that all the assets included in the portfolio are on the positive side of a hyper-plane defined by asset features. Intuitively, this means that the investor seeks a portfolio of minimum risk subject to a return constraint such that the assets in the portfolio fit within their investment policy. Given a set of time-adapted asset attributes (such as sales growth, for example), a mixed binary quadratic program is introduced to simultaneously solve for a separating hyperplane in the assets and an asset allocation that minimizes the portfolio variance subject to return constraints. The proposed portfolio optimizer with investment policy constraints is explored in relation to the well-tested portfolio norm penalization technique proposed by Ho et al. (2015) for the particular instance of enforcing a volatility limit on the assets that are eligible for investment. The proposed optimization problem is challenging to solve in the case of a large number of assets and features per asset. Two variants of a partial alternating direction method (PADM) are proposed to address performance concerns. The deterministic variant of the PADM arrives at a partial minimum solution in less than half the time of an exact commercial MIP solver on average but does not find a solution that can be entirely separated by a hyper-plane as often as the commercial solver. A randomized restart variant of the PADM obtains separable solutions more often than its deterministic counterpart, only failing in 2/244 cases but requiring more time to terminate. Still, the randomized PADM shows a $\approx 25\%$ improvement in solution times over the commercial MIP solver on average. Lastly, it is demonstrated by a monthly rebalancing experiment consisting of 286 stocks that depending on the choice of asset attributes, adding a constraint-based on investment fundamentals results in improved risk-adjusted returns.

Structure of the paper

The remainder of this manuscript is organized as follows. Section 2 gives background on support vector machines, cardinality constrained portfolio optimization, and alternating direction methods to heuristically solve cardinality constrained portfolio optimization problems. Section 3 introduces the proposed representative investor's allocation problem, relates the problem to regularized portfolio optimization, and introduces an alternating direction method to approximately solve the proposed optimization problem. Section 4 presents the methodology and experiments conducted to demonstrate the proposed framework's use and effectiveness. Lastly, Section 5 concludes the paper.

2 Background

2.1 Cardinality Constrained Portfolio Optimization

First, suppose there are N assets available in a long-only setting. A portfolio is represented by $x \in \mathbb{R}^N_+$ where its elements denote the proportions of the portfolio invested in a particular asset. Let $R \in \mathbb{R}^N$ denote the random single-period return of all the assets. Then, also let $\mu \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N \times N}$ denote the mean and covariance of the random returns variable. The traditional mean-variance approach with a cardinality constraint is of the form:

$$\begin{aligned} \min_{x} \quad & x^\intercal \Sigma x \\ \text{s.t.} \quad & \mathbf{1}^\intercal x = 1 \quad \mu^\intercal x \geq \bar{R} \\ & |\text{supp}(x)| \leq K \quad x \geq 0 \end{aligned} \tag{Card-MVO}$$

where $\bar{R}$ denotes the investor's minimum required expected return, $\text{supp}(x) \coloneqq \{i \in [N] : x_i \neq 0\}$, $[N] \coloneqq \{1, 2, \ldots, N\}$, and $K \in [N]$ representing a cardinality limit. Introducing $z_i \in \{0, 1\}$ allows the support constraint to be rewritten as $\sum_{i \in [N]} z_i \leq K$, $x_i \leq z_i \quad i = 1, 2, \ldots, N$. In general, Card-MVO is an NP-hard problem and been tackled by various heuristics and exact methods. Exact branch and bound methods have been used to solve cardinality constrained optimization problems to global optimality. Bertsimas and Cory-Wright (2021) provide an overview of exact solution methods for portfolio optimization with cardinality constraints and review the literature to find that there has not been a scalable exact algorithm to solve cardinality constrained optimization problems with greater than $N = 500$ assets. Bertsimas and Cory-Wright (2021) proceeds by using a heuristic to warm start an exact method to solve larger problem instances with up to 3200 assets. Therefore heuristics are important to solve large instances even in the case of searching for global optima. With regards to heuristic methods to solve Card-MVO, there are a numerous approaches that have been taken including: i) dropping the cardinality constraint in favour of l_1 norm penalization to obtain sparsity (see Wu and Yang (2014) for example) and, ii) greedy heuristics (see Chang et al. (2000) for example). One primal heuristic of interest is the alternating direction method proposed by Costa et al. (2020) whereby auxillary variables are introduced to the nominal problem Card-MVO and improvement steps are taken alternately between the two sets of variables. This approach is shown to obtain several desirable properties such as convergence, partial minimization, good solution quality, and is quick relative to commercial solvers. Given all the aforementioned custom schemes to solve Card-MVO, one would be inclined to think that commercial solvers are unable to solve instances of Card-MVO in an effective manner, however, Anis and Kwon (2020) note that commercial solvers have experienced speed-ups that are on the order of $1.5 - 2 \times 10^6$ over the period between 1991 and 2009, and therefore are able to directly solve their cardinalty constrained portfolio optimization problem with Gurobi 9.0.

2.2 Support Vector Machines

Given a dataset of n pairs $\{(y_1, u_1), (y_2, u_2), \ldots, (y_n, u_n)\}$, with $y_i \in \mathbb{R}^p$ and $u_i \in \{-1, 1\}$. One can define a classification rule by determining which side a given input lies on a hyperplane (taking the sign of a hyperplane):

$$G(x) = \text{sign}[y^\intercal w + b] \tag{1}$$

Friedman (2017) show that for the given classification rule and under the assumption of linear separability, the hyperplane that creates the most margin between the positive and negative classes is given by the following optimization problem:

$$\begin{aligned}\min_{w,b} \quad & \frac{1}{2}\|w\|^2 \\ \text{s.t.} \quad & u_i(w^\intercal y_i + b)) \geq 1 \quad \forall i = 1, 2, \ldots, n\end{aligned} \tag{SVM-1}$$

Furthermore, in the non-separable case by introducing $\xi_i \geq 0$, and relaxing the constraint $u_i(w^\intercal y_i + b)) \geq 1$ to $u_i(w^\intercal y_i + b)) \geq 1 - \xi_i$. It follows that SVM-1 is equivalent to the following problem:

$$\begin{aligned}\min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \\ \text{s.t.} \quad & u_i(w^\intercal y_i + b)) \geq 1 - \xi_i \quad \forall i = 1, 2, \ldots, n \\ & \xi_i \geq 0 \quad \forall i = 1, 2, \ldots, n\end{aligned} \tag{SVM-2}$$

Where C is a parameter that penalizes miss-classification. Menon (2009), argues that the primal is of interest when tackling large-scale problems because its lack of constraints allows it to be tackled by gradient descent algorithms. The dual of SVM-2 has historically been of interest because it lends itself easier to the "kernel trick" since it explicitly makes use of dot products between transformed input vectors [1]. Furthermore, the dual formulation also has simple box constraints for the dual variables. In this work, the primal problem is used because it has a natural interpretation within the context of MVO. This will be discussed in the next section.

2.3 Alternating Direction Methods

This section gives the necessary background on alternating direction methods (ADM) that will be used to obtain heuristic solutions for the proposed allocation problem. The goal of ADMs is to find a partial minimum:

Definition 1. Let $(u^*, v^*) \in \mathcal{U} \times \mathcal{V}$ be a feasible point of ADM-1 such that $h(u, v) \geq 0$. Then (u^*, v^*) is a *partial minimum* of problem ADM-1 iff:

$$f(u^*, v^*) \leq f(u, v^*) \quad \forall u \in \mathcal{U}$$

$$f(u^*, v^*) \leq f(u^*, v) \quad \forall v \in \mathcal{V}$$

[1]The kernel trick is a technique used to project the input data space onto a higher dimension. It is more likely that the data will be linearly separable in a transformed higher-dimensional space. See Friedman (2017) for more details. The dual formulation of SVM-2 does not require the specific transformation mapping but only requires an expression for the dot product of the transformed vectors.

In general, ADM's proceed by partitioning the set of decision variables into two groups and iteratively optimizing one group of variables while keeping the other variables fixed. The optimization problems resulting from partitioning the optimization problem into two sub-problems are usually easier to solve than the joint problem that considers all decision variables simultaneously. However, the sub-problem's tractability largely depends on the structure of the joint optimization problem. Geißler et al. (2017), explore the application of ADM style methods to mixed-integer programs by showing the equivalence between feasibility pumps and ADMs [2]. Stemming from Geißler et al. (2017)'s original work, several applied mixed-integer application problems have been solved using ADMs; see e.g., (Burgard et al., n.d.; Geißler et al., 2015; Geißler et al., 2018; Goettlich et al., 2021; Kleinert and Schmidt, 2020; Schewe et al., 2020a; Schewe et al., 2020b; Yu et al., 2020).

ADMs have also been used to solve portfolio optimization problems. Costa et al. (2020) use ADMs to solve the cardinality constrained portfolio optimization problem. The following optimization problem is considered to provide an exposition on ADM algorithms:

$$\begin{aligned} \min_{u,v} \quad & f(u,v) \\ \text{s.t.} \quad & h(u,v) \geq 0 \\ & u \in \mathcal{U},\ v \in \mathcal{V} \end{aligned} \tag{ADM-1}$$

where $\mathcal{U} \subseteq \mathbb{R}^{n_u}$, $\mathcal{V} \subseteq \mathbb{R}^{n_v}$, $f : \mathbb{R}^{n_u+n_v} \to \mathbb{R}$, and $g : \mathbb{R}^q \to \mathbb{R}$ The standard ADM proceeds by initializing u and v and then iteratively optimizing each set of variables while holding the other set fixed. Algorithm 1 shows the steps for the standard ADM algorithm applied to problem ADM-1.

Algorithm 1 Standard ADM

1: Initialize $(u,v) = (u^{(0)}, v^{(0)}) \in \mathcal{U} \times \mathcal{V}$ and $k = 1$
2: **while** Stopping Criteria Not Met **do**
3: Compute $u^{(k)} \in \arg\min_u \{f(u, v^{(k-1)}) \mid h(u, v^{(k-1)}) \geq 0,\ u \in \mathcal{U}\}$
4: Compute $v^{(k)} \in \arg\min_v \{f(u^{(k)}, v) \mid h(u^{(k)}, v) \geq 0,\ v \in \mathcal{V}\}$
5: Update iteration counter/ stopping criteria
6: **end while**

Costa et al. (2020) notes that the coupling constraints induced by $h(u,v)$ may lead to poor convergence of the algorithm. Geißler et al. (2017) introduces a modified variant of the ADM algorithm, which avoids stalling the algorithm. The modified ADM proceeds by moving the coupling constraints to the objective

[2] Feasibility pumps are a primal heuristic for mixed-integer programs that work by constructing two sequences. The first sequence contains integer-feasible points; the second one contains points that are feasible w.r.t. the continuous relaxation. If the sequences converge, then an optimal point has been found

function via penalization. This yields the following objective function:

$$\phi(u, v; \rho) \coloneqq f(u, v) + \rho^{\intercal}[h(u, v)]_{-}$$

with $[a]_{-} = \max\{0, -a\}$ and $\rho \in \mathbb{R}^q_+$ representing the penalty parameters. The proposed penalty ADM (PADM) proceeds by computing a partial minimum of $\min_{(u,v)\in\mathcal{U}\times\mathcal{V}} \phi(u, v; \rho)$, and checking if feasibility is achieved. If feasibility is not achieved then the penalty parameters are incremented and a partial minimum is computed again until some stopping condition is met. Algorithm 2 formally states the PADM.

Algorithm 2 Penalty ADM

Initialize $(u, v) = (u^{(0,0)}, v^{(0,0)}) \in \mathcal{U} \times \mathcal{V}$, $\rho^{(0)} \geq 0$, and $l, k = (0, 0)$
while Infeasible : $h(u^{(l,k)}, v^{(l,k)}) < 0$ and $l \leq$ IterLim **do**
    $k = 0$
    while $(u^{(l,k)}, v^{(l,k)})$ is not a partial minimum with $\rho = \rho^{(l)}$ **do**
        Compute $u^{(l,k+1)} \in \arg\min_u\{\phi(u, v^{(l,k)}; \rho^{(l)}) \mid u \in \mathcal{U}\}$
        Compute $v^{(l,k+1)} \in \arg\min_v\{\phi(u^{(l,k+1)}, v; \rho^{(l)}) \mid v \in \mathcal{V}\}$
        $k \leftarrow k + 1$
        Update other stopping criteria; e.g. $\sqrt{||u^{(l,k)} - u^{(l,k-1)}||^2 + ||v^{(l,k)} - v^{(l,k-1)}||^2}$
    end while
    Increment $\rho^{(l+1)}$ such that $\rho^{(l+1)} \geq \rho^{(l)}$
    $l \leftarrow l + 1$
end while

Geißler et al. (2017) show that the PADM is equivalent to a feasibility pump with deterministic restart rules in place of random restart rules, whereas the ADM is equivalent to a feasibility pump that does not consider stalling or cycling. The use of deterministic restarts allows the PADM to obtain higher quality solutions than the ADM because of its implicit use of restarts. Costa et al. (2020) provide an overview of the convergence theory of both the ADM and PADM methods. For the PADM, it is the case that the following theorem holds:

Theorem 1. *if i) $\mathcal{U}$ and $\mathcal{V}$ are non-empty compact sets, ii) f and h are continuous and iii) $\rho_i^{(l)} \nearrow \infty$ $\forall i \in [q]$. Furthermore, let $(u^{(l)}, v^{(l)}) \to (u^*, v^*)$ be the sequence of partial minima generated by Algorithm 2, then the following holds:*

1. *there exists $\bar{\rho}$ such that (u^*, v^*) is a partial minimizer of the feasibility measure: $\rho^{\intercal}[h(u, v)]_{-}$*
2. *if (u^*, v^*) are feasible, then (u^*, v^*) is a partial minimum of ADM-1*
3. *if (u^*, v^*) are feasible and f is also continuously differentiable then (u^*, v^*) is a stationary point of ADM-1*

4. *if (u^*, v^*) are feasible, f is continuously differentiable and f and $\mathcal{U} \times \mathcal{V}$ is convex then (u^*, v^*) is a global optimum of ADM-1*

Although, mixed integer programs do not satisfy part 4 of Theorem 1 since the feasible sets are typically non-convex. However, it is the case that parts 1-3 are possible to satisfy. A PADM is used to efficiently obtain approximate solutions to the problem proposed in the next section.

3 Proposed Methodology

3.1 Integrated Mean-Variance and Support Vector Machines

As discussed in Section 1, the representative investor seeks a portfolio of minimum risk subject to a return constraint such that the assets in the portfolio fit within their investment policy whereby a linear hyperplane can separate the investable assets. Assume that the investor's set-up is as given in Section 2. Also, assume that the investor is subject to the constraint that they must select the assets such that asset i's inclusion in the portfolio denoted by z_i is given by:

$$z_i = \begin{cases} 1, & y_i^\intercal w + b \geq \epsilon \\ 0, & y_i^\intercal w + b \leq -\epsilon \\ \text{undefined} & \text{otherwise} \end{cases} \tag{2}$$

where ϵ is a given positive scalar and $y_i \in \mathbb{R}^p$ is a vector of attributes for asset i such as most recent earnings per share, quarterly income growth, debt to equity ratio, ..., etc. Equivalently, $z_i = \frac{1}{2} + \frac{1}{2}\text{sign}_\epsilon[y_i^\intercal w + b]$ where $\text{sign}_\epsilon[x]$ is the sign function strictly defined on the domain $|x| \geq \epsilon$. Therefore, one of the two inequalities in equation 2 must hold for the assets to be ϵ separable by w and b. One way to model this is to use a big-M formulation such that $M_w \geq y_i w + b + \epsilon$ and then it follows that one of the two inequalities will always be true if both of the following equalities hold:

$$y_i^\intercal w + b \leq M_w z_i - \epsilon \tag{3}$$

$$-M_w(1 - z_i) + \epsilon \leq y_i^\intercal w + b \tag{4}$$

Therefore, the investor who is cardinality constrained and requires that assets be separable, jointly looks for the portfolio x and maximum margin hyper-plane (w, b) such that it separates all the assets to be included in the portfolio from those that are not to be included. i.e they solve the following optimization problem:

$$
\begin{aligned}
\min_{x,z,w,b} \quad & x^\intercal \Sigma x + \frac{1}{2}||w||^2 \\
\text{s.t. (i)} \quad & \mu^\intercal x \geq \bar{R}, \text{ (ii)} \quad \sum_i z_i \leq K \\
\text{(iii)} \quad & x_i \leq z_i && \forall i = 1, 2, ..., N \\
\text{(iv)} \quad & y_i^\intercal w + b \leq M_w z_i - \epsilon && \forall i = 1, 2, ..., N \\
\text{(v)} \quad & -M_w(1 - z_i) + \epsilon \leq y_i^\intercal w + b && \forall i = 1, 2, ..., N \\
& x \in \Delta_N \quad z \in \{0, 1\}^N
\end{aligned}
\tag{SVM-MVO 1}
$$

Where Δ_N is the N dimensional simplex, constraints (i) and (ii) ensure that the expected return exceeds the target and that the cardinality is less than K. Furthermore, constraint (iii) ensures that if asset i is not eligible to be included in the portfolio, it will receive zero weight. Lastly, constraints (iv) and (v) ensure that all assets eligible for investment are on the positive side (exceeding ϵ) of the hyper-plane (w, b) and SVM-MVO 1 and that all assets on the negative side of the hyper-plane (less than $-\epsilon$). SVM-MVO 1 may be infeasible, i.e. there does not exist a hyperplane that can separate the assets. However, similar to the case of SVM-2, the problem can be made feasible by relaxing constraints (iv) and (v) by introducing the slack variables $\xi \in \mathbb{R}_+^N$ and penalizing them in the objective function as follows:

$$
\begin{aligned}
\min_{x,z,w,b} \quad & x^\intercal \Sigma x + \frac{1}{2}||w||^2 + \frac{C}{N}\sum_i \xi_i \\
\text{s.t. (i)} \quad & \mu^\intercal x \geq \bar{R}, \text{ (ii)} \quad \sum_i z_i \leq K \\
\text{(iii)} \quad & x_i \leq z_i && \forall i = 1, 2, ..., N \\
\text{(iv)} \quad & y_i^\intercal w + b \leq M_w z_i - \epsilon + \xi_i && \forall i = 1, 2, ..., N \\
\text{(v)} \quad & -M_w(1 - z_i) + \epsilon - \xi_i \leq y_i^\intercal w + b && \forall i = 1, 2, ..., N \\
& x \in \Delta_N \quad z \in \{0, 1\}^N \quad \xi \geq \mathbf{0}
\end{aligned}
\tag{SVM-MVO 2}
$$

where C is an external hyper-parameter describes the investors preference for the assets to be separable by their features. SVM-MVO 2 is a mixed-binary quadratic program with $p+1+2N$ continuous decision variables and N binary variables. ϵ provides a scaling for the importance of maximum margin and hyperplane separability relative to risk because the maximum margin multiplied by epsilon between the hyperplane (w, b) and the $i = 1, 2, \ldots, N$ assets is equal to Margin $= \frac{|y_i^\intercal w+b|}{||w||^2}$ for some asset $i \in [N]$. In the linearly separable case, $|y_i^\intercal w + b| = \epsilon$ which means that a larger ϵ will be a higher margin solution for a given w and ϵ. Although daunting at first glance, SVM-MVO 2 can be tackled for problems with a large number of assets N and exogenous factors p due to advances in algorithmic and hardware capabilities. Although Bixby (2012) notes that advances in solvers and hardware have made larger instances solvable, it is still advantageous to obtain approximate solutions to the proposed optimization problems quickly. Problems SVM-MVO 1 and SVM-MVO 2 will be jointly referred to as SVM-MVO problems going forward. In the next sub-section, a PADM is proposed to solve SVM-MVO 2.

3.2 SVM-MVO as a Regularization Penalty

This section motivates the use the SVM-MVO optimization problems by considering a particular class of SVM-MVO and relating it regularized weighted elastic net portfolio optimization proposed by Ho et al. (2015). In the work by Ho et al. (2015), the optimal portfolio solves:

$$\begin{aligned} \min_{x} \quad & x^\intercal \Sigma x + x^\intercal \mathrm{diag}\{\alpha_i\} x + \beta^\intercal |x| \\ \text{s.t.} \quad & \mu^\intercal x \geq \bar{R},\ x \in \Delta_N \end{aligned} \tag{ELASTIC-NET}$$

where $\beta \in \mathbb{R}_+^N$ and $\alpha_i \in \mathbb{R}_+ \ \forall i \in [N]$ denote the penalty parameters. Ho et al. (2015) show that α_i and β_i correspond to the sizes of box-uncertainty sets used in the robust counterpart of the mean-variance optimization problem. Specifically, α_i and β_i correspond to uncertainty in the diagonal of the **estimated** covariance matrix Σ_{ii} and expected return μ_i respectively for for asset $i = 1, 2, \ldots, N$. Ho et al. (2015) use a bootstrapping approach to select the values of β_i and α_i; they proceed as follows: Let $\{X^{(t)} \in \mathbb{R}^N\}_{t \in [T_{train}]}$ denote T_{train} observations of the random variable X used to available to estimate Σ and μ. The estimation errors are computed by re-sampling T_{train} observations with replacement from the original set of observations B times. Each re-sampling yields estimates of the mean and covariance $\hat{\mu}^{(b)}, \hat{\Sigma}^{(b)}$ $b \in [B]$. Let p_1 and p_2 denote the investor's aversion to estimation risk of the mean and variance respectively. The values for β_i and α_i are defined as the quantiles of the bootstrapped deviations

$$\alpha_i = \inf\left\{t \middle| \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left[|\hat{\Sigma}_{ii}^{(b)} - \Sigma_{ii}| < t\right] > p_1\right\}$$

and

$$\beta_i = \inf \left\{ t \Big| \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\left[|\hat{\mu}_i^{(b)} - \mu_i| < t \right] > p_2 \right\}$$

Under the assumption that the observations are independent and identically sampled from a distribution (iid) with finite mean $\bar{\mu}_i$ and variance σ_i^2 one can apply the central limit theorem and the delta method to obtain the following expressions for the quantiles:

$$\alpha_i = \frac{1}{\sqrt{T_{train}}} \Phi^{-1}\Big(\frac{p_1+1}{2}\Big) \sqrt{\mathrm{CM}_4(X_i) - \sigma_i^4} = \sqrt{\frac{\kappa_i - 1}{T_{train}}} \Phi^{-1}\Big(\frac{p_1+1}{2}\Big) \sigma_i^2 \tag{5}$$

and

$$\beta_i = \frac{\sigma_i}{\sqrt{T_{train}}} \Phi^{-1}\Big(\frac{p_2+1}{2}\Big) \tag{6}$$

where Φ^{-1} denotes the inverse cdf for the normal distribution, $\mathrm{CM}_4(X_i)$ denotes the fourth central moment of the random return for asset i, and κ_i denotes the kurtosis of the random return for asset i. For practical use, the sample estimates denoted by $\hat{\kappa}$ and Σ_{ii} of κ and σ^2 can be used. See Appendix 5.3 for the derivation of the equations 5 and 6. Equations 5 and 6 both show that the penalties decrease when more iid observations are available and when the aversions to estimation error (p_1 and p_2) decrease. Equation 5 shows that the size of the uncertainty set for Σ_{ii} is proportional to the value of Σ_{ii} and proportional heaviness of the tails of the return distribution as measured by κ_i. Similarly, equation 6 shows that the size of the expected return's uncertainty set is proportional to the standard deviation $\sqrt{\Sigma_{ii}}$.

Now suppose that the imposed separating hyper-plane in the SVM-MVO formulation takes the following form:

$$(-\Sigma_{ii})w + b = 0 \text{ where } b - \epsilon \in \mathbb{R}_+,\ w - \epsilon \geq \mathbb{R}_+$$

then this separating hyper-plane corresponds to a threshold on the standard deviation of each asset's returns such that an asset is guaranteed not to be included ($z_i = 0$) if $\Sigma_{ii} \geq$ threshold. This behaviour can be modelled by adding a large penalty to the objective function as follows:

$$\begin{aligned} \min_x \quad & x^\intercal \Sigma x + M \sum_{i=1}^{N} \mathbf{1}[\sqrt{\Sigma_{ii}/T_{train}} \geq \sqrt{\text{threshold}/T_{train}}]^\intercal |x_i| \\ \text{s.t.} \quad & \mu^\intercal x \geq \bar{R},\ x \in \Delta_N \end{aligned} \tag{7}$$

where M is a large constant such that it is never optimal to have $x_i > 0$ if $\Sigma_{ii} \geq$ threshold. As such, adding a threshold on the volatility can be interpreted as having 100% aversion to assets with mean estimation errors that exceed a threshold. Furthermore, the work by Ho et al. (2015) suggests that it may be beneficial to enforce a threshold limit on the product of the $\sqrt{\hat{\kappa}-1}\Sigma_{ii}$.

3.3 SVM-MVO Penalized Alternating Direction Method

This section outlines the PADM used to obtain heuristic solutions for SVM-MVO 1. In problem SVM-MVO 2, the term $\frac{C}{N}\sum_i \xi_i$ represents a penalty for violating the separating hyper-plane feasibility. Therefore, $\frac{C}{N}\sum_i \xi_i$ shares a similar purpose as the coupling constraint penalty term $\rho^\intercal[h(u,v)]_-$ from (2.3). Furthermore, the proposed SVM-MVO optimization problems lend themselves naturally to partitioning the decision variables into two sets - one set for the mean variance problem and the other for the SVM problem. Partitioning the decision variables such that $u = (z, x)$ and $v = (w, b)$, the nominal problem becomes minimizing $f(u, v) = x^\intercal \Sigma x + \frac{1}{2}||w||^2$ with:

$$u \in \mathcal{U} \coloneqq \{(x, z) \in \Delta_N \times \{0,1\}^N \mid \mu^\intercal x \geq R,\ x_i \leq z_i\ \forall i \in [N],\ \mathbf{1}^\intercal z \leq K\} \text{ and}$$
$$v \in \mathcal{V} \coloneqq \{(w, b) \in \mathbb{R}^{m+1}\}$$

Instead of using constraints (iv) and (v) present in problem SVM-MVO 1, one can also use the complementary type constraints:

$$(2z_i - 1)(y_i^\intercal w + b) \geq \epsilon \quad i = 1, 2, ..., N \tag{8}$$

Within the context of the proposed algorithm, the complementary formulation is advantageous over the big-M type constraints. Although the complementary constraints are non-convex and non-linear when considering both u and v jointly, they are linear in u and v when holding the other variable set fixed as done in sub-problems for ADM-type methods. Therefore, the SVM-type constraints can be represented in the form of ADM-1 by setting $h_i(u, v) = (2z_i - 1)(y_i^\intercal w + b) - \epsilon\ \forall i \in [N]$. It then follows that introducing the slack variables such that $h_i(u, v) \geq -\xi_i$ and $\xi_i \geq 0\ \forall i \in [N]$ and penalizing ξ_i in the objective ensures that $\xi_i = [h_i(u, v)]_-$. Therefore, the penalized objective function for use in a PADM can be written as:

$$\phi(u, v) = x^\intercal \Sigma x + \frac{1}{2}||w||^2 + \frac{1}{N}\sum_i \rho_i \xi_i \tag{9}$$

Each of the aforementioned ingredients can be directly substituted into Algorithm 2 to obtain heuristic solutions to SVM-MVO 1. The implementation considered here initializes (x, z) by solving Card-MVO, and then subsequently initializes (w, b) by solving SVM-2 with the labels $u_i = 2z_i - 1$. Essentially, the algoirthm works as follows: given the previous iteration's SVM sub-problem results: (w^l, b^l), to obtain x^{l+1}, z^{l+1} one solves $\min_{x,z} \{x^\intercal \Sigma x + \sum_i \mu_i \xi_i : \xi_i \geq 1 - (2z_i - 1)(y^\intercal w^l + b^l),\ (x, z) \in \mathcal{U}\}$ After computing x^{l+1}, z^{l+1} it follows that w^{l+1}, b^{l+1} are computed by solving $\min_{w,b} \{\frac{1}{2}||w||^2 + \sum_i \mu_i \xi_i : \xi_i \geq 1 - (2z_i^{l+1} - 1)(y^\intercal w + b),\ (w, b) \in \mathcal{V}\}$. This sub-procedure is repeated until convergence to a partial minima, after which the penalty parameters are updated and the sub-procedure to search for a partial minima is repeated until feasibility is obtained. Since $\mathcal{U}$ and $\mathcal{V}$ are compact, non empty sets, and f and h are continuous it follows that the solution generated by Algorithm 2 is a partial minimum of $\rho^\intercal \xi$. If the solution generated by Algorithm 2 satisfies the linear separability constraints then the solution is partial minimum of SVM-MVO 1. Lastly, the PADM takes advantage of the structure of the proposed optimization problem by leveraging the complementary constraints and their linearity for each sub-problem in u and v. Big-M constraints are avoided and the number of constraints required is halved.

4 Experiments

This section demonstrates the proposed capital allocation problem. First, datasets, data requirements and parameter mean and covariance estimation procedures are described. Secondly, the financial performance of the proposed SVM-MVO problem is presented. The financial analysis first looks at a single-date instance to compare the proposed approach with more traditional approaches by comparing different implied portfolio structures and efficient frontiers. The separating hyperplane for a two-dimensional instance is also presented to gain intuition. Second, in addition to a single problem instance, a rebalancing experiment is presented where the portfolio is rebalanced monthly, and traditional mean-variance rebalancing is compared with the proposed SVM-MVO problem as the rebalancing rule. Lastly, The speed and solution quality of the proposed PADM algorithm is presented and is compared to exact branch and bound solutions obtained via Gurobi - 9.12 for a set of test instances. All the experiments were performed using Python on google colaboratory (colab). The colab notebook was hosted on a Linux machine with Intel(R) Xeon(R) CPU @ 2.20GHz and 16GB of RAM.

4.1 Data Preparation

4.1.1 Data Cleaning

Two datasets were required to solve the SVM-MVO problem. A dataset of daily returns is required to estimate the mean μ and covariance Σ. Another dataset was also needed to provide a fundamentals vector y_i for each asset. The dataset of returns was obtained via Yahoo Finance using the yfinance python package (Aroussi, 2021). The universe of assets was taken to be the tickers comprising the S&P 500 as of April 5th, 2020, taken from (*List of S&P 500 Companies* 2020) and the dates spanned from September 30th, 1999 to October 31st, 2020. The daily price series for each asset was obtained with dates on the vertical axis and tickers on the horizontal axis. First, rows that were missing every value for every ticker were dropped. Also, tickers with one or more missing values were excluded from the dataset. The Compustat Daily Updates - Fundamentals Quarterly database from Wharton Research Data Services (WRDS) was used to obtain vectors of balance sheet information (e.g. the company's cash on hand) y_i for each ticker in the and each available quarter (Wharton Research Data Services, 2021). The balance sheet information is stored as columns, and ticker - quarter pairs represent the rows of the WRDS dataset. Three tickers in the daily price series were not in the WRDS dataset, and their corresponding columns were removed from the daily price series. The beginning date was set such that 98% of the tickers had a report before the start date. The remaining tickers that did not have a quarterly report available before the start date was removed from the daily price series and the WRDS dataset. Also, if a column in the WRDS dataset was missing for over 35% of ticker-quarter pairs or was entirely missing for at least one hundred tickers, the column was dropped from the WRDS dataset, resulting in seven out of forty-eight company fundamentals variables being removed. After removing "bad" columns from the WRDS dataset, tickers in the WRDS dataset were removed if they were missing all the balance sheet info for more than 5% of the ticker's reports, resulting in seventy tickers being removed. Lastly, for the remaining missing values in the WRDS dataset, a simple KNN imputation scheme was used with $k = 2$ (Pedregosa et al., 2011). The use of imputation is justified because of the small percentage of missing values after the aforementioned data cleansing process. Please see the Appendix 5.2 for details on the code repository – specifically the Data Processing.ipynb notebook. Please see Appendix 5.1 for the data dictionary of WRDS variables. Appendix 5.1.1 shows the definitions for the raw data taken from Wharton Research Data Services, 2021. Appendix 5.1.2 shows the definitions of variables that are derived from the WRDS dataset. After this entire process there are 244 periods for optimization with 243 out of sample months available over a universe of $N = 286$ assets and $p = 38$

fundamental factors. Of course, removing historical data of stocks that do not have continuous streams of data introduces survivor-ship bias; however, for this work, comparisons will be across capital allocation algorithms and thus is not expected to impact the presented conclusions.

4.1.2 Implementation and Estimation

In this work, the trading is assumed to be completed on the first trading day of the month. The first trading day of the month corresponds to the first date of the month present in the daily price series. The out-of-sample return for a given month's trade equals the percentage difference between the price at the first trading day of the following month and the price at the first trading day of the given month. Each trading day is mapped to a month label to conduct the experiments. The WRDS dataset is also assigned to a label for a trade month by joining the `datadate` field with the trade dates. `datadate` must be before the trade date (to avoid looking into the future), and the most recent quarterly report is taken to be the feature vector for the trade date/month.

The final preparatory piece is to, for each month, estimate the mean and covariance matrix for the return distribution at the horizon. Meucci (2010) points out the common pitfall of simply estimating covariance matrices and means based on daily linear returns and then scaling them up to a particular time horizon (e.g. a month); this is incorrect since linear returns are not additive. Following the advice by Meucci (2010), the daily difference in the logarithm of the prices is computed. For each trading time, the log-returns up to that point (making sure that the log-return does not depend on the trade date plus one day) are used to compute estimates of the mean and covariance of the log returns. Following Longerstaey and Spencer (1996), exponentially smoothed estimators of the mean and covariance of the log-returns are used. An exponential smoothing parameter $\lambda = 0.985$ is used as opposed to 0.94 (as per (Longerstaey and Spencer, 1996)) because of the work done by Pafka et al. (2004) showing that a higher λ is more suited to portfolio optimization. Assuming that the log-returns are independent and normally distributed over the horizon implies that scaling the estimated parameters by the number of days in the horizon gives estimates of the covariance and mean of the monthly log-returns. Lastly, the exponential transformations of log-normal variables were used to obtain the estimated mean and covariance of the linear returns (Tarmast, 2001).

4.2 Model Demonstration

This section presents the SVM-MVO model for a single instance. In this section, the trading month is July 2000. Figures and computations for sections 4.2.1 – 4.2.2 are done in the notebook: Joint Optimization - 2 Factor Single Month.ipynb using exact MIP solves via Gurobi - 9.12. Figure and computations for section 4.2.3 are done in the notebook ADM Comparison NoteBook.ipynb. For this section, the cardinality limit K is set to be $\lfloor 0.8N \rfloor$. For the SVM-MVO problems, y_i $i \in [N]$ is set to be the vector of the most recent sales growth and the most recent earnings by assets ratio: $y_i = [\text{LEV1}_{(i,t)} \ \text{GRW1}_{(i,t)}]$ where t is such that it is the closest report date to the trade date such that the reported date is before the trade date. LEV1 and GRW1 correspond to the most recently reported debt to equity ratios and the most recently reported sales growth metrics, respectively. For problem SVM-MVO 2, C is set to 5000 to ensure that the SVM constraints (iv) and (v) (from problems SVM-MVO 1 and SVM-MVO 2) are obeyed if possible.

4.2.1 Portfolio Comparison

First, Card-MVO, SVM-MVO 1, and SVM-MVO 2 are all solved directly for the problem instance such that no return constraint is imposed. A direct comparison of the optimal portfolios resulting from the optimizations is presented in Figure 1. The resulting portfolios obtained using SVM-MVO 1 and SVM-MVO 2 are identical as $\xi_i^* = 0 \ \forall i$. Furthermore, the SVM-MVO portfolios do not differ much from the cardinality constrained minimum variance portfolio, with the main difference being that the cardinality constrained portfolio has an allocation of $\approx 2\%$ in HES.

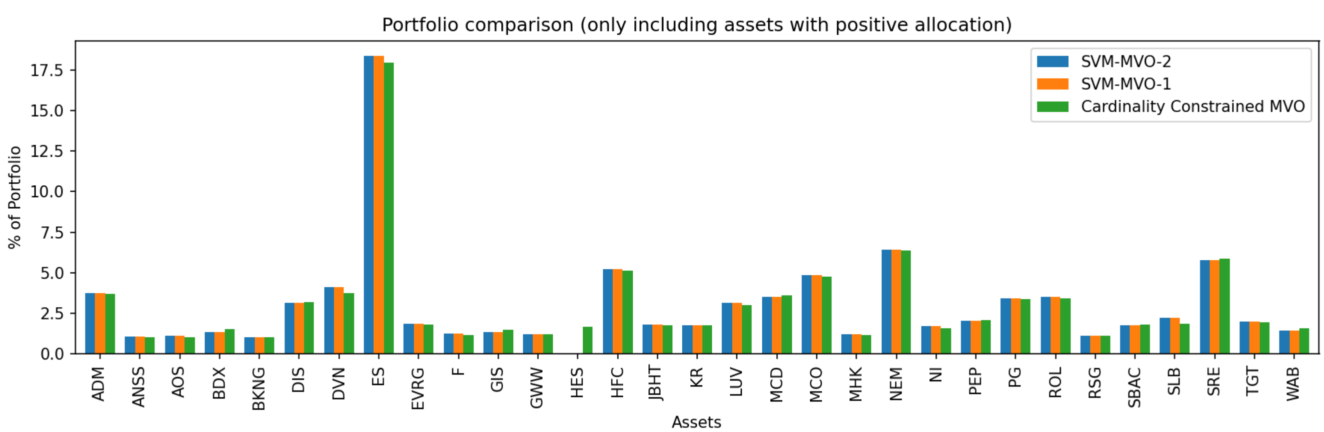

Figure 1: Portfolio composition for SVM-MVO 2 (blue), SVM-MVO 1 (orange), and Card-MVO (green)

The optimal separating hyper-plane obtained via SVM-MVO 2 has the following parameters:

$$w = [0.03027239,\ 0.05956292]$$

$$b = 0.00620327$$

which means: if $0.03027239\text{LEV1} + 0.05956292\text{GRW1} + 0.00620327 \geq 0.001$ then the asset is eligible for capital allocation by mean variance optimization. For example, the HES ticker does not satisfy the eligibility criterion as it's feature vector is $y = [0.064120\ 0.202965]$.

4.2.2 Efficient Frontier

One may ask a remaining question: "How does the investment criterion (defined by the hyper-plane) differ as the investors risk tolerance changes?" To answer this, one should first look at the efficient frontier. Figure 2 varies the target return from the minimum expected return to the maximum expected return and computes the corresponding in-sample risk for both the SVM-MVO and the cardinality constrained MVO. Formulation SVM-MVO 2 was used to compute the efficient frontier. Figure 2 shows that the frontier corresponding to the SVM-MVO offers slightly higher in-sample risk for the same level of expected return - this is as expected since the SVM-MVO model has additional constraints that prevent the portfolio from attaining the same minimum as the optimal cardinality constrained portfolio.

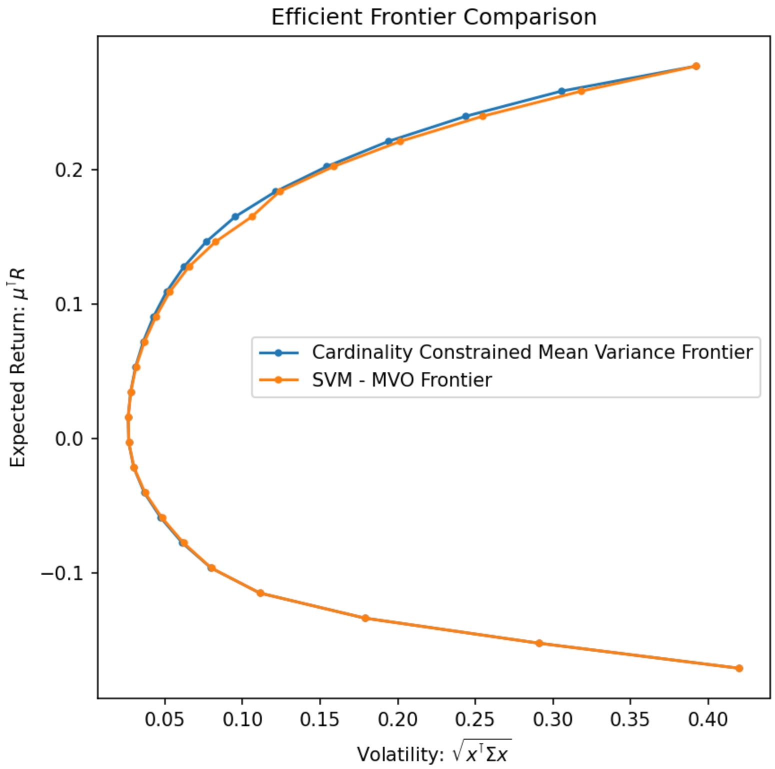

Figure 2: Efficient frontiers for Card-MVO (blue) and SVM-MVO 2 (orange)

Each value of the targeted return in the efficient frontier corresponds to an optimal separating hyper-plane. The hyper-planes are shown for given return targets in Figure 3. Between the target return range of -15% and 16% the hyper-planes remain stable and are equal to the hyper plane defined by $0.03027239\text{LEV1} + 0.05956292\text{GRW1} + 0.00620327 = 0$. When setting the return target above 16% the optimal separating hyper-plane essentially becomes equivalent to a threshold on sales growth (GRW1).

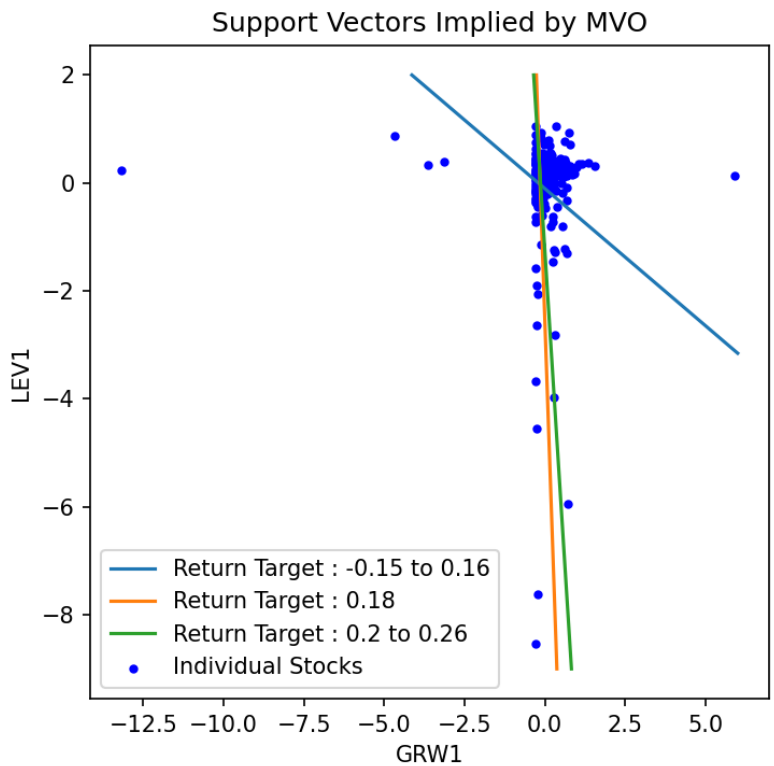

Figure 3: Separating hyper-planes for different return targets

4.2.3 Convergence Penalized Alternating Direction Method

All the results presented have relied on exact MIP solving routines until this point. However, as the number of features p used to separate the assets into eligibility groups increases, the number of combinations of points that can be separated increases, thereby enlarging the set of feasible solutions. Furthermore, increasing p also increases the number of continuous decision variables. The combined effect of the two points mentioned above is that adding more factors makes SVM-MVO 2 more challenging to solve to optimality. For demonstration purposes, the two-factor problem considered in this section is solved using the proposed PADM method of Section 2.3. This subsection focuses on the observed convergence for the proposed algorithm. A more detailed analysis of the PADM's performance properties concerning the exact solution is explored in Section 4.3. Since the solution is separable by (w, b) iff $\sum_i \xi_i = 0$ the feasibility penalty function can also be taken as $\rho \sum_i \xi_i$ where ρ is a positive scalar. In this case, the penalized objective takes the following form:

$$\phi(u, v) = x^\intercal \Sigma x + \frac{1}{2}||w||^2 + \frac{\rho}{N}\sum_i \xi_i \tag{10}$$

The initial value of $\rho^{(0)}$ is taken to be $\frac{50000}{2^{20}}$. After computing the partial minima ρ is updated by $\rho^{(l+1)} = 2\rho^{(l)}$. The stopping condition for the inner loop on outer iteration l is to check that $||w^{(l,k+1)} - w^{(l,k)}|| \leq 10^{(-6)}$ or if all $w_k^{(l,k)} \geq 10^{(-12)}$ check that the relative difference $||w^{(l,k+1)}./w^{(l,k)} - 1|| \leq 5\%$. The following condition is used to evaluate convergence of the outer loop: $||w^{(l,k(l+1))} - w^{(l,k(l))}|| \leq 10^{(-12)}$, where $k(l)$ is the last iteration count for the inner loop evaluated on the outer iteration l. Figure 4 shows the components of w for each iteration (l, k), indexed by integers starting at 0 for initialization.

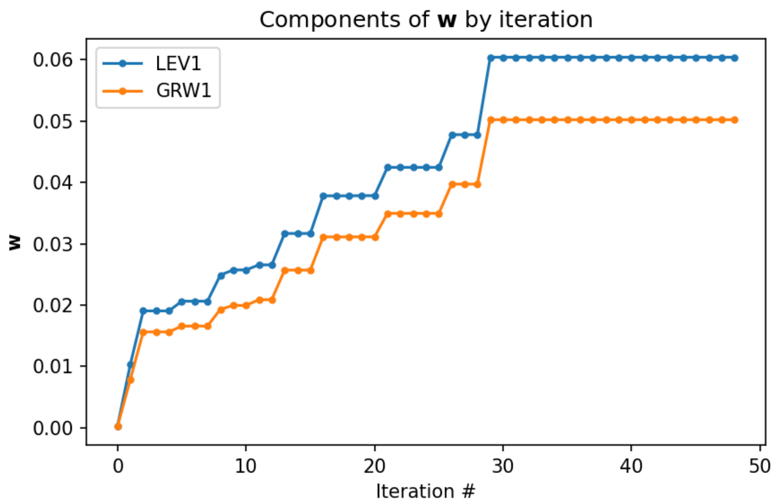

Figure 4: Convergence of w: After ≈ 50 iterations the PADM converges to a solution that is SVM feasible

During each iteration, a cardinality constrained optimization problem is solved with a penalization for including assets in the portfolio that are on the negative side of the previous iterations hyper-plane (w, b). The current iteration's cardinality constrained penalized MVO problem implies values for $\xi_i = \xi_i^{MVO}$. Likewise, after computing the current iteration's (x, z), a SVM problem is solved with the labels defined via z. The current iteration's SVM problem also implies values for ξ_i. For fixed ρ, at convergence to a partal minima, $\xi_i = \xi_i^{SVM}$. For fixed ρ, convergence to a partial minima occurs if $|\xi_i^{SVM} - \xi_i^{MVO}| \to 0 \ \forall i$ Furthermore, for a point (x, z, w, b) to be feasible in SVM-MVO 1, it must be the case that $\xi = \xi^{MVO} = \xi^{SVM} = 0$. Figure 5 shows the values of $\sum_i \xi_i^{(l,k),MVO}$ and $\sum_i \xi_i^{(l,k),SVM}$ plotted against iteration number. The points where the lines interest indicate points in which a partial minima was obtained. Figure 5 also shows the convergence of $\xi \to 0$ and thus also shows the convergence of (x, z, w, b) to a SVM-feasible solution.

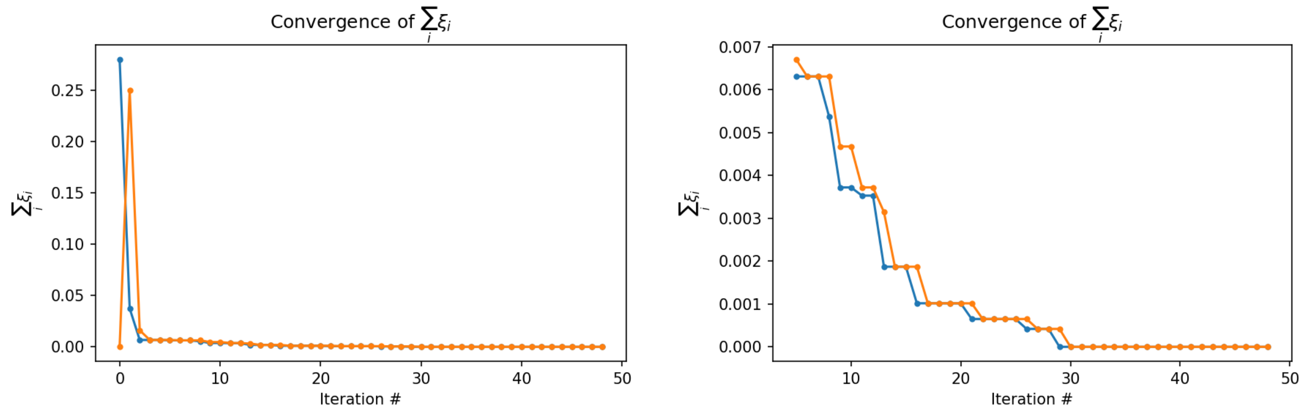

Figure 5: Convergence of ξ. Orange denoted ξ^{MVO} abd blue denotes ξ^{SVM}. As the iteration number increases, the PADM converges to a SVM feasible partial minimum

4.3 Effectiveness of the Alternating Direction Method

This section evaluates the proposed PADM on all 244 test available test instances by considering a minimum variance setting. The feature vector y_i comprises three components: LEV1, GRW1, and Momentum. The cardinality of the portfolio is constrained to be less than $\lfloor 0.8N \rfloor$. SVM-MVO 2 is used as the baseline problem to obtain the exact solutions directly from Gurobi, and C is set to 5000. Four methods are evaluated - three of which are variants of the PADM, and the other is the exact solution method. The first PADM uses only a single penalty parameter ρ and considers the penalty function given by the total constraint violation i.e $[h(u,v)]_- \sum_i \xi_i$. The initial value of ρ is $\frac{5000}{2^{(15)}}$, and the stopping conditions and penalty increment update are the same as those outlined in Section 4.2.3. The second configuration of PADM considered is identical to the first except a vector of penalty parameters $\rho \in \mathbb{R}_+^N$ are used for each ξ_i resulting in the constraint violation $\rho^\intercal [h_i(u,v)]_-$ with $[h_i(u,v)]_- = \xi_i$. The penalty parameters are initialized such that $\rho_i = \frac{5000}{2^{(15)}}$. The third variant of the PADM algorithm is a slight modification added to the algorithm. After observing the results for the PADM, it became apparent that even after increasing ρ, it was possible that the next iterate would be nearly identical to the current iterate while the next iterate was still not feasible – the algorithm was stalling. The idea to avoid stalling is to add a random restart after detecting a stalled iterate while keeping track of the best solution seen thus far. This approach is formally stated in Algorithm 3

Algorithm 3 Randomized Penalty ADM

Initialize $(u, v) = (u^{(0,0)}, v^{(0,0)}) \in \mathcal{U} \times \mathcal{V}$, $\rho^{(0)} \geq 0$, and $l, k = (0, 0)$
while Infeasible : $h(u^{(l,k)}, v^{(l,k)}) < 0$ and $l \leq$ IterLim **do**
    while $(u^{(l,k)}, v^{(l,k)})$ is not a partial minimum with $\rho = \rho^{(l)}$ **do**
        Compute $u^{(l,k+1)} \in \arg\min_u \{\phi(u, v^{(l,k)}; \rho^{(l)}) \mid u \in \mathcal{U}\}$
        Compute $v^{(l,k+1)} \in \arg\min_v \{\phi(u^{(l,k+1)}, v; \rho^{(l)}) \mid v \in \mathcal{V}\}$
        $k \leftarrow k + 1$
        Update other stopping criteria; e.g. $\sqrt{||u^{(l,k)} - u^{(l,k-1)}||^2 + ||v^{(l,k)} - v^{(l,k-1)}||^2}$
    end while
    $k = 0$
    Store partial minima: $(\tilde{u}^{(l)}, \tilde{v}^{(l)}) = (u^{(l,k)}, v^{(l,k)})$
    if $l > 0$, $\sqrt{||u^{(l+1)} - u^{(l)}||^2 + ||v^{(l+1)} - v^{(l)}||^2} \leq 10^{(-12)}$ and $h(\tilde{u}^{(l)}, \tilde{v}^{(l)}) < 0$ **then**
        Generate random $(u^{(l,k)}, v^{(l,k)}) = (u^*, v*)$
        Keep track of best start and use it if $l =$ IterLim $- 1$
    end if
    Increment $\rho^{(l+1)}$ such that $\rho^{(l+1)} \geq \rho^{(l)}$
    $l \leftarrow l + 1$
end while

To generate a random restarted solution, a random matrix $A \in \mathbb{R}^{N \times N}$ with uniform random entries is generated and the covariance matrix is replaced with the random convex combination $uAA^{\intercal} + (1 - u)I_{N \times N}$, where $u \sim \text{Unif}[0, 1]$. The random solution is initialized by solving the penalized cardinality constrained MVO problem defined by the current (w, b) iterate with the covariance matrix replaced by the aforementioned random matrix yielding $(\tilde{x}, \tilde{z})$. After solving the random penalized MVO instance, the SVM problem is solved with the labels defined by $\tilde{z}$ to obtain $(\tilde{w}, \tilde{b})$. After generating the random solution, the covariance matrix is restored to its true value and the algorithm proceeds from the new random start point. The purpose of this adjustment is to reduce the likelihood of the algorithm getting stuck in a partial minimum that is not feasible. All four of the approaches above are investigated in this section and are evaluated by comparing:

1. Feasibility
2. Relative objective gaps
3. Selection accuracy
4. Run time

The computations for the results in this section were all generated in the notebook: ADM Comparison NoteBook Timing.ipynb. The figures were generated in the notebook: Results.ipynb.

4.3.1 Feasibility

With regards to feasibility, fortunately, exact branch and bound solves applied to SVM-MVO 2 resulted in solutions such that $\sum_i \xi < 10^{(-9)}$ for all 244 test instances. All instances were solved to global optimality with the default gap tolerance of $10^{(-4)}$, hence providing a good baseline for evaluating the PADM algorithms. Figure 6 shows $\sum_i \xi_i$ for each test instance and solution method. It is visually clear from Figure 6 that the randomized PADM obtains feasible solutions much more often than its deterministic counterparts.

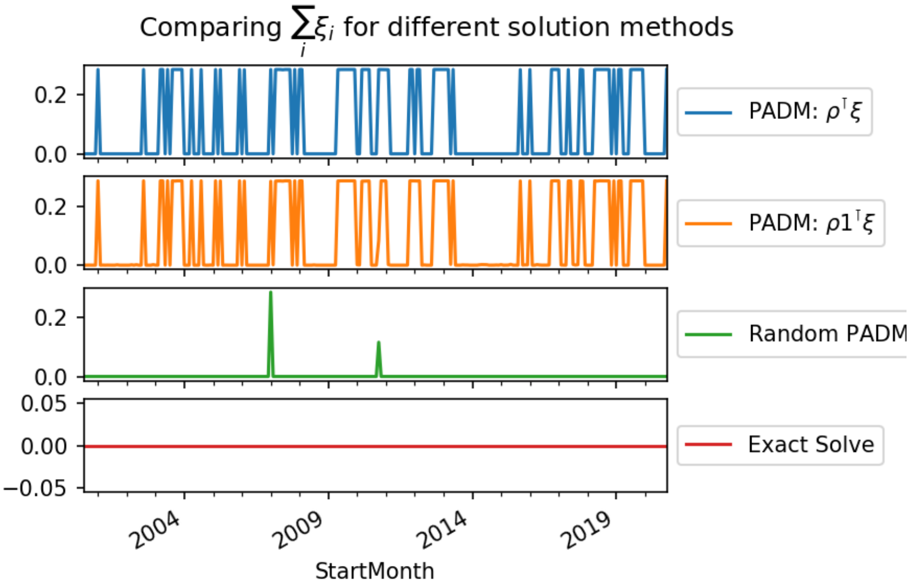

Figure 6: Infeasibility metric: $\sum_i \xi_i$ for different solution methods. PADM: $\rho^\intercal \xi$ (blue) denotes case where the penalty parameter $\rho \in \mathbb{R}^N_+$. PADM: $\rho \mathbf{1}^\intercal \xi$ (orange) denotes the single parameter case. Random PADM (green) denotes Algorithm 3. Exact Solve (red) denotes the B&B solution via Gurobi

PADM: $\rho \mathbf{1}^\intercal \xi$ with a single penalty parameter performed the worst concerning feasibility; only attaining SVM feasibility for 107 out of 244 instances. PADM: $\rho^\intercal \xi$ with a vector penalty parameter performed better with 162 instances attaining SVM feasibility. Lastly, the randomized PADM attained feasibility for all but two instances, clearly dominating the other two PADM methods. From this point on PADM: $\rho \mathbf{1}^\intercal \xi$ will not be considered in the analysis because, in addition to attaining feasibility for the smallest number of problem instances, it was also i) slower than the vectorized penalty PADM for 237 out of 244 instances and ii) when it did achieve feasibility it resulted in solutions that were identical with the vectorized penalty ADM. The non-randomized PADM will be referred to as PADM.

4.3.2 Objective Function

The relative gap between the heuristic solution and the exact solution was used to measure the performance of the PADM algorithms. For a method $m \in \{\text{PADM}, \text{Random PADM}\}$, the relative gap is given by

$$\text{Solution Gap} = \frac{f^{(m)} - f^{(\text{Exact})}}{f^{(\text{Exact})}}$$

where $f(u,v) = f(x,z,w,b) = x^\intercal \Sigma x + \frac{1}{2}||w||^2$. For both, methods the solution gaps are computed based only on their respective feasible instances. The empirical cdf (ECDF) of the solution gap is computed for both the PADM and the randomized PADM in Figure 7. It is clear from Figure 7 that the Randomized PADM has a higher probability of having a lower solution gap.

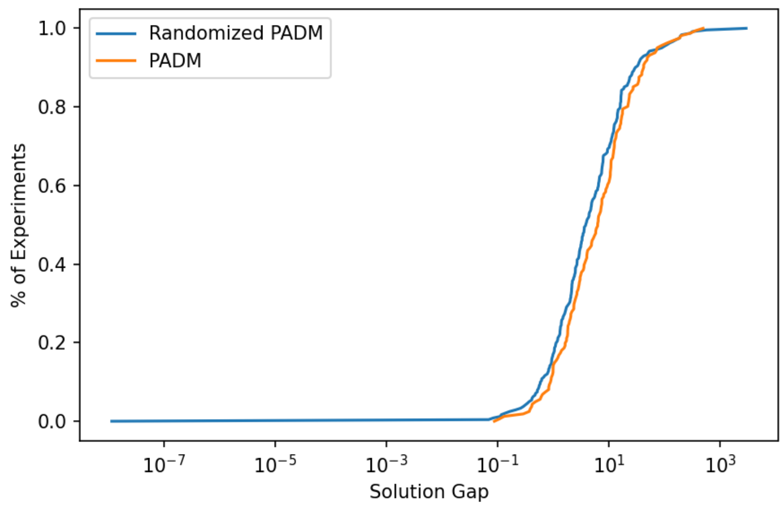

Figure 7: Relative Solution Gap: Randomized PADM (blue), PADM (orange)

Although the solution gaps are significant, the randomized PADM has a definite advantage over the PADM. In the case of the PADM, 5.1% of instances have an optimality gap less than 50% whereas, in the case of the random PADM, 7.3% of instances have an optimality gap under 50%.

4.3.3 Selection Accuracy

In addition to the solution gap, one can also calculate the selection accuracy. The selection accuracy is the percentage of correctly chosen assets in the portfolio. The selection accuracy is given by:

$$\text{Selection Accuracy} = 1 - \frac{|z^m - z^{\text{Exact}}|}{\sum_i z_i^m + z_i^{\text{Exact}}} \quad m \in \{\text{PADM}, \text{Random PADM}\}$$

As evidenced by Figure 8, the selection accuracy obtained by the randomized PADM yields a more accurate selection over its deterministic counterpart. The median selection accuracy for the random PADM is 76% versus 61% for the PADM.

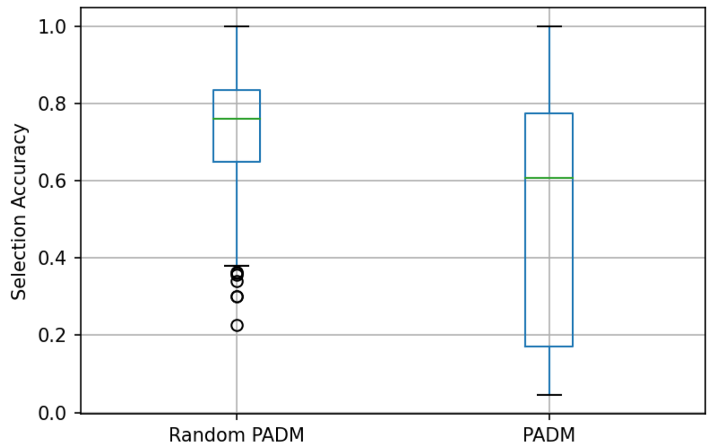

Figure 8: Selection Accuracy: Random PADM vs. PADM

4.3.4 Timing

Proceeding in a similar fashion to Costa et al. (2020), to compare the running time between the random PADM, the PADM and the commercial MIP solver the performance ratio is calculated (Dolan and Moré, 2002):

$$\frac{t_q^m}{\min\{t_q^a : a \in M\}} \quad \text{where} \quad M = \{\text{Random PADM},\ \text{PADM},\ \text{Exact}\}$$

for the problem instance q, with running times denoted by t. Sorting the performance ratios and plotting the ECDF yields Figure 9. From this, it is clear that the PADM is the fastest method most of the time $\approx 83\%$. The exact method seems to be the fastest $\approx 16\%$ of the time, and the random method is rarely the fastest. Still, the random PADM has desirable characteristics over the exact solution. For example, $\approx 85\%$ of the random PADM running times are within a factor of 4 of the fastest running time, whereas only $\approx 68\%$ of the exact solution times are within a factor of 4 of the fastest running time.

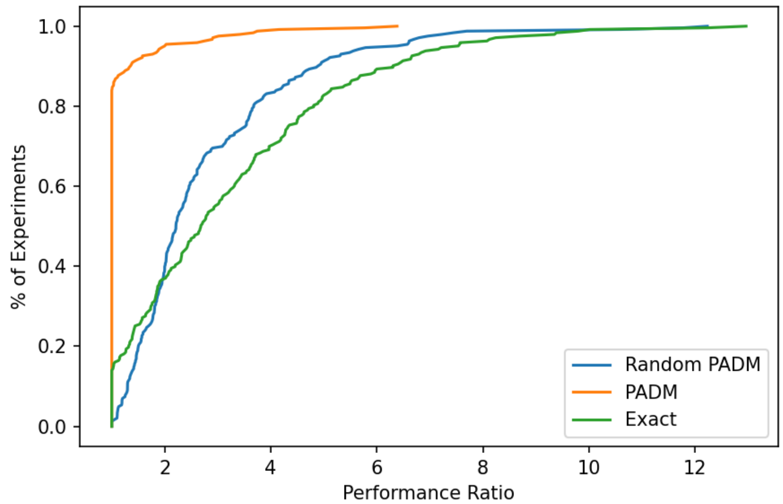

Figure 9: Performance Ratio ECDF

Furthermore, even though the Random PADM is not as fast as the PADM, it still has the desired quality that it is on average faster than the exact method and has less variable running times. Table 1 highlights this phenomenon.

	95th Percentile	**Mean**	**Median**
Exact	172.18	77.75	70.50
PADM	39.54	26.96	28.29
Randomized PADM	79.96	57.85	59.92

Table 1: Run-time statistics

4.4 Monthly Rebalancing Experiment

This section aims to evaluate the out-of-sample performance of the proposed SVM-MVO approach by testing it as a monthly rebalancing scheme. All but one problem instance is considered in this section because the last instance does not have an out-of-sample return. In general, on the first trade date of the month, the portfolio is rebalanced according to its specified optimization scheme. Let $t = 1, 2, \dots, 244 - 1$ denote the index for the rebalancing times. For all monthly rebalancing points, the target return is given by the average market expected market return as of that time marked up by a percentage $\kappa \in \mathbb{R}$:

$$(\frac{1+\kappa}{N})(\mathbf{1}^\intercal \mu_\mathbf{t})$$

κ is set to 10% for each rebalancing point. Also, the asset limit K for all rebalancing points is set to be $\lfloor 0.9N \rfloor$. The *benchmark model* in this section is the cardinality constrained mean-variance problem: Card-MVO. Three different SVM-MVO model specifications will be introduced and tested as monthly rebalancing policies. In addition to the constraints placed on the rebalancing optimization problem, at each rebalancing point, a constraint on the allowable portfolio turnover is placed:

$$||x^{(t)} - x^{(t-1)}||_1 \leq C_0 \quad t = 2, 3, \ldots, 244 - 1$$

Similar to Costa and Kwon (2019), for each experiment, C_0 will be taken to be one of the values in the set $\{0.75, 1.00, 1.25\}$ for each of the optimization procedures and will be fixed throughout the testing period. In addition to a turnover constraint on x, it also makes sense to place a turnover constraint on the investment policy defined by the separating hyperplane. Put simply, one month of realized returns should not change an entire investment strategy. For some η, the policy constraint is set so that:

$$|w^{(t)} - w^{(t-1)}| \leq \eta ||w^{(t-1)}||_1$$

which restricts $w^{(t)}$ from completely changing at each rebalancing point t. η is taken to be 0.1. Lastly, one must decide what factors y_i should be used in the SVM-MVO 2 model. For simplicity, the factors selected will be inspired by empirical finance and business knowledge. The first SVM-MVO model considered will be inspired by observations from empirical research in finance that claim that volatility has a negative impact on asset returns. Bae et al. (2007) summarizes the literature on the volatility effect and states that as volatility increases, the risk premium on equities should also increase and thus, the required return on stocks should increase, resulting in a price correction - this effect is called volatility feedback. Given the volatility feedback hypothesis, one can argue that introducing a limit on volatility may offer better risk-adjusted returns because volatility will be less (by nature of introducing a threshold), and the volatility feedback loop will less negatively impact the returns. Furthermore, as discussed and shown in Section 3.2, introducing a volatility limit on the select assets is equivalent to having full aversion to assets with estimation error of the mean return μ_i exceeding a threshold. Thus, a natural hyperplane to use satisfies the following:

$$(-\Sigma_{ii})w + b = 0 \text{ where } b \in \mathbb{R}_+,\ w \geq \mathbb{R}_+, \text{ and } \sigma_i \text{ is the volatility of asset } i$$

Adding non-negativity constraints on w and b and using $-\sigma_i$ ensures that the hyperplane models a threshold on volatility such that $z_i = \mathbf{1}_{\{\sigma_i \leq \frac{b}{w}\}}$. This model will use the exact MIP solver with $C = 2500$ at each rebalancing point since $y_i = -\sigma_i$ is a scalar. This model is hence coined *Exact - Vol*.

Although *Exact-Vol* is parsimonious, loosely speaking, it is not overly expressive; there are many other empirical observations from the finance literature. A higher dimensional hyperplane with more factors may offer more nuanced and better-performing investment criteria. For example, flexible criteria could consider popular investment ratios such as book/price ratios, price to earnings ratios and cash-flow to market capitalization. Dechow et al. (2001) show that historically, companies with lower market-cap to price ratios (PB), lower price to earnings ratios (PE), and higher cash-flow to market capitalization are not targeted by short-sellers. Hou et al. (2006) provide a review of studies that verify a common variation between the ratios above and stock returns. Based on the empirical evidence regarding financial ratios and equity returns, one may be inclined to select the separating hyperplane such that

$$[-\text{INV1},\ \text{INV3},\ \text{PRO6}]w + b = 0 \text{ with } b \in \mathbb{R}^3_+,\ w \geq \mathbb{R}^3_+$$

where INV1, INV3, and PRO6, correspond to the PE ratio, the inverse PB ratio and the cash-flow to market capitalization ratio. The hyperplane above encourages small PE ratios, small PB ratios and, high cash-flow to market capitalization companies to be eligible for capital allocation. In this case, since the exact solve takes $\approx 120s$, the deterministic PADM is used to rebalance the portfolio at each trade time with initialization: $\rho_i = \frac{2500}{2^{(15)}}\ \forall i \in [N]$. The third model will be referred to as *PADM-Fundamentals* All figures and results in this section were generated from the notebook: Results.ipynb.

4.4.1 Portfolio Metrics

This section presents the out of sample results for the SVM-MVO portfolios. First, it is assumed that each one of the portfolios starts with \$1.00. For each model
$m \in \{\text{Benchmark}, \text{Exact - Vol}, \text{PADM - Fundamental}\}$ and each turnover limit $C_0 \in \{0.75, 1.00, 1.25\}$ the returns over the trade dates are denoted by $r_t^{(m,C_0)} \quad t \in \{2, 3, \ldots, 244\}$. The cumulative wealth up to time t, for a particular turnover and model combination is then given by $W_t(C_0, m) = \prod_{t^* \in \{2,3,\ldots,t\}}(1 + r_{t^*}^{(m,C_0)})$. Figures 10, 11, and 12, show the relative wealth $W_t(C_0, m)/W_t(C_0, \text{Benchmark})$ for each SVM-MVO model $m \in \{\text{Exact - Vol}, \text{PADM - Fundamental}\}$ for values of $C_0 = 0.75, 1.00,$ and 1.25 respectively. For the lower turnover constraint $C_0 = 0.75$, the SVM-MVO models do not display much out performance relative

to the Card-MVO. However, for the higher turnover allowances, the SVM-MVO models do out-perform the baseline model. Post - 2009 the volatility thresholded optimizer exhibits strong returns. The Fundamentals strategy displayed out-performance pre-2009 relative to both of the other models. Overall, the volatility thresholded model performs the best when measured by wealth.

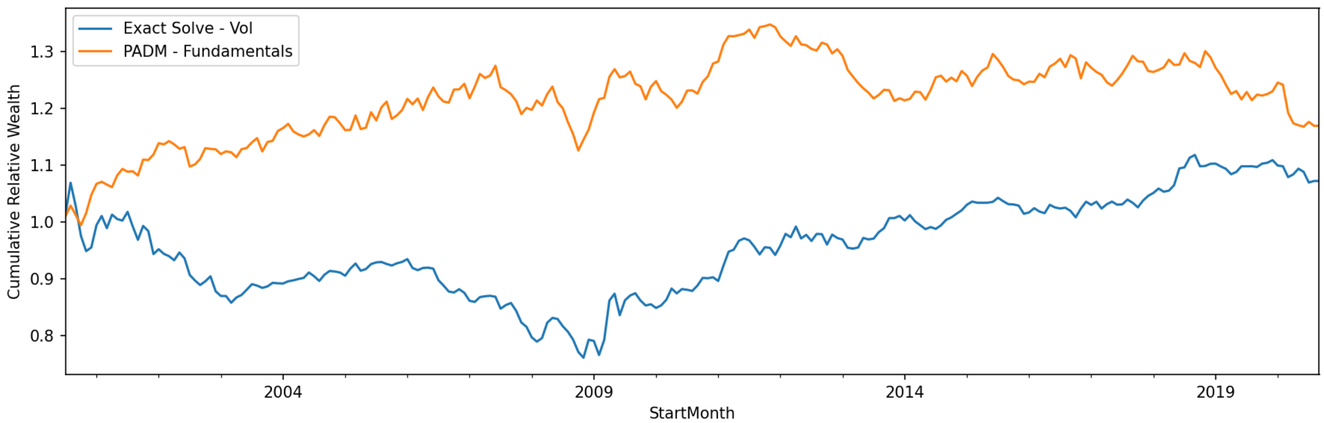

Figure 10: Relative portfolio wealth compared to benchmark with $C_0 = 0.75$

Figure 11: Relative portfolio wealth compared to benchmark with $C_0 = 1.00$

Figure 12: Relative portfolio wealth compared to benchmark with $C_0 = 1.25$

The following portfolio statistics are computed and shown in Table 2.

- Annualized arithmetic return: μ (ann.) $= 12\ \frac{1}{243} \sum_t r_t^{(m,C_0)}$
- Annualized volatility: σ (ann.) $= \sqrt{12}\ \mathbb{S}d\{r_t^{(m,C_0)}\}$ where $\mathbb{S}d$ is the standard deviation.
- Sharpe Ratio = μ (ann.) $/\sigma$ (ann.)

$\lVert x^{(t)} - x^{(t-1)} \rVert_1 \leq 0.75$	μ (ann.)	σ (ann.)	**Sharpe Ratio**
Card-MVO	0.141941	**0.123282**	**1.151354**
Exact Solve - Vol	0.148542	0.146378	1.014782
PADM - Fundamentals	**0.151041**	0.132752	1.137772
$\lVert x^{(t)} - x^{(t-1)} \rVert_1 \leq 1.00$	μ (ann.)	σ (ann.)	**Sharpe Ratio**
Card-MVO	0.134525	**0.122279**	1.10015
Exact Solve - Vol	**0.159118**	0.140469	**1.132761**
PADM - Fundamentals	0.151743	0.134802	1.125673
$\lVert x^{(t)} - x^{(t-1)} \rVert_1 \leq 1.25$	μ (ann.)	σ (ann.)	**Sharpe Ratio**
Card-MVO	0.131436	**0.12042**	1.091488
Exact Solve - Vol	**0.153209**	0.13913	1.101195
PADM - Fundamentals	0.144856	0.131187	**1.104195**

Table 2: Portfolio statistics: Annualized arithmetic return, Annualized Volatility, Sharpe Ratio. Best properties in bold font

Similar to the case of the terminal wealth, the nominal problem performed preferably in the case of a turnover limit of 0.75. However, in the case of higher turnover limits both *Exact - Vol* and *PADM - Fundamentals* out-perform Card-MVO in terms of the realized Sharpe ratio.

5 Conclusion

This work proposes a novel approach to the problem of capital allocation by enforcing that investors determine an asset's eligibility for capital based on factors other than expected returns and covariances. It is proposed that investors determine a thresholding rule for allocation eligibility based on fundamental factors, i.e. the asset eligibility logic takes the form of a support vector machine. The motivations for this approach are three-fold: i) It has been documented that investors rely on specific criteria other than expected returns and covariances to evaluate potential investments, ii) constraints on traditional portfolio optimization techniques have been shown theoretically and empirically to improve allocation outcomes, and iii) many existing approaches that combine portfolio optimization and learning techniques do not fully treat portfolio risk and focus exclusively on return. The representative investor simultaneously solves for a mean-variance optimal portfolio, with a cardinality constraint such that all the assets included in the portfolio are on the positive side of a hyperplane defined by asset features. Intuitively, this means that the investor seeks a portfolio of minimum risk subject to a return constraint such that the assets in the portfolio fit within their investment policy.

A mixed binary quadratic program is introduced to simultaneously solve a separating hyperplane and an asset allocation that minimizes the portfolio variance subject to return constraints. The proposed optimization model addresses all three of the aforementioned considerations and furthermore is related to well tested portfolio norm penalization techniques for the particular instances of enforcing a volatility limit on the assets that are eligible for investment. The proposed optimization problem is challenging to solve in the case of a large number of assets and features per asset – to address this; two variants of a PADM are proposed to arrive at a partial minimum solution. The deterministic variant arrives at a partial minimum solution in less than half the time of an exact commercial MIP solver on average but does not find a solution that can be completely separated by a hyper-plane as often as the commercial solver. The second variant of the proposed PADM: a randomized restart PADM obtains SVM feasible solutions more often than its deterministic counterpart but in more time. Still, the randomized PADM shows a $\approx 25\%$ improvement in solution times over the commercial MIP solver on average. Furthermore, it is shown by a monthly rebalancing experiment consisting of 286 stocks that depending on the choice of asset attributes, adding a constraint-based on investment fundamentals can result in improved risk-adjusted returns. Further directions for future work include parallel implementations of the randomized PADM algorithm and statistical learning frameworks for factor selection.

References

Anis, H., & Kwon, R. (2020). Cardinality constrained risk parity portfolios. *Available at SSRN 3805592*.

Aroussi, R. (2021). *Yfinance*. https://pypi.org/project/yfinance/

Bae, J., Kim, C.-J., & Nelson, C. (2007). Why are stock returns and volatility negatively correlated? *Journal of Empirical Finance*, *14*, 41–58. https://doi.org/10.1016/j.jempfin.2006.04.005

Behr, P., Guettler, A., & Miebs, F. (2013). On portfolio optimization: Imposing the right constraints. *Journal of Banking & Finance*, *37*(4), 1232–1242.

Bertsimas, D., & Cory-Wright, R. (2021). A scalable algorithm for sparse portfolio selection.

Bixby, R. E. (2012). A brief history of linear and mixed-integer programming computation.

Black, F., & Litterman, R. B. (1991). Asset allocation. *The Journal of Fixed Income*, *1*(2), 7–18.

Blanchet, J., Chen, L., & Zhou, X. Y. (2018). Distributionally robust mean-variance portfolio selection with wasserstein distances. *arXiv preprint arXiv:1802.04885*.

Burgard, J. P., Costa, C. M., & Schmidt, M. (n.d.). Decomposition methods for robustified k-means clustering problems: If less conservative does not mean less bad.

Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, *27*(13), 1271–1302.

Cho, E., & Cho, M. J. (2008). Variance of sample variance. *Section on Survey Research Methods–JSM*, *2*, 1291–1293.

Clarke, R. G., de Silva, H., & Murdock, R. (2005). A factor approach to asset allocation. *The Journal of Portfolio Management*, *32*(1), 10–21.

Costa, C. M., Kreber, D., & Schmidt, M. (2020). An alternating method for cardinality-constrained optimization: A computational study for the best subset selection and sparse portfolio problems.

Costa, G., & Kwon, R. H. (2019). Risk parity portfolio optimization under a markov regime-switching framework. *Quantitative Finance*, *19*(3), 453–471.

Dechow, P., Hutton, A., Meulbroek, L., & Sloan, R. (2001). Short sellers, fundamental analysis and stock returns. *Journal of Financial Economics*, *61*, 77–106. https://doi.org/10.1016/S0304-405X(01)00056-3

Dolan, E. D., & Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*, *91*(2), 201–213.

Fan, A., & Palaniswami, M. (2001). Stock selection using support vector machines. *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*, *3*, 1793–1798.

Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction.* springer open.

Geißler, B., Morsi, A., Schewe, L., & Schmidt, M. (2015). Solving power-constrained gas transportation problems using an mip-based alternating direction method. *Computers & Chemical Engineering*, *82*, 303–317.

Geißler, B., Morsi, A., Schewe, L., & Schmidt, M. (2017). Penalty alternating direction methods for mixed-integer optimization: A new view on feasibility pumps.

Geißler, B., Morsi, A., Schewe, L., & Schmidt, M. (2018). Solving highly detailed gas transport minlps: Block separability and penalty alternating direction methods. *INFORMS Journal on Computing*, *30*, 309–323.

Goettlich, S., Hante, F., Potschka, A., & Schewe, L. (2021). Penalty alternating direction methods for mixed-integer optimal control with combinatorial constraints. *Mathematical Programming*, *188*.

Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Math. Oper. Res.*, *28*, 1–38. https://doi.org/10.1287/moor.28.1.1.14260

Ho, M., Sun, Z., & Xin, J. (2015). Weighted elastic net penalized mean-variance portfolio design and computation. *SIAM Journal on Financial Mathematics*, *6*(1), 1220–1244.

Hou, K., Karolyi, G., & Kho, B.-C. (2006). What factors drive global stock returns? *Ohio State University, Charles A. Dice Center for Research in Financial Economics, Working Paper Series*, *24*. https://doi.org/10.2139/ssrn.908345

Jagannathan, R., & Ma, T. (2002). *Risk reduction in large portfolios: Why imposing the wrong constraints helps* (Working Paper No. 8922). National Bureau of Economic Research.

Kleinert, T., & Schmidt, M. (2020). Computing feasible points of bilevel problems with a penalty alternating direction method. *INFORMS Journal on Computing*, *33*.

Lai, T. L., Xing, H., & Chen, Z. (2011). Mean–variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, *5*(2A), 798–823.

List of s&p 500 companies. (2020). Retrieved April 5, 2020, from https://en.wikipedia.org/wiki/List_of_S%5C&P_500_companies

Longerstaey, J., & Spencer, M. (1996). Riskmetricstm—technical document.

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

Menon, A. K. (2009). Large-scale support vector machines: Algorithms and theory.

Meucci, A. (2010). Quant nugget 2: Linear vs. compounded returns – common pitfalls in portfolio management. *GARP Risk Professional*.

Olivares-Nadal, A. V., & DeMiguel, V. (2018). Technical note—a robust perspective on transaction costs in portfolio optimization. *Operations Research*, *66*(3), 733–739.

Pafka, S., Potters, M., & Kondor, I. (2004). Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization. *arXiv.org, Quantitative Finance Papers*.

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655.

Pedersen, L. H. (2015). *Efficiently inefficient: How smart money invests and market prices are determined*. Princeton University Press. http://www.jstor.org/stable/j.ctt1287knh

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Schewe, L., Schmidt, M., & Weninger, D. (2020a). A decomposition heuristic for mixed-integer supply chain problems. *Operations Research Letters*, *48*.

Schewe, L., Schmidt, M., & Weninger, D. (2020b). A decomposition heuristic for mixed-integer supply chain problems. *Operations Research Letters*, *48*(3), 225–232.

Tarmast, G. (2001). Multivariate log-normal distribution.

Tillmann, A. M., Bienstock, D., Lodi, A., & Schwartz, A. (2021). Cardinality minimization, constraints, and regularization: A survey.

Wharton Research Data Services, W. (2021). *Fundamentals quarterly compustat - capital iq from standard poor's*. https://wrds-www.wharton.upenn.edu/

Wu, L., & Yang, Y. (2014). Nonnegative elastic net and application in index tracking. *Applied Mathematics and Computation*, *227*, 541–552.

Yu, X., Cui, G., Yang, J., Li, J., & Kong, L. (2020). Quadratic optimization for unimodular sequence design via an adpm framework. *IEEE Transactions on Signal Processing*, *68*, 3619–3634.

Appendix

5.1 Variable Definitions

5.1.1 Fundamentals Quarterly Compustat

WRDS CODE	Description
actq	ACTQ – Current Assets - Total (ACTQ)
ancq	ANCQ – Non-Current Assets - Total (ANCQ)
atq	ATQ – Assets - Total (ATQ)
ceqq	CEQQ – Common/Ordinary Equity - Total (CEQQ)
cheq	CHEQ – Cash and Short-Term Investments (CHEQ)
cogsq	COGSQ – Cost of Goods Sold (COGSQ)
cshiq	CSHIQ – Common Shares Issued (CSHIQ)
cshopq	CSHOPQ – Total Shares Repurchased - Quarter (CSHOPQ)
cshoq	CSHOQ – Common Shares Outstanding (CSHOQ)
cstkq	CSTKQ – Common/Ordinary Stock (Capital) (CSTKQ)
dlcq	DLCQ – Debt in Current Liabilities (DLCQ)
dlttq	DLTTQ – Long-Term Debt - Total (DLTTQ)
dpq	DPQ – Depreciation and Amortization - Total (DPQ)
epspi12	EPSPI12 – Earnings Per Share (Basic) - Including Extraordinary Items - 12 Months Movi (EPSPI12)
epspiq	EPSPIQ – Earnings Per Share (Basic) - Including Extraordinary Items (EPSPIQ)
epspxq	EPSPXQ – Earnings Per Share (Basic) - Excluding Extraordinary Items (EPSPXQ)
intanq	INTANQ – Intangible Assets - Total (INTANQ)
invtq	INVTQ – Inventories - Total (INVTQ)
lctq	LCTQ – Current Liabilities - Total (LCTQ)
lltq	LLTQ – Long-Term Liabilities (Total) (LLTQ)
niq	NIQ – Net Income (Loss) (NIQ)
oibdpq	OIBDPQ – Operating Income Before Depreciation - Quarterly (OIBDPQ)
opepsq	OPEPSQ – Earnings Per Share from Operations (OPEPSQ)
pncepsq	PNCEPSQ – Core Pension Adjustment Basic EPS Effect (PNCEPSQ)
rectq	RECTQ – Receivables - Total (RECTQ)
req	REQ – Retained Earnings (REQ)
saleq	SALEQ – Sales/Turnover (Net) (SALEQ)
teqq	TEQQ – Stockholders Equity - Total (TEQQ)
capxy	CAPXY – Capital Expenditures (CAPXY)
dltisy	DLTISY – Long-Term Debt - Issuance (DLTISY)
dltry	DLTRY – Long-Term Debt - Reduction (DLTRY)
niy	NIY – Net Income (Loss) (NIY)
oancfy	OANCFY – Operating Activities - Net Cash Flow (OANCFY)
txty	TXTY – Income Taxes - Total (TXTY)
dvpspq	DVPSPQ – Dividends per Share - Pay Date - Quarter (DVPSPQ)
mkvaltq	MKVALTQ – Market Value - Total (MKVALTQ)
prccq	PRCCQ – Price Close - Quarter (PRCCQ)

5.1.2 Derived Fundamental Factors

Variable ID	Compustat Formula	Description
ROC1	(txty+req)/atq	Earnings before tax by Assets
ROC2	(txty + req)/(dlttq + ceqq)	Earnings before tax by Capital = Debt plus Equity
ROC3	niq/(dlttq + ceqq)	Net Income by Capital
ROC4	oancfy/atq	Cashflow by assets
ROC5	oancfy/(dlttq + ceqq)	Cashflow by capital
INV1	prccq/epspxq	PE Ratio
INV2	dvpspq/prccq	debt to Equity
INV3	ceqq/(prccq*choq)	Equity to Market Cap
INV4	(atq - intanq - lltq - lctq)/(mkvaltq/prccq)	Net Tangible Assets per Share
PRO1	txty+req/saleq	Earnings before tax by Sales
PRO2	req/saleq	Earnings after tax by Sales
PRO3	niq/saleq	Net Income by Sales
PRO4	oancfy/saleq	Cashflow/Sales
PRO5	req/ceqq	Earnings by Equity
PRO6	oancfy/(prccq*choq)	Cashflow by Market Cap
PRO7	req/oancfy	Earnings by Equity
LEV1	dlttq/ceqq	Debt to Equity
LEV2	(LLTQ + LCTQ) /(dlttq+ceqq)	Liabilities/Capital
LEV3	(LLTQ + LCTQ) /(ceqq)	Liabilities/Equity
LEV4	atq/ceqq	Assets by Equity
LEV5	atq/(prccq*choq)	Assets by Market Value
RET1	req/atq	Return on Assets
LIQ1	actq/lctq	Current Assets to Current Liabilities
LIQ2	lctq/atq	Current Liabilities to Total Assets
LIQ3	lctq/ceqq	Current Liabilities to Equity
LIQ4	dlttq/lltq	Long term Debt to Total Liabilities
LIQ5	cheq/lctq	Cash to current liabilities
RISK1	txty+req/LCTQ	Pre tax profit by current liabilities
RISK2	req/LCTQ	Profit by current liabilities
RISK3	oancfy/LCTQ	Cashflow by Current Liabilities
GRW1	saleq	Sales growth
GRW2	atq	Asset growth
GRW3	req	Retained earnings growth
GRW4	epspxq	EPS growth
GRW5	oancfy	Cashflow growth
GRW6	teqq	Shareholders equity growth
GRW7	capxy	Capex growth
GRW8	rectq	Receivables

5.2 Code Repository

https://github.com/davidislip/SVM-and-MVO/

5.3 Bootstrapping and Central Limit Theorems

This section provides the rationale for the equations for quantifying the estimation errors associated with the mean and variance of asset returns. The equations are presented below for convenience:

$$\alpha = \frac{1}{\sqrt{T_{train}}}\Phi^{-1}\Big(\frac{p_1+1}{2}\Big)\sqrt{\mathrm{CM}_4(X)-\sigma^4} = \sqrt{\frac{\kappa-1}{T_{train}}}\Phi^{-1}\Big(\frac{p_1+1}{2}\Big)\sigma^2 \tag{11}$$

and

$$\beta = \frac{\sigma}{\sqrt{T_{train}}}\Phi^{-1}\Big(\frac{p_2+1}{2}\Big) \tag{12}$$

For simplicity the asset index i is dropped. The central limit theorem states that if $\{X_t\}_{t\in[T]}$ are T samples drawn iid from a distribution with true mean $\bar{\mu}$ and variance σ^2, then $Z = \lim_{T\to\infty}\sqrt{T}(\frac{1}{T}\sum X_T - \mu)/\sigma$ follows a standard normal distribution. The mean and variance of the sample estimate $\frac{1}{T}\sum_t X_t = \frac{1}{T}\sum_t X_t^2$ are $\bar{\mu}$ and $\frac{\sigma}{\sqrt{T}}$, therefore, for large T:

$$\frac{1}{T}\sum_t X_t \overset{d}{\to} N(\bar{\mu}, \frac{\sigma}{\sqrt{T}}) \tag{13}$$

The case of determining an approximation for the distribution of Σ is more complex because obtaining the asymptotic distribution of $\Sigma = \frac{1}{T}\sum X_i^2 - (\frac{1}{T}\sum_t X_t)^2$ requires the use of the delta method. Let W_2 and W_1 denote $\frac{1}{T}\sum X_i^2$ and $\frac{1}{T}\sum_t X_t$ respectively. W_1 and W_2 converge to a joint normal distribution by the multivariate central limit theorem with covariance $\Xi_{W_j,W_k} = \mathbb{E}[W_jW_k] - \mathbb{E}[W_j]\mathbb{E}[W_k]$ and true mean $\bar{\mathbf{W}}$ i.e $\sqrt{T}(\mathbf{W}-\bar{\mathbf{W}}) \overset{\mathbf{d}}{\to} \mathbf{N}(\mathbf{0},\Xi)$. In general, for any differentiable $g(W_1,W_2)$ one can write the following first-order taylor expansion $g(W_1,W_2) \approx g(\bar{\mathbf{W}}) + \nabla g(\bar{\mathbf{W}})(\mathbf{W}-\bar{\mathbf{W}})$. Using this approximation $\mathbb{V}[g(\mathbf{W})] \approx \nabla\mathbf{g}(\mathbf{W})^\intercal\Xi\nabla\mathbf{g}(\mathbf{W})$ and as such $\sqrt{T}(g(\mathbf{W}) - \mathbf{g}(\bar{\mathbf{W}})) \overset{\mathbf{d}}{\to} \mathbf{N}(\mathbf{0}, \nabla\mathbf{g}(\mathbf{W})^\intercal\Xi\nabla\mathbf{g}(\mathbf{W}))$. One can set $g(W_1,W_2) = W_2 - W_1^2$ and use the first order delta method as described above to obtain asymptotic convergence condition.

The first order delta method does not always work. Convergence to normality does not hold for Bernoulli variables with $p=0.5$ because the variance of g ends up being zero. To combat this, the second order taylor approximation must be used. For simplicity, in the case of Bernoulli random variables g can be expressed as a function of a single random parameter: $\hat{p}$ since $\frac{1}{T}\sum_t X_t = \frac{1}{T}\sum_t X_t^2$. In this case, $g(\hat{p}) = \hat{p}(1-\hat{p})$ and the true mean is $p = 1/2$. It follows by taking the second order taylor expansion around the true mean i.e $\hat{p} = 0.5$ that, $\hat{p}(1-\hat{p}) = 1/4 + 0*(\hat{p}-0.5) + 1/2(-2)(\hat{p}-1/2)^2$. Since $\hat{p} = \frac{1}{T}\sum_t X_t$ is asymptotically normal by the central limit theorem and since $\hat{p}(1-\hat{p})$ is a constant minus an asymptotically normal variable squared, it is then true that $\hat{p}(1-\hat{p})$ is asymptotically negative χ^2 and is not normal as it would be in the case that $g'(p) \neq 0$. In the case that the first order delta method is applicable and the first order taylor series introduces uncertainty in g, it follows:

$$\frac{1}{T}\sum X_i^2 - (\frac{1}{T}\sum_t X_t)^2 \xrightarrow{d} N\big(\sigma^2, \frac{1}{\sqrt{T}}\sqrt{\mathrm{CM}_4(X) - \sigma^4 + \mathbf{O}(T^{-2})}\big) \tag{14}$$

where the expression for variance of the sample variance is derived by Cho and Cho (2008).

Lastly, one can express the quantile of the absolute deviation of a normal variables in a closed form equation. The results can be derived as follows: letting $Z \sim N(0,1)$ then the quantile of the absolute deviation satisfies $P(|Z| \leq x) = p$ which is the same as $2\Phi(x) - 1 = p$ which implies

$$x = \Phi^{-1}(\frac{p+1}{2}) \tag{15}$$

Applying equation 15 to equations 13 and 14 yields equations 12 and 11. The notebook titled Bootstrapping.ipynb verifies these results for various distributions.