Selection Confidence Sets for Equally Weighted Portfolios

Davide Ferrari[2], Alessandro Fulci*[1], and Sandra Paterlini[1]

[1]Department of Economics and Management, University of Trento

[2]Faculty of Economics and Management, Free University of Bozen-Bolzano

October 20, 2025

Abstract

Given a universe of N assets, investors often form equally weighted portfolios (EWPs) by selecting subsets of assets. EWPs are simple, robust, and competitive out-of-sample, yet the uncertainty about which subset truly performs best is largely ignored. Traditional approaches typically rely on a single selected portfolio, but this fails to consider alternative investment strategies that may perform just as well when accounting for statistical uncertainty. To address this selection uncertainty, we introduce the Selection Confidence Set (SCS) for EWPs: the set of all portfolios that, under a given loss function and at a specified confidence level, contains the unknown set of optimal portfolios under repeated sampling. The SCS quantifies selection uncertainty by identifying a range of plausible portfolios, challenging the idea of a uniquely optimal choice. Like a confidence set, its size reflects uncertainty – growing with noisy or

*Corresponding author: E-mail: alessandro.fulci@unitn.it. Address: Department of Economics and Management, University of Trento, Via Inama 5, 38122 Trento, Italy

limited data, and shrinking as the sample size increases. Theoretically, we establish that the SCS covers the unknown optimal selection with high probability and characterize how its size grows with underlying uncertainty, corroborating these results through Monte Carlo experiments. Applications to the French 17-Industry Portfolios and Layer-1 cryptocurrencies underscore the importance of accounting for selection uncertainty when comparing equally weighted strategies.

Keywords: Equally Weighted Portfolios, Selection Confidence Set, Selection Uncertainty, Subset Selection, Wald Test.

1 Introduction

The Mean-Variance theory developed by Markowitz (1952) is the cornerstone of portfolio theory as it provides a rigorous mathematical framework for identifying portfolios that maximize expected return for a given level of risk. Despite subsequent advances in portfolio selection (e.g., Sharpe (1964); Ross (1976); Fama and French (1992); Efron et al. (2004); Goto and Xu (2015); Chen et al. (2021)), a central challenge remains the uncertainty surrounding the asset selection process. Even small estimation errors in expected returns or covariances can lead to large variability in the chosen allocation, making it difficult to deem any single portfolio superior. This sensitivity can lead to high turnover (Chopra and Ziemba, 2013) and make allocations untradeable once transaction costs are accounted for (DeMiguel et al., 2009). Numerous studies have sought to mitigate the effects of estimation error from various perspectives, including Bayesian approaches, shrinkage estimators, factor-based models, and the combination of tangent, minimum-variance, and risk-free portfolios (e.g., see DeMiguel et al. (2013), Bauder et al. (2021), Bodnar et al. (2022a), Bodnar et al. (2024), Bodnar et al. (2025)).

A growing body of research shows that simple allocation rules can outperform optimized strategies. Particularly, the Equally Weighted Portfolio (EWP), or $1/N$ rule, stands out

as a robust benchmark due to its simplicity, strong out-of-sample performance, resilience to estimation error, and built-in diversification (DeMiguel et al., 2009; Plyakha et al., 2015; Bodnar et al., 2022b). By avoiding estimation of return moments, EWPs sidestep instability, limit shorting, and exploit mean-reversion when rebalancing (Malladi and Fabozzi, 2017). Motivated by these advantages, recent work has explored data-driven methods for selecting optimal EWP subsets. Lee (2020) propose a deep learning algorithm that identifies top-performing assets, yielding substantial out-of-sample gains. Caparrini et al. (2024) apply tree-based classifiers to select equally weighted subsets of S&P 500 constituents, achieving superior risk-adjusted performance relative to the full EWP. Other learning approaches – such as support vector machines, gradient boosting, and nature-inspired heuristics – have also been investigated, but their effectiveness is often limited by noise and the existence of many similarly performing portfolios, making it difficult to identify a single dominant EWP (e.g., see Paiva et al. (2019), Chen et al. (2021), Asawa (2021)).

Motivated by the need to characterize uncertainty in EWP selection, this paper adopts a multi-portfolio perspective rather than focusing on identifying a single best portfolio. To formalize selection uncertainty, we introduce the Selection Confidence Set (SCS) for EWPs: the set of all asset selections whose performance under a chosen loss function is statistically indistinguishable from that of the unknown optimal EWP at the $(1-\alpha)\%$ confidence level, with $0 < \alpha < 1$. To construct the SCS, we use a Wald-type screening test to retain all the feasible portfolios not distinguishable from the unknown optimum. Under regularity conditions, the SCS achieves asymptotic coverage of at least $1-\alpha$, offering a principled basis for multi-portfolio selection and quantifying uncertainty in portfolio decisions.

Analogous to confidence intervals in parameter estimation, the SCS offers a statistical framework for quantifying uncertainty in portfolio selection by identifying statistically plausible candidates. It allows practitioners to assess whether a proposed portfolio is empirically supported, helping to avoid overconfident or unwarranted choices. The size and composition of the SCS serve as diagnostics for selection uncertainty: large sets reflect limited information

and difficulty in distinguishing among portfolios, often due to noise or small samples. Empirically, we find that an unexpectedly large number of EWPs are statistically indistinguishable from the empirical optimum in common situations, revealing substantial uncertainty often overlooked in practice. As sample size increases, the SCS naturally concentrates around the true optimum. As an illustration, Figure 1 shows the sample mean and standard deviation of EWP returns over a 3-year window for the French 17-Industry Portfolios data described in Section 6. At the 5% significance level, 408 portfolios have mean-variance losses $L(\mu, \sigma^2) = \sigma^2 - 0.5\mu$ statistically indistinguishable from the empirical optimum based on the test in Section 2; these portfolios form the SCS. In contrast, the remaining portfolios perform significantly worse. The size and diversity of the SCS underscore the extent of selection uncertainty and the risk of overinterpreting any single estimated optimum.

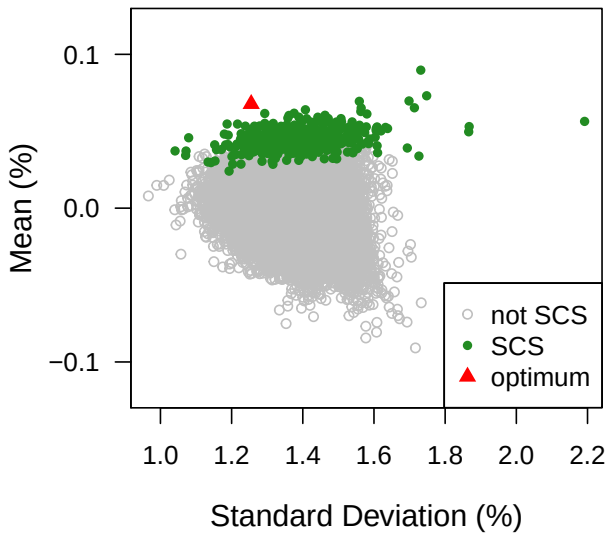

Figure 1: Sample means and standard deviations of EWPs from the French 17-Industry Portfolios data over a 3-year window. The red triangle marks the empirical optimum minimizing $L(\mu, \sigma^2) = \sigma^2 - 0.5\mu$. Solid green dots represent portfolios in the 95% SCS based on the Wald-type test in Section 2; open gray dots indicate portfolios outside the SCS.

While confidence sets for model selection have been studied in statistics, this paper is the first to develop a practical inferential framework for equally weighted financial portfolios. Ferrari and Yang (2015) introduced selection confidence sets for linear regression using F-

test screening, and Zheng et al. (2019) extended the approach to general parametric models via likelihood ratio tests. Previously, Hansen et al. (2011) proposed the Model Confidence Set (MCS), which retains models that are not significantly worse than the best in-sample performer while controlling for multiple comparisons. This method has been widely applied in empirical finance (e.g., Amendola et al. (2020), Liang et al. (2022), Chen et al. (2023)). While sharing the same multi-model rationale as the SCS, the MCS does not necessarily aim to cover the population-optimal model. In contrast, our SCS is inferential in nature and aligns with the approach of Ferrari and Yang (2015), targeting coverage of the true optimal set of portfolios in the population with a specified confidence level under repeated sampling.

The rest of the paper is organized as follows. Section 2 introduces the EWP selection framework, the SCS construction via Wald test screening, and examples under independent sampling. Section 3 presents theoretical coverage guarantees and asymptotic size analysis. Section 4 introduces post-selection metrics to summarize selection uncertainty and the composition of the SCS. Section 5 reports Monte Carlo results on coverage and size. Section 6 applies the method to the French 17-Industry Portfolios and Layer-1 cryptocurrency datasets. Section 7 concludes. Proofs are provided in the Appendix, while additional numerical results are collected in the Supplementary Materials.

2 Confidence Sets for the optimal EWPs

2.1 Setup and Objectives

Let $\boldsymbol{X} = (X_1, \ldots, X_N)^\top \in \mathbb{R}^N$, denote a random vector of asset returns, $N < \infty$ is the finite number of assets. A portfolio is a linear combination of the components of $\boldsymbol{X}$. We consider portfolios constructed from N assets, defined as

$$Y_{\mathbf{s}} = \frac{\sum_{i=1}^{N} s_i X_i}{\sum_{j=1}^{N} s_j}, \quad \mathbf{s} = (s_1, \ldots, s_N) \in \mathcal{S} := \{0,1\}^N \setminus \{\mathbf{0}\}, \tag{1}$$

where $\mathbf{s}$ is a binary vector indicating asset selection, i.e., $s_i = 1$ if asset i is selected, and $s_i = 0$ otherwise, and $\mathcal{S}$ represents the set of all feasible selections. The mean and variance of $Y_{\mathbf{s}}$ are denoted by $\mu_{\mathbf{s}} := \mathbb{E}[Y_{\mathbf{s}}]$ and $\sigma^2_{\mathbf{s}} := \mathrm{var}[Y_{\mathbf{s}}]$, respectively. While we focus on the general case where $\mathcal{S}$ includes all asset subset combinations, in practice restrictions may be introduced to reflect investment policies, transaction costs, or operational requirements.

We define the loss function as a continuously differentiable mapping $L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, $(\mu, \sigma^2) \mapsto L(\mu, \sigma^2)$, assigning a real-valued performance criterion to each portfolio based on its mean return μ and variance σ^2. The loss incurred by portfolio $Y_{\mathbf{s}}$ depends on the asset selection $\mathbf{s}$ only through its mean and variance, $\mu_{\mathbf{s}}$ and $\sigma^2_{\mathbf{s}}$; for instance, under a mean–variance criterion, one may take $L(\mu_{\mathbf{s}}, \sigma^2_{\mathbf{s}}) = \sigma^2_{\mathbf{s}} - \gamma\mu_{\mathbf{s}}$ for some $\gamma > 0$. Based on L, we define the set of optimal asset selections as

$$\mathcal{S}_0 := \arg\min_{\mathbf{s}\in\mathcal{S}} L(\mu_{\mathbf{s}}, \sigma^2_{\mathbf{s}}). \tag{2}$$

Since $\mathcal{S}$ is finite, the minimum is always attained and $\mathcal{S}_0$ contains at least one selection. In the rest of the paper, we use $L_0 := L(\mu_{\mathbf{s}_0}, \sigma^2_{\mathbf{s}_0})$, $\mathbf{s}_0 \in \mathcal{S}_0$ to denote the minimum loss.

Given a sample $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(T)}$, where each $\boldsymbol{X}^{(t)}$ represents the returns of N assets at time t, and a user-specified confidence level $0 < \alpha < 1$ (e.g., 0.01 or 0.05), our goal is to construct a SCS, denoted by $\widehat{\mathcal{S}}_\alpha$, defined as a subset of $\mathcal{S}$ satisfying the coverage property

$$\mathbb{P}(\mathcal{S}_0 \subseteq \widehat{\mathcal{S}}_\alpha) \geq 1 - \alpha, \tag{3}$$

asymptotically, as $T \to \infty$. This condition ensures that, with approximate probability of at least $1 - \alpha$ under repeated sampling, the true optimal selections $\mathcal{S}_0$ are included in the constructed set. In analogy to classical confidence intervals for parameter estimation, the SCS quantifies uncertainty in the portfolio selection process by identifying a set of EWPs whose empirical performance is statistically indistinguishable from L_0.

The SCS supports two main tasks in portfolio selection. 1) It enables formal plausibility

assessment of a given candidate portfolio $\mathbf{s}$: by checking whether $\mathbf{s} \in \widehat{\mathcal{S}}_\alpha$, practitioners can formally test whether its performance is statistically indistinguishable from that of an unknown optimum. This allows for rigorous validation of alternative portfolio configurations. Moreover, comparing the asset composition of $\mathbf{s}$ with those of other plausible elements in $\widehat{\mathcal{S}}_\alpha$ reveals robust inclusion patterns, highlighting complementary investment strategies and asset combinations that consistently contribute to strong performance. 2) The SCS quantifies the degree of selection uncertainty. Its size and structure reflect the informativeness of the data: a singleton SCS indicates a uniquely optimal allocation; a small but non-singleton points to manageable ambiguity; and a large SCS suggests high uncertainty, with many portfolios exhibiting comparable performance.

Overall, the SCS is a statistically grounded tool for assessing model uncertainty in portfolio selection. By identifying the full set of portfolios with performance statistically indistinguishable from a portfolio with minimal loss, it complements traditional EWPs optimization with formal inferential guarantees, helping decision-makers navigate the uncertainty inherent in financial allocation.

2.2 Construction Based on Wald Test Screening

To construct $\widehat{\mathcal{S}}_\alpha$, we aim to compare the population loss associated with each candidate selection $\mathbf{s}$ to that of any optimal selection $\mathbf{s}_0 \in \mathcal{S}_0$, using a Wald-type test. To quantify the performance discrepancy between any selections $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, we define the performance differential as $\delta(\mathbf{s}; \mathbf{s}') := L(\mu_\mathbf{s}, \sigma^2_\mathbf{s}) - L(\mu_{\mathbf{s}'}, \sigma^2_{\mathbf{s}'})$. To identify selections whose performance is statistically indistinguishable from the optimal, we test the hypotheses

$$H_0 : \delta(\mathbf{s}) = 0 \quad \text{vs.} \quad H_1 : \delta(\mathbf{s}) > 0, \tag{4}$$

where $\delta(\mathbf{s}) := \delta(\mathbf{s}; \mathbf{s}_0) = L(\mu_\mathbf{s}, \sigma^2_\mathbf{s}) - L_0$, and $\mathbf{s}_0 \in \mathcal{S}_0$ is a fixed reference portfolio achieving the minimal loss L_0. If H_0 is not rejected – i.e., the data do not provide sufficient evidence

to distinguish the candidate portfolio $Y_{\mathbf{s}}$ from the reference Y_0 – we include $\mathbf{s}$ in the SCS.

Since the true optimal set $\mathcal{S}_0$ is unknown, we use an empirical proxy obtained by minimizing a sample-based loss over the feasible support space $\mathcal{S}$. Given return observations $\boldsymbol{X}^{(t)} = (X_1^{(t)}, \ldots, X_N^{(t)})^\top \in \mathbb{R}^N$, $t = 1, \ldots, T$, corresponding to EWPs returns $Y_{\boldsymbol{s}}^{(t)}$, $t = 1, \ldots, T$, we consider the sample mean and variance of portfolio returns for $\boldsymbol{s} \in \mathcal{S}$,

$$\hat{\mu}_{\boldsymbol{s}} := \frac{1}{T} \sum_{t=1}^{T} Y_{\boldsymbol{s}}^{(t)}, \quad \hat{\sigma}_{\boldsymbol{s}}^2 := \frac{1}{T-1} \sum_{t=1}^{T} \left(Y_{\boldsymbol{s}}^{(t)} - \hat{\mu}_{\boldsymbol{s}} \right)^2, \tag{5}$$

and define the empirical optimum as

$$\hat{\mathbf{s}}_0 := \arg\min_{\mathbf{s} \in \mathcal{S}} L(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}_{\mathbf{s}}^2). \tag{6}$$

Since the empirical loss is computed over a finite set and portfolio moments are continuous functions of the data, $\hat{\mathbf{s}}_0$ is unique with probability one under mild conditions. In contrast, the population-optimal set $\mathcal{S}_0$ may contain more than one element. Consequently, in regular settings, $\hat{\mathbf{s}}_0$ converges in probability to some element of $\mathcal{S}_0$, although not necessarily uniquely. This follows from the consistency of the sample means and variances used to evaluate the loss (see proof of Proposition 1), justifying the use of $\hat{\mathbf{s}}_0$ as a reference. However, since all elements in $\mathcal{S}_0$ attain the same minimal population loss L_0, the presence of multiple optima is not a concern and does not affect the validity of using $\hat{\mathbf{s}}_0$ as a reference, as any such limit point represents an equally optimal selection.

For a fixed α, the SCS $\widehat{\mathcal{S}}_\alpha$ is defined as

$$\widehat{\mathcal{S}}_\alpha := \{ \mathbf{s} \in \mathcal{S} : \; Z(\mathbf{s}; \hat{\mathbf{s}}_0) \leq q_{1-\alpha} \}, \tag{7}$$

where $Z(\mathbf{s}; \mathbf{s}')$ is the studentized differential for selections $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, defined by

$$Z(\mathbf{s}; \mathbf{s}') := \frac{\hat{\delta}(\mathbf{s}; \mathbf{s}')}{\sqrt{\hat{\tau}(\mathbf{s}; \mathbf{s}')/T}} := \frac{L(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}_{\mathbf{s}}^2) - L(\hat{\mu}_{\mathbf{s}'}, \hat{\sigma}_{\mathbf{s}'}^2)}{\sqrt{\nabla \hat{\delta}(\mathbf{s}; \mathbf{s}')^\top \hat{\mathbf{V}}(\mathbf{s}; \mathbf{s}') \nabla \hat{\delta}(\mathbf{s}; \mathbf{s}')/T}}, \tag{8}$$

where $\hat{\delta}(\mathbf{s};\mathbf{s}') = L(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}}) - L(\hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'})$ represents the empirical performance differential, $\hat{\tau}(\mathbf{s};\mathbf{s}')^2$ is its estimated variance and $q_{1-\alpha}$ is an $(1-\alpha)$-quantile specified depending on the distribution of $Z(\mathbf{s};\hat{\mathbf{s}}_0)$. In the denominator, $\nabla\hat{\delta}(\mathbf{s};\mathbf{s}')$ represents 4×1 gradient vector of the loss difference evaluated at the moment estimates $\hat{\boldsymbol{\theta}}(\mathbf{s};\mathbf{s}') := (\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}}, \hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'})^\top$, that is,

$$\nabla\hat{\delta}(\mathbf{s};\mathbf{s}') := \left(\frac{\partial}{\partial\hat{\mu}_{\mathbf{s}}}\hat{\delta}(\mathbf{s};\mathbf{s}'), \frac{\partial}{\partial\hat{\sigma}^2_{\mathbf{s}}}\hat{\delta}(\mathbf{s};\mathbf{s}'), \frac{\partial}{\partial\hat{\mu}_{\mathbf{s}'}}\hat{\delta}(\mathbf{s};\mathbf{s}'), \frac{\partial}{\partial\hat{\sigma}^2_{\mathbf{s}'}}\hat{\delta}(\mathbf{s};\mathbf{s}')\right)^\top, \tag{9}$$

and $\hat{\mathbf{V}}(\mathbf{s};\mathbf{s}')$ is a consistent estimator of the asymptotic variance $\mathbf{V}(\mathbf{s};\mathbf{s}')$ of the moment estimator $\hat{\boldsymbol{\theta}}(\mathbf{s};\mathbf{s}')$ defined in (11).

The selection confidence set $\widehat{\mathcal{S}}_\alpha$ is constructed by comparing each candidate $\mathbf{s}\in\mathcal{S}$ to the empirical benchmark $\hat{\mathbf{s}}_0$. Since $\hat{\mathbf{s}}_0$ converges in probability to an element of the population-optimal set $\mathcal{S}_0$, the set $\widehat{\mathcal{S}}_\alpha$ includes, with high probability, all selections whose performance is statistically indistinguishable from the optimal loss L_0. Selections with higher loss than $\hat{\mathbf{s}}_0$ may still be included if their difference in performance is small compared to its uncertainty. Rather than committing to the single best selection $\hat{\mathbf{s}}_0$, $\widehat{\mathcal{S}}_\alpha$ offers a robust alternative by retaining a number of well-performing selections, which is particularly useful when small performance differences do not justify distinct portfolio allocations.

The null distribution of $Z(\mathbf{s};\hat{\mathbf{s}}_0)$ can be specified in various ways. Here, we rely on a normal approximation justified by the central limit theorem. Particularly, for each $\mathbf{s},\mathbf{s}'\in\mathcal{S}$, we assume that the empirical moments $\hat{\boldsymbol{\theta}}(\mathbf{s},\mathbf{s}')$ satisfy

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}(\mathbf{s};\mathbf{s}') - \boldsymbol{\theta}(\mathbf{s};\mathbf{s}')\right) \xrightarrow{d} \mathcal{N}_4(\mathbf{0}, \mathbf{V}(\mathbf{s};\mathbf{s}')), \tag{10}$$

where $\mathbf{V}(\mathbf{s};\mathbf{s}')$ is the 4×4 matrix

$$\mathbf{V}(\mathbf{s};\mathbf{s}') := \begin{pmatrix} \mathbf{V}_{\mathbf{ss}} := \mathrm{Cov}[(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}})] & \mathbf{V}_{\mathbf{ss}'} := \mathrm{Cov}[(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}}), (\hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'})] \\ \mathbf{V}_{\mathbf{s}'\mathbf{s}} := \mathrm{Cov}[(\hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'}), (\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}})] & \mathbf{V}_{\mathbf{s}'\mathbf{s}'} := \mathrm{Cov}[(\hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'})] \end{pmatrix}. \tag{11}$$

This relation holds under mild regularity conditions. For example, when the data are assumed i.i.d., it is sufficient to have finite fourth moment $\mathbb{E}[Y_{\mathbf{s}}^4] < \infty$ for all $\mathbf{s} \in \mathcal{S}$. In the time series case, it is sufficient to have finite $4 + \epsilon$ moments, with ϵ being a small positive constant, together with an appropriate mixing condition; e.g., see (Andrews, 1991).

If consistency of the sample moments in $\hat{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}')$ applies uniformly for $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, the empirical loss converges to the population loss, and the event $\hat{\mathbf{s}}_0 \in \mathcal{S}_0$ occurs with probability tending to one. Therefore, the test statistic $Z(\mathbf{s}; \hat{\mathbf{s}}_0)$ under the null hypothesis $H_0 : \delta(\mathbf{s}) = 0$ converges in distribution to a standard normal distribution by the Delta method, that is $Z(\mathbf{s}; \hat{\mathbf{s}}) \xrightarrow{d} \mathcal{N}(0, 1)$. Therefore, to retain $1 - \alpha$ coverage asymptotically, the quantile $q_{1-\alpha}$ in (7) is set as the $(1 - \alpha)$ quantile of the standard normal distribution.

While the above test serves as our default screening tool due to its flexibility in accommodating a wide class of loss functions and its reliability in large samples, the SCS framework is compatible with any alternative test specification for the null $H_0 : \delta(\mathbf{s}) = 0$. Examples include the finite-sample Sharpe ratio test of Memmel (2003) and the robust bootstrap procedure of Ledoit and Wolf (2008). Extensions to variance (e.g., Ledoit and Wolf (2011), implemented in Ding et al. (2021)) and smooth functionals of higher-order moments (e.g., Ledoit and Wolf (2018) implemented in Leippold and Yang (2023)) further broaden the applicability. While these tests are typically used for ad hoc pairwise comparisons, our framework incorporates them to construct an SCS that covers the population-optimal selection with the prescribed confidence level.

2.3 Screening under i.i.d. sampling and loss-specific criteria

If portfolio returns $Y_{\mathbf{s}}^{(1)}, \ldots, Y_{\mathbf{s}}^{(T)}$ form an i.i.d. sequence for all $\mathbf{s} \in \mathcal{S}$, then the sample moments associated with any pair of selections $(\mathbf{s}, \mathbf{s}')$ admit the limiting distribution in (10),

with the following blocks for $\mathbf{V}(\mathbf{s};\mathbf{s}')$:

$$\mathbf{V}_{\mathbf{ss}} = \begin{pmatrix} \sigma_{\mathbf{s}}^2 & \mu_{3,\mathbf{s}} \\ \mu_{3,\mathbf{s}} & \mu_{4,\mathbf{s}} - \sigma_{\mathbf{s}}^4 \end{pmatrix}, \mathbf{V}_{\mathbf{ss}'} = \begin{pmatrix} \sigma_{\mathbf{ss}'} & \mu_{3,\mathbf{ss}'} \\ \mu_{3,\mathbf{s}'\mathbf{s}} & \mu_{4,\mathbf{ss}'} - \sigma_{\mathbf{ss}'}^2 \end{pmatrix}, \mathbf{V}_{\mathbf{s}'\mathbf{s}'} = \begin{pmatrix} \sigma_{\mathbf{s}'}^2 & \mu_{3,\mathbf{s}'} \\ \mu_{3,\mathbf{s}'} & \mu_{4,\mathbf{s}'} - \sigma_{\mathbf{s}'}^4 \end{pmatrix}$$

and $\mathbf{V}_{\mathbf{s}'\mathbf{s}} = \mathbf{V}_{\mathbf{ss}'}^\top$. This structure captures both marginal and joint variation up to the fourth moment, which is necessary to account for tail risk and asymmetries in performance differentials. Specifically, the third-order moments include the marginal skewness terms $\mu_{3,\mathbf{s}} = \mathbb{E}[(Y_{\mathbf{s}} - \mu_{\mathbf{s}})^3]$ and $\mu_{3,\mathbf{s}'} = \mathbb{E}[(Y_{\mathbf{s}'} - \mu_{\mathbf{s}'})^3]$, as well as the mixed skewness terms $\mu_{3,\mathbf{ss}'} = \mathbb{E}[(Y_{\mathbf{s}} - \mu_{\mathbf{s}})(Y_{\mathbf{s}'} - \mu_{\mathbf{s}'})^2]$ and $\mu_{3,\mathbf{s}'\mathbf{s}} = \mathbb{E}[(Y_{\mathbf{s}'} - \mu_{\mathbf{s}'})(Y_{\mathbf{s}} - \mu_{\mathbf{s}})^2]$. The fourth-order moments include the marginal kurtoses $\mu_{4,\mathbf{s}} = \mathbb{E}[(Y_{\mathbf{s}} - \mu_{\mathbf{s}})^4]$, $\mu_{4,\mathbf{s}'} = \mathbb{E}[(Y_{\mathbf{s}'} - \mu_{\mathbf{s}'})^4]$, and the joint kurtotic term $\mu_{4,\mathbf{ss}'} = \mathbb{E}[(Y_{\mathbf{s}} - \mu_{\mathbf{s}})^2(Y_{\mathbf{s}'} - \mu_{\mathbf{s}'})^2]$. In this case, $\mathbf{V}(\mathbf{s};\mathbf{s}')$ is estimated by substituting the sample means and variances into the analytical form derived above.

The special case where $Y_{\mathbf{s}}^{(t)}$ is normally distributed serves as a benchmark for EWP selection under idealized conditions. Under normality, all third-order moments vanish; namely, $\mu_{3,\mathbf{s}} = \mu_{3,\mathbf{s}'} = \mu_{3,\mathbf{ss}'} = \mu_{3,\mathbf{s}'\mathbf{s}} = 0$. Moreover, fourth-order moments simplify as: $\mu_{4,\mathbf{s}} = 3\sigma_{\mathbf{s}}^4$, $\mu_{4,\mathbf{s}'} = 3\sigma_{\mathbf{s}'}^4$, and $\mu_{4,\mathbf{ss}'} = \sigma_{\mathbf{ss}'}^2 + 2\sigma_{\mathbf{s}}^2\sigma_{\mathbf{s}'}^2$, thus leading to simple expressions for screening statistics under common loss functions $L(\mu, \sigma^2)$. For example, for the mean-variance loss $L(\mu, \sigma^2) = -\gamma\mu + \sigma^2$, $\gamma > 0$, where γ denotes the risk aversion coefficient, the SCS is constructed by testing $H_0 : \delta(\mathbf{s};\mathbf{s}_0) = \gamma(\mu_{\mathbf{s}_0} - \mu_{\mathbf{s}}) + \sigma_{\mathbf{s}}^2 - \sigma_{\mathbf{s}_0}^2 = 0$. This leads to

$$\widehat{\mathcal{S}}_\alpha = \left\{ \mathbf{s} \in \mathcal{S} : Z(\mathbf{s};\hat{\mathbf{s}}_0) = \frac{\sqrt{T}\left[\gamma(\hat{\mu}_{\hat{\mathbf{s}}_0} - \hat{\mu}_{\mathbf{s}}) + \hat{\sigma}_{\mathbf{s}}^2 - \hat{\sigma}_{\hat{\mathbf{s}}_0}^2\right]}{\sqrt{\gamma^2\left(\hat{\sigma}_{\mathbf{s}}^2 - 2\hat{\sigma}_{\mathbf{s}\hat{\mathbf{s}}_0} + \hat{\sigma}_{\hat{\mathbf{s}}_0}^2\right) + 2\left(\hat{\sigma}_{\mathbf{s}}^4 - 2\hat{\sigma}_{\mathbf{s}\hat{\mathbf{s}}_0}^2 + \hat{\sigma}_{\hat{\mathbf{s}}_0}^4\right)}} \leq q_{1-\alpha} \right\}.$$

For the Sharpe ratio loss $L(\mu, \sigma^2) = -\mu/\sigma$, the SCS is constructed by testing $H_0 : \delta(\mathbf{s};\mathbf{s}_0) = \mu_{\mathbf{s}_0}/\sigma_{\mathbf{s}_0} - \mu_{\mathbf{s}}/\sigma_{\mathbf{s}} = 0$. Thus, we obtain

$$\widehat{\mathcal{S}}_\alpha = \left\{ \mathbf{s} \in \mathcal{S} : Z(\mathbf{s};\hat{\mathbf{s}}_0) = \frac{\sqrt{T}(\hat{r}_{\mathbf{s}} - \hat{r}_{\hat{\mathbf{s}}_0})}{\sqrt{2(1 - \hat{\rho}_{\mathbf{s}\hat{\mathbf{s}}_0}) + \frac{1}{2}(\hat{r}_{\mathbf{s}}^2 + \hat{r}_{\hat{\mathbf{s}}_0}^2) - \hat{\rho}_{\mathbf{s}\hat{\mathbf{s}}_0}^2 \hat{r}_{\mathbf{s}}\hat{r}_{\hat{\mathbf{s}}_0}}} \leq q_{1-\alpha} \right\}, \tag{12}$$

where $\hat{r}_{\mathbf{s}} = \hat{\mu}_{\mathbf{s}}/\hat{\sigma}_{\mathbf{s}}$ and $\hat{\rho}_{\mathbf{s}\hat{\mathbf{s}}_0} = \hat{\sigma}_{\mathbf{s}\hat{\mathbf{s}}_0}/(\hat{\sigma}_{\mathbf{s}}\hat{\sigma}_{\hat{\mathbf{s}}_0})$.

3 Asymptotic Validity and Size

In this section, we study the asymptotic coverage probability for the set of optimal selections $\mathcal{S}_0$ and the SCS size as $T \to \infty$, assuming a fixed number of assets $N < \infty$. To establish the asymptotic properties of the SCS, we proceed under high-level conditions that ensure valid inference without requiring restrictive model assumptions. For all $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ we assume:

(A1) The vector of estimators $(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}}, \hat{\mu}_{\mathbf{s}'}, \hat{\sigma}^2_{\mathbf{s}'})$ jointly satisfies the central limit theorem as in Equation (10), with asymptotic covariance matrix $\mathbf{V}(\mathbf{s}, \mathbf{s}')$.

(A2) Each element of the estimated covariance matrix $\hat{\mathbf{V}}(\mathbf{s}, \mathbf{s}')$ converges in probability to its population counterpart $\hat{\mathbf{V}}_{jk}(\mathbf{s}, \mathbf{s}') \xrightarrow{p} \mathbf{V}_{jk}(\mathbf{s}, \mathbf{s}')$, for all $j, k \in \{1, 2, 3, 4\}$.

Next, we show that the SCS defined in (7) contains the true set of optimal selections $\mathcal{S}_0$ with probability of at least $1 - \alpha$ in large samples.

Proposition 1 (Asymptotic coverage of the optimal set)**.** *Under Assumptions (A1)–(A2), the Selection Confidence Set $\widehat{\mathcal{S}}_\alpha$ defined in* (7) *satisfies the asymptotic coverage property:*

$$\lim_{T\to\infty} \mathbb{P}\left(\mathcal{S}_0 \subseteq \widehat{\mathcal{S}}_\alpha\right) \geq 1 - \alpha.$$

The result relies on two key aspects implied by Assumptions (A1) and (A2): (i) convergence in probability of sample moments $(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}})$ to their population counterparts, ensuring that $\hat{\mathbf{s}}_0$ converges to $\mathbf{s}_0 \in \mathcal{S}_0$; and (ii) convergence in distribution of the test statistic $Z(\mathbf{s}, \hat{\mathbf{s}}_0)$ under the null, ensuring that our test has asymptotic size α. As already mentioned, these hold under general dependence and heteroskedasticity in asset returns, requiring only that sample moments of portfolio returns admit a joint asymptotic normal distribution. In finite samples, $\widehat{\mathcal{S}}_\alpha$ may include many suboptimal portfolios, especially when loss differentials are

small or variance estimates are noisy, reducing the power of pairwise tests. A large SCS is not a flaw but a reflection of data uncertainty, necessary for valid coverage. Hence, characterizing its size and structure is crucial for assessing selection ambiguity. To this end, we introduce the standardized loss differential

$$\gamma(\mathbf{s}) := \frac{\delta(\mathbf{s})}{\tau(\mathbf{s})} = \frac{L(\mu_{\mathbf{s}}, \sigma_{\mathbf{s}}^2) - L(\mu_0, \sigma_0^2)}{\sqrt{\nabla\delta(\mathbf{s}, \mathbf{s}_0)^\top \mathbf{V}(\mathbf{s}, \mathbf{s}_0) \nabla\delta(\mathbf{s}, \mathbf{s}_0)}}, \tag{13}$$

where $\nabla\delta(\mathbf{s}, \mathbf{s}_0)$ is the gradient of $\delta(\mathbf{s}, \mathbf{s}_0)$ with respect to $(\mu_{\mathbf{s}}, \sigma_{\mathbf{s}}^2, \mu_{\mathbf{s}_0}, \sigma_{\mathbf{s}_0}^2)^\top$. This ratio captures the signal strength relative to estimation uncertainty and governs the probability that a non-optimal support is erroneously included in the SCS. The next proposition makes this relationship precise by characterizing the limiting expected size of the SCS.

Proposition 2 (Asymptotic expected size of the SCS)**.** *Under assumptions (A1)–(A2), the expected size of the Selection Confidence Set $\widehat{\mathcal{S}}_\alpha$ satisfies*

$$\mathbb{E}[|\widehat{\mathcal{S}}_\alpha|] = |\mathcal{S}_0|(1-\alpha) + \sum_{\mathbf{s} \in \mathcal{S} \setminus \mathcal{S}_0} \Phi\left(q_{1-\alpha} - \sqrt{T}\gamma(\mathbf{s})\right) + o(1) \tag{14}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and $q_{1-\alpha}$ is its $(1-\alpha)$ quantile.

Proposition 2 shows the relationship between the inclusion probability of a suboptimal asset selection $\mathbf{s} \notin \mathcal{S}_0$ and its standardized differential $\gamma(\mathbf{s})$. If $\sqrt{T}\gamma(\mathbf{s}) \to \infty$, then the inclusion probability $\Phi(z_{1-\alpha} - \sqrt{T}\gamma(\mathbf{s}))$ decays exponentially, and $\mathbf{s}$ is eventually excluded from $\widehat{\mathcal{S}}_\alpha$. Thus, the cardinality of the SCS is driven by the number of suboptimal supports for which $\gamma(\mathbf{s}) = o(T^{-1/2})$, as these remain empirically indistinguishable from optimal ones. In this sense, the SCS quantifies the set of portfolios that are statistically non-separable from the best-performing configurations, and its size reflects the inherent difficulty of model selection under limited information.

To further elaborate, let $\gamma_{\min} := \min_{\mathbf{s} \notin \mathcal{S}_0} \delta(\mathbf{s})/\tau(\mathbf{s})$ denote the minimal standardized signal among selections outside the optimal set. Since the function $x \mapsto \Phi(q_{1-\alpha} - \sqrt{T}x)$ is

monotonically decreasing in x, from (14), we immediately obtain the approximate bounds

$$|\mathcal{S}_0|(1-\alpha)+o(1) \leq \mathbb{E}[|\widehat{\mathcal{S}}_\alpha|] \leq |\mathcal{S}_0|(1-\alpha)+\left(2^N-|\mathcal{S}_0|-1\right)\Phi\left(z_{1-\alpha}-\sqrt{T}\gamma_{\min}\right)+o(1). \quad (15)$$

If $\gamma_{\min}$ is bounded away from zero – i.e., all suboptimal models are well-separated from the optimal set – then the expected SCS size concentrates around $|\mathcal{S}_0|(1-\alpha)$, reflecting the expected retention of optimal supports. In contrast, when $\gamma_{\min}$ is small, the inclusion probability $\Phi(z_{1-\alpha}-\sqrt{T}\gamma_{\min})$ remains large, and many suboptimal models may enter $\mathcal{S}_\alpha$.

4 Post-selection Metrics

We introduce descriptive statistics derived from the SCS to summarize uncertainty in portfolio selection. These include global measures of multiplicity, performance variability, and asset-level inclusion patterns.

Global selection uncertainty. Basic summaries are the cardinality $|\widehat{\mathcal{S}}_\alpha|$, the relative cardinality $|\widehat{\mathcal{S}}_\alpha|/|\mathcal{S}|$ and its scale-adjusted version, the Relative Multiplicity Index (RMI):

$$\text{RMI} := 1 - \frac{\log(|\widehat{\mathcal{S}}_\alpha|)}{\log(|\mathcal{S}|)}, \quad (16)$$

which ranges from 0 (maximal uncertainty) to 1 (unique selection). The logarithmic transformation accounts for the natural scale of growth in $|\mathcal{S}|$, which is typically exponential in the number of assets. In addition, to quantify the empirical range of acceptable losses we consider the observed loss interval $(\hat{L}_0, \hat{L}_{\max})$, where $\hat{L}_0 = L(\hat{\mu}_{\hat{\mathbf{s}}_0}, \hat{\sigma}^2_{\hat{\mathbf{s}}_0})$ and $\hat{L}_{\max} := \max_{\mathbf{s}\in\mathcal{S}_\alpha} L(\hat{\mu}_{\mathbf{s}}, \hat{\sigma}^2_{\mathbf{s}})$ and the associated performance spread $\hat{\delta}_\alpha := \hat{L}_{\max} - \hat{L}_0$. These are especially relevant in applied risk-sensitive settings; for example a manager may accept a candidate allocation $\mathbf{s} \in \mathcal{S}_\alpha$ only if this metrics falls below a tolerable threshold.

Lower-boundary (LB) portfolios. The lower boundary $\underline{\widehat{\mathcal{S}}_\alpha} \subseteq \mathcal{S}_\alpha$ contains the most parsimonious selections in the SCS: $\underline{\widehat{\mathcal{S}}_\alpha} := \left\{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha : \nexists\, \mathbf{s}' \in \widehat{\mathcal{S}}_\alpha \text{ with } \text{supp}(\mathbf{s}') \subset \text{supp}(\mathbf{s})\right\}$. These

are minimal configurations that cannot be further reduced without losing information, revealing which assets are essential for achieving acceptable risk-return trade-offs. As such, they provide a data-driven foundation for lean investment strategies – especially valuable when transaction costs, regulatory limits, or interpretability concerns favor compact allocations.

Asset inclusion and co-inclusion importance. The inclusion importance of asset $j \in \{1, \ldots, N\}$ measures how often it appears across portfolios in the SCS:

$$\mathrm{II}_\alpha(j) := \frac{1}{|\widehat{\mathcal{S}}_\alpha|} \sum_{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha} s_j,$$

with value close to 1 indicates that asset j appears in nearly all selected portfolios, while a value near 0 indicates rare or no inclusion. To quantify how often assets i and j appear together in the SCS, we define their co-inclusion importance as

$$\mathrm{CII}_\alpha(i,j) := \frac{\sum_{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha} s_i s_j}{\sum_{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha} s_i + \sum_{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha} s_j - \sum_{\mathbf{s} \in \widehat{\mathcal{S}}_\alpha} s_i s_j},$$

if the denominator is strictly positive and $\mathrm{CII}_\alpha(i,j) := 1$ otherwise. This index measures how frequently assets i and j co-occur relative to their marginal inclusion frequencies. High values indicate complementarity, while low values reflect mutual exclusivity. To visualize these patterns, one can construct an undirected asset graph where nodes represent assets and edge weights proportional to $\mathrm{CII}_\alpha(i,j)$, revealing clusters, substitution patterns, and diversification structures.

5 Numerical Experiments

5.1 Simulation Setup

Asset returns $\boldsymbol{X} \in \mathbb{R}^N$ are drawn from a multivariate normal distribution with mean $\boldsymbol{\eta} \in \mathbb{R}^N$ and covariance $\boldsymbol{\Sigma} = \mathbf{DRD}$, where $\mathbf{R} \in \mathbb{R}^{N \times N}$ is the correlation matrix and

$\mathbf{D} = \text{diag}(\sqrt{\Sigma_{11}}, \ldots, \sqrt{\Sigma_{NN}})$ is a diagonal matrix of marginal standard deviations. To reflect realistic market conditions with a weak positive risk–return association, we set $\eta_j = -0.002 + \Sigma_{jj}/10 + \varepsilon_j$, where marginal variances Σ_{jj} are drawn independently from a uniform distribution on $[0.01, 0.03]$, and $\varepsilon_j \sim \mathcal{N}(0, 0.02)$ for $j = 1, \ldots, N$. The correlation matrix $\mathbf{R}$ is set according to the following asset dependence models.

Model 1) Graph structure. We generate an inverse covariance matrix $\mathbf{\Omega}^{-1}$ based on a scale-free graph structure, reflecting empirically observed dependencies in asset returns (Kritzman et al., 2010). Using the `huge.generator` function from the `R` package `huge` (Jiang et al., 2021), we construct a sparse adjacency matrix $\mathbf{\Theta}$, where off-diagonal entries of $\mathbf{\Omega}$ corresponding to edges are set to a constant $v > 0$. To ensure positive definiteness, diagonal entries are adjusted as $\mathbf{\Omega}_{jj} = |e| + 0.2$, with e being the smallest eigenvalue of $v\mathbf{\Theta}$. We consider $v \in \{0.2, 1\}$ to account for mild and strong conditional dependence. The resulting precision matrix is inverted and rescaled to obtain the correlation matrix $\mathbf{R}$.

Model 2) Exchangeable structure. We set $\{\mathbf{R}\}_{jk} = \rho$ for $j \neq k$ with $\rho \in \{0.25, 0.75\}$ corresponding to mild and strong correlation.

For various values of T and N, we compute Monte Carlo estimates of the expected SCS cardinality $\kappa_\alpha := \mathbb{E}[|\widehat{\mathcal{S}}_\alpha|]$, that of its lower boundary $\underline{\kappa}_\alpha := \mathbb{E}[|\underline{\widehat{\mathcal{S}}}_\alpha|]$, and the coverage probability $p_\alpha := \mathbb{P}(\mathbf{s}_0 \in \widehat{\mathcal{S}}_\alpha)$, based on 300 simulation runs. For each run, we construct SCSs at confidence levels 90%, 95%, and 99%, using the Sharpe ratio loss $(-\mu/\sigma)$, mean-variance loss $(-0.5\,\mu + \sigma^2)$, and the expected shortfall (ES) loss $(-\mu + \sigma\,\phi(z_{0.1})/0.1)$, where $\phi(\cdot)$ is the pdf of the standard normal distribution. In all scenarios, the true and empirical optimal EWPs are identified via exhaustive search.

5.2 Results

In Table 1, we report Monte Carlo estimates of κ_α, $\underline{\kappa}_\alpha$, and p_α for Model 1 under strong conditional dependence ($v = 1$) with fixed universe size ($N = 10$). Table 2 shows analogous results for Model 2, under strong correlation ($\rho = 0.75$). Analogous results for Model 1 (with

$v = 0.2$) and Model 2 (with $\rho = 0.25$) are reported in Table S1 and S2 in the Supplementary Materials. Across all settings, the size of the SCS decreases rapidly with T. As expected, higher confidence levels $1 - \alpha$ lead to more inclusive screening, yielding larger SCSs and improved coverage. For sufficiently large T, the coverage approaches or exceeds the nominal level $1 - \alpha$, with the only exception being expected shortfall (ES) loss with $v = 0.2$, which requires $T > 1000$ to reach the nominal coverage, confirming the asymptotic validity and size properties established in Section 3. For smaller T, coverage often falls below the nominal level, reflecting the limitations of asymptotic approximations and increased test statistic variance in finite samples.

$L(\mu, \sigma^2)$	T	$1-\alpha = 0.90$			$1-\alpha = 0.95$			$1-\alpha = 0.99$		
		κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$
$-\mu/\sigma$	100	125.7	71.3	3.4	294.5	84.7	5.1	630.9	97.3	7.6
		(7.0)	(2.6)	(0.1)	(13.0)	(2.1)	(0.1)	(14.8)	(1.0)	(0.1)
	250	67.5	82.3	2.8	143.2	91.7	4.1	418.8	99.3	6.4
		(4.7)	(2.2)	(0.1)	(6.8)	(1.6)	(0.1)	(13.2)	(0.6)	(0.1)
	1000	13.1	89.0	1.6	25.4	98.0	2.2	81.2	99.3	3.6
		(0.7)	(1.8)	(0.1)	(1.4)	(0.8)	(0.1)	(3.8)	(0.6)	(0.1)
$-0.5\,\mu + \sigma^2$	100	290.4	71.0	4.4	523.6	87.7	5.9	864.5	96.3	8.3
		(19.5)	(2.6)	(0.1)	(21.5)	(2.0)	(0.1)	(16.5)	(1.2)	(0.1)
	250	167.6	84.3	3.6	364.0	91.0	4.7	727.9	100.0	6.9
		(12.3)	(2.3)	(0.1)	(19.1)	(1.5)	(0.1)	(18.8)	(0.0)	(0.1)
	1000	19.7	92.3	2.2	56.0	96.3	2.7	205.5	100.0	3.9
		(1.6)	(1.4)	(0.1)	(5.3)	(0.8)	(0.1)	(13.4)	(0.0)	(0.1)
$-\mu + \sigma\,\frac{\phi(z_{0.1})}{0.1}$	100	8.9	71.0	1.6	14.6	81.0	2.2	33.1	90.7	3.8
		(0.3)	(2.6)	(0.1)	(0.5)	(2.3)	(0.1)	(1.1)	(1.7)	(0.2)
	250	5.9	80.0	1.3	9.2	88.3	1.3	15.2	96.3	1.6
		(0.2)	(2.3)	(0.0)	(0.3)	(1.9)	(0.0)	(0.4)	(1.1)	(0.1)
	1000	3.0	95.3	1.1	4.4	97.7	1.2	6.6	100.0	1.2
		(0.1)	(1.2)	(0.0)	(0.1)	(0.9)	(0.0)	(0.2)	(0.0)	(0.0)

Table 1: Monte Carlo estimates of SCS size κ_α, coverage probability p_α (in %), and lower boundary size $\underline{\kappa}_\alpha$ under the Sharpe ratio loss ($-\mu/\sigma$), mean-variance loss ($-0.5\,\mu + \sigma^2$), and expected shortfall loss ($-\mu + \sigma\,\phi(z_{0.1})/0.1$). Data are generated from Model 1 with $v = 1$ and $N = 10$. Monte Carlo standard errors are shown in parentheses.

The size of the lower boundary subset $\underline{\widehat{\mathcal{S}}}_\alpha$ is substantially smaller than that of the SCS

$L(\mu, \sigma^2)$	T	$1-\alpha = 0.90$			$1-\alpha = 0.95$			$1-\alpha = 0.99$		
		κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$
$-\mu/\sigma$	100	74.9	90.7	3.6	196.4	95.7	4.8	557.3	98.7	6.9
		(7.2)	(1.7)	(0.1)	(14.2)	(1.2)	(0.1)	(20.5)	(0.7)	(0.1)
	250	17.8	95.3	2.6	60.4	98.0	3.6	223.3	99.3	5.3
		(2.3)	(1.2)	(0.1)	(5.8)	(0.8)	(0.1)	(13.2)	(0.5)	(0.1)
	1000	3.3	97.0	1.8	4.7	99.7	2.1	10.8	100.0	2.5
		(0.1)	(1.0)	(0.0)	(0.2)	(0.3)	(0.0)	(0.8)	(0.0)	(0.0)
$-0.5\,\mu + \sigma^2$	100	116.4	90.7	4.0	306.6	96.0	5.5	645.2	99.7	7.8
		(11.1)	(1.7)	(0.1)	(18.8)	(1.1)	(0.1)	(20.7)	(0.3)	(0.1)
	250	32.4	94.0	2.9	73.6	98.3	3.5	321.9	99.7	5.7
		(3.5)	(1.4)	(0.1)	(7.5)	(0.7)	(0.1)	(17.2)	(0.3)	(0.1)
	1000	3.4	97.3	1.8	5.8	99.3	2.0	14.5	100.0	2.5
		(0.1)	(0.9)	(0.0)	(0.4)	(0.5)	(0.0)	(1.2)	(0.0)	(0.0)
$-\mu + \sigma\,\frac{\phi(z_{0.1})}{0.1}$	100	2.2	91.0	1.0	2.6	97.3	1.1	4.1	98.3	1.3
		(0.1)	(1.7)	(0.0)	(0.1)	(0.9)	(0.0)	(0.1)	(0.7)	(0.0)
	250	1.8	97.0	1.0	2.1	99.0	1.0	2.8	99.7	1.0
		(0.1)	(1.0)	(0.0)	(0.1)	(0.6)	(0.0)	(0.1)	(0.3)	(0.0)
	1000	1.2	99.7	1.0	1.4	100.0	1.0	1.6	100.0	1.0
		(0.0)	(0.3)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 2: Monte Carlo estimates of SCS size κ_α, coverage probability p_α (in %), and lower boundary size $\underline{\kappa}_\alpha$ under the Sharpe ratio loss ($-\mu/\sigma$), mean-variance loss ($-0.5\,\mu + \sigma^2$), and expected shortfall loss ($-\mu + \sigma\,\phi(z_{0.1})/0.1$). Data are generated from Model 2 with $\rho = 0.75$ and $N = 10$. Monte Carlo standard errors are shown in parentheses.

$\widehat{\mathcal{S}}_\alpha$, often comprising fewer than ten portfolio selections. Its constituent selections typically correspond to portfolios concentrated on a small number of assets. These sparse combinations have higher variance in estimated performance and are thus more likely to attain extreme empirical performance near the decision boundary. Similarly to the full SCS, the size of $\underline{\widehat{\mathcal{S}}_\alpha}$ decreases with T and both sets eventually concentrate on the true optimal set $\mathcal{S}_0$ as $T \to \infty$.

We observe that both the asset dependence structure and loss function affect selection uncertainty, as reflected in the size of the SCS. For example, in Model 1, the stronger conditional dependence with $v = 1$ reduces the SCS size under the mean-variance loss compared to $v = 0.2$. Stronger dependence increases portfolio variances $\sigma_{\mathbf{s}}$, amplifying the loss differentials $\hat{\delta}(\mathbf{s})$ across selections. As this signal gain outweighs the rise in estimation noise,

the power of the screening test improves, resulting in smaller SCSs. We find that the choice of loss function leads to heterogeneous SCS sizes, primarily due to different sensitivity to estimation error. For example, the mean–variance loss is more sensitive than the ES loss, resulting in larger confidence sets. By contrast, the ES loss yields more stable inference and requires smaller samples to achieve comparable precision.

We examine how the number of assets N influences selection uncertainty. Figure 2 shows that the expected size of the SCS increases with N. As N grows, the number of feasible EWPs, $2^N - 1$, rises exponentially, leading to a sharp increase in κ_α. For example, in Model 2 with $\rho = 0.25$ and under the Sharpe ratio loss, κ_α increases from about 8 to 231 as N rises from 6 to 14, considering the 95% confidence level. This growth honestly reflects the intrinsic uncertainty of a larger asset universe: as the number of feasible portfolios explodes, many display nearly identical empirical performance, making it harder to confidently identify a single superior strategy.

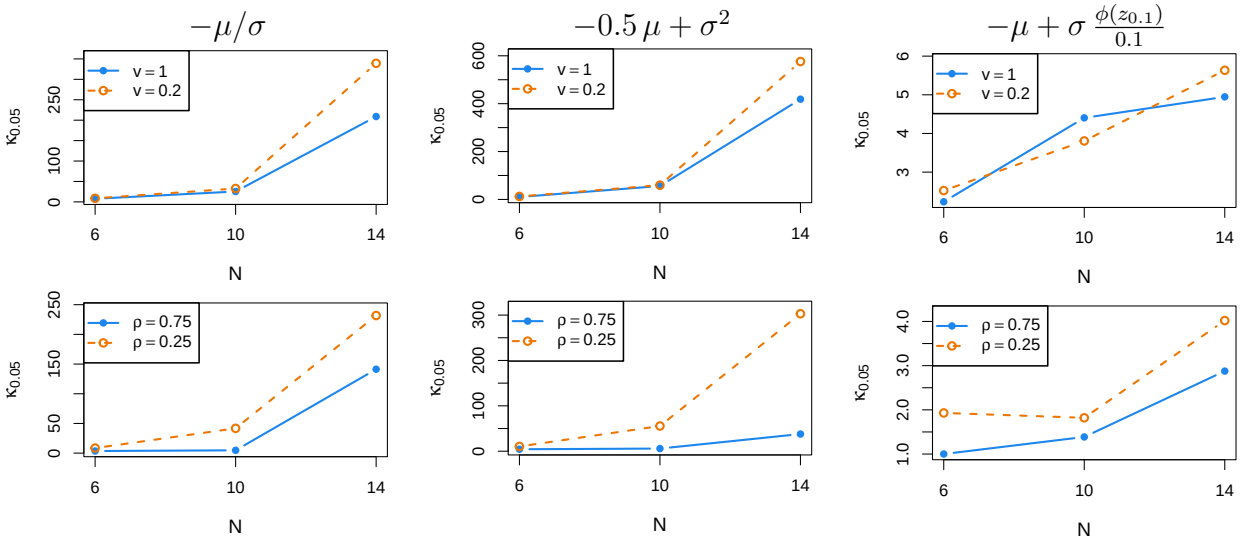

Figure 2: Monte Carlo estimates of SCS size κ_α at the 95% confidence level. Columns correspond to Sharpe ratio loss $(-\mu/\sigma)$, mean-variance loss $(-0.5\,\mu + \sigma^2)$, and expected shortfall loss $(-\mu + \sigma\,\phi(z_{0.1})/0.1)$. Top and bottom rows correspond, respectively, to Model 1 ($v \in \{0.2, 1\}$) and Model 2 ($\rho \in \{0.25, 0.75\}$).

v	T	$1-\alpha=0.90$			$1-\alpha=0.95$			$1-\alpha=0.99$		
		κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$	κ_α	$p_\alpha\%$	$\underline{\kappa}_\alpha$
1	100	310.7	91.0	5.2	447.4	94.7	6.3	747.4	100.0	8.2
		(13.4)	(1.7)	(0.1)	(15.2)	(1.3)	(0.1)	(14.0)	(0.0)	(0.1)
	250	133.8	93.0	3.9	248.0	97.7	5.2	515.0	100.0	7.1
		(6.5)	(1.5)	(0.1)	(10.3)	(0.9)	(0.1)	(12.9)	(0.0)	(0.1)
	1000	25.7	95.7	2.2	42.5	98.7	2.8	118.6	99.7	4.3
		(1.5)	(1.2)	(0.1)	(2.0)	(0.7)	(0.1)	(5.0)	(0.3)	(0.1)
0.2	100	294.5	84.0	5.3	441.3	91.0	6.4	755.7	100.0	8.4
		(13.1)	(2.1)	(0.1)	(14.5)	(1.7)	(0.1)	(13.3)	(0.0)	(0.1)
	250	164.3	92.0	4.4	279.1	95.3	5.5	543.3	99.7	7.3
		(8.6)	(1.6)	(0.1)	(11.3)	(1.2)	(0.1)	(13.8)	(0.3)	(0.1)
	1000	31.9	96.3	2.5	51.8	98.7	3.0	126.1	99.7	4.5
		(1.7)	(1.1)	(0.1)	(2.6)	(0.7)	(0.1)	(5.1)	(0.3)	(0.1)

Table 3: Monte Carlo estimates of SCS size κ_α, coverage probability p_α (in %), and lower boundary size $\underline{\kappa}_\alpha$ for SCSs constructed using the Ledoit and Wolf bootstrap test for Sharpe ratio differences (Ledoit and Wolf, 2008). Monte Carlo standard errors are in parentheses.

5.3 Construction via Ledoit and Wolf Sharpe ratio test

We replicate the simulation setup from Section 5.1, using the Ledoit and Wolf (LW) bootstrap test for Sharpe ratio differences (Ledoit and Wolf, 2008) to construct the SCS. The LW procedure aims to improve finite-sample accuracy of the studentized statistic $Z(\mathbf{s}, \mathbf{s}')$ by bootstrapping. The analysis is conducted using the R function `sharpeTesting` in the package `PeerPerformance` (Ardia and Boudt, 2018). In Table 3, we report Monte Carlo estimates of κ_α, $\underline{\kappa}_\alpha$, and p_α under Model 1 for $v \in \{0.2, 1\}$. The LW-based SCS is typically larger than that obtained via the Wald test under asymptotic normality, with differences becoming more pronounced in larger samples. For example, with $1-\alpha = 0.95$ and $v = 1$, the LW procedure yields an expected SCS size of about 447 for $T = 100$ and 43 for $T = 1000$, compared to 295 and 25, respectively, under the asymptotic approach. This reflects a greater conservativeness of LW and potentially lower power. However, for smaller samples, the LW test achieves coverage closer to the nominal level, better controlling type I error when asymptotic approximations break down.

6 Analysis of Cryptocurrency and Industry Portfolios

In this section, we illustrate the SCS methodology by analyzing two datasets: (i) the *L1-Crypto* dataset, containing daily log-returns for 16 Layer-1 cryptocurrencies from January 1, 2022, to January 1, 2025 ($T = 1095$) (source: Yahoo Finance, `https://finance.yahoo.com`); and (ii) the French *17 Industry* dataset, comprising daily returns for 17 equal-weighted U.S. industry portfolios from November 1, 2021, to October 31, 2024 ($T = 755$). Equal-weighted series are used to avoid large-cap bias, providing a more balanced view of market performance (source: K. French data library `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/`). Tables S3 and S4 in the Supplementary Materials report additional information on the asset composition. The two datasets represent distinct market environments: the L1-Crypto data reflect high volatility and speculative dynamics, while the 17 Industry portfolios capture the stability of mature equity markets.

We study selection uncertainty under two loss functions for each dataset: the Sharpe ratio loss, suited to the reward-to-variability focus of crypto markets, and the mean–variance loss, which is more aligned with the risk-sensitive nature of traditional equities. Point estimates of the optimal EWP obtained via exhaustive search on the L1-Crypto data are {TRON , Mantra} under the Sharpe ratio loss, and {Bitcoin, TRON} under the mean–variance loss. For the 17 Industry data, the corresponding estimates are {Steel, Utilities} under both the Sharpe ratio and the mean–variance losses. Next, we apply our methodology to assess the selection uncertainty surrounding these point estimates.

Table 4 reports the global post-selection metrics defined in Section 4 for each dataset: SCS size ($|\widehat{\mathcal{S}}_\alpha|$), lower boundary size ($|\underline{\widehat{\mathcal{S}}_\alpha}|$), Relative Multiplicity Index (RMI), loss bounds ($\hat{L}_0$, $\hat{L}_{\max}$), and performance spread ($\hat{\delta}_\alpha$). The size of the SCS directly reflects the degree of selection uncertainty in each dataset. For example, at the 95% confidence level, the SCS under the Sharpe ratio for the L1-Cryptocurrency data includes 122 equally weighted portfolios (EWPs) that are statistically indistinguishable from an optimal EWP, compared to 430 in the 17 Industry dataset. This suggests that uncertainty is more contained in the

L1-Cryptocurrency case. Although this may seem counterintuitive given the high volatility of cryptocurrencies, it reflects the fact that a few dominant assets drive Sharpe ratio performance, limiting the number of near-optimal configurations. In contrast, the larger number of equivalent strategies in the 17 Industry dataset indicates greater redundancy among sectors, with many combinations yielding comparable risk-adjusted performance and thus inflating the selection ambiguity. When accounting for the total size of the selection space $\mathcal{S}$ in each dataset, the RMI index confirms that selection uncertainty is lower for the L1-Cryptocurrency data (RMI= 56.68) compared to that of the 17 Industry data (RMI= 48.54).

As expected, higher confidence levels yield larger SCSs. The increase is substantial – for instance, under the Sharpe ratio loss, changing the confidence level from 95% to 99% increases the SCS size from 122 to 8,128 for the L1-Crypto dataset, and from 430 to 10,178 for the 17 Industry dataset. The RMI decreases as the confidence level increases, reflecting the dilution of explanatory power across a larger set of selected portfolios. This highlights a fundamental trade-off: higher confidence improves coverage but reduces selectivity, making it harder to identify clearly dominant investment strategies. On the other hand, the LB set remains small – just 12 for cryptocurrencies and 9 for equities at the 99% confidence level under the Sharpe ratio loss. These lean allocations cannot be statistically rejected and show remarkable stability across confidence levels, suggesting a robust core of individually strong-performing assets.

In Figure 3, we plot the marginal inclusion importance (II_α) for each asset over $\alpha \in [0, 0.05]$ for the L1-Crypto and 17-Industry datasets, under the Sharpe ratio and mean–variance loss, respectively. Some assets maintain high and stable inclusion importance as α increases, indicating frequent selection in top-performing portfolios across confidence levels. Others show a rapid decline, with their trajectories quickly clustering below the rest, signaling weak and unstable contributions to portfolio performance. In the L1-Crypto data, Bitcoin (BTC) and Mantra (OM) consistently exhibit large inclusion importance, reflecting their central role under the Sharpe ratio criterion. In the 17-Industry data, Oil & Gas, Steel, and Con-

Dataset	$L(\mu, \sigma)$	$(1-\alpha)\%$	$\lvert\widehat{\mathcal{S}}_\alpha\rvert$	$\lvert\underline{\widehat{\mathcal{S}}_\alpha}\rvert$	RMI (%)	$\hat{L}_0$ (%)	$\hat{L}_{\max}$ (%)	$\hat{\delta}_\alpha$ (%)
L1-Crypto	$-\mu/\sigma$	90	20	3	72.99	-5.09	1.68	6.77
		95	122	9	56.68	-5.09	3.37	8.46
		99	8,128	12	18.82	-5.09	3.37	8.46
	$-\mu + 0.5\sigma^2$	90	85	4	59.94	0.03	0.17	0.14
		95	288	5	48.94	0.03	0.24	0.21
		99	3,068	9	27.61	0.03	0.24	0.21
17 Industry	$-\mu/\sigma$	90	115	5	59.73	-2.96	2.27	5.23
		95	430	7	48.54	-2.96	2.43	5.39
		99	10,178	9	21.68	-2.96	5.02	7.98
	$-\mu + 0.5\sigma^2$	90	112	4	59.96	-0.02	0.03	0.05
		95	408	7	48.99	-0.02	0.07	0.09
		99	8,291	8	23.43	-0.02	0.07	0.09

Table 4: Post-selection metrics for the L1-Crypto and 17 Industry data under the Sharpe ratio ($-\mu/\sigma$) and mean–variance ($-\mu + 0.5\sigma^2$) loss functions, evaluated at confidence levels $(1-\alpha)\%$: SCS size ($|\widehat{\mathcal{S}}_\alpha|$), lower boundary size ($|\underline{\widehat{\mathcal{S}}_\alpha}|$), percent Relative Multiplicity Index (RMI), loss bounds ($\hat{L}_0$, $\hat{L}_{\max}$), and performance spread ($\hat{\delta}_\alpha = \hat{L}_{\max} - \hat{L}_0$).

struction Materials stand out dominate the other sectors in terms of inclusion importance under the mean–variance criterion.

These plots illustrate a key strength of our approach: important assets may be excluded from the empirical optimum due to sampling variability, yet still emerge as influential within the full Selection Confidence Set (SCS). In the L1-Crypto data under the mean-variance loss, the two assets composing the empirical optimum (TRX and OM) consistently receive high inclusion importance across all confidence levels. However, other key assets – such as Bitcoin (BTC) – are not part of the best empirical EWP despite their relevance. Similar considerations apply to the 17 Industry data. While Steel and Utilities are part of the empirically optimal portfolio under the minimum-loss, there are several sectors with high inclusion importance – such as Oil & Gas, Construction Materials Drugs and Mines – entirely ignored by the single selected EWP. By examining the entire SCS, we identify systematically important assets that a single optimal EWP would overlook, highlighting how the full SCS reveals structural relevance of assets beyond a single empirically selected EWP.

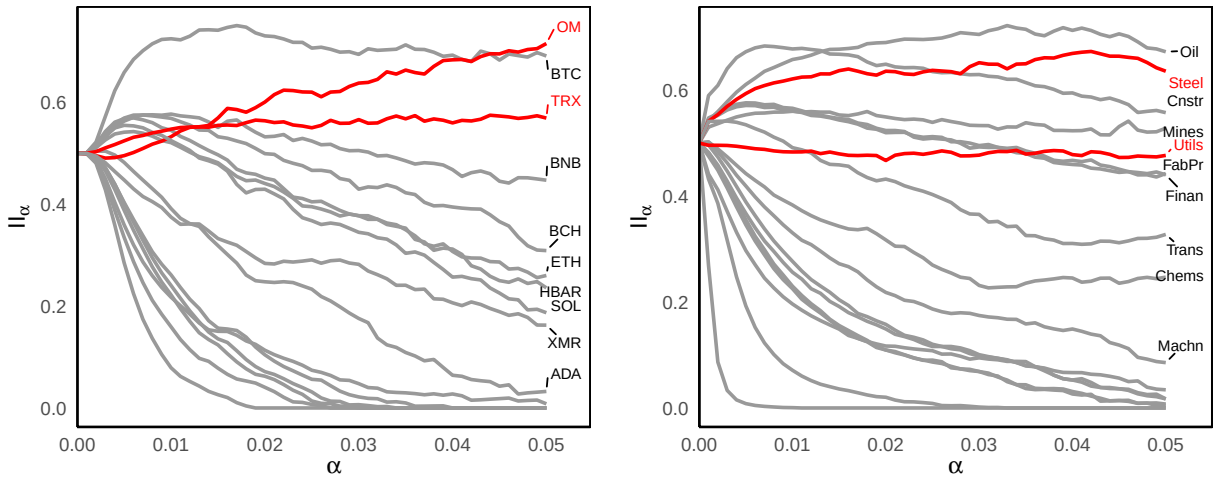

Figure 3: Inclusion importance (Π_α) as a function of α for the L1-Crypto dataset under the Sharpe ratio loss (left) and for the 17 Industry dataset under the mean-variance loss (right). Red lines mark assets included in the empirical optimal portfolio. Labels are reported for the 10 assets with the largest Π_α value at $\alpha = 0.05$.

Figure 4 shows graphs representing the co-inclusion importance metric (CII_α) at the 95% confidence level. Both datasets exhibit dense connectivity, indicating that many asset pairs frequently co-appear in EWPs within the SCS. For the L1-Crypto data, a few assets such as Bitcoin (BTC) and Mantra (OM) are highly central with many edges concentrated around these assets, reflecting their frequent joint inclusion in high-Sharpe portfolios. For the 17-Industry data, Oil & Gas, Steel, Utilities, and Construction Materials dominate in centrality, indicating that these sectors are consistently selected alongside others under the mean–variance criterion. This pattern reflects strong complementarities, where their inclusion typically enhances overall portfolio performance. By contrast, isolated nodes represent assets with limited complementarity, contributing little when combined with others and often playing a marginal or standalone role in portfolio construction.

Overall, although cryptocurrencies and industries differ markedly in their return and volatility profiles, both yield similarly large confidence sets – highlighting the inherent uncertainty in identifying a single optimal EWP. Rather than relying solely on a unique selection, our method complements it with richer information by uncovering a range of statistically indistinguishable alternatives and systematically important assets, offering a more robust

foundation for portfolio decisions.

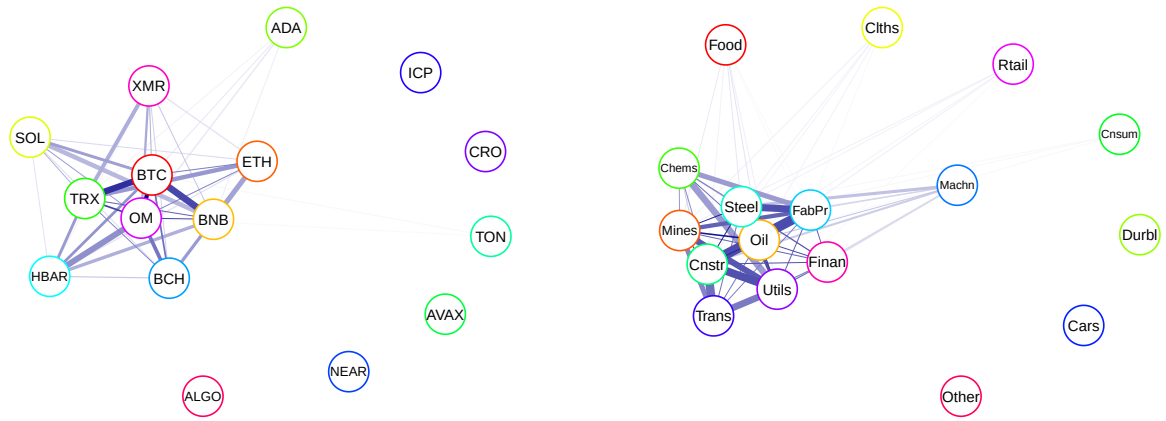

Figure 4: Co-inclusion importance (CII_α) graphs at the 95% confidence level for the L1-Crypto dataset under the Sharpe ratio loss (left) and for the 17 Industry dataset under the mean-variance loss (right). The edges represent pairs of assets such that $\text{CII}_\alpha > 0.01$; thicker edges correspond to larger co-inclusion values.

7 Conclusion

Statistically grounded confidence sets provide a rigorous framework for addressing a long-standing but often unquantified issue in portfolio theory – the abundance of nearly optimal solutions arising from uncertainty in expected return and risk estimates. While estimation risk for a single portfolio is well documented, our SCS approach addresses statistical uncertainty by identifying all portfolios that are indistinguishable from the true optimum at a specified confidence level. Using a flexible framework compatible with a broad class of performance losses, we establish asymptotic coverage and show that the expected size of the SCS depends on the signal-to-noise ratios of competing portfolios—shrinking exponentially in T as identification strengthens.

We remark that the test used to construct the SCS is asymptotically most powerful among those based on normally distributed estimators (see Lehmann and Romano (2005)). As such, the resulting SCS serves as a best-case benchmark for our pairwise approach: its size reflects

the minimal inclusion needed to control type I error at level α. Despite this optimality, the SCS can still be large when uncertainty is high, particularly when performance differences are faint. This largeness is not a flaw, but a statistical necessity, reflecting the indistinguishability between the true optimum and many near-optimal portfolios. Without additional structural information on the space of feasible portfolios $\mathcal{S}$, such uncertainty cannot be further reduced. Future work may explore incorporating shape constraints, economic priors, or regularization to refine the SCS without compromising coverage.

While our analysis assumes a fixed number of assets N, extending the methodology to high-dimensional settings presents nontrivial theoretical and computational challenges. Nevertheless, the asymptotic upper bound in Equation (15) shows that the SCS size decays exponentially with T. This suggests that, under appropriate sparsity or separability conditions on the loss function, selection complexity may remain tractable even as N increases, offering a promising direction for both theoretical and computational extensions to high-dimensional portfolio problems.

Our numerical experiments confirm the theoretical behavior of the SCS, with coverage approaching the nominal level as sample size increases. However, in high-dimensional settings with weakly identified loss differentials, coverage may fall short of $1-\alpha$ due to limitations of asymptotic approximations. Based on the observed finite sample behavior, future work may consider bootstrap or higher-order corrections for $Z(\mathbf{s}, \hat{\mathbf{s}}_0)$, or ad hoc tests tailored to specific loss functions. However, we observed that while the bootstrapped test of (Ledoit and Wolf, 2008) slightly improves coverage in small samples, it does not fundamentally reduce selection uncertainty under weak signals and often yields larger SCSs than those obtained using asymptotic normality (see Table 3). The proof of Proposition 1 shows that SCS validity relies on the consistency of the selection procedure—specifically, a high probability of correct selection $\mathbb{P}(\hat{\mathbf{s}}_0 \in \mathcal{S}_0)$. While this holds asymptotically, finite-sample performance may suffer, particularly in high-dimensional or weak-signal settings, leading to reduced coverage. Incorporating uncertainty in $\hat{\mathbf{s}}_0$ – e.g., by adjusting the variance estimator or using bootstrap

methods – could improve the finite-sample accuracy of the SCS.

Through our real-data illustrations, we demonstrate that selection uncertainty is not an abstract concept but a concrete feature of portfolio construction. In both cryptocurrency and equity markets, a large number of EWPs cannot be statistically distinguished from the optimum under standard loss functions. This finding highlights the limitations of approaches relying solely on point estimates and motivates a shift toward distributional characterizations of portfolio selection. The diagnostic tools introduced (such as the RMI, loss intervals, inclusion importance, and co-inclusion networks) are interpretable metrics to assess the stability and structure of the SCS. These tools provide practitioners with a principled basis for evaluating portfolio robustness, guiding diversification choices, and understanding the extent of selection ambiguity.

Disclosure statement

The authors declare that there are no conflicts of interest to disclose.

Data Availability Statement

The data that support the findings of this study are openly available in Harvard Dataverse at `https://doi.org/10.7910/DVN/VG0FU6`, reference number V1.

References

A. Amendola, M. Braione, V. Candila, and G. Storti. A model confidence set approach to the combination of multivariate volatility forecasts. *International Journal of Forecasting*, 36(3):873–891, 2020. doi: 10.1016/j.ijforecast.2019.10.001.

D. W. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858, 1991.

D. Ardia and K. Boudt. The peer performance ratios of hedge funds. *Journal of Banking & Finance*, 87:351–368, 2018.

Y. S. Asawa. Modern machine learning solutions for portfolio selection. *IEEE Engineering Management Review*, 50(1):94–112, 2021.

D. Bauder, T. Bodnar, N. Parolya, and W. Schmid. Bayesian mean–variance analysis: Optimal portfolio selection under parameter uncertainty. *Quantitative Finance*, 21(2): 221–242, 2021.

O. Bodnar, T. Bodnar, and V. Niklasson. Incorporating different sources of information for bayesian optimal portfolio selection. *Journal of Business & Economic Statistics*, 43(2): 365–377, 2025. doi: 10.1080/07350015.2024.2379361.

T. Bodnar, M. Lindholm, V. Niklasson, and E. Thorsén. Bayesian portfolio selection using var and cvar. *Applied Mathematics and Computation*, 427:127120, 2022a.

T. Bodnar, Y. Okhrin, and N. Parolya. Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, 41(1):140–156, 2022b. doi: 10.1080/07350015.2021.2004897.

T. Bodnar, N. Parolya, and E. Thorsén. Two is better than one: Regularized shrinkage of large minimum variance portfolios. *Journal of Machine Learning Research*, 25(173):1–32, 2024.

A. Caparrini, J. Arroyo, and J. Escayola Mansilla. S&p 500 stock selection using machine learning classifiers: A look into the changing role of factors. *Research in International Business and Finance*, 70:102336, June 2024. doi: 10.1016/j.ribaf.2024.102336.

W. Chen, H. Zhang, M. K. Mehlawat, and L. Jia. Mean–variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, 100:106943, 2021.

Z. Chen, L. Zhang, and C. Weng. Does climate policy uncertainty affect chinese stock market volatility? *International Review of Economics & Finance*, 84:369–381, 2023. doi: 10.1016/j.iref.2022.11.030.

V. K. Chopra and W. T. Ziemba. *The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice*, pages 365–373. 2013.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2009.

V. DeMiguel, A. Martin-Utrera, and F. J. Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

Y. Ding, Y. Li, and X. Zheng. High dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics*, 222(1):502–515, 2021. doi: 10.1016/j.jeconom.2020.07.013.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32:407–499, 2004.

E. Fama and K. French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.

D. Ferrari and Y. Yang. Condence sets for model selection by f-testing. *Statistica Sinica*, pages 1637–1658, 2015.

S. Goto and Y. Xu. Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415–1441, 2015. doi: 10.1017/S0022109015000526.

P. Hansen, A. Lunde, and N. J.M. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

H. Jiang, X. Fei, H. Liu, K. Roeder, J. Lafferty, L. Wasserman, X. Li, and T. Zhao. *huge: High-Dimensional Undirected Graph Estimation*, 2021. URL `https://CRAN.R-project.org/package=huge`. R package version 1.3.5.

M. Kritzman, S. Page, and D. Turkington. In defense of optimization: the fallacy of 1/n. *Financial Analysts Journal*, 66:31–39, 2010.

O. Ledoit and M. Wolf. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859, 2008.

O. Ledoit and M. Wolf. Robust performances hypothesis testing with the variance. *Wilmott*, (55):86–89, 2011.

O. Ledoit and M. Wolf. Robust performance hypothesis testing with smooth functions of population moments. University of Zurich, Department of Economics, Working Paper, (305), 2018.

S. I. Lee. Deeply equal-weighted subset portfolios. *arXiv preprint arXiv:2006.14402*, 2020.

E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer, 2005.

M. Leippold and H. Yang. Mixed-frequency predictive regressions with parameter learning. *Journal of Forecasting*, 42(8):1955–1972, 2023. doi: 10.1002/for.2999.

C. Liang, M. Umar, F. Ma, and T. L. D. Huynh. Climate policy uncertainty and world renewable energy index volatility forecasting. *Technological Forecasting and Social Change*, 182:121810, 2022. doi: 10.1016/j.techfore.2022.121810.

R. Malladi and F. J. Fabozzi. Equal-weighted strategy: Why it outperforms value-weighted strategies? theory and evidence. *Journal of Asset Management*, 18:188–208, 2017.

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

C. Memmel. Performance hypothesis testing with the sharpe ratio. *Finance Letters*, 1(1): 21–23, 2003. URL `https://ssrn.com/abstract=412588`.

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

Y. Plyakha, R. Uppal, and G. Vilkov. Why do equal-weighted portfolios outperform value-weighted portfolios. *SSRN Electronic Journal*, 2015.

S. A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13 (3):341–360, 1976.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

C. Zheng, D. Ferrari, and Y. Yang. Model selection confidence sets by likelihood ratio testing. *Statistica Sinica*, 29(2):827–851, 2019.

8 Appendix: Proofs of Propositions

Proof of Proposition 1. Fix any $\mathbf{s}_0 \in \mathcal{S}_0$. Using the law of total probability we obtain the following lower bound for inclusion probability $\mathbb{P}(\mathbf{s}_0 \in \widehat{\mathcal{S}}_\alpha)$

$$\mathbb{P}(\mathbf{s}_0 \in \widehat{\mathcal{S}}_\alpha) = \mathbb{P}(Z(\mathbf{s}_0, \hat{\mathbf{s}}_0) \leq q_{1-\alpha}) \geq \mathbb{P}(Z(\mathbf{s}_0, \hat{\mathbf{s}}_0) \leq q_{1-\alpha} \mid \hat{\mathbf{s}}_0 \in \mathcal{S}_0)\, \mathbb{P}(\hat{\mathbf{s}}_0 \in \mathcal{S}_0). \tag{17}$$

Step 1. We first show $\mathbb{P}(\hat{\mathbf{s}}_0 \in \mathcal{S}_0) \to 1$ as $T \to \infty$. Let $L_Z(\mathbf{s}) := L(\hat{\mu}_\mathbf{s}, \hat{\sigma}^2_\mathbf{s})$ and $L(\mathbf{s}) := L(\mu_\mathbf{s}, \sigma^2_\mathbf{s})$. By Assumption (A2), for each fixed $\mathbf{s} \in \mathcal{S}$ $(\hat{\mu}_\mathbf{s}, \hat{\sigma}^2_\mathbf{s}) \xrightarrow{p} (\mu_\mathbf{s}, \sigma^2_\mathbf{s})$, and by continuity of L, which implies $L_Z(\mathbf{s}) \xrightarrow{p} L(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}$ and $\max_{\mathbf{s}\in\mathcal{S}} |L_Z(\mathbf{s}) - L(\mathbf{s})| \xrightarrow{p} 0$. Since $\mathcal{S}$ is finite

and $L(\mathbf{s}) > L(\mathbf{s}_0)$ for all $\mathbf{s} \notin \mathcal{S}_0$, define the minimal separation margin between any non-optimal and optimal selections as: $\epsilon := \min_{\mathbf{s} \notin \mathcal{S}_0} \min_{\mathbf{s}_0 \in \mathcal{S}_0} \{L(\mathbf{s}) - L(\mathbf{s}_0)\} > 0$, For sufficiently large T, with probability tending to one, we have $|L_Z(\mathbf{s}) - L(\mathbf{s})| < \epsilon/4 \quad$ for all $\mathbf{s} \in \mathcal{S}$. Now fix any $\mathbf{s}_0 \in \mathcal{S}_0$ and any $\mathbf{s} \notin \mathcal{S}_0$. On this high-probability event, we have:

$$L_Z(\mathbf{s}) > L(\mathbf{s}) - \epsilon/4 \geq L(\mathbf{s}_0) + \epsilon - \epsilon/4 = L(\mathbf{s}_0) + 3\epsilon/4 > L_Z(\mathbf{s}_0) + \epsilon/2.$$

This shows that, with probability tending to one, all non-minimizers have strictly larger empirical loss than any population minimizer. Therefore, Step 1 is complete.

Step 2. We now show that under the event $\{\hat{\mathbf{s}}_0 \in \mathcal{S}_0\}$, the test statistic

$$Z(\mathbf{s}_0, \hat{\mathbf{s}}_0) = \frac{L_Z(\mathbf{s}_0) - L_Z(\hat{\mathbf{s}}_0)}{\hat{\tau}(\mathbf{s}_0, \hat{\mathbf{s}}_0)/\sqrt{T}}.$$

converges in distribution to a standard normal under the null hypothesis $H_0 : L(\mathbf{s}_0) = L(\hat{\mathbf{s}}_0)$. For fixed $\hat{\mathbf{s}}_0$, by Assumption (A2), the empirical moment estimators are jointy asymptotically normal. Hence, by the delta method, $\sqrt{T}\tau(\mathbf{s}_0, \hat{\mathbf{s}}_0)[L_Z(\mathbf{s}_0) - L_Z(\hat{\mathbf{s}}_0)] \xrightarrow{d} \mathcal{N}(0, 1)$. By Assumption (A3), the estimated standard error in the denominator satisfies $\hat{\tau}(\mathbf{s}_0, \hat{\mathbf{s}}_0) \xrightarrow{p} \tau(\mathbf{s}_0, \hat{\mathbf{s}}_0)$. By Slutsky's theorem, the test statistic satisfies $Z(\mathbf{s}_0, \hat{\mathbf{s}}_0) \xrightarrow{d} \mathcal{N}(0, 1)$ under the null. Therefore, the probability of not rejecting under the null satisfies

$$\mathbb{P}(Z(\mathbf{s}_0, \hat{\mathbf{s}}_0) \leq q_{1-\alpha} \mid \hat{\mathbf{s}}_0 \in \mathcal{S}_0) \to 1 - \alpha.$$

From Step 1 and Step 2, taking the limit for $T \to \infty$ for the right hand side in (17) gives $(1 - \alpha)$, which concludes the proof. □

Proof of Proposition 2. Let $\hat{\delta}(\mathbf{s}) = \hat{L}(\mathbf{s}) - \hat{L}(\hat{\mathbf{s}}_0)$ denote the empirical loss difference, and let $\delta(\mathbf{s}) = L(\mathbf{s}) - L_0$ denote the population counterpart, where $L_0 = L(\mathbf{s}_0)$ for some $\mathbf{s}_0 \in \mathcal{S}_0$. Conditioning on the event $\hat{\mathbf{s}}_0 = \mathbf{s}_0 \in \mathcal{S}_0$, the Delta method and Assumptions (A1)–(A2),

imply that the test statistic satisfies

$$Z(\mathbf{s}, \hat{\mathbf{s}}_0) \mid \hat{\mathbf{s}}_0 = \mathbf{s}_0 \xrightarrow{d} \mathcal{N}\left(\frac{\delta(\mathbf{s}; \mathbf{s}_0)}{\tau(\mathbf{s}; \mathbf{s}_0)}, 1\right), \quad \text{as } T \to \infty.$$

Therefore, conditioning on $\hat{\mathbf{s}}_0 = \mathbf{s}_0$, the expected size of the selection confidence set satisfies :

$$\mathbb{E}[|\widehat{\mathcal{S}}_\alpha|] = \sum_{\mathbf{s} \in \mathcal{S}_0} \mathbb{P}(Z(\mathbf{s}, \mathbf{s}_0) \leq q_{1-\alpha}) + \sum_{\mathbf{s} \notin \mathcal{S}_0} \mathbb{P}(Z(\mathbf{s}, \mathbf{s}_0) \leq q_{1-\alpha}) + o(1).$$

For $\mathbf{s} \in \mathcal{S}_0$, we have $\delta(\mathbf{s}; \mathbf{s}_0) = 0$, so $Z(\mathbf{s}, \mathbf{s}_0) \xrightarrow{d} \mathcal{N}(0, 1)$ and $\mathbb{P}(Z(\mathbf{s}, \mathbf{s}_0) \leq q_{1-\alpha}) = (1 - \alpha) + o(1)$. For $\mathbf{s} \notin \mathcal{S}_0$, the population loss difference is positive, and the test statistic satisfies $\sqrt{T} Z(\mathbf{s}, \mathbf{s}_0) \xrightarrow{d} \mathcal{N}(\gamma(\mathbf{s}), 1)$ where $\gamma(\mathbf{s}) := \delta(\mathbf{s})/\tau(\mathbf{s})$. Combining both parts, we obtain:

$$\mathbb{E}[|\widehat{\mathcal{S}}_\alpha| \mid \hat{\mathbf{s}}_0 \in \mathcal{S}_0] = |\mathcal{S}_0|(1 - \alpha) + \sum_{\mathbf{s} \notin \mathcal{S}_0} \Phi(q_{1-\alpha} - \sqrt{T}\gamma(\mathbf{s})) + o(1). \tag{18}$$

By the law of total expectation,

$$\mathbb{E}[|\widehat{\mathcal{S}}_\alpha|] = \mathbb{E}[|\widehat{\mathcal{S}}_\alpha| \mid \hat{\mathbf{s}}_0 \in \mathcal{S}_0] \cdot \mathbb{P}(\hat{\mathbf{s}}_0 \in \mathcal{S}_0) + \mathbb{E}[|\widehat{\mathcal{S}}_\alpha| \mid \hat{\mathbf{s}}_0 \notin \mathcal{S}_0] \cdot \mathbb{P}(\hat{\mathbf{s}}_0 \notin \mathcal{S}_0). \tag{19}$$

We already showed in the proof of Proposition 1 that $\mathbb{P}(\hat{\mathbf{s}}_0 \in \mathcal{S}_0) \to 1$ under (A1). Thus, the first term in (19) is $\mathbb{E}[|\widehat{\mathcal{S}}_\alpha| \mid \hat{\mathbf{s}}_0 \in \mathcal{S}_0](1 + o(1))$. Since $\widehat{\mathcal{S}}_\alpha \subseteq \mathcal{S}$, its size is uniformly bounded above by $|\mathcal{S}| < \infty$, and we already showed in the proof of Proposition 1 that $\mathbb{P}(\hat{\mathbf{s}}_0 \notin \mathcal{S}_0) \to 0$. The second term of (19) is upper bounded by $|\mathcal{S}| \cdot \mathbb{P}(\hat{\mathbf{s}}_0 \notin \mathcal{S}_0) = o(1)$. Hence, by substituting (18) into ((19)) we conclude

$$\mathbb{E}[|\widehat{\mathcal{S}}_\alpha|] = |\mathcal{S}_0|(1 - \alpha) + \sum_{\mathbf{s} \notin \mathcal{S}_0} \Phi(q_{1-\alpha} - \sqrt{T}\gamma(\mathbf{s})) + o(1), \tag{20}$$

and concludes the proof. □