Integration of Support Vector Machines and Mean-Variance Optimization for Capital Allocation

David Islip [1] Roy H. Kwon [1] Seongmoon Kim [2]

[1]Department of Mechanical and Industrial Engineering, University of Toronto
[2]School of Business, Yonsei University

September 13th, 2024

Cardinality Constrained Portfolio Optimization

Markowitz (1952) was the first to propose selecting a portfolio of assets to optimize a trade-off between risk and expected return.

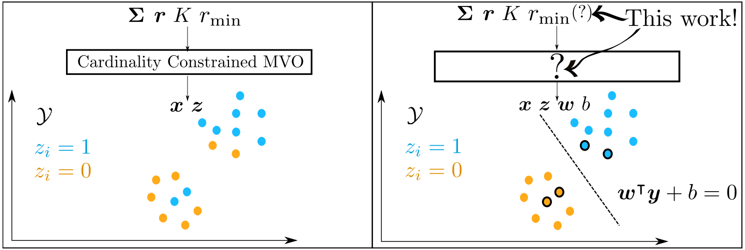
Investors generally prefer portfolios with fewer assets (lower cardinality) due to frictions such as transaction costs, yielding the following optimization:

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{1}^\top \boldsymbol{x} = 1, \quad \boldsymbol{r}^\top \boldsymbol{x} \geq r_{\min}, \quad |\mathrm{supp}(\boldsymbol{x})| \leq K, \quad \boldsymbol{x} \geq 0 \tag{Card-MVO}$$

where

- A portfolio is represented by $\boldsymbol{x} \in \mathbb{R}_+^N$
- $\boldsymbol{r} \in \mathbb{R}^N$ denote the mean of the random returns
- $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ denote the covariance of the random returns
- $r_{\min}$ denotes the investor's minimum required expected return
- $K \in [N]$ representing a cardinality limit
- $|\mathrm{supp}(\boldsymbol{x})|$ denotes the number assets receiving capital for investment
- Mixed integer quadratic formulation: Let $\boldsymbol{z} \in \{0,1\}^N$ denote binary variables such that $z_i = 0 \implies x_i = 0 \; \forall i \in [N]$

Objective of this work



- Card-MVO includes assets that do not respect the structure of the feature space
- The desired outcome (right) showcases how an asset screener in the form of a hyperplane would capture the inherent structure of the feature space
- In the desired outcome, the circled assets differ from the decision made by Card-MVO because they violate an investor's sense of eligibility based on an asset-screening hyperplane

Model Rationale: jointly identify a portfolio $\boldsymbol{x}$ and hyperplane $H(\boldsymbol{w}, b)$ such that the portfolio has desirable risk-return properties and the hyperplane accurately classifies the assets.

Mixed Integer Programming Approach

- An asset screener defined by a hyperplane $(\boldsymbol{w}, b)$ with features $\boldsymbol{y}^{(i)} \in \mathcal{Y} \subset \mathbb{R}^p$ classifies assets as eligible $(z_i = 1)$ or ineligible $(z_i = 0)$
- Binary variables $\boldsymbol{t} \in \{0,1\}^p$ are introduced such that $t_j = 0 \implies w_j = 0 \ \forall j \in [p]$

$$\min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{t}} \quad \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} + \epsilon \left(\frac{1}{2} \|\boldsymbol{w}\|_2^2 + \frac{C}{N} \mathbf{1}^\top \boldsymbol{\xi} \right) \tag{SVM-MVO}$$

$$\text{s.t.} \quad -M(1 - z_i) + 1 - \xi_i \le (\boldsymbol{y}^{(i)})^\intercal \boldsymbol{w} + b \le M z_i - 1 + \xi_i, \quad \forall i \in [N] \tag{1}$$

$$\boldsymbol{r}^\top \boldsymbol{x} \ge r_{\min}, \ \mathbf{1}^\top \boldsymbol{z} \le K, \ \boldsymbol{x} \le \boldsymbol{z} \tag{2}$$

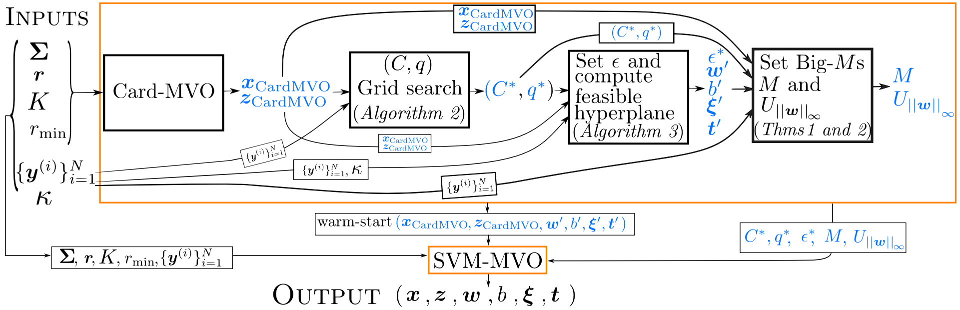
$$\mathbf{1}^\top \boldsymbol{t} \le q, \ -U_{\|\boldsymbol{w}\|_\infty} \boldsymbol{t} \le \boldsymbol{w} \le U_{\|\boldsymbol{w}\|_\infty} \boldsymbol{t} \tag{3}$$

$$\boldsymbol{x} \in \Delta_N, \ \boldsymbol{z} \in \{0,1\}^N, \ (\boldsymbol{w}, b, \boldsymbol{\xi}) \in \mathbb{R}^{p+1} \times \mathbb{R}_+^N, \ \boldsymbol{t} \in \{0,1\}^p \tag{4}$$

- (1) - (3) model the hyperplane separation, portfolio and feature-selection constraints, and Δ_N is simplex.
- $U_{\|\boldsymbol{w}\|_\infty}$ and M are big-M constants, and C represents the relative preference for separation over margin, and ϵ weighs the asset-screening objective against the portfolio objective.

SVM-MVO determines $H(\boldsymbol{w}, b)$ by solving the support vector machine (SVM) problem with features $\{y^{(i)}\}_{i=1}^N$ and labels $2\boldsymbol{z} - 1$ (Cortes and Vapnik, 1995).

Unified Parameter Selection Strategy



The proposed strategy

- only requires κ and $\{\boldsymbol{y}^{(i)}\}_{i=1}^{N}$ as additional inputs (compared to Card-MVO)
- ensures the risk of the resulting portfolio is within $1+\kappa$ of the smallest possible risk
- selects big-Ms so that no optimal solutions are excluded
- aims to avoid degeneracy ($\boldsymbol{w} = \mathbf{0}$)

In-Sample Experiments

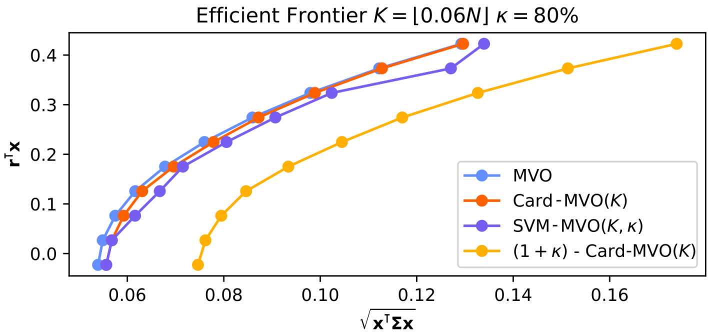


Figure: Efficient Frontier for MVO, Card-MVO, and the SVM-MVO model

- ϵ is set so that the risk of any SVM-MVO solution is within $(1+\kappa)$ of Card-MVO's risk

Out-of-Sample Experiments

Figure: Wealth relative to Card-MVO by turnover limit C_0 and cardinality limit K

Conclusion

- **A new portfolio selection model**:
  - Augments cardinality-constrained optimization with a preference for portfolios where a low-dimensional hyperplane separates eligible and ineligible assets.
  - Presents convex mixed-integer quadratic programming models for joint portfolio and hyperplane selection that are amenable to commercial solvers.
- **A principled parameter selection strategy**: Ensures valid big-M values, risk control, and informative hyperplanes.
- **Empirical demonstration of financial performance**: Results show improved out-of-sample risk-adjusted returns compared to cardinality-constrained mean-variance optimization.

Thank you!
I look forward to your questions

Integrated SVM and MVO – Degeneracy

- $\boldsymbol{w} = \mathbf{0}$ is referred to as SVM degeneracy (Rifkin et al., 1999)
- when $K << N$, SVM-MVO will tend to set $(\boldsymbol{w}, b) = (\mathbf{0}, -1)$ yielding a hyperplane that correctly classifies the majority of the assets
- To combat this issue we introduce a class weighted model (WSVM-MVO):

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z},\boldsymbol{w},b,\boldsymbol{\xi}^+,\boldsymbol{\xi}^-,\boldsymbol{t}} \quad & \boldsymbol{x}^\intercal \boldsymbol{\Sigma} \boldsymbol{x} + \epsilon \left(\frac{1}{2} \|\boldsymbol{w}\|_2^2 + \frac{C}{N} \left(\alpha^+ \mathbf{1}^\intercal \boldsymbol{\xi}^+ + \alpha^- \mathbf{1}^\intercal \boldsymbol{\xi}^- \right) \right) \\ s.t. \quad & -M(1 - z_i) + 1 - \xi_i^+ \leq (\boldsymbol{y}^{(i)})^\intercal \boldsymbol{w} + b \leq M z_i - 1 + \xi_i^-, \quad \forall i \in [N] \\ & (2), (3), (4) \end{aligned} \tag{5}$$

- where α^+ and α^- are such that $\alpha^+ = (N-K)/N$ and $\alpha^- = K/N$

Integrated SVM and MVO – Best Subset SVM

Let the best-subset selection SVM be given by

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{t}\in\mathbb{R}^{p+1}\times\mathbb{R}_+^N\times\{0,1\}^p} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \frac{C}{N}\mathbf{1}^\intercal\boldsymbol{\xi} \tag{BSS-SVM}$$

$$\text{s.t.} \qquad u_i\left((\boldsymbol{y}^{(i)})^\intercal\boldsymbol{w} + b\right) \geq 1 - \xi_i \ \forall i \in [N], \ (3)$$

- where $u_i = 2z_i - 1 \in \{-1, 1\}$
- A class-weighted version can also be considered

Integrated SVM and MVO – Parameter Selection Strategy

- Guiding principle: Select C, q, ϵ such that the resulting hyperplane is non-degenerate, and the portfolio has risk comparable to a baseline Card-MVO

Proposition 1

Consider the versions of Card-MVO and SVM-MVO where the constraint $\mathbf{1}^\intercal \boldsymbol{z} \leq K$ is replaced with the constraint $\mathbf{1}^\intercal \boldsymbol{z} = K$. Let $\boldsymbol{z}_{\text{CardMVO}}$ be part of an optimal solution to Card-MVO and let $\epsilon > 0$. If BSS-SVM with labels defined by $2\boldsymbol{z}_{\text{CardMVO}} - 1$ does not admit any degenerate optimal solutions, then, SVM-MVO does not admit any degenerate optimal solutions.

Proposition 2

Given a feasible solution to SVM-MVO denoted by $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \tilde{\boldsymbol{w}}, \tilde{b}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{t}})$ and a value of κ, setting

$$\epsilon = \kappa \; \tilde{\boldsymbol{x}}^\intercal \boldsymbol{\Sigma} \tilde{\boldsymbol{x}} / \left(\frac{1}{2} \|\tilde{\boldsymbol{w}}\|_2^2 + \frac{C}{N} \mathbf{1}^\intercal \tilde{\boldsymbol{\xi}} \right) \tag{6}$$

implies any solution (including any optimal solution) $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{t})$ with an SVM-MVO objective value that is no larger than $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \tilde{\boldsymbol{w}}, \tilde{b}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{t}})$'s objective, satisfies $\boldsymbol{x}^\intercal \boldsymbol{\Sigma} \boldsymbol{x} \leq (1 + \kappa) \tilde{\boldsymbol{x}}^\intercal \boldsymbol{\Sigma} \tilde{\boldsymbol{x}}$

Integrated SVM and MVO – Big-M's

- Guiding principle: Select M and $U_{\|\boldsymbol{w}\|_\infty}$ so that no optimal solutions to SVM-MVO (WSVM-MVO) are excluded

Theorem 1

Given a feasible solution to SVM-MVO and associated objective function value $\text{Obj}_{\text{SVM-MVO}}$, *the following inequalities hold for any optimal solution to SVM-MVO*

$$\|\boldsymbol{w}\|_\infty \leq \|\boldsymbol{w}\|_2 \leq \min\left\{\sqrt{\frac{2\text{Obj}_{\text{SVM-MVO}}}{\epsilon}}, \sqrt{2C}\right\} := UB_{\boldsymbol{w}} \quad \textit{and} \quad |b| \leq M_{\mathcal{Y}}\ UB_{\boldsymbol{w}} + 1 \tag{7}$$

where $M_{\mathcal{Y}} := \max_{i\in[N]}\ \max_{\mathcal{T}\subset[p]:|\mathcal{T}|\leq q} \left\|\boldsymbol{y}_{\mathcal{T}}^{(i)}\right\|_2$. *Thus, yielding valid values for* $U_{\|\boldsymbol{w}\|_\infty}$, *and* $U_{\|\boldsymbol{w}\|_2}$, $U_{|b|}$ *for any optimal solution to SVM-MVO*

where $U_{(\cdot)}$ denotes an upper bound on $(\cdot)$

Theorem 2

Given $U_{\|\boldsymbol{w}\|_2}$ *and* $U_{|b|}$, *setting* $M = 1 + U_{|b|} + U_{\|\boldsymbol{w}\|_2} M_{\mathcal{Y}}$, *does not exclude any optimal solutions of SVM-MVO*

Integrated SVM and MVO – Parameter Selection Strategy

Propositions 1 and 2 suggest:

- Select C and q so that the hyperplane parameters $(\boldsymbol{w}, b)$ obtained by solving BSS-SVM with labels $2\boldsymbol{z}_{\text{CardMVO}} - 1$ are such that $\boldsymbol{w} \neq \boldsymbol{0}$
    - Cross-validation on $2\boldsymbol{z}_{\text{CardMVO}} - 1$, does this well in practice (using a heuristic for BSS-SVM that produces feasible solutions)
    - If we find a non-degenerate $\boldsymbol{w}$, then it is guaranteed that $\boldsymbol{w} = 0$ is not optimal for SVM-MVO
    - Solving Card-MVO then BSS-SVM yields a feasible solution to SVM-MVO
- Given κ, setting ϵ via Equation (6) guarantees that the risk of any portfolio $\boldsymbol{x}$, forming part of an optimal solution to SVM-MVO satisfies $\boldsymbol{x}^\intercal \boldsymbol{\Sigma} \boldsymbol{x} \leq (1+\kappa)\boldsymbol{x}_{\text{CardMVO}}^\intercal \boldsymbol{\Sigma} \boldsymbol{x}_{\text{CardMVO}}$

Integrated SVM and MVO – New Experiments

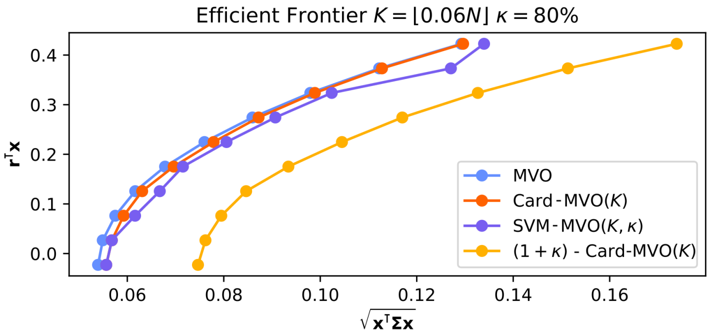


Figure: Efficient Frontier for MVO, Card-MVO, and the SVM-MVO model

- ϵ is set so that the risk of any SVM-MVO solution is within $(1+\kappa)$ of Card-MVO's risk

References I

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

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

Rifkin, R., Pontil, M., and Verri, A. (1999). A note on support vector machine degeneracy. In *International Conference on Algorithmic Learning Theory*, pages 252–263. Springer.