Multi-hypothesis prediction for portfolio optimization: a structured ensemble learning approach to risk diversification

Expert Systems With Applications 292 (2025) 128633

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multi-hypothesis prediction for portfolio optimization: A structured ensemble learning approach to risk diversification

Alejandro Rodriguez Dominguez [a,b], Muhammad Shahzad [a,*], Xia Hong [a]

[a] Department of Computer Science, University of Reading Whiteknights House, RG6 6UR, Reading, United Kingdom
[b] Miralta Finance Bank S.A, 28020, Madrid, Spain

ARTICLE INFO

Keywords:
Diversity
Ensemble learning
Multiple hypotheses prediction
Portfolio optimization
Predict-then-optimize
Structured models.

ABSTRACT

This work proposes a unified framework for portfolio allocation, covering both asset selection and optimization, based on a multiple-hypothesis predict-then-optimize approach. The portfolio is modeled as a structured ensemble, where each predictor corresponds to a specific asset or hypothesis. Structured ensembles formally link predictors' diversity, captured via ensemble loss decomposition, to out-of-sample risk diversification. A structured data set of predictor output is constructed with a parametric diversity control, which influences both the training process and the diversification outcomes. This data set is used as input for a supervised ensemble model, the target portfolio of which must align with the ensemble combiner rule implied by the loss. For squared loss, the arithmetic mean applies, yielding the equal-weighted portfolio as the optimal target. For asset selection, a novel method is introduced which prioritizes assets from more diverse predictor sets, even at the expense of lower average predicted returns, through a diversity-quality trade-off. This form of diversity is applied before the portfolio optimization stage and is compatible with a wide range of allocation techniques. Experiments conducted on the full S&P 500 universe and a data set of 1300 global bonds of various types over more than two decades validate the theoretical framework. Results show that both sources of diversity effectively extend the boundaries of achievable portfolio diversification, delivering strong performance across both one-step and multi-step allocation tasks.

1. Introduction

The problem of portfolio allocation with a quantitative approach was introduced by Harry Markowitz in 1952 in the so-called Modern Portfolio Theory (MPT) (Markowitz, 1952). It consists of two phases: asset selection and portfolio weight optimization. The asset selection phase decides which assets from a larger set, such as a market, will be included in the portfolio, which has a fixed number of assets but not predetermined choices or weights. The portfolio weight optimization phase determines the percentage of capital allocated to each selected asset. It also has the merit of introducing the concept of diversification as a fundamental component in risk management for investments. In this specific case, diversification is induced by the covariance matrix (Markowitz, 1952). Since then, a large number of methods known as "plug-in" (or data-driven) approaches have been developed which use optimization inputs with improved estimates of future returns (Best & Grauer, 1991; Broadie, 1993; Giglio, Kelly, & Xiu, 2022).

The standard data-driven approach to portfolio choice optimization replaces population parameters with their sample estimates, which leads to poor out-of-sample performance due to parameter uncertainty (Zhang, Li, & Guo, 2018). This issue is typically addressed by deriving expected loss functions that quantify the impact of using sample means and covariance matrices to estimate the optimal portfolio (Kan & Zhou, 2007). Other researchers have focused on portfolio optimization using robust estimators, which can be computed by solving a single non-linear programming problem (DeMiguel & Nogales, 2007; Lassance, Martin-Utrera, & Simaan, 2023). Shrinkage techniques adjust extreme coefficients toward more central values in the sample covariance matrix to reduce estimation error, often incorporating a parameter to control the level of shrinkage (Bodnar, Parolya, & Thorsen, 2024; Ledoit & Wolf, 2003).

Some researchers have also employed the use of predictive models to optimize decision-making processes, often with the objective of minimizing decision-related costs. Such a conventional "predict-then-optimize" paradigm decouples prediction and decision-making into two distinct stages: firstly, a predictive model is developed to maximize predictive accuracy; secondly, decisions are derived based on the model's outputs and associated cost functions. A significant limitation

* Corresponding author.
E-mail addresses: arodriguez@miraltabank.com, a.j.rodriguezdominguez@pgr.reading.ac.uk (A. Rodriguez Dominguez), m.shahzad2@reading.ac.uk (M. Shahzad), xia.hong@reading.ac.uk (X. Hong).

https://doi.org/10.1016/j.eswa.2025.128633
Received 22 January 2025; Received in revised form 28 May 2025; Accepted 13 June 2025
Available online 18 June 2025

of this approach is its failure to incorporate cost considerations during the predictive modeling phase (Donti & Kolter, 2017). To address this shortcoming, recent advances have introduced integrated methodologies such as "predict-and-optimize" (Vanderschueren, Verdonck, Baesens, & Verbeke, 2022; Wilder, Dilkina, & Tambe, 2019) and Smart "Predict, then Optimize" (SPO) (Elmachtoub & Grigas, 2017). These approaches bridge the gap between structured prediction and cost-sensitive optimization, offering a robust solution for enhancing decision-making processes (Kotary, Vito, Christopher, Hentenryck, & Fioretto, 2023). These approaches, however, are computationally expensive and mostly focus on solving the classification problems (Goh & Jaillet, 2016).

While the aforementioned solutions reduce parameter uncertainty and can improve out-of-sample performance, they fail to establish a connection between in-sample and out-of-sample diversification, which is a key feature addressed in this work. In this context, a framework is proposed for portfolio optimization using structured ensembles for prediction with multiple hypotheses. Each portfolio component corresponds to a hypothesis, with its return series modeled by an individual predictor. This approach differs from return forecasting (Ma, Han, & Wang, 2021) or scoring models (Nguyen & Lo, 2012; Zhao, Liu, Wu, Zhang, & Zhang, 2022), where models are used to score assets and predict their returns before applying optimization techniques for portfolio allocation. Building on prior research, the degree of diversity in ensemble learning is parametrically controlled during predictor training and encoded in a structured data set of predictions (Dominguez, Shahzad, & Hong, 2025). This data set serves as input for an ensemble model designed to predict portfolio returns. The ensemble model is optimized as a supervised prediction problem, using as a target a portfolio whose weights are chosen to match the ensemble combiner rule derived from the Bias-Variance-Diversity decomposition (Wood et al., 2024). In the case of squared loss, the corresponding combiner is the arithmetic mean, implying that the equal-weighted portfolio should be selected. The optimal portfolio weights are then obtained by normalizing the optimal ensemble parameters. To the best of the authors' knowledge, this is the first method which connects ensemble learning diversity with portfolio risk diversification, enabling its parametric control before the optimization stage.

Additionally, this work also investigates the hypothesis that out-of-sample portfolio diversification relates to the asset or hypothesis selection. Diversity in asset selection is controlled parametrically, independently of the learning phase. A parameter regulates the diversity of asset sets-randomly selected from ranked out-of-sample return predictions-across various portfolio sizes and parameters. Results show that greater diversity in asset sets increases out-of-sample portfolio diversification and risk-adjusted returns, even when the sets have lower average returns. This work links both sources of diversity (i.e., asset selection and portfolio weight optimization) to out-of-sample portfolio diversification and demonstrates that the diversification levels from both sources can be parametrically controlled prior to decision-making, offering valuable insights for practitioners.

Following are the main contributions proposed in this work:

- A consistent framework is introduced for applying structured ensemble models in multi-hypothesis prediction settings for portfolio optimization, where the ensemble combiner rule defines the portfolio target. Specifically, in the case of squared loss, this corresponds to the equal-weighted portfolio.
- It presents a "plug-in" and robust method enabling parametric a priori control over out-of-sample portfolio risk diversification by linking ensemble diversity, derived from the Bias-Variance-Diversity loss decomposition (Wood et al., 2024), to portfolio risk diversification.
- A strategy is proposed for parametrically controlling diversity in asset or hypothesis selection, leveraging a diversity-quality trade-off in predicted returns to enhance out-of-sample portfolio diversification.
- The expansion of diversification limits (Dalio, 2018) is validated by the proposed framework, as well as through the incorporation of diversity in the asset selection stage across other methodologies, demonstrated empirically in both one-step and multi-step decision settings.

The paper is organized as follows: Section 2 provides a review of previous work on data-driven and robust optimization, Predict-then-Optimize frameworks, structured ensemble learning, and diversity. Section 3 describes the proposed framework and methodology. Section 4 presents the experimental results and discussions. Finally, in Section 5 the final remarks and future outlook are offered.

2. Literature review

Portfolio optimization solutions can be broadly categorized into data-driven approaches and robust optimization methods. Data-driven methods typically rely on plug-in optimization, where the objective function is predefined by frameworks such as Modern Portfolio Theory (MPT), and data is used solely to estimate model parameters. In contrast, robust optimization modifies the objective function and relaxes certain framework assumptions to improve adaptability while tackling uncertainty. This approach was among the first to incorporate loss functions, which were later extended to predictive models and predict-then-optimize frameworks, albeit with computational challenges. Ensemble-based methods further introduce loss functions which account for diversity in predictions. This work integrates ensemble models with multiple hypothesis prediction, establishing a connection between diversity in learning and risk diversification, while also introducing computationally efficient analytical solutions.

2.1. Data-driven and robust optimization

Reducing uncertainty in the optimization stage is critical in many applications, and two prominent approaches are data-driven optimization and robust optimization. Data-driven optimization relies on data to estimate uncertainties, using statistical models or machine learning to incorporate these estimates into decision-making (Bertsimas, Gupta, & Kallus, 2013; Shapiro, Dentcheva, & Ruszczyński, 2009). It aims to optimize expected performance metrics such as cost or profit by leveraging historical or real-time data. On the other hand, robust optimization focuses on worst-case scenarios, ensuring decisions remain feasible and effective for all outcomes within a predefined uncertainty set, such as polyhedral or ellipsoidal bounds (Ben-Tal, Ghaoui, & Nemirovski, 2009). This approach prioritizes stability and reliability, often at the expense of conservatism if the uncertainty set is poorly calibrated (Bertsimas, Brown, & Caramanis, 2011). Recent advances combine the strengths of both approaches, such as distributionally robust optimization (DRO), which optimizes for the worst-case distribution within plausible distributions, bridging probabilistic and robust methodologies (Esfahani & Kuhn, 2015; Rahimian & Mehrotra, 2022). In the context of portfolio optimization, data-driven approaches leverage historical and real-time market data to model uncertainties, employing methods such as stochastic programming, machine learning, and empirical scenario generation to predict asset returns and covariance structures (Giglio et al., 2022; Ma et al., 2021; Ozelim, Ribeiro, Schiavon, Domingues, & Queiroz, 2023; Roncalli, 2013). Robust portfolio optimization solutions include fuzzy approaches (Wu & Liu, 2012), genetic algorithms (Akopov, 2014), multiobjective particle swarm optimization (Chen & Zhou, 2018), and learning-guided multiobjective evolutionary algorithms (Lwin, Qu, & Kendall, 2014). Building on these, this work presents a mixed method which combines both approaches, featuring a plug-in facet but robust in that the risk diversification level can be adjusted parametrically before optimization. Therefore, this addresses the disconnection between out-of-sample diversification and input data in the optimization process.

2.2. Predictive modeling integrated with decision optimization

Predictive models have become essential tools for optimizing decision-making processes, particularly when the objective is to minimize decision-related costs. However, this approach overlooks cost considerations during the predictive modeling phase, leading to suboptimal decision quality (Donti & Kolter, 2017). Recent advancements have sought to address this issue through integrated frameworks like "Predict-and-optimize" (PO), Smart Predict-then-Optimize (SPO) and End-to-End Predict-Then-Optimize (EPO) (Elmachtoub & Grigas, 2017, 2022; Vanderschueren et al., 2022). These frameworks use surrogate loss functions to manage computational challenges associated with original loss functions and link structured prediction outputs to decision variables (Elmachtoub & Grigas, 2017). The SPO+ framework further advances this integration by defining loss functions based on the objective costs of nominal optimization problems, improving computational efficiency and decision-aware modeling (Elmachtoub & Grigas, 2022). However, most existing studies in structured prediction focus on classification problems (Goh & Jaillet, 2016) and are computationally intensive, as they require backpropagation through optimization solutions, which often involves handcrafted rules, and they face challenges in handling non-convex or discrete optimization models (Kotary et al., 2023).

2.3. Diversity in multi-hypothesis learning ensembles for portfolio diversification

A unified theory of ensemble diversity explains that diversity is a hidden dimension in the bias-variance decomposition of the ensemble loss (Wood et al., 2024). Moreover, studies on diversity measures in ensemble learning show their significant impact on model accuracy, generalization, and robustness (Ortega, Cabañas, & Masegosa, 2022).

Researchers link portfolio estimation to statistical decision theory by treating the difference in investor utility between the "true" optimal portfolio and the plug-in portfolio as an economic loss function (Kan & Zhou, 2007). Outcomes are enhanced by incorporating the investor's utility objective directly into the statistical weight estimation process as in Bayesian decision theoretic approaches, rather than addressing estimation and utility maximization as separate issues (Avramov & Zhou, 2010; Kan & Zhou, 2007). Research has demonstrated the economic implications of using Mean Square Error (MSE) in portfolio optimization and financial forecasting, showing that integrating MSE into decision-making processes can enhance portfolio performance by minimizing estimation errors and optimizing risk-return trade-offs (Cai, Cui, Lassance, & Simaan, 2024; Lasse Heje Pedersen & Levine, 2021; Mahadi, Ballal, Moinuddin, & Al-Saggaf, 2022). This work extends this idea by incorporating diversity into the MSE loss function within an ensemble learning framework. Furthermore, for the first time, it establishes a connection between this notion of diversity and portfolio diversification, proposing parametric methods to manage diversification prior to decision-making.

3. Framework description

3.1. Preliminary

Given $\boldsymbol{x} \in \mathbb{R}^{N\times M}$ as a vector of returns with $M \geq 2$ assets (each with N time-stamps) having mean $\boldsymbol{\mu}$ and a positive-definite covariance matrix $\boldsymbol{\Sigma}$, then the optimal portfolio $\boldsymbol{w} \in \mathbb{R}^M$ can be found by solving the following optimization problem (Markowitz, 1952):

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \left(\boldsymbol{w}^\top \boldsymbol{\mu} - r\right)^2 + \boldsymbol{w}^\top \Sigma \boldsymbol{w} \\ & \boldsymbol{w}^\top \boldsymbol{\mu} \geq r \\ & \boldsymbol{w}^\top \mathbf{1} = 1 \end{aligned} \tag{1}$$

where $\left(\boldsymbol{w}^\top \boldsymbol{\mu} - r\right)^2$ represents the bias of the expected return of the portfolio $\boldsymbol{w}^\top \boldsymbol{\mu}$ compared to the target return r, and $\boldsymbol{w}^\top \Sigma \boldsymbol{w}$ represents the variance of the portfolio. The constraint $\boldsymbol{w}^\top \boldsymbol{\mu} \geq r$ ensures that the expected return of the portfolio is above the target return. The expression in (1) generalizes the standard MSE reflecting the bias-variance trade-off (Cai et al., 2024).

If the aforementioned portfolio optimization can be formulated as a prediction task, then the objective utility function of the mean-variance in (1) can be represented as an MSE loss function (Cai et al., 2024). This setup allows for finding the optimal weights which minimize the total prediction error of the portfolio, accounting for both bias and variance, while ensuring that the expected return remains above the target value. In order to incorporate diversity into such an optimization, this work applies ensemble models to harness diversity in ensemble learning and link it to portfolio diversification.

In a multiple-hypothesis prediction setup in which each hypothesis is focused on a particular portfolio constituent, the portfolio can be modeled as an ensemble of hypotheses. Following that, the diversity term in the Bias-Variance-Diversity decomposition with the MSE Loss can be connected to diversification and controlled parametrically, while the centroid combiner rule is applied as the portfolio target. To elaborate, in a supervised setting with the data set of time series returns $(\boldsymbol{x}, \boldsymbol{r})$, the Bias-variance-diversity decomposition can be expressed in terms of the MSE loss $\mathcal{L}(\boldsymbol{r}, \bar{\boldsymbol{r}})$ as Wood et al. (2024):

$$\underbrace{\frac{1}{M}\sum_{j=1}^{M}(\mathring{f}_j(\boldsymbol{x}_j) - \boldsymbol{r})^2}_{\text{Average Bias}} + \underbrace{\frac{1}{M}\sum_{j=1}^{M}\mathbb{E}\left[(f_j(\boldsymbol{x}_j) - \mathring{f}_j(\boldsymbol{x}_j))^2\right]}_{\text{Average Variance}} - \underbrace{\mathbb{E}\left[\frac{1}{M}\sum_{j=1}^{M}(f_j(\boldsymbol{x}_j) - \bar{\boldsymbol{r}})^2\right]}_{\text{Diversity}} \tag{2}$$

where $f_j(\boldsymbol{x}_j)$ denote the temporal predictions and $\mathring{f}_j(\boldsymbol{x}_j) = \mathbb{E}[f_j(\boldsymbol{x}_j)]$ represents the centroid prediction of the j-th model in the portfolio ensemble respectively, while $\bar{\boldsymbol{r}} = \frac{1}{M}\sum_{j=1}^{M} f_j(\boldsymbol{x}_j)$ represents the centroid combiner rule.

In the next section, the Portfolio-Structured Ensemble Model (PSEM) is introduced, where each individual hypothesis predicts a specific portfolio constituent. The diversity term in Eq. (2) is linked to both portfolio generalization performance and risk diversification. The predict-then-optimize setting with PSEM for portfolio optimization is then described, in which the target portfolio corresponds to the combiner rule. Under MSE loss, this rule becomes the arithmetic combiner, that is, the equal-weighted portfolio. In this context, the optimal ensemble parameters are equivalent to the optimal portfolio weights.

3.2. Proposed portfolio structured ensemble model (PSEM)

The PSEM is a portfolio model formulated as a structured ensemble for prediction across multiple hypotheses. Formally, it can be defined as a map $\mathcal{E} : \boldsymbol{x} \in \mathbb{R}^{N\times M} \rightarrow \boldsymbol{r} \in \mathbb{R}^N$ which aggregates M predictions from multiple individual predictors or hypotheses $\{f_{\theta_j}(\boldsymbol{x}_j)\}_{j=1}^{M}$, parameterized by $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_j\}_{j=1}^{M}$, to produce portfolio predictions for a given asset selection time series return $\boldsymbol{x}$:

$$\mathcal{E}(\boldsymbol{x}) = g\left(f_{\theta_1}(\boldsymbol{x}_1), f_{\theta_2}(\boldsymbol{x}_2), \ldots, f_{\theta_M}(\boldsymbol{x}_M)\right) = \hat{\boldsymbol{r}} \tag{3}$$

where $g(\cdot)$ denotes the centroid combiner rule (e.g., averaging, voting, or weighted combination) for the ensemble (Wood et al., 2024). A diversity parameter $\varepsilon \in [0, 1]$ modulates each predictor's update at step i, such that the update rule for the j-th predictor is given by:

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_j - \eta_j\left(\frac{\partial \mathcal{L}(f_{\theta_j}(x_{ij}), y_{ij})}{\partial \boldsymbol{\theta}_j} + \frac{\lambda_p}{N}\boldsymbol{\theta}_j\right)\delta\left(f_{\theta_j}(x_{ij})\right) \tag{4}$$

where $y_{ij} = x_{(i+\tau)j}$ with time lag $\tau > 0$ is the target, η_j is the learning rate, and the loss function $\mathcal{L}$ combines a squared error term and regularization with parameter λ_p. The indicator function $\delta(f_{\theta_j}(x_{ij}))$, which adjusts the degree to which each predictor's parameters are updated, is defined as:

$$\delta(f_{\theta_j}(x_{ij})) = \begin{cases} 1-\varepsilon & \mathcal{L}(f_{\theta_j}(x_{ij}), y_{ij}) < \mathcal{L}(f_{\theta_k}(x_{ik}), y_{ik}) \;\; \forall k \neq j \\ \frac{\varepsilon}{M-1} & \text{otherwise} \end{cases} \tag{5}$$

$\delta(f_{\theta_j}(x_{ij}))$ indicates whether the output of the j-th predictor is the top prediction for the i-th instance. A smaller ε places more emphasis on the predictors which perform the best, while a larger ε encourages updates to less accurate predictors, fostering greater diversity in the ensemble. Algorithm 1 provides the pseudo-code for the structured data set formation with stochastic gradient descent.

After training, the predictions of all predictors form a structured data set:

$$\hat{\boldsymbol{x}}(\varepsilon) = \begin{bmatrix} f_{\theta_1}(x_{11}) & \cdots & f_{\theta_1}(x_{N1}) \\ \vdots & \ddots & \vdots \\ f_{\theta_M}(x_{1M}) & \cdots & f_{\theta_M}(x_{NM}) \end{bmatrix} \tag{6}$$

Algorithm 1 Structured data set Formation with Gradient Descent.

Require: Individual predictors $\{f_{\theta_j}(\boldsymbol{x}_j)\}_{j=1}^{M}$ with parameters $\{\boldsymbol{\theta}_j\}_{j=1}^{M}$, input returns $\{x_{i,j}\}^{N,M}$, targets $\{y_{i,j}\}^{N,M}$ with $y_{i,j} = x_{(i+\tau),j}$, learning rates η_j, diversity parameter ε, regularization parameter λ_p, loss function $\mathcal{L}$.

Ensure: Structured data set $\hat{\boldsymbol{x}}(\varepsilon)$, optimal parameters $\hat{\boldsymbol{\Theta}}$.

Randomly initialize $\boldsymbol{\theta}_j$ for all $j = 1$ to M
for $i = 1$ to N **do**
  for $j = 1$ to M **do**
    if $M = 1$ **then**
      $\delta(f_{\theta_j}(x_{i,j})) \leftarrow 1$
    else
      for $k = 1$ to M **do**
        if $j \neq k$ **then**
          if $\mathcal{L}(f_{\theta_j}(x_{i,j}), y_{i,j}) < \mathcal{L}(f_{\theta_k}(x_{i,k}), y_{i,k})$ **then**
            $\delta(f_{\theta_j}(x_{i,j})) \leftarrow 1 - \varepsilon$
          else
            $\delta(f_{\theta_j}(x_{i,j})) \leftarrow \frac{\varepsilon}{M-1}$
          end if
        end if
      end for
    end if

$$\boldsymbol{\theta}_j \leftarrow \boldsymbol{\theta}_j - \eta_j \left(\frac{\partial \mathcal{L}(f_{\theta_j}(x_{i,j}), y_{i,j})}{\partial \boldsymbol{\theta}_j} + \frac{\lambda_p}{N} \boldsymbol{\theta}_j \right) \delta(f_{\theta_j}(x_{i,j}))$$

    $\hat{x}_{i,j}(\varepsilon) \leftarrow f_{\theta_j}(x_{i,j})$
  end for
end for
return $\hat{\boldsymbol{x}}(\varepsilon)$ and $\hat{\boldsymbol{\Theta}}$

which serves as input to a structured ensemble model. The PSEM predictions are given by:

$$\hat{\boldsymbol{r}} = g(\hat{\boldsymbol{x}}(\varepsilon)^T) = \hat{\boldsymbol{x}}(\varepsilon)^T \boldsymbol{w} \tag{7}$$

with ensemble parameters $\boldsymbol{w}$, which, when properly scaled, are equivalent to portfolio weights. By incorporating the diversity parameter ε, the ensemble achieves a balance between prediction accuracy and diversity, enhancing its generalization capabilities in portfolio forecasting tasks. Interestingly, this diversity aligns with the concept of portfolio diversification, as originally proposed in MPT (Markowitz, 1952).

3.3. Predict-then-optimize method with PSEM

In the predict-then-optimize setting with PSEM, the optimal portfolio weights $\boldsymbol{w}$ are obtained by minimizing the loss function $\mathcal{L}(\hat{\boldsymbol{r}}, \bar{\boldsymbol{r}})$, formulated as a supervised learning problem between the PSEM prediction $\hat{\boldsymbol{r}}$, as defined in Section 3.2, and a portfolio target $\bar{\boldsymbol{r}}$ whose weights reflect the ensemble combiner rule in Eq. (2). This combiner rule depends on the choice of loss function, following the bias-variance-diversity decomposition framework proposed by Wood et al. (2024). For the squared loss, the combiner rule reduces to the arithmetic combiner. When the target portfolio time series is defined using the arithmetic combiner, it is equivalent to the equal-weighted portfolio representing the naive diversification strategy (DeMiguel, Garlappi, & Uppal, 2009). The target portfolio returns are given by $\bar{\boldsymbol{r}} = \boldsymbol{x}^T \boldsymbol{w}_{\text{target}} = \boldsymbol{x}^T \frac{1}{\|\boldsymbol{x}^T\|}$, and the optimization is formulated as:

$$\min_{\boldsymbol{w}} \left\| \bar{\boldsymbol{r}} - \hat{\boldsymbol{x}}(\varepsilon)^T \boldsymbol{w} \right\|^2 + \lambda_s \|\boldsymbol{w}\|_p \tag{8}$$

Here, λ_s denotes the PSEM regularization parameter. The predict-then-optimize approach can be implemented in a "plug-in" format, where the structured data set $\hat{\boldsymbol{x}}(\varepsilon)^T$ is constructed independently prior to training the structured ensemble and deriving the portfolio weights. In the experimental section, portfolio weight constraints are applied post-optimization, including non-negativity constraints, $w_i \geq 0, \ \forall i \in \{1, \ldots, M\}$, and normalization, $\sum_{i=1}^{M} w_i = 1$. However, the method can also be extended to a fully constrained plug-in optimization framework using quadratic solvers, which is left for future work.

Although the ensemble prediction $\bar{\boldsymbol{r}} = \frac{1}{M} \sum_{j=1}^{M} \hat{\boldsymbol{x}}_j(\varepsilon)$ is computed as an equally weighted average of base learner outputs, the optimal ensemble weights $\boldsymbol{w}$ derived from Eq. (8) do not generally coincide with uniform or equal-weighting. This discrepancy arises due to induced diversity via the parameter ε, variation in base learner initialization and regularization, and inherent dependencies among predictors. These factors collectively break the symmetry of the hypothesis space, resulting in non-uniform projections of the ensemble mean onto the span of predictions. As a result, the final ensemble weights adapt to capture these structural and informational asymmetries, ultimately improving loss alignment while diverging from naive equal weighting.

Fig. 1 shows a representation of the PSEM optimization, illustrated as a shooting gallery where the shooters represent individual predictors. Each shooter receives inputs such as wind conditions, distance, and target position, aiming to hit their respective targets. In each iteration, the trainer penalizes the best shooter by limiting how much their parameters update during gradient descent, while the remaining shooters update equally. This process introduces diversity in learning through ε. Once the training of the individual predictors is completed, all shots are gathered into a diversified data set, which can be used as input for structured prediction models, like the Structured Radial Basis Function Network (s-RBFN) (Dominguez et al., 2025), shown in this figure and trained using least-squares.

In the next section, the PSEM optimization is detailed for the case of the s-RBFN model.

3.4. Example with structured radial basis function network (s-RBFN)

In this case, the structured data set is used as input to a radial basis function network, where each j-th predictor $f_{\theta_j}(\boldsymbol{x}_j)$ is associated with a specific basis function $\phi_j(\hat{x}_{ij}(\varepsilon), \mu_j, \sigma_j)$. The basis function ϕ_j is applied element-wise to the entries of the j-th input vector of predictions $\hat{\boldsymbol{x}}_j(\varepsilon) \in \mathbb{R}^N$, with μ_j and σ_j representing the center and scale parameters specific to the j-th predictor. This defines a transformation map $\boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon)) : \mathbb{R}^{N \times M} \rightarrow \mathbb{R}^{N \times M}$, where $\boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon))_{ij} = \phi_j(\hat{x}_{ij}(\varepsilon), \mu_j, \sigma_j)$. For example, the Gaussian basis function $\phi_j(\cdot) = \exp\left(\frac{-1}{2\sigma_j^2} \left| \hat{x}_{ij}(\varepsilon) - \mu_j \right|^2 \right)$ can be used where the parameters of the centers μ_j and the scales σ_j for the basis functions are calculated from each column of $\hat{\boldsymbol{x}}(\varepsilon)$ as the mean and standard deviation, respectively. The s-RBFN formulation can now be expressed in matrix form as follows:

$$\boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon))^T \boldsymbol{w} = \begin{bmatrix} \phi(\hat{x}_{11}(\varepsilon), \mu_1, \sigma_1) & \cdots & \phi(\hat{x}_{1M}(\varepsilon), \mu_M, \sigma_M) \\ \vdots & \ddots & \vdots \\ \phi(\hat{x}_{N1}(\varepsilon), \mu_1, \sigma_1) & \cdots & \phi(\hat{x}_{NM}(\varepsilon), \mu_M, \sigma_M) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_M \end{bmatrix} \tag{9}$$

The optimal s-RBFN parameters are obtained by least-squares with the regularization parameter λ_s and target portfolio, the arithmetic combiner or equal-weighted portfolio, $\bar{\boldsymbol{r}}$:

$$\boldsymbol{w} = \left(\boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon))^{\mathrm{T}} \boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon)) + \lambda_s * \boldsymbol{I}_{(mxm)} \right)^{-1} \boldsymbol{\Phi}(\hat{\boldsymbol{x}}(\varepsilon))^{\mathrm{T}} \bar{\boldsymbol{r}} \tag{10}$$

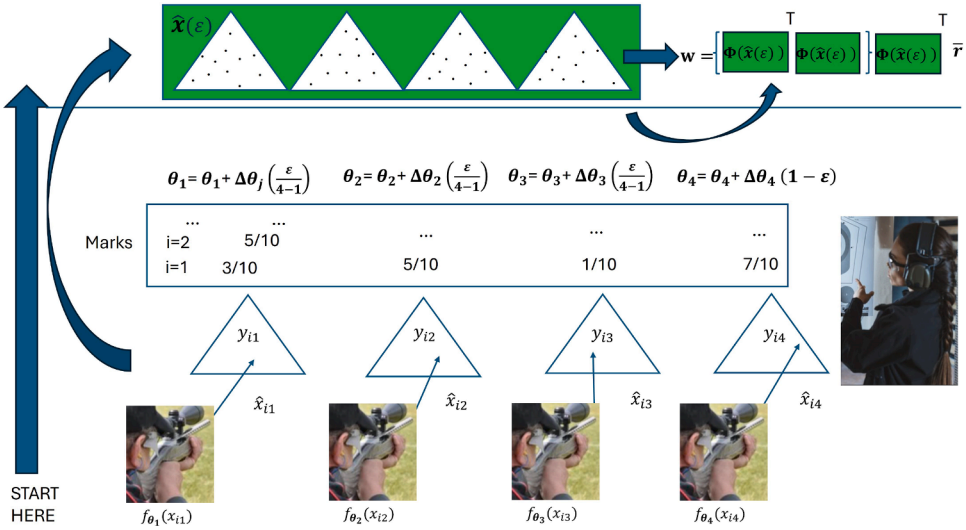


Fig. 1. Structured data set formation, including diversity (ε), and analytical ensemble training of the s-RBFN model, described in Section 3.4. $f_{\theta_j}(x_{ij})$ is the j^{th} shooter (predictor) with input data x_{ij} (for example wind, distance, etc.) and $\boldsymbol{\theta}_j$ her shooting skills, $\hat{x}_{ij}$ is the i^{th} shoot from j^{th} shooter and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\varepsilon)$ is the full set of shoots after training, $\bar{\boldsymbol{r}}$ is the portfolio target for the training set (equal-weighted portfolio), and $\boldsymbol{w}$ are the s-RBFN model parameters optimized by least-squares.

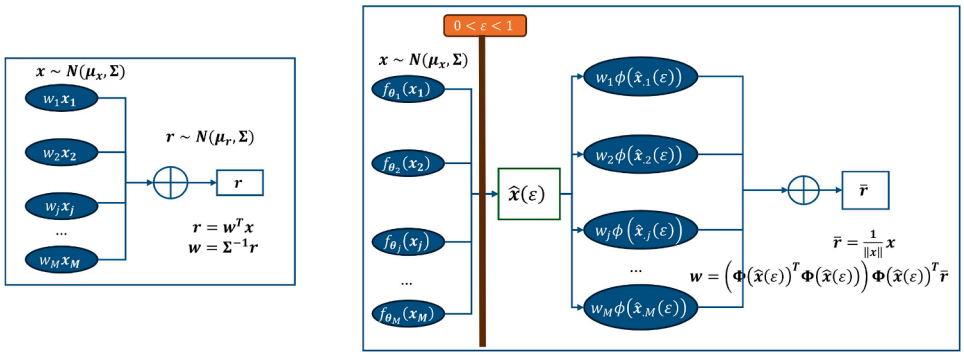


Fig. 2. Systems representation comparison between the Mean-Variance framework for portfolio optimization from Modern Portfolio Theory (Markowitz, 1952)(left blue box) and the predict-then-optimize PSEM framework (right blue box). In the Mean-Variance case, the portfolio constituents' returns, $\boldsymbol{x}$, and the target portfolio returns, $\boldsymbol{r}$, are assumed to be normally distributed. In contrast, in the PSEM framework, the target is the equal-weighted portfolio $\bar{\boldsymbol{r}}$. Additional parameters and variables include the portfolio weights $\boldsymbol{w}$, the diversity parameter $0 < \varepsilon < 1$ used for forming the structured data set $\hat{\boldsymbol{x}}(\varepsilon)$, the individual predictors $\{f_{\theta_j}(\boldsymbol{x}_j)\}_{j=1}^{M}$, and the s-RBFN model basis functions represented in matrix form by $\boldsymbol{\Phi}$.

The optimal portfolio weights are equivalent to the s-RBFN optimal parameters $\boldsymbol{w}$, after applying weight normalization and other relevant constraints.

In Fig. 2, a comparative analysis between two portfolio optimization frameworks is presented: the classical mean-variance approach from Modern Portfolio Theory (MPT) (Markowitz, 1952), shown in the blue box on the left, and the s-RBFN model trained via the least-squares algorithm (Dominguez et al., 2025), as the PSEM model, shown in the blue box on the right. Despite their structural differences, both methodologies exhibit notable conceptual and computational parallels.

In the mean-variance framework, the key plug-in component is the covariance matrix, which is based exclusively on the input data set $\boldsymbol{x}$ and does not provide a guarantee of achieving portfolio diversification outside the sample. In contrast, the PSEM framework incorporates portfolio diversification directly into the modeling process by treating assets as hypotheses and the portfolio as an ensemble predictor, leveraging the diversity among individual predictors during training. The structured data set $\hat{\boldsymbol{x}}(\varepsilon)$, constructed from predictions generated by models using the same input data as the mean-variance model, enables the use of a continuous diversification parameter $0 < \varepsilon < 1$. This allows for the creation of multiple diversified plug-in data sets from the same input and for portfolio diversification to be adjusted prior to weight optimization.

Both approaches allow closed-form solutions via the least-squares algorithm, rendering them computationally equivalent in terms of efficiency. In the PSEM framework, the input data $\boldsymbol{x}$ is replaced with the diversified structured data set of predictions $\hat{\boldsymbol{x}}(\varepsilon)$, which is then augmented using basis functions within the s-RBFN model. These basis functions-such as radial, Gaussian, or other types-introduce additional flexibility to the s-RBFN model, enabling it to capture more complex non-linear relationships within the data.

The focus of this work is on portfolio allocation, which encompasses both asset selection and portfolio optimization. The latter has been addressed in the previous sections. The former is introduced in the next section through a novel method which leverages the diversity of asset return predictions as an asset selection tool. This approach seeks to uncover connections between diversity in asset selection and out-of-sample portfolio diversification, identify new limits on

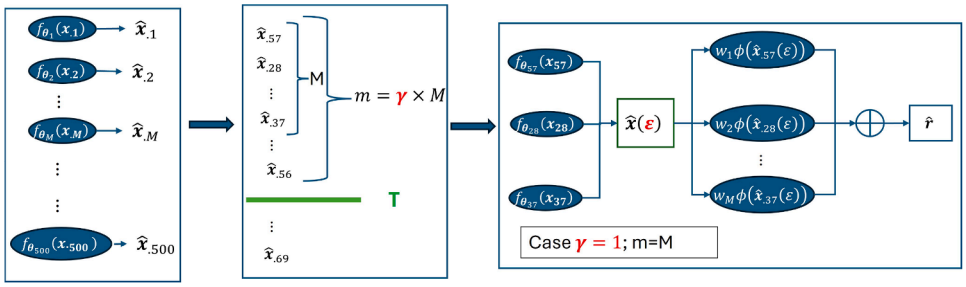


Fig. 3. Assets or hypotheses selection including forecasting (see left block). A 1-month cumulative return ranking with threshold T, random selection of M constituents from a sample of m candidates ($m = \gamma \times M \geq M$) and γ the diversity parameter in the hypothesis selection stage (see middle block). Model architecture for s-RBFN for $m = M$ and $\gamma = 1$, and ε the diversity parameter for the learning stage (see right block).

portfolio diversification introduced by this additional source of variability, and evaluate the ability to control these effects parametrically prior to weight optimization.

3.5. Parametric control of out-of-sample diversification at the asset or hypothesis selection stage

The portfolio allocation is structured into two stages (Markowitz, 1952): asset selection and portfolio optimization. In the asset selection, the cumulative 1-month forecasts for the m members of a hypothetical market are computed and ranked based on their predicted cumulative returns. A threshold T is applied to the ranking, and the top-performing assets exceeding this threshold are selected based on rational investor behavior, prioritizing those with the highest expected or predicted returns (Markowitz, 1952). To incorporate diversity into the asset selection process, a parameter γ is introduced. This parameter acts as a multiplier of the number of portfolio constituents M, generating a sample of size $m = \gamma \times M$ consisting of assets with predictions exceeding the threshold T, from which a set of M portfolio constituents is randomly selected.

This can be seen in Fig. 3, where the forecast is first computed (see left block). Then, depending on whether $\gamma = 1$, no diversity is added; if $\gamma > 1$, diversity is introduced into the asset selection, provided the sample size is such that all predictions in the sample exceed T. From this sample, M portfolio constituents are randomly selected, with the parameter γ controlling the degree of diversity in asset or hypothesis selection (see central block). The right block refers to the structured prediction model with multiple hypotheses, where the hypotheses have already been selected. At this stage, the parameter ε regulates the diversity in the learning process of individual predictors and the ensemble. Once the prediction data set with the selected diversity is obtained, the portfolio weights are optimized as described in Section 3.4 (see right block in Fig. 3).

This setup allows the testing of the hypothesis that there is a trade-off between the diversity of expected or predicted future returns and their average value as a group. A more diverse group of predictions is preferred, even if it includes lower returns which reduce the group average, highlighting a balance between diversity and the quality of return predictions for asset or hypothesis selection. This trade-off is verified by the experiments.

In conclusion, modeling the portfolio allocation problem as a structured prediction task over multiple hypotheses allows the separation of the process into two key components: hypotheses selection and weight optimization. This approach provides parametric methodologies to induce diversity in hypothesis selection through the parameter γ and to regulate diversity in the learning process of individual predictors based on the selected hypotheses through the parameter ε. These two sources of diversity are directly connected to out-of-sample portfolio risk diversification and can be controlled and adjusted prior to the optimization and decision-making process.

4. Numerical results

4.1. Data and experimental setup

The time series of daily prices for the S&P 500 index members (S&P, 2005) (excluding weekends) was obtained from Bloomberg (S&P, 2024). Returns were calculated as the percentage change in daily prices.

To empirically demonstrate the connection between diversity in the structured prediction setting with multiple hypotheses and out-of-sample portfolio diversification, numerous out-of-sample Sharpe ratios are computed using the predictive setting. The Sharpe ratio is the quotient between return and volatility (risk) for a portfolio of M stocks over a time period Δt. The Sharpe ratio is computed for different values of portfolio size M and diversity parameter ε, both for the case without diversity in asset selection and for the scenario where the portfolio allocation decision is made once, performing 100 simulations for each trial, as in Modern Portfolio Theory (MPT) (Markowitz, 1952). This allows comparison with traditional methods and reveals the pattern of portfolio diversification behavior in relation to the parameters M and ε.

Next, a more realistic case not considered in MPT is included, where the problem becomes sequential over 10-months, making it more practical and useful for practitioners (in this case, the Sharpe ratio over the 10-month period is reported). For these cases, the asset selection is performed by ranking 1-month cumulative return predictions and choosing the top M assets based on different threshold values T. In the one-step decision-making cases, the same assets are usually selected across simulations, whereas in the sequential cases, the non-stationarity in the data may alter prediction ranking and asset selection. Finally, a second source of diversity is introduced through asset or hypothesis selection, controlled by the parameter γ. The experiments are performed for both sources of diversity in one-step as well as in the multi-step decision-making cases.

The structured ensemble models used are the s-RBFN optimized by least-squares and two-layer networks as individual predictors, with the same set of hyperparameter configurations for all cases as described in Dominguez et al. (2025). The equal-weighted model and the MSE-weighted model are also included, where the weights are either equally distributed across all constituents or inversely proportional to their generalization prediction error. For asset selection, the ranking is constructed by calculating the 1-month recurrent prediction of the daily time series and then ranking them based on the accumulated monthly return. The prediction is performed using the same model as the individual predictors with identical hyperparameters, but each of the 500 index members has its own network.

To evaluate the generalization of the proposed framework across different market regimes and asset classes, additional experiments are presented in Appendix D. These include 1336 global bond time series from 2014 to 2018, with the distribution by sector, seniority, rating, tenor (curve), and country (ISO) shown in Fig. D.15, as well as all S&P 500

Table 1
Sets of values for the individual predictors and s-RBFN hyperparameters, (Dominguez et al., 2025).

(a) M number of hypotheses, κ number of neurons per layer, η learning rates for the predictors, χ is a multiplicative factor for random initial predictors' parameters Θ.

M	κ	η	χ
[2, 5, 10, 20, 35]	[20, 200, 2000]	[0.03, 0.3]	[0.0001, 0.01, 0.1, 1]

(b) ε is the diversity parameter in the optimization and γ in the asset selection, λ_p is the regularization parameter for the predictors and λ_s for the s-RBFN.

ε	λ_p	λ_s	γ
[0, 0.1, 0.35, 0.5]	[0, 0.0001, 0.01, 0.07]	[0, 3, 5]	[1, 2, 3, 5]

constituents during the Global Financial Crisis (2007–2009) and the U.S. equity rally (2009–2016).

4.2. Model evaluation and performance metrics

One of the most commonly used portfolio performance metrics is the Sharpe Ratio (Sharpe, 1994), defined as the ratio of the average daily return over a period to its standard deviation. In this work, return predictions are evaluated using 1-month cumulative returns. Accordingly, a slight variation of the Sharpe Ratio is considered, where the cumulative return over one month is divided by the standard deviation of daily returns during that same period. This modified Sharpe Ratio is used to assess out-of-sample portfolio performance in the one-step decision scenario. For the multi-step decision setting, the modified Sharpe Ratio is computed over a 10-month sequential investment horizon.

For the individual predictors, a two-layer multi-layer perceptron (MLP) is used with hyperbolic tangent activation functions. The number of neurons in each layer is denoted by κ, the learning rate by η, and a multiplicative factor for the random initial weights during gradient descent by χ, which affects model performance. Regularization parameters are represented by λ_p. These individual predictors are applied during the asset selection stage, where RMSE is used as the metric for time-series forecasting performance. Assets are selected from a pool of diversified assets based on their predicted values, with diversity scaled by the parameter γ.

The s-RBFN is employed as the PSEM model in the experiments, where the number of predictors or hypotheses is denoted by M, and the regularization parameter is represented by λ_s. Once the optimal portfolio weights are obtained-by normalizing the ensemble parameters, they are used to evaluate out-of-sample portfolio performance using the modified Sharpe Ratio: 1-month for the one-step decision case and 10-month for the multi-step decision case. Portfolio performance serves as the criterion for selecting the optimal network configuration for both the individual predictors and the s-RBFN. Multiple hyperparameter combinations are tested in the experiments, and all hyperparameter values used are listed in Table 1.

Additional evaluation metrics-Maximum Drawdown, Sortino Ratio, and Omega Ratio, alongside Sharpe Ratio-are reported in Appendix D. These are applied to both the regime-based analyses and the global bond data set, supporting the sensitivity study in Section D.1 and the comparative performance analysis in Section D.2.

4.3. Parametric portfolio risk diversification

Experiments validate the hypothesis that out-of-sample portfolio risk diversification can be controlled prior to the decision making process if the portfolio is modeled as a structured ensemble in a multiple-hypothesis prediction setting.

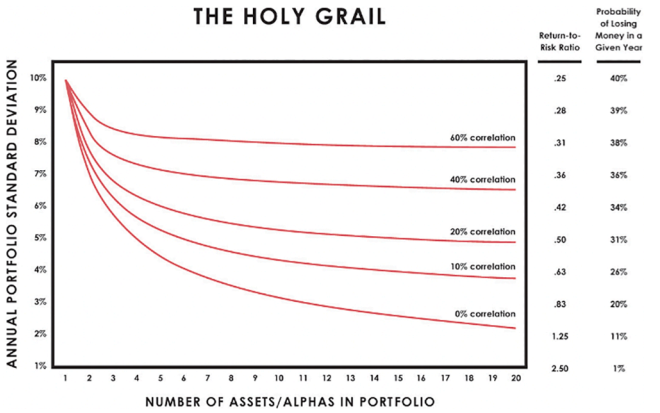


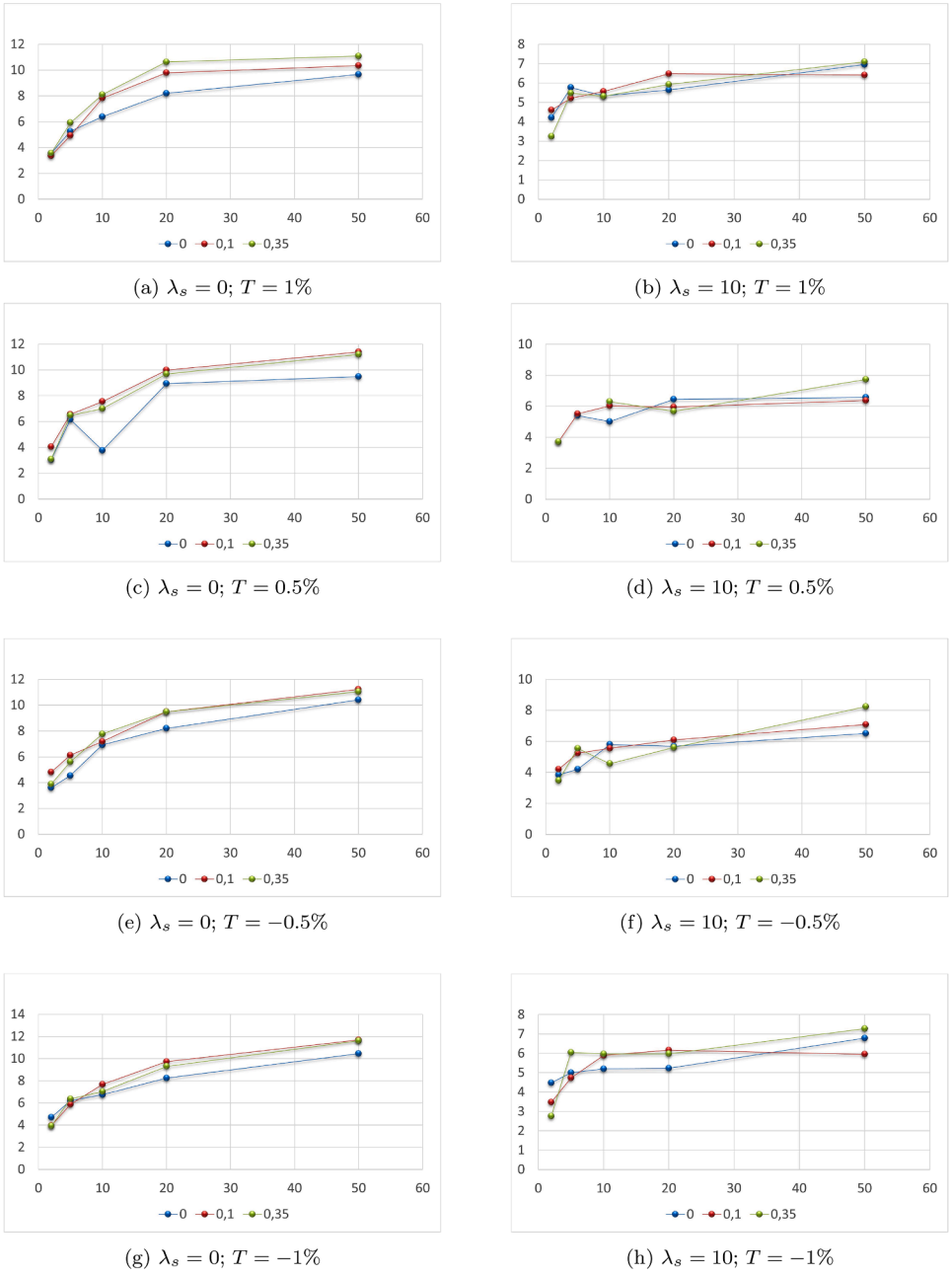
Fig. 4. The limits of diversification, (Dalio, 2018). Sharpe and Return-to-Risk ratios are the same (inverted axis). The standard deviation is reduced up to a diversification limit.

4.3.1. Diversity in the ensemble learning stage

In this section, asset selection is performed for each portfolio allocation as described previously, taking portfolio constituents as the top M of the ranking based on the 1-month accumulated return that are also higher than the threshold T. Two types of experiments are carried out: one-step decision-making (1-month and 100 simulations per trial) and sequential multi-step decision-making (10-months and 1 simulation per trial). The structured prediction model s-RBFN with radial and Gaussian basis functions is implemented, both with and without regularization. The experiments are repeated for different threshold values $T = -1\,\%, -0.5\,\%, 0.0\,\%, 0.5\,\%, 1\,\%$ and diversity parameter $\varepsilon = 0, 0.1, 0.35$.

The result for the one-step decision-making case can be seen in Fig. 5, showing a strong similarity with the results from the MPT setup (Markowitz, 1952), which is illustrated in Fig. 4. Fig. 4 shows the diversification limits according to MPT, where the Sharpe ratios are presented with an inverted axis compared to Fig. 5, with the same horizontal axes referring to the number of assets or stocks M. In the case of MPT, the Sharpe ratios are displayed for different values of the correlation between assets, and it is observed that the negative growth of the correlation has the same effect as the positive growth of the diversity parameter in ensemble learning in the multi-hypothesis prediction setting, as shown in Fig. 5. Thus, portfolio diversification, which depends on correlation (the lower, the better) and the number of portfolio constituents, is to some extent equivalent to diversity in the learning of structured ensembles in a setting like the one discussed (regulated by ε) and the number of predictors. The experiments empirically validate the equivalence between out-of-sample portfolio diversification and diversity in the learning of a structured ensemble in a multi-hypothesis prediction setting when used to model a portfolio, according to Section 3.2.

In Fig. 6, the same set of experiments is carried out, but for the sequential multi-step decision-making case is presented. In this case, the resemblance is different from that shown in MPT and Fig. 4, which makes sense since those setups are one-step, whereas the results here are sequential multi-step cases. Nevertheless, the conclusions remain unchanged, and the same patterns can be observed regarding the relationship between diversity in ensemble learning and out-of-sample portfolio diversification. A consistent increase in Sharpe ratios can be seen as the parameter ε and the number of predictors, hypotheses, or assets M increase. On the other hand, the regularization parameter has a greater impact in this case than in the one-step scenario shown in Fig. 5. It can be observed that the out-of-sample risk-adjusted portfolio performance improves even in the case where the diversity parameter ε is not applied. This may indicate that performance in multi-step settings depends not only on diversification but also on the complexity of the data. The same



(a) $\lambda_s = 0$; $T = 1\%$ (b) $\lambda_s = 10$; $T = 1\%$

(c) $\lambda_s = 0$; $T = 0.5\%$ (d) $\lambda_s = 10$; $T = 0.5\%$

(e) $\lambda_s = 0$; $T = -0.5\%$ (f) $\lambda_s = 10$; $T = -0.5\%$

(g) $\lambda_s = 0$; $T = -1\%$ (h) $\lambda_s = 10$; $T = -1\%$

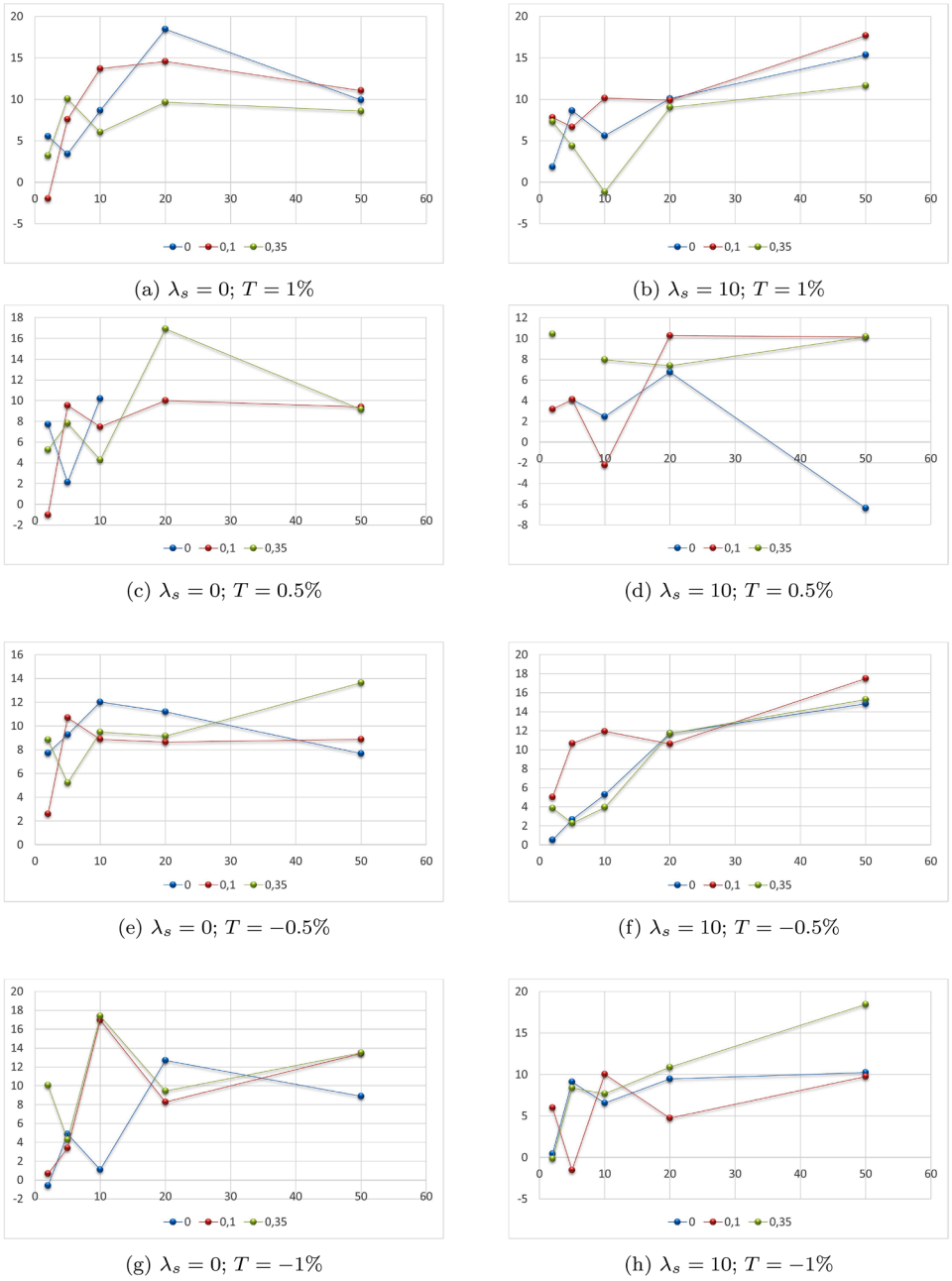
Fig. 5. One-step (1-month) 100 simulations Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), s-RBFN regularization parameter λ_s, and $-1\,\% \leq T \leq 1\,\%$. s-RBFN with Gaussian basis functions.

experiments can be seen for the case of radial basis functions with the same conclusions in the Appendix in Figs. A.10 and A.11.

4.3.2. Diversity-quality trade-off of returns predictions: Including diversity in the asset selection stage

In this section, the same experiments are performed, but here the M portfolio constituents or hypotheses are randomly selected from the top $m = M \times \gamma$ stocks based on the 1-month cumulative forecast ranking. The parameter of asset selection diversity γ is a multiplier which scales the candidate pool relative to the final selection size M. This approach seeks to test the hypothesis that out-of-sample diversification in portfolio allocation is achieved not only during the optimization process-whether through the inclusion of the parameter ϵ in this structured prediction model framework or through other plug-in methods without such aid-but also at the stage of hypothesis selection, when choosing assets or stocks prior to optimizing portfolio weights. For this purpose, the experiments are repeated for different values of the multiplier, including in the no-diversity case, which is equivalent to the results in the previous section, with $\gamma = 1, 2, 3, 5$ for both one-step-ahead and multi-step-ahead cases, using the s-RBFN model with radial basis functions.

Figs. 7 and 8 present the out-of-sample Sharpe ratios for different numbers of stocks M; various values of the asset selection diversity parameter γ; the learning stage diversity parameter ϵ; and threshold values T used in the asset selection process based on forecast rankings, for the multi-step sequential (10-month strategy) case. Figs. C.13 and C.14 in the Appendix show the same results for the one-step decision case (1-month, 100 simulations).



(a) $\lambda_s = 0$; $T = 1\%$ (b) $\lambda_s = 10$; $T = 1\%$

(c) $\lambda_s = 0$; $T = 0.5\%$ (d) $\lambda_s = 10$; $T = 0.5\%$

(e) $\lambda_s = 0$; $T = -0.5\%$ (f) $\lambda_s = 10$; $T = -0.5\%$

(g) $\lambda_s = 0$; $T = -1\%$ (h) $\lambda_s = 10$; $T = -1\%$

Fig. 6. Multi-step (10-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\varepsilon = 0, 0.1, 0.35$ (colours), s-RBFN regularization parameter λ_s, and $-1\,\% \leq T \leq 1\,\%$. s-RBFN with Gaussian basis functions.

Within the proposed PSEM framework, the Sharpe ratios, along with the average Sharpe ratio for all levels of diversity determined by the structured ensemble diversity parameter (ϵ), can also be analyzed to examine the impact of diversity in asset or hypothesis selection on out-of-sample portfolio diversification. These results are presented in Table 2 for the one-step-ahead case (1-month) and Table 3 for the multi-step experiments (10-month strategy), both using the s-RBFN model with radial basis functions.

In the one-step ahead case, a generalized increase in out-of-sample performance, represented by the Sharpe ratio, can be observed when selecting more diverse assets or hypotheses during the stock selection process prior to weight optimization, even if they have worse average predictions. This can be seen in Table 2, where the average Sharpe ratio improves as the values of m grow, driven by the increasing value of γ, allowing for greater diversity of predictors with different out-of-sample performances as part of the stock selection.

This indicates that a higher number of candidates with greater diversity in predictions, and worse average 1-month out-of-sample return predictions (with relatively similar RMSE in their predictions), leads to portfolios with better out-of-sample Sharpe ratios and greater diversification, not due to the learning stage, but due to better asset selection. This pattern holds for both thresholds $T = 0.5\,\%$ and $T = 0.0\,\%$, meaning that the variety of 1-month cumulative return predictions from the sample of m candidates includes positive predictions above 0.5 % and 0.0 %, respectively.

In the case of negative thresholds $T = -0.5\,\%$ and $T = -1\,\%$, it can be observed that the pattern is more pronounced, which surprisingly indicates that, by adding negative return predictions to the M

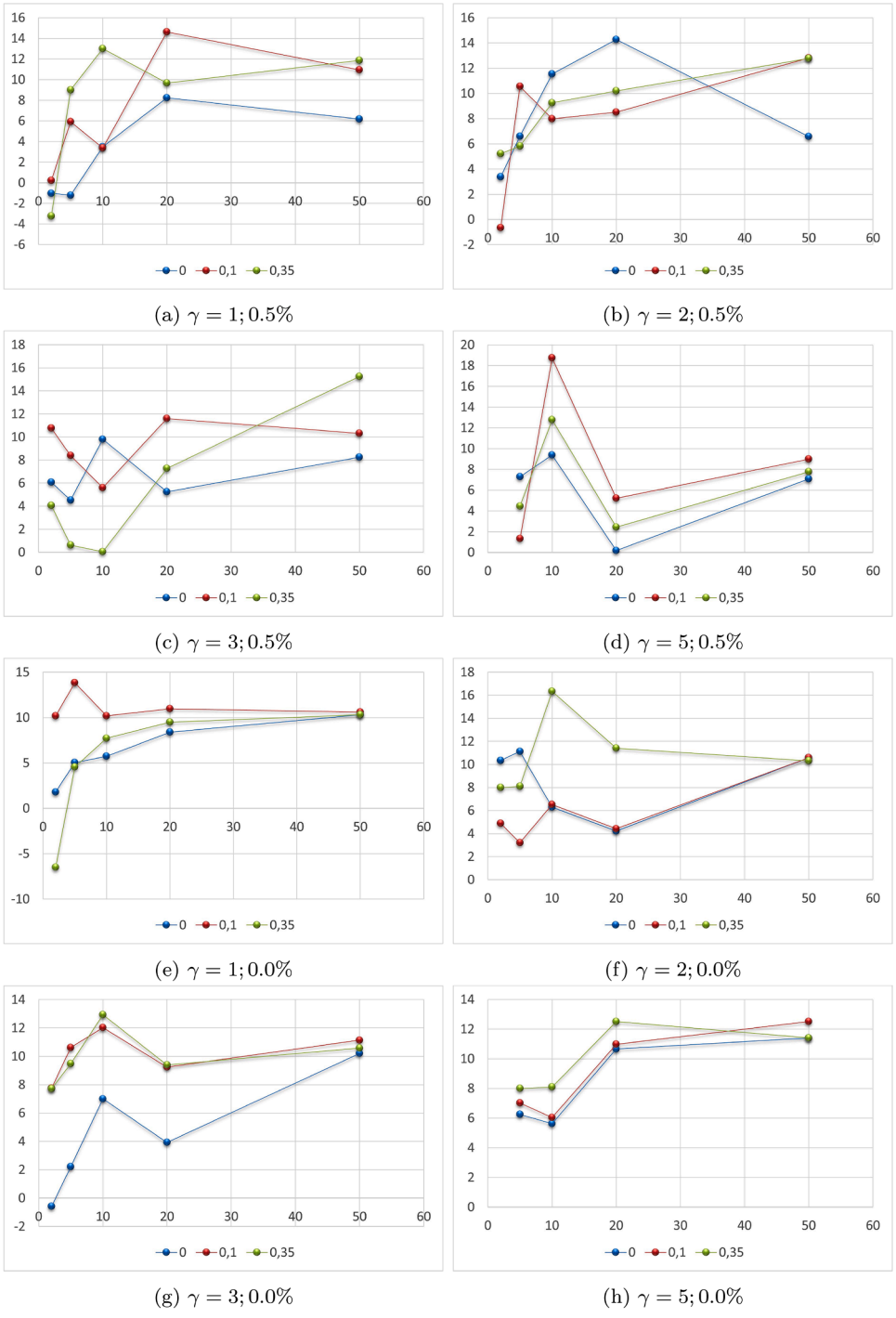


Fig. 7. Multi-step (10-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), multiplicative factor γ, and $0\,\% \leq T \leq 0.5\,\%$. s-RBFN with radial basis functions.

portfolio constituents improve portfolio generalization performance and out-of-sample diversification. This occurs because it adds diversity to the hypothesis selection, even if it downgrades the average performance of the group, compared to other selections with more positive predictions but less diversity. This represents the diversity-quality trade-off of return predictions in asset selection discussed in Section 3.5. It should be highlighted that $\gamma = 2, 3, 5$ and $M = 20, 50$ portfolio cases have constituents selected randomly from samples of $m = 40$ to $m = 250$. As the index consists of 500 stocks, this indicates that the spectrum of negative predictions is not small at all.

Table 2 also shows that, as T decreases, increasing diversity improves the average Sharpe ratio up to a limit, with $T = -0.5\,\%$ yielding better values than $T = -1.0\,\%$ and the other thresholds for the one-step-ahead case (1-month). This is consistent with the limits of diversification, as shown in Fig. 4 (Dalio, 2018).

Table 3 shows similar results, confirming the same hypothesis for sequential multi-step 10-month cases. It is important to note that this scenario is much more restrictive, as it involves making ten consecutive decisions over 10-months instead of just one. One might expect to prioritize the positiveness of predictions (quality) over the diversity of the

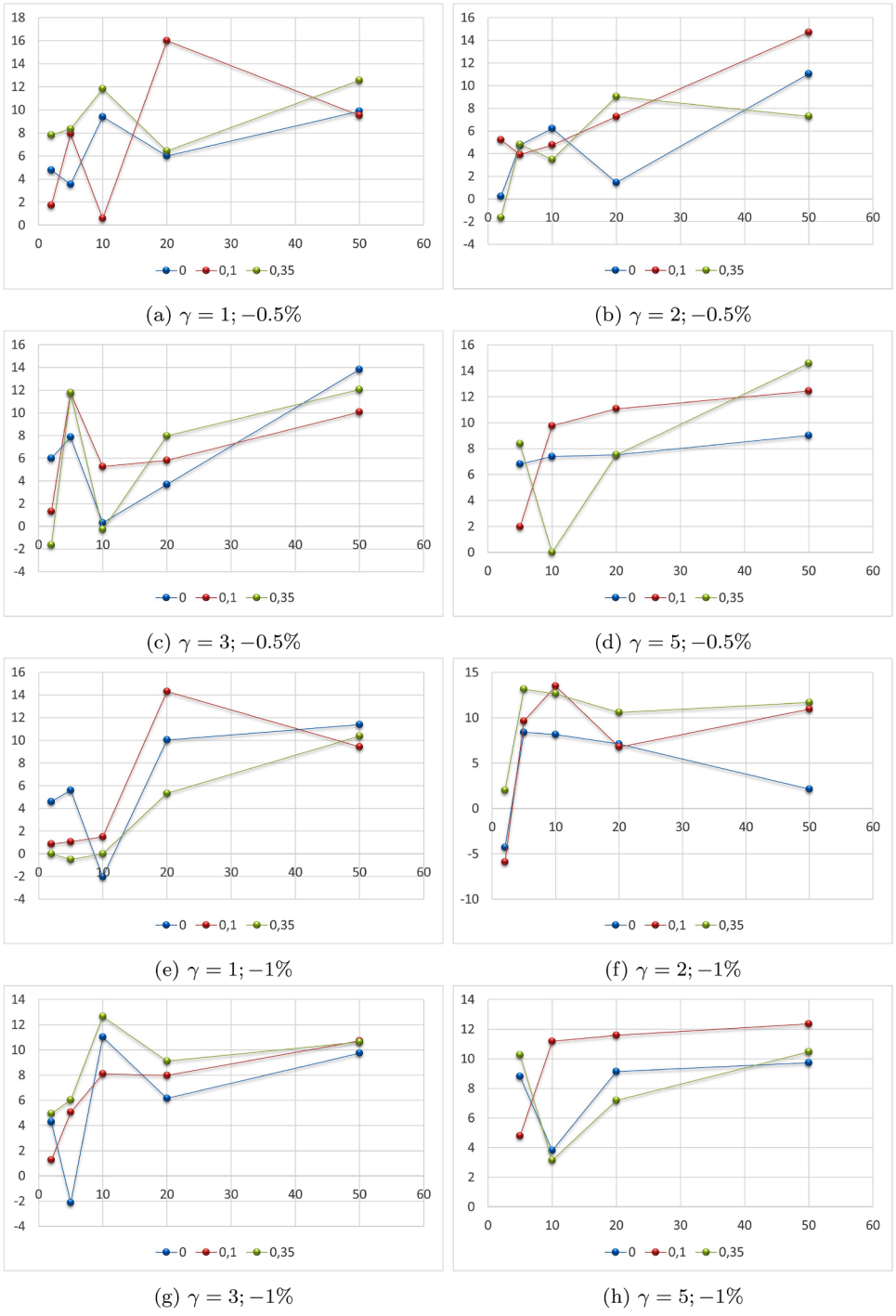


Fig. 8. Multi-step (10-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), multiplicative factor γ, and $-1\,\% \leq T \leq -0.5\,\%$. s-RBFN with radial basis functions.

prediction set during asset selection in such problems, as this would seem the most logical approach. However, the experiments validate the initial hypothesis and confirm the existence of this trade-off, with a significant impact on generalization performance and out-of-sample risk diversification. In Table 3 it can be observed that with a threshold $T = -0.5\,\%$ for ranking predictions versus $T = 0.5\,\%$, the average out-of-sample Sharpe ratios are better for experiments with negative T in most cases. Furthermore, in the case of a positive T, the results are negative for portfolios with fewer assets, indicating the need for a minimum number of stocks in the portfolio (Table 3, case $T = 0.5\,\%$). This issue can be addressed easily without increasing M by adding diversity in asset or hypothesis selection through γ. This is very useful for practitioners, as with constraints on the number of constituents, they can increase the level of diversification of their portfolio with the asset selection diversity component.

Finally, it can be shown that both in the one-step ahead case (1-month), which is comparable to the MPT setting and other plug-in methods, and in the multi-step case (10-month), the diversity parameter in the ensemble ($\epsilon = 0, 0.1, 0.35$), which relates to weight optimization but not asset selection, shows improvement in all tables for ($\gamma = 1$). Without applying diversity in asset selection and only in optimization, the ratios improve with the parameter and the number of

Table 2
One-step-ahead experiments (1-month; 100 simulations): Sharpe ratios for different diversity parameters in the learning stage $\epsilon = 0, 0.1, 0.35$, number of stocks M, and multiplicative factor or diversity parameter in the asset selection stage $\gamma = 1, 2, 3, 5$. T represents the threshold in the ranking of return predictions. Avg. Sh. refers to the average Sharpe ratio of the elements in the corresponding column.

Div. (ϵ)	($T = 0.5\,\%$)					($T = 0.0\,\%$)					($T = -0.5\,\%$)					($T = -1\,\%$)				
	2	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**
Div. ($\gamma = 1$); $m = M \times \gamma = \{2, 5, 10, 20, 50\}$																				
0	3	5	6	8	10	3	4	6	8	10	4	5	7	8	9	3	5	6	8	10
0.1	4	6	6	9	10	4	7	7	8	10	4	5	8	8	11	1	6	7	8	10
0.35	5	6	7	9	10	5	5	6	8	10	3	6	7	8	10	0	6	8	9	10
Avg. Sh.	4	6	6	8	10	4	5	6	8	10	4	5	7	8	10	1	6	7	8	10
Div. ($\gamma = 2$); $m = M \times \gamma = \{4, 10, 20, 40, 100\}$																				
0	4	5	6	8	9	4	6	6	7	10	4	5	6	8	10	4	5	5	7	9
0.1	4	5	8	9	11	4	5	7	9	10	4	5	7	8	10	3	5	8	9	11
0.35	3	5	8	9	11	6	8	9	10	10	4	6	7	8	10	3	6	7	9	11
Avg. Sh.	4	5	7	9	11	5	6	7	8	10	4	5	7	8	10	3	5	7	8	10
Div. ($\gamma = 3$); $m = M \times \gamma = \{6, 15, 30, 60, 150\}$																				
0	4	5	6	7	9	4	6	7	8	9	4	5	6	8	10	5	5	6	8	10
0.1	5	5	7	9	11	6	7	8	11	3	4	6	7	9	1	6	7	8	11	
0.35	4	5	7	9	11	6	6	7	9	10	3	5	7	9	11	3	5	7	9	11
Avg. Sh.	4	5	7	9	10	5	6	8	9	7	3	5	7	9	10	3	5	7	9	10
Div. ($\gamma = 5$); $m = M \times \gamma = \{25, 50, 100, 250\}$																				
0		6	8	7	11		5	6	8	10		5	7	8	10		5	6	8	8
0.1		6	7	3	6		5	7	9	10		6	8	9	11		5	8	9	12
0.35		5	11	9	11		6	8	10	10		6	8	8	11		6	7	9	11
Avg. Sh		5	9	6	9		5	7	9	10		6	8	8	11		5	7	9	10

Table 3
Multi-step-ahead experiments (10-month sequential strategy): Sharpe ratios for different diversity parameters in the learning stage $\epsilon = 0, 0.1, 0.35$, number of stocks M, and multiplicative factor or diversity parameter in the asset selection stage $\gamma = 1, 2, 3, 5$. T represents the threshold in the ranking of return predictions. Avg. Sh. refers to the average Sharpe ratio of the elements in the corresponding column.

Div. (ϵ)	($T = 0.5\,\%$)					($T = 0.0\,\%$)					($T = -0.5\,\%$)					($T = -1\,\%$)				
	2	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**	**2**	**5**	**10**	**20**	**50**
Div. ($\gamma = 1$); $m = M \times \gamma = \{2, 5, 10, 20, 50\}$																				
0	-1	-1	3	8	6	2	5	6	8	10	5	4	9	6	10	5	6	-2	10	11
0.1	0	6	3	15	11	10	14	10	11	11	2	8	1	16	10	1	1	1	14	9
0.35	-3	9	13	10	12	-7	5	8	9	10	8	8	12	6	13	0	-1	0	5	10
Avg. Sh.	-1	5	7	11	10	2	8	8	10	10	5	7	7	9	11	2	2	0	10	10
Div. ($\gamma = 2$); $m = M \times \gamma = \{4, 10, 20, 40, 100\}$																				
0	3	7	12	14	7	10	11	6	4	11	0	5	6	1	11	-4	8	8	7	2
0.1	-1	11	8	9	13	5	3	7	4	11	5	4	5	7	15	-6	10	14	7	11
0.35	5	6	9	10	13	8	8	16	11	10	-2	5	3	9	7	2	13	13	11	12
Avg. Sh.	3	8	10	11	11	8	7	10	7	10	1	4	5	6	11	-3	10	11	8	8
Div. ($\gamma = 3$); $m = M \times \gamma = \{6, 15, 30, 60, 150\}$																				
0	6	5	10	5	8	-1	2	7	4	10	6	8	0	4	14	4	-2	11	6	10
0.1	11	8	6	12	10	8	11	12	9	11	1	12	5	6	10	1	5	8	8	11
0.35	4	1	0	7	15	8	9	13	9	11	-2	12	8	12	4	-2	11	6	10	
Avg. Sh.	7	5	5	8	11	7	7	11	8	11	4	8	8	9	11	-1	3	8	8	10
Div. ($\gamma = 5$); $m = M \times \gamma = \{25, 50, 100, 250\}$																				
0		7	9	0	7		6	6	11	11		7	7	8	9		9	4	9	10
0.1		1	19	5	9		7	6	11	13		2	10	11	12		5	11	12	12
0.35		4	13	2	8		8	8	13	11		0	8	8	15		10	3	7	10
Avg. Sh.		4	14	3	8		7	7	11	12		6	6	9	12		8	6	9	11

assets, consistent with the results shown in more detail in the previous section.

However, in cases where diversity is included in asset selection prior to optimization with ($\gamma = 2, 3, 5$), it can be seen that the parameter ϵ still has a strong impact, although slightly reduced, on the out-of-sample Sharpe ratio and portfolio diversification. This can only be explained as diversification resulting from a combination of diversity in hypothesis selection, given by γ, and the parametrically introduced diversity in the optimization process through ϵ during the learning of individual predictors.

Appendix D.1 presents additional experiments across different market regimes and a fixed income data set comprising 1336 globally diversified bonds from 2014 to 2018, as shown in Fig. D.15. Specifically, out-of-sample portfolio performance metrics-including the Sharpe, Sortino, and Omega ratios, as well as maximum drawdown-are reported for the full range of the diversity parameter ϵ.

Results for U.S. equity portfolios under bull and bear market regimes are shown in Tables D.5 and D.6, respectively. Table D.7 presents the corresponding results for the fixed income data set. Across all cases, the diversity parameter ϵ improves performance metrics up to an optimal value. In the bull market regime, this optimal value is around 0.5, while in the bear market regime (2008 Credit Crunch), two optimal ranges are observed, typically around 0.5 and 0.95. This suggests that, while bull markets exhibit a well-defined optimal level of diversification, bear markets may benefit from multiple diversity regimes, indicating a more complex relationship between ensemble diversity and out-of-sample portfolio diversification under crisis conditions.

For the fixed income dataset, a pattern similar to the U.S. equity bull market case is observed. In both scenarios, the diversity parameter ϵ exhibits a clearly defined optimal range, and deviations from this range are primarily influenced by the number of portfolio constituents. As reported in Table D.7, the optimal value of ϵ is approximately 0.35 for small and large portfolios (5 and 100 assets or more), and around 0.1 for medium-sized portfolios (20–60 assets).

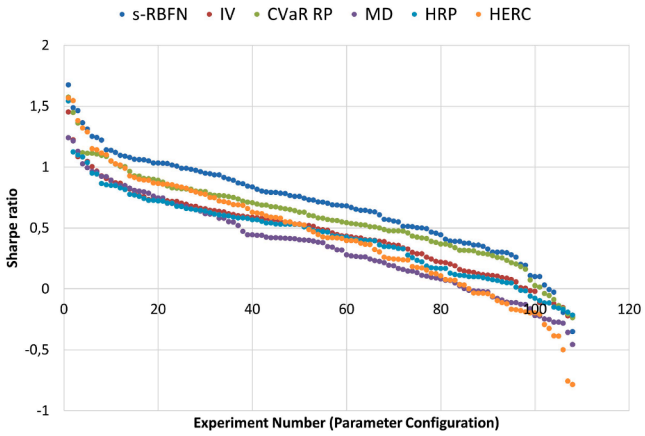


Fig. 9. Sharpe ratios distributions across all experiments (108 configurations for ϵ,γ,M,T) for each portfolio optimization model.

4.4. Comparative performance analysis

This section presents a comparative analysis between the proposed method, implemented with the s-RBFN as the PSEM model, and several well-established, robust and data-driven "plug-in" portfolio optimization methods. These consist of Inverse Volatility (IV) (Ang, Hodrick, Xing, & Zhang, 2006), CVaR Risk Parity (CVaR RP) (Kapsos, Christofides, & Rustem, 2018), Maximum Diversification (MD) (Choueifaty & Coignard, 2008), Hierarchical Risk Parity (HRP) (Lopez de Prado, 2016) and Hierarchical Equal Risk Contribution (HERC) (Raffinot, 2018). The comparison is conducted over a two-year investment horizon with monthly portfolio reallocation. The performance metric used is the monthly standard Sharpe Ratio, calculated using a time series of 1-month cumulative returns. This is computed as the average return over the 24-month period divided by its standard deviation and then annualized. It is important to note that the Sharpe ratios reported in this section are generally lower because they are calculated using average returns, whereas in previous sections, a modified version of the Sharpe ratio was used, based on cumulative returns over 1-month and 10-month periods.

Table 4 reports the average Sharpe Ratios across various threshold levels T and values of the asset selection diversity parameter γ. It can be shown how the presented method to select assets based on prediction diversity (diversify-quality trade-off of predictions as in Section 4.3.2) makes all the methods improve generalization performance, due to increased out-of-sample diversification measured by the Sharpe ratios. Also, the s-RBFN is the top-performing method, as can be seen in the average value across all the experiments at the bottom of the same table. For selections of assets with less than -0.5 % cumulative monthly return, as in Table 4(a), diversity is not of clear value, but for selections above 0.0 % and 0.3 % it is a clear winning selection choice for all methods.

Table 4
Average Sharpe ratios across all configurations of the optimization diversity parameter $\epsilon \in \{0, 0.1, 0.35\}$ and number of portfolio constituents $M \in \{5, 10, 20, 35\}$, grouped by threshold T and the asset selection diversity parameter (multiplicative factor) $\gamma \in \{1, 2, 3\}$.

(a) Threshold $T = -0.5$						
γ	s-RBFN	IV	CVaR RP	MD	HRP	HERC
1	0.79	0.59	0.74	0.41	0.54	0.55
2	0.64	0.40	0.50	0.29	0.36	0.28
3	0.55	0.41	0.46	0.36	0.37	0.42
(b) Threshold $T = 0.0$						
γ	s-RBFN	IV	CVaR RP	MD	HRP	HERC
1	0.70	0.40	0.59	0.31	0.36	0.39
2	0.85	0.56	0.63	0.55	0.55	0.41
3	0.69	0.43	0.49	0.29	0.43	0.51
(c) Threshold $T = 0.3$						
γ	s-RBFN	IV	CVaR RP	MD	HRP	HERC
1	0.55	0.45	0.64	0.35	0.45	0.42
2	0.58	0.40	0.57	0.32	0.37	0.52
3	0.83	0.49	0.65	0.33	0.48	0.47
Avg.	**0.69**	**0.46**	**0.58**	**0.35**	**0.44**	**0.44**

Fig. 9 displays the distribution of Sharpe ratios, ordered from highest to lowest, across all experiments conducted in this section (108 in total). These experiments include various configurations of the diversity parameters ϵ and γ, threshold T, and the number of portfolio constituents M. The horizontal axis represents individual experiments, sorted by decreasing Sharpe ratio. Fig. 9 clearly demonstrates that the s-RBFN consistently outperforms all other models-often by a significant margin-highlighting the robustness and accuracy of the proposed method.

The results indicate that the s-RBFN method consistently outperforms the other portfolio optimization approaches across all threshold levels and values of the asset selection diversity parameter (γ), achieving the highest average Sharpe ratios in most cases. Furthermore, performance generally improves for all methods as γ increases, suggesting that greater diversity in asset selection positively contributes to out-of-sample performance.

In Appendix D.2, the proposed method is compared against alternative portfolio optimization approaches across distinct market regimes-specifically, the bear market during the Global Financial Crisis (2007-2009) and the bull market of the subsequent U.S. equity rally (2009-2016). Additionally, a diversified global bond data set comprising 1336 instruments from 2014 to 2018, characterized in Fig. D.15, is used to evaluate performance across varying portfolio sizes.

As shown in Tables D.8 and D.9, the proposed s-RBFN model consistently achieves the highest average performance across all portfolio sizes under both bull and bear market regimes. Table D.10 presents the results for the fixed income data set, where the s-RBFN attains superior Sharpe, Sortino, and Omega ratios, along with the second-lowest maximum drawdown among all evaluated methods. Notably, the s-RBFN model with 10 portfolio constituents delivers the best performance across all metrics for both the U.S. equity portfolios during the Credit Crunch (Table D.9) and the fixed income data set (Table D.10).

Overall, the proposed solution demonstrates significantly stronger performance for small-sized portfolios compared to other methods, which can be attributed to the enhanced diversification induced by the diversity parameters.

5. Conclusion and future work

This work explored the relationship between the diversity of individual predictors in structured ensembles and portfolio diversification, offering a novel perspective within a multiple-hypothesis predict-then-optimize framework. One key contribution was the use of the ensemble combiner rule as the portfolio target during optimization-for example, the equal-weighted portfolio with the MSE loss. To the best of the authors' knowledge, this represents the first learning-based plug-in method that enables parametric control of out-of-sample diversification prior to decision-making. Experimental results confirmed that out-of-sample diversification and generalization performance depend not only on the optimization procedure but also on the asset selection stage. Notably, when RMSE ranges were similar, asset selections with greater diversity in return predictions consistently outperformed those with less diversity, highlighting the importance of the diversity-quality trade-off. This forms a third contribution, and both sources of diversity were shown to extend the boundaries of portfolio diversification beyond previously suggested limits (Dalio, 2018).

For future work, it would be valuable to explore deeper architectures for individual predictors, as well as the integration of multimodal data sets, enabling hypotheses to incorporate information beyond stock time series. Furthermore, extending the methodology to settings where each hypothesis represents a portfolio or a group of assets could open new avenues for application. Another promising direction involves investigating the interplay between ensemble complexity, prediction diversity, and portfolio risk. Additionally, exploring alternative loss functions and ensemble combiners tailored to various portfolio risk objectives may further enhance the flexibility and effectiveness of the proposed framework.

CRediT authorship contribution statement

Alejandro Rodriguez Dominguez: Conceptualization, Methodology, Data curation, Writing- Original draft preparation, Visualization, Validation. **Muhammad Shahzad:** Writing- Reviewing and Editing, Methodology, Visualization, Supervision, Validation. **Xia Hong:** Writing- Reviewing and Editing, Supervision, Validation

Data availability

Data will be made available on request.

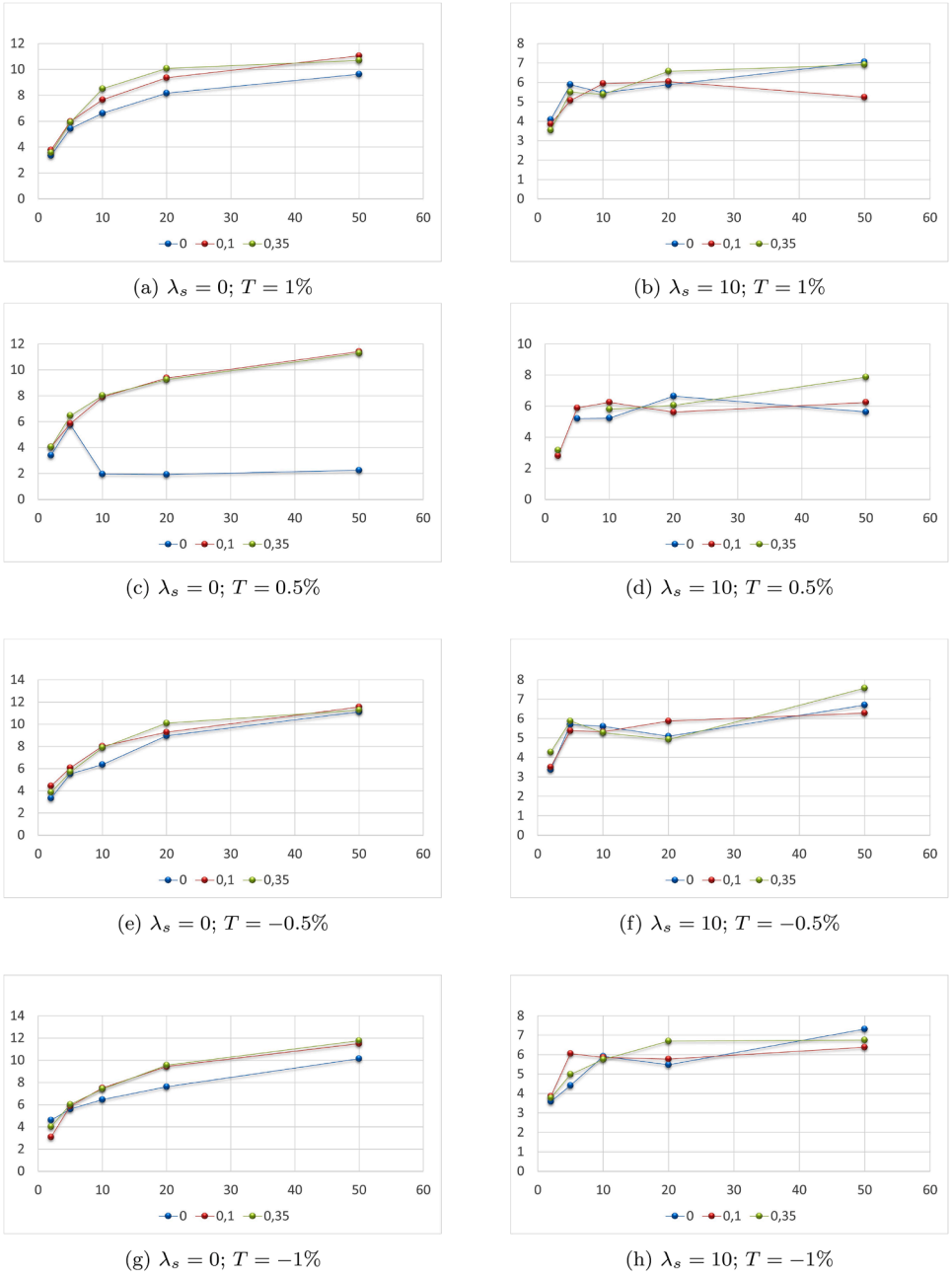
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge Miraltabank Asset Management, and in particular CIO Ignacio Fuertes Aguirre, for providing the fixed income data used in the experiments.

Appendix A. Diversity in Learning Stage: Radial Basis Functions



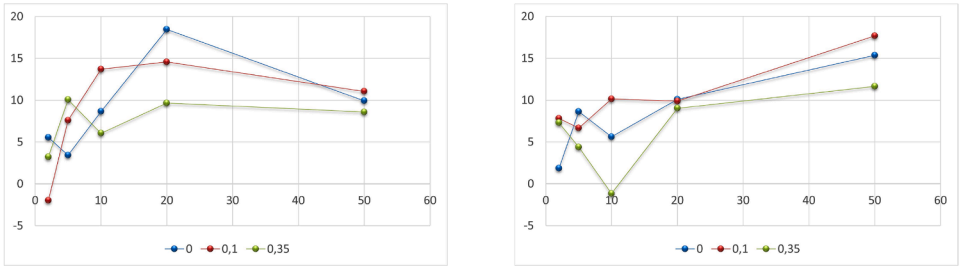
(a) $\lambda_s = 0$; $T = 1\%$ (b) $\lambda_s = 10$; $T = 1\%$

(c) $\lambda_s = 0$; $T = 0.5\%$ (d) $\lambda_s = 10$; $T = 0.5\%$

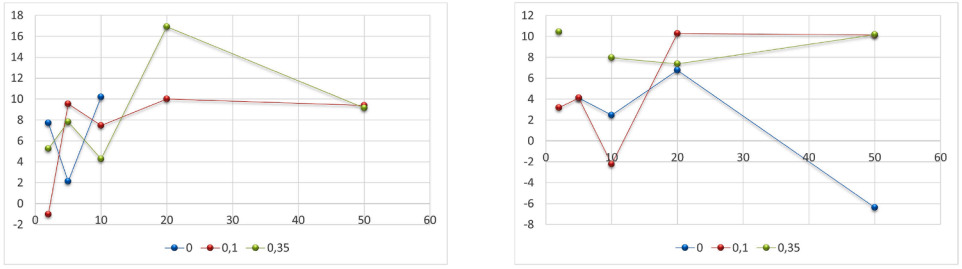
(e) $\lambda_s = 0$; $T = -0.5\%$ (f) $\lambda_s = 10$; $T = -0.5\%$

(g) $\lambda_s = 0$; $T = -1\%$ (h) $\lambda_s = 10$; $T = -1\%$

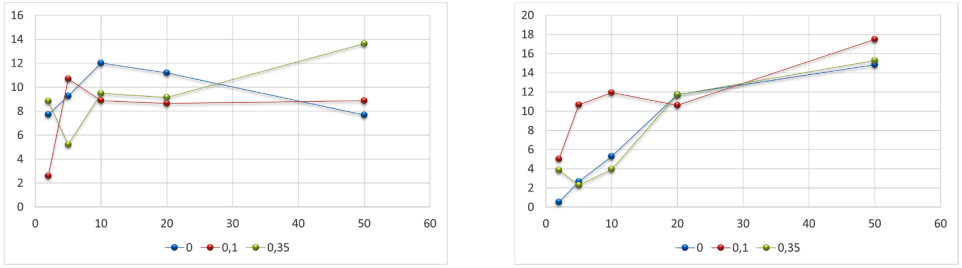
Fig. A.10. One-step (1-month) 100 simulations Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\varepsilon = 0, 0.1, 0.35$ (colours), s-RBFN regularization parameter λ_s, and $-1\,\% \leq T \leq 1\,\%$. s-RBFN with radial basis functions.



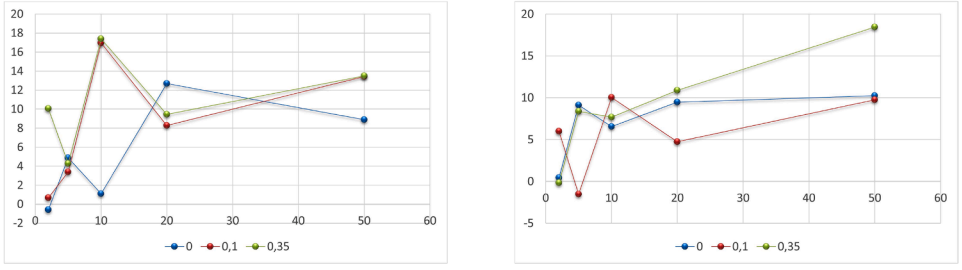
(a) $\lambda_s = 0$; $T = 1\%$ (b) $\lambda_s = 10$; $T = 1\%$



(c) $\lambda_s = 0$; $T = 0.5\%$ (d) $\lambda_s = 10$; $T = 0.5\%$



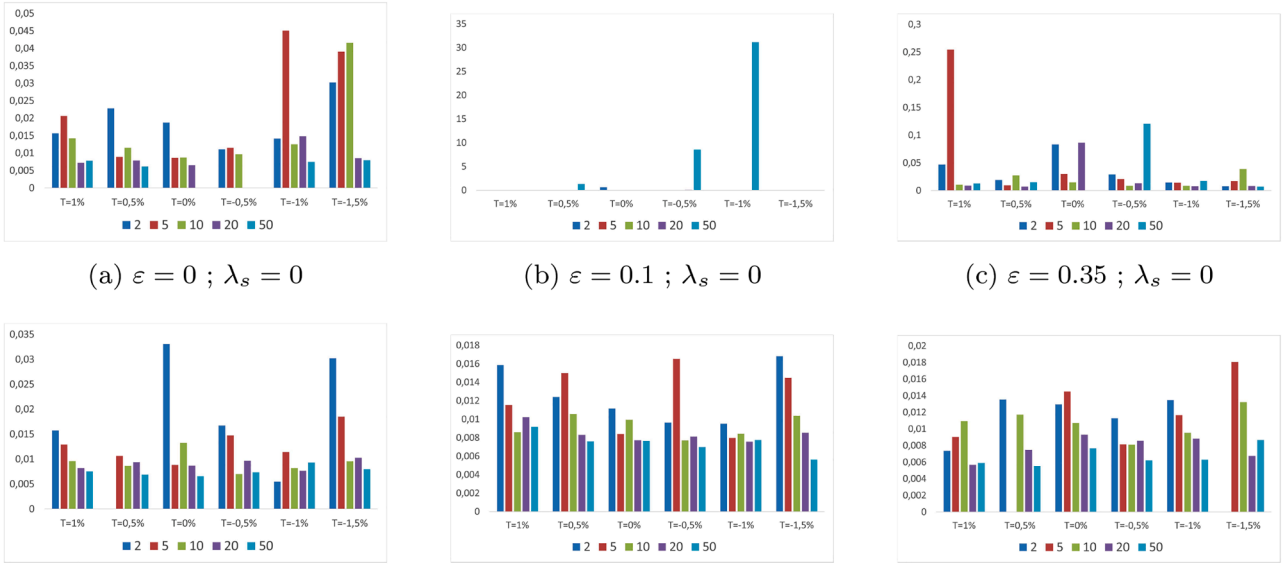
(e) $\lambda_s = 0$; $T = -0.5\%$ (f) $\lambda_s = 10$; $T = -0.5\%$



(g) $\lambda_s = 0$; $T = -1\%$ (h) $\lambda_s = 10$; $T = -1\%$

Fig. A.11. Multi-step (10-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), s-RBFN regularization parameter λ_s, and $-1\,\% \leq T \leq 1\,\%$. s-RBFN with radial basis functions.

Appendix B. Ensemble Regularization in the Gaussian Basis Functions Experiments (Multi-step ahead case)

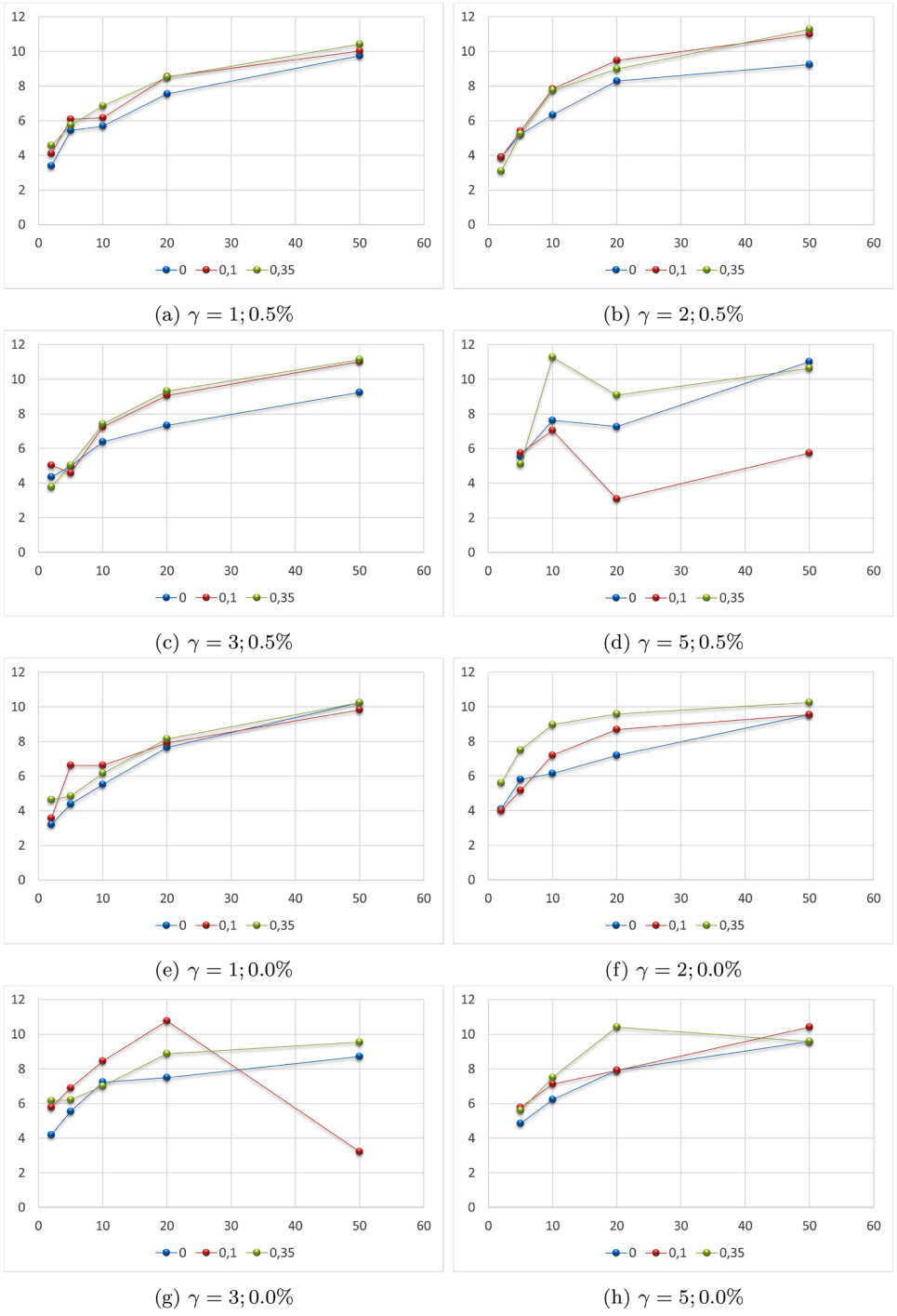


(a) $\varepsilon = 0$; $\lambda_s = 0$ (b) $\varepsilon = 0.1$; $\lambda_s = 0$ (c) $\varepsilon = 0.35$; $\lambda_s = 0$

(d) $\varepsilon = 0$; $\lambda_s = 10$ (e) $\varepsilon = 0.1$; $\lambda_s = 10$ (f) $\varepsilon = 0.35$; $\lambda_s = 10$

Fig. B.12. Portfolio ensemble 10-month (multi-step) average test RMSE with different diversity parameter ε, ensemble (s-RBFN) regularization parameter λ_s, threshold for asset selection T, number of portfolio constituents or predictors (2, 5, 10, 20, 50), and Gaussian basis functions. To perform a closer analysis of the regularization parameter, it can be seen how for $\lambda_s = 10$ the generalization error is more stable for all model hyperparameters, and in the case of no regularization parameter, there are some cases in which the results are quite unstable (like in Figs. B.12(b) and B.12(c)) where the test RMSE are displayed. It can be concluded that the regularization parameter reduces the uncertainty of the s-RBFN hyperparameters in the case of the Gaussian basis function. This instability in hyperparameter selection was not observed with the s-RBFN using radial basis functions.

Appendix C. Including Diversity in the Asset or Hypothesis Selection Stage: One-step Decision-making



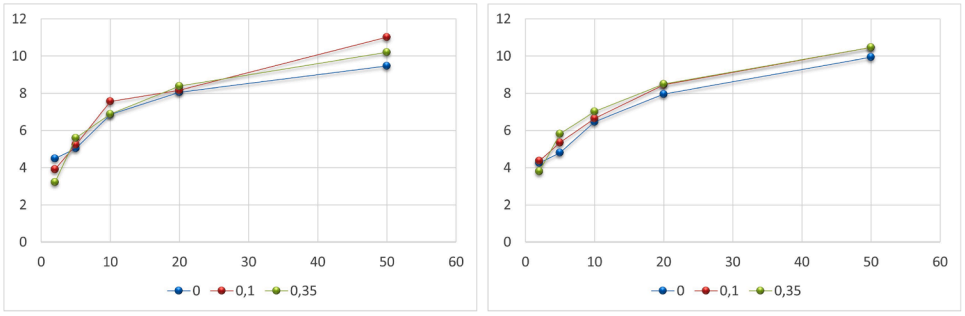
(a) $\gamma = 1; 0.5\%$ (b) $\gamma = 2; 0.5\%$

(c) $\gamma = 3; 0.5\%$ (d) $\gamma = 5; 0.5\%$

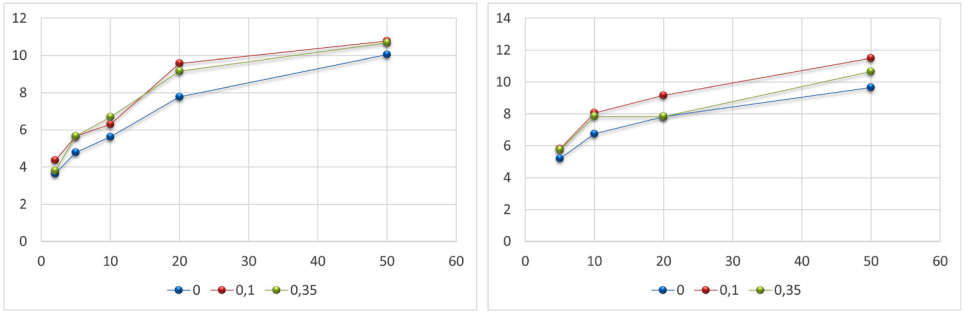
(e) $\gamma = 1; 0.0\%$ (f) $\gamma = 2; 0.0\%$

(g) $\gamma = 3; 0.0\%$ (h) $\gamma = 5; 0.0\%$

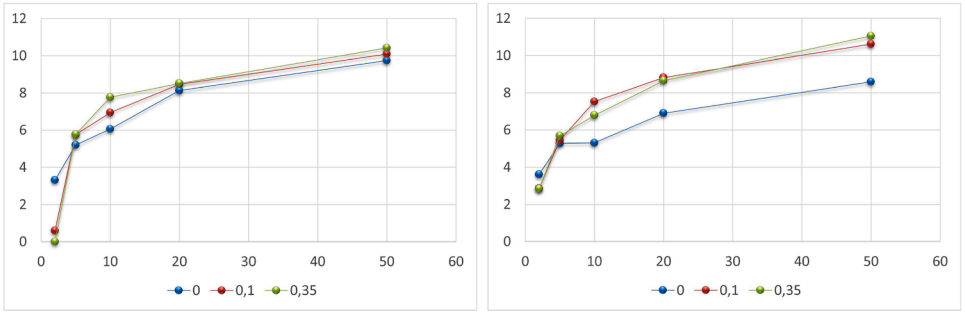
Fig. C.13. One-step (1-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), multiplicative factor or diversity parameter in the asset selection stage γ, and $0\,\% \leq T \leq 0.5\,\%$. s-RBFN with radial basis functions.



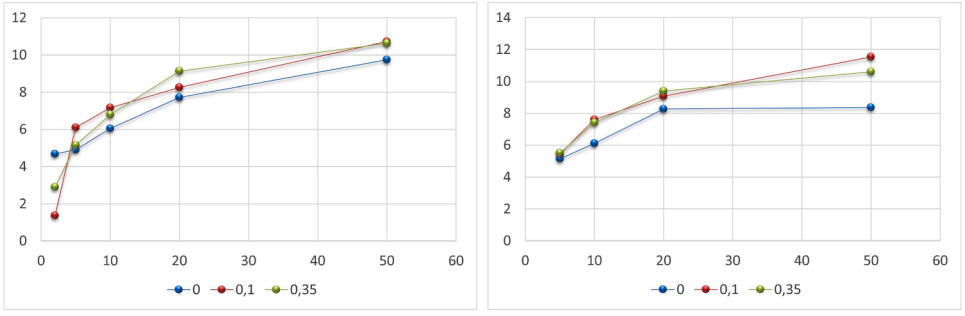
(a) $\gamma = 1; T = -0.5\%$ (b) $\gamma = 2; T = -0.5\%$



(c) $\gamma = 3; T = -0.5\%$ (d) $\gamma = 5; T = -0.5\%$



(e) $\gamma = 1; T = -1\%$ (f) $\gamma = 2; T = -1\%$



(g) $\gamma = 3; T = -1\%$ (h) $\gamma = 5; T = -1\%$

Fig. C.14. One-step (1-month) Sharpe ratios (Y-axis) for different number of predictors M (X-axis), diversity parameter $\epsilon = 0, 0.1, 0.35$ (colours), multiplicative factor or diversity parameter in the asset selection stage γ, and $-1\,\% \leq T \leq -0.5\,\%$. s-RBFN with radial basis functions.

Appendix D. Sensitivity Analysis of Diversity Parameters across different Regimes, Asset Classes and Performance Metrics

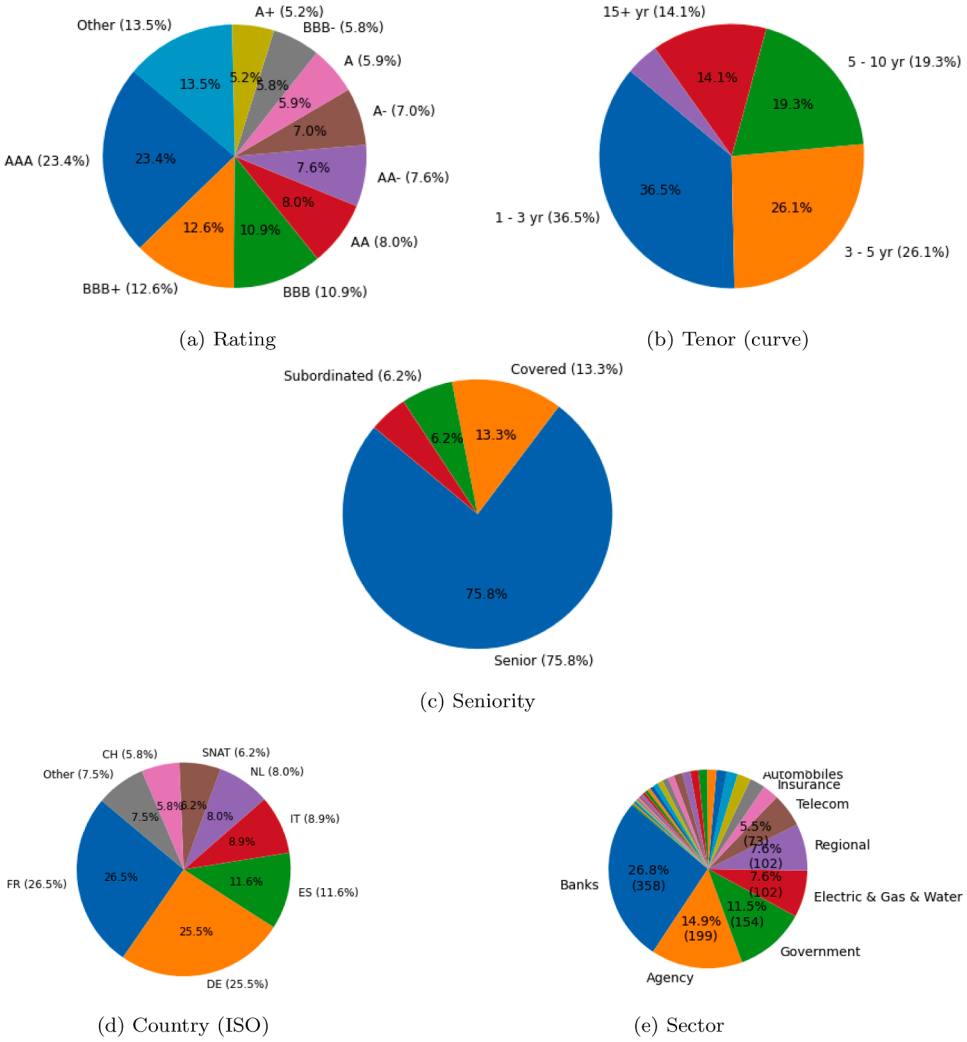
D.1. Performance analysis of the s-RBFN model across different regimes and asset classes as a function of the diversity parameter ϵ and the number of portfolio constituents n

Table D.5
Performance metrics across varying values of ϵ and number of assets n during the U.S. equities market rally spanning 2009 to 2016.

Metric	n	**0**	**0.1**	**0.35**	**0.5**	**0.75**	**0.95**
Sharpe	5	0.645	0.629	0.399	**1.035**	0.685	0.691
	10	1.093	**1.208**	0.960	0.687	1.023	1.220
	20	**1.369**	1.135	1.135	1.452	1.135	0.852
	100	1.020	1.182	1.061	1.231	**1.229**	1.186
	Avg.	1.032	1.039	0.889	1.101	1.018	0.987
Max Drawdown	5	-0.307	**-0.286**	-0.219	-0.328	-0.336	-0.285
	10	-0.352	-0.252	-0.258	-0.348	-0.322	**-0.215**
	20	-0.261	-0.235	**-0.220**	-0.259	-0.248	-0.289
	100	-0.260	-0.259	-0.261	**-0.231**	-0.258	-0.252
	Avg.	-0.295	-0.258	-0.240	-0.292	-0.291	-0.260
Sortino	5	0.923	0.871	0.536	**1.417**	0.893	0.983
	10	1.403	**1.626**	1.336	0.934	1.397	1.763
	20	1.897	1.532	1.541	**1.900**	1.453	1.125
	100	1.328	1.541	1.385	**1.520**	1.611	1.557
	Avg.	1.388	1.393	1.200	1.443	1.339	1.357
Omega	5	1.123	1.118	1.073	**1.205**	1.130	1.132
	10	1.221	**1.242**	1.185	1.129	1.198	1.237
	20	1.270	1.224	1.224	**1.296**	1.211	1.165
	100	1.201	1.238	1.211	1.231	**1.246**	1.237
	Avg.	1.204	1.206	1.173	1.215	1.196	1.193

Table D.6
Performance metrics across varying values of the diversity parameter ϵ and number of assets n, averaged over M, during the U.S. Credit Crunch Crisis (2007–2009).

Metric	n	0	0.1	0.35	0.5	0.75	0.95
Sharpe	5	-0.454	0.007	**0.380**	-0.027	-0.744	0.585
	10	-0.457	-0.027	-0.225	-0.203	-0.484	**-0.112**
	20	**0.129**	-0.041	-0.034	0.215	-0.284	-0.226
	100	-0.186	-0.170	-0.209	**-0.124**	-0.264	-0.298
	Avg.	-0.242	-0.058	-0.022	-0.035	-0.429	**-0.013**
Max Drawdown	5	-0.646	**-0.690**	-0.605	-0.645	-0.790	-0.488
	10	-0.713	-0.564	-0.604	-0.607	-0.640	**-0.534**
	20	**-0.498**	-0.570	-0.551	-0.382	-0.578	-0.575
	100	-0.532	-0.540	-0.577	-0.545	-0.581	**-0.606**
	Avg.	-0.597	-0.591	-0.584	**-0.545**	-0.647	-0.551
Sortino	5	-0.623	**0.008**	0.542	-0.036	-1.077	0.846
	10	-0.625	-0.038	-0.319	-0.270	-0.749	**-0.167**
	20	**0.191**	-0.053	-0.047	0.309	-0.401	-0.313
	100	-0.267	-0.238	-0.291	**-0.173**	-0.372	-0.412
	Avg.	-0.331	-0.080	-0.029	-0.043	-0.900	**-0.011**
Omega	5	0.921	1.001	1.073	0.995	0.878	**1.111**
	10	0.912	**0.995**	0.960	0.964	0.916	0.981
	20	1.024	0.992	0.994	**1.039**	0.949	0.960
	100	0.967	0.970	0.963	**0.978**	0.954	0.947
	Avg.	0.956	0.990	0.998	**0.994**	0.924	**1.000**



(a) Rating (b) Tenor (curve)

(c) Seniority

(d) Country (ISO) (e) Sector

Fig. D.15. Distribution of the 1336 bonds in the data set across key categories: rating, curve (tenor), seniority, country (ISO), and sector.

Table D.7
Performance metrics across varying values of ϵ and number of assets n on a fixed income data set covering the global bond market (public, private, institutional, corporate, municipal sectors) from 2014 to 2018.

Metric	*n*	**0**	**0.1**	**0.35**	**0.5**	**0.75**	**0.95**
Sharpe	5	-0.25	-0.28	**0.01**	-0.68	-0.31	0.03
	10	0.48	-0.57	-0.20	**0.51**	0.15	-0.44
	20	-0.31	**0.51**	-0.54	-0.55	-0.33	-0.20
	100	0.41	0.31	**0.52**	-0.48	0.39	0.35
	Avg.	**0.08**	-0.01	-0.05	-0.30	-0.02	-0.06
Max Drawdown	5	-0.10	-0.21	**-0.07**	-0.19	-0.17	-0.09
	10	-0.14	-0.10	**-0.06**	-0.15	-0.07	-0.10
	20	-0.06	-0.12	-0.10	-0.08	-0.09	**-0.05**
	100	-0.08	-0.06	-0.06	-0.07	**-0.05**	-0.06
	Avg.	-0.10	-0.12	**-0.07**	-0.12	-0.10	-0.08
Sortino	5	-0.29	-0.29	**0.01**	-0.75	-0.32	0.04
	10	4.52	-0.67	-0.24	**9.34**	0.18	-0.49
	20	-0.36	**12.24**	-0.61	-0.64	-0.43	-0.24
	100	2.97	1.36	**16.16**	-0.62	2.17	1.17
	Avg.	1.21	3.16	**3.33**	1.83	0.40	0.12
Omega	5	0.93	0.91	**1.00**	0.84	0.91	1.01
	10	2.84	0.88	0.96	**5.01**	1.03	0.90
	20	0.94	**4.76**	0.89	0.90	0.91	0.96
	100	1.44	1.20	**3.79**	0.91	1.31	1.18
	Avg.	1.54	**1.94**	1.66	1.91	1.29	1.26

D.2. Comparative analysis of different portfolio optimization methods

Table D.8
Performance metrics by strategy and number of assets during the U.S. Equity Rally regime.

n	s-RBFN	1_N	CVaR RP	HERC	HRP	IV	MD
Sharpe Ratio							
5	1.11	0.51	0.58	0.08	0.48	0.58	0.49
10	1.03	1.11	1.10	1.20	1.17	1.10	1.16
20	1.16	1.12	1.20	1.00	1.18	1.20	1.11
100	1.15	1.13	1.19	0.97	1.23	1.19	0.83
Avg	1.11	0.97	1.02	0.81	1.02	1.02	0.90
Maximum Drawdown							
5	-0.30	-0.31	-0.28	-0.41	-0.29	-0.28	-0.26
10	-0.31	-0.34	-0.27	-0.27	-0.25	-0.27	-0.27
20	-0.26	-0.22	-0.20	-0.23	-0.20	-0.20	-0.19
100	-0.25	-0.26	-0.21	-0.23	-0.21	-0.21	-0.32
Avg	-0.28	-0.28	**-0.24**	-0.29	**-0.24**	**-0.24**	-0.26
Sortino Ratio							
5	1.25	0.71	0.80	0.11	0.65	0.80	0.68
10	1.46	1.39	1.41	1.67	1.52	1.41	1.51
20	1.64	1.50	1.61	1.26	1.57	1.61	1.52
100	1.55	1.47	1.53	1.26	1.58	1.53	1.07
Avg	1.48	1.27	1.34	1.08	1.33	1.34	1.20
Omega Ratio							
5	1.00	0.94	0.94	0.98	0.94	0.94	0.94
10	1.17	0.91	0.90	0.92	0.89	0.90	0.87
20	1.02	0.96	0.95	0.93	0.94	0.95	0.93
100	0.99	0.97	0.95	0.95	0.95	0.95	0.95
Avg	**1.04**	0.95	0.93	0.95	0.93	0.93	0.92

Table D.9
Performance metrics by strategy and number of assets during the U.S. Credit Crunch (2007–2009).

Assets	s-RBFN	1/N	CVaR RP	HERC	HRP	IV	MD
Sharpe Ratio							
5	-0.64	-0.33	-0.37	-0.11	-0.34	-0.37	-0.35
10	0.82	-0.53	-0.59	-0.42	-0.65	-0.59	-0.72
20	-0.15	-0.22	-0.28	-0.38	-0.34	-0.28	-0.38
100	-0.07	-0.18	-0.28	-0.23	-0.31	-0.28	-0.28
Avg.	**-0.01**	-0.32	-0.38	-0.29	-0.41	-0.38	-0.43
Maximum Drawdown							
5	-0.80	-0.60	-0.57	-0.54	-0.58	-0.57	-0.56
10	-0.60	-0.68	-0.61	-0.69	-0.62	-0.61	-0.62
20	-0.64	-0.55	-0.52	-0.59	-0.54	-0.52	-0.53
100	-0.56	-0.53	-0.51	-0.50	-0.51	-0.51	-0.48
Avg.	-0.65	-0.59	**-0.55**	-0.58	-0.56	-0.55	-0.55
Sortino Ratio							
5	-0.12	-0.49	-0.55	-0.17	-0.50	-0.55	-0.52
10	0.07	-0.72	-0.80	-0.52	-0.86	-0.80	-0.91
20	-0.21	-0.31	-0.39	-0.50	-0.48	-0.39	-0.50
100	-0.11	-0.25	-0.40	-0.31	-0.44	-0.40	-0.37
Avg.	**-0.09**	-0.44	-0.54	-0.38	-0.57	-0.54	-0.58
Omega Ratio							
5	0.89	0.94	0.94	0.98	0.94	0.94	0.94
10	1.01	0.91	0.90	0.92	0.89	0.90	0.87
20	0.97	0.96	0.95	0.93	0.94	0.95	0.93
100	0.99	0.97	0.95	0.95	0.95	0.95	0.95
Avg.	**0.97**	0.95	0.94	0.95	0.93	0.94	0.92

Table D.10
Performance metrics by strategy and number of assets on a data set of 1336 global bonds, including public, corporate, institutional, municipal, and other sectors (2014–2018).

Assets	s-RBFN	1/N	IV	MD	HRP	HERC
Sharpe Ratio						
5	-0.25	-0.48	-0.80	-0.65	-0.73	-0.78
10	0.48	0.47	0.29	-0.14	0.32	0.19
20	-0.31	-0.37	-0.42	0.08	-0.16	-0.37
100	0.41	0.39	0.32	0.06	0.36	-0.02
Avg.	**0.08**	0.00	-0.15	-0.16	-0.05	-0.25
Maximum Drawdown						
5	-0.10	-0.11	-0.12	-0.10	-0.11	-0.12
10	-0.14	-0.12	-0.12	-0.13	-0.13	-0.14
20	-0.06	-0.06	-0.09	-0.17	-0.11	-0.09
100	-0.08	-0.08	-0.15	-0.07	-0.12	-0.18
Avg.	**-0.10**	**-0.09**	-0.12	-0.12	-0.12	-0.13
Sortino Ratio						
5	-0.29	-0.54	-0.73	-0.59	-0.67	-0.73
10	4.52	4.42	1.91	-0.25	2.19	0.80
20	-0.36	-0.48	-0.58	0.11	-0.23	-0.50
100	2.97	2.85	4.09	0.09	3.87	-0.13
Avg.	**1.21**	1.06	1.17	-0.16	1.29	0.11
Omega Ratio						
5	0.93	0.90	0.80	0.82	0.81	0.81
10	2.84	2.62	1.53	0.91	1.69	1.20
20	0.94	0.91	0.79	1.11	0.88	0.84
100	1.44	1.40	1.65	1.07	1.79	0.98
Avg.	**1.53**	1.46	1.19	0.98	1.29	0.96

References

Akopov, A. (2014). Parallel genetic algorithm with fading selection. *International Journal of Computer Applications in Technology*, *49*, 325–331. https://doi.org/10.1504/IJCAT.2014.062368

ANG, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259–299. https://doi.org/10.1111/j.1540-6261.2006.00836.x

Avramov, D., & Zhou, G. (2010). Bayesian portfolio analysis. *Review of Financial Economics*, *2*, 25–47. https://api.semanticscholar.org/CorpusID:45787980.

Ben-Tal, A., Ghaoui, L., & Nemirovski, A. (2009). Robust optimization. https://doi.org/10.1515/9781400831050

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, *53*(3), 464–501. https://doi.org/10.1137/080734510.

Bertsimas, D., Gupta, V., & Kallus, N. (2013). Data-driven robust optimization. *Mathematical Programming*, *167*. https://doi.org/10.1007/s10107-017-1125-8

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, *4*(2), 315–342. http://www.jstor.org/stable/2962107.

Bodnar, T., Parolya, N., & Thorsen, E. (2024). Two is better than one: Regularized shrinkage of large minimum variance portfolios. *Journal of Machine Learning Research*, *25*(173), 1–32. http://jmlr.org/papers/v25/22-1337.HTML.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, *45*(1-4), 21-58. https://doi.org/10.1007/BF02282040

Cai, Z., Cui, Z., Lassance, N., & Simaan, M. (2024). The economic value of mean squared error: Evidence from portfolio selection. https://doi.org/10.2139/ssrn.4856139

Chen, C., & Zhou, Y.-s. (2018). Robust multiobjective portfolio with higher moments. *Expert Systems with Applications*, *100*, 165–181.

Choueifaty, Y., & Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management*, *35*, 40–51. https://doi.org/10.3905/JPM.2008.35.1.40

Dalio, R. (2018). Principles. Simon & Schuster. https://books.google.es/books?id=qNNmDwAAQBAJ.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, *22*, 1915–1953. https://api.semanticscholar.org/CorpusID:1073674.

DeMiguel, V., & Nogales, F. J. (2007). Portfolio selection with robust estimation. *European Finance eJournal*. https://api.semanticscholar.org/CorpusID:263885541.

Dominguez, A. R., Shahzad, M., & Hong, X. (2025). Structured radial basis function network: Modelling diversity for multiple hypotheses prediction. In M. Bramer, & F. Stahl (Eds.), *Artificial intelligence XLI* (pp. 88–101). Cham: Springer Nature Switzerland.

Donti, P., & Kolter, J. (2017). Task-based end-to-end model learning. https://doi.org/10.48550/arXiv.1703.04529

Elmachtoub, A. N., & Grigas, P. (2017). Smart "predict, then optimize". ArXiv, *abs/1710.08005*. https://api.semanticscholar.org/CorpusID:38399700.

Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, *68*(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922

Esfahani, P. M., & Kuhn, D. (2015). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, *171*, 115–166. https://api.semanticscholar.org/CorpusID:14542431.

Giglio, S., Kelly, B., & Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics*, *14*(Volume 14, 2022), 337–368. https://doi.org/10.1146/annurev-financial-101521-104735

Goh, C. Y., & Jaillet, P. (2016). Structured prediction by least squares estimated conditional risk minimization. ArXiv, abs/1611.07096. https://api.semanticscholar.org/CorpusID:2579082.

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*(3), 621-656. https://doi.org/10.1017/S0022109000004129

Kapsos, M., Christofides, N., & Rustem, B. (2018). Robust risk budgeting. *Annals of Operations Research*, *266*(1), 199–221. https://doi.org/10.1007/s10479-017-2469-4

Kotary, J., Vito, V. D., Christopher, J., Hentenryck, P. V., & Fioretto, F. (2023). Predict-then-optimize by proxy: Learning joint models of prediction and optimization. https://arxiv.org/abs/2311.13087.

Lassance, N., Martin-Utrera, A., & Simaan, M. (2023). The risk of expected utility under parameter uncertainty. *Management Science, forthcoming*. https://doi.org/10.1287/mnsc.2023.00178

Lasse Heje Pedersen, A. B., & Levine, A. (2021). Enhanced portfolio optimization. *Financial Analysts Journal*, *77*(2), 124–151. https://doi.org/10.1080/0015198X.2020.1854543

Ledoit, O., & Wolf, M. (2003). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*. https://doi.org/10.2139/ssrn.433840

Lwin, K., Qu, R., & Kendall, G. (2014). A learning-guided multi-objective evolutionary algorithm for constrained portfolio optimization. *Applied Soft Computing*, *24*, 757–772. https://doi.org/10.1016/j.asoc.2014.08.026

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, 113973. https://doi.org/10.1016/j.eswa.2020.113973

Mahadi, M., Ballal, T., Moinuddin, M., & Al-Saggaf, U. M. (2022). Portfolio optimization using a consistent vector-based MSE estimation approach. *IEEE Access*, *10*, 86635–86645. https://doi.org/10.1109/ACCESS.2022.3197896

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Nguyen, T.-D., & Lo, A. W. (2012). Robust ranking and portfolio optimization. *European Journal of Operational Research*, *221*(2), 407–416. https://doi.org/10.1016/j.ejor.2012.03.023

Ortega, L. A., Cabañas, R., & Masegosa, A. (2022). Diversity and generalization in neural network ensembles. In G. Camps-Valls, F. J. R. Ruiz, & I. Valera (Eds.), *Proceedings of the 25th international conference on artificial intelligence and statistics* (pp. 11720–11743). PMLR (vol. *151*). Proceedings of Machine Learning Research. https://proceedings.mlr.press/v151/ortega22a.html.

Ozelim, L., Ribeiro, D., Schiavon, J., Domingues, V., & Queiroz, P. (2023). Hposs: A hierarchical portfolio optimization stacking strategy to reduce the generalization error of ensembles of models. *PloS ONE*, *18*, e0290331. https://doi.org/10.1371/journal.pone.0290331

Lopez de Prado, M. (2016). Building diversified portfolios that outperform out of sample:. *The Journal of Portfolio Management*, *42*, 59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Raffinot, T. (2018). The hierarchical equal risk contribution portfolio. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3237540

Rahimian, H., & Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, *3*, 1–85. https://doi.org/10.5802/ojmo.15

Roncalli, T. (2013). Introduction to risk parity and budgeting. *CGN: Sovereign wealth funds as investors (sub-topic)*. https://api.semanticscholar.org/CorpusID:154472565.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). Lectures on stochastic programming. Modeling and theory. https://doi.org/10.1137/1.9780898718751

Sharpe, W. F. (1994). The sharpe ratio. https://api.semanticscholar.org/CorpusID:55394403.

S&P (2005). S&p 500 — Wikipedia, the free encyclopedia. [Online; accessed 02-September-2024].

S&P (2024). S&p 500 — bloomberg l.p. [Online; accessed 02-June-2024].

Vanderschueren, T., Verdonck, T., Baesens, B., & Verbeke, W. (2022). Predict-then-optimize or predict-and-optimize? an empirical evaluation of cost-sensitive learning strategies. *Information Sciences*, *594*, 400–415. https://api.semanticscholar.org/CorpusID:247061868.

Wilder, B., Dilkina, B., & Tambe, M. (2019). Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 1658–1665. https://doi.org/10.1609/aaai.v33i01.33011658

Wood, D., Mu, T., Webb, A. M., Reeve, H. W. J., Luján, M., & Brown, G. (2024). A unified theory of diversity in ensemble learning. *Journal of Machine Learning Research : JMLR*, *24*(1).

Wu, X., & Liu, Y.-K. (2012). Optimizing fuzzy portfolio selection problems by parametric quadratic programming. *Fuzzy Optimization and Decision Making*, *11*. https://doi.org/10.1007/s10700-012-9126-9

Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with markowitz's mean-variance framework: A review of literature. *Fuzzy Optimization and Decision Making*, *17*. https://doi.org/10.1007/s10700-017-9266-z

Zhao, W., Liu, X., Wu, Y., Zhang, T., & Zhang, L. (2022). A learning-to-rank-based investment portfolio optimization framework for smart grid planning. *Frontiers in Energy Research*, *10*. https://doi.org/10.3389/fenrg.2022.852520