AIMS Mathematics, 11(3): 6162–6216.
DOI: 10.3934/math.2026255
Received: 16 January 2026
Revised: 17 February 2026
Accepted: 02 March 2026
Published: 11 March 2026

https://www.aimspress.com/journal/Math

Research article

Wavelet–PCA mirror-flow portfolios: Multiscale risk geometry and reduced-order dynamics on the simplex

Muhammad Hilal Alkhudaydi*

Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

* **Correspondence:** Email: mh.ayedh@tu.edu.sa.

Abstract: Financial markets exhibit heterogeneous investment horizons and nonstationary dependence, so portfolio risk and comovement are intrinsically multiscale and time varying. Classical single-horizon covariance models and repeated static reoptimization can therefore be fragile in high-dimensional universes, especially considering transaction costs.

We proposed a reduced-order portfolio optimization method in which long-only allocations evolved dynamically through simplex-preserving mirror flow. The strategy was parameterized by a small number of time-varying decision variables, and exposures to a larger asset universe were obtained by combining a few investable basis portfolios. The basis was updated online from the principal components of a wavelet-based multiscale covariance estimator and mapped to admissible weights through an entropy-based softmax construction. A wavelet multiscale layer was integrated as a structural component of the model rather than as a preprocessing step, inducing both a multiscale quadratic risk geometry and a risk-normalized multiscale return signal that drove allocation dynamics.

We provided a self-contained formulation and established core properties of the scheme, including positive definiteness and stability of the multiscale covariance updates, stability of the principal component analysis (PCA)–softmax basis under perturbations, simplex invariance of the mirror update, and explicit turnover control that transferred from reduced allocations to traded allocations, including the effect of periodic basis refresh. Empirical experiments were then conducted on real equity data, including transaction costs, compared against equal-weight and classical mean–variance baselines, including sensitivity and ablation analyses. The results illustrated that the multiscale and reduced-order components produced materially different dynamic allocations and contributed to risk-adjusted performance and trading stability. We referred to the proposed method as WPROD-R, a shorthand for a wavelet-driven product (multiplicative mirror-flow) scheme with a reduced-order representation.

Keywords: multiscale portfolios; wavelet covariance; mirror flow; reduced-order optimization
Mathematics Subject Classification: 62P05

1. Introduction

Financial prices reflect the interaction of heterogeneous market participants operating at different investment horizons. Under the fractal market hypothesis, this heterogeneity implies that predictability, comovement, and risk may vary across time scales rather than being well-described by a single sampling frequency; recent empirical evidence shows that multiscale perspectives can improve pricing and cash-and-carry-style strategies in futures markets [1]. In addition, the intensity and structure of risk linkages can change across horizons, so a single-scale covariance can conceal scale-dependent comovements that matter for portfolio construction [2]. Nonstationary regimes further complicate risk estimation: interdependencies may reorganize abruptly, as documented in high-frequency markets where dependence is regime-specific [3]. These features motivate portfolio rules that (i) model risk at multiple scales and (ii) adapt dynamically over time.

Most classical portfolio frameworks remain predominantly static. Even when richer objectives (e.g., alternative risk measures or environmental, social, and governance (ESG) considerations) are incorporated, the resulting allocations are typically derived from repeated solutions of a static optimization problem [4]. More broadly, dynamic-systems approaches in economics emphasize that resilience and responses to shocks evolve over time and can be meaningfully analyzed only through explicit dynamics [5]. Our goal is to bring this dynamical viewpoint into a mathematically controlled portfolio optimization framework while keeping the decision dimension small and the trading constraints explicit.

Zhao et al. formulated a joint scheduling model that explicitly couples interactions among runways, taxiways, and gates and solved it with a two-stage hybrid optimization procedure that combines differential evolution with ant colony optimization (ACO), illustrating how structured problem coupling and algorithmic design can be developed in tandem—A perspective that aligns with our joint construction of multiscale risk geometry and reduced-order portfolio dynamics [6]. Deng et al. proposed a privacy protection-enhanced, vertical-horizontal, federated learning (FL) framework that incorporates a dual differential privacy (DP) mechanism together with a sharpness-aware minimization optimizer, offering theoretical privacy and convergence analysis for multisource heterogeneous data and motivating stability-aware learning when inputs are distributed or heterogeneous across sources [7]. In a complementary signal-processing setting, Deng et al. introduced an antinoise diagnostic pipeline based on the time-reassigned multisynchrosqueezing transform (TMSST) to obtain a high-resolution time–frequency representation (TFR) and a complex sparse dictionary-learning scheme that preserves phase information under interference, reinforcing the value of multiscale time–frequency operators for robust feature extraction in noisy environments, consistent with our wavelet-based feature construction for risk and signal estimation [8].

Portfolio risk can be defined beyond a single-horizon covariance by constructing a quadratic form that aggregates variance and dependence across multiple time scales or frequency bands, such that each scale contributes an interpretable component of the total risk. One approach decomposes asset returns into scale-specific components and then defines portfolio risk using the variance and covariance at each scale, which yields a multiscale quadratic form whose terms can be interpreted as horizon-dependent contributions to total risk [9]. A closely related frequency-domain perspective represents returns in terms of frequencies and therefore measures portfolio risk across different frequency bands, allowing risk to be analyzed and managed independently of a single sampling horizon [10]. Building on this

idea, multiscale quadratic risk can be operationalized through multiple spectral risk constraints, where separate quadratic constraints are imposed on different frequency ranges and combined within one optimization problem to control exposures at several horizons simultaneously [11]. A more direct time-domain formulation is the mean squared variance (MSV), defined as the average of portfolio variance computed over multiple horizons, which can be written as a quadratic form and provides a clear interpretation as a measure of variance stability across scales, while also revealing practical limitations such as concentration and sensitivity to estimation errors [12]. Robustness can be incorporated by defining risk under mean-covariance ambiguity using the Gelbrich risk model, which preserves a quadratic structure with a clear interpretation and reduces to a regularized mean-variance framework under specific assumptions, although it may not admit closed forms for more general loss functions [13]. Extending beyond Gaussian assumptions, non-Gaussian multivariate models motivate multiscale-aware quadratic risk definitions that better capture tail behavior and dependence across scales, even if they may still underestimate extreme risks at high confidence levels [14]. Finally, practical studies have shown that such multiscale quadratic risk measures can be embedded into real-world risk management and hedging systems, while noting that implementation choices, market frictions, and model approximations can materially affect the outcomes [15].

A practical way to reduce a high-dimensional portfolio decision to a small set of dynamic variables while maintaining simplex feasibility (i.e., weights remain nonnegative and sum to one) is to replace the original decision with a lower-dimensional representation that still maps back into the simplex with a controlled error. Random projection techniques follow this idea by embedding a large quadratic program into a smaller space, solving the reduced problem, and then lifting the solution back, which can substantially decrease the number of effective decision variables while offering an explicit approximation viewpoint, even though the lifted solution can be slightly infeasible, and the projection error may be difficult to characterize in some instances [16]. A complementary route uses structured reformulations that tighten or approximate mixed-integer portfolio models while preserving the simplex structure of the problem. The approximated projected perspective reformulation (AP2R) introduces an approximation that reduces decision complexity and can be strengthened using dual information, yielding a clearer approximation mechanism but simultaneously facing trade-offs such as potentially weaker bounds and sensitivity of performance to formulation choices [17]. When transaction costs are central, a fast Löwner–John ellipsoid approximation can be combined with dynamic modeling to reduce the computational burden and provide an approximation theory that explains why the simplified representation can still deliver strong performance, even though the approach remains challenged by dimensionality and does not fully cover all real market frictions [18]. More broadly, the review of linear programming (LP) and mixed-integer linear programming (MILP) formulations emphasizes that simplex-feasible portfolio models can be made more tractable by reformulating constraints and using approximations to limit the effective degrees of freedom. However, incorporating realistic features can increase complexity, and large-scale instances remain difficult, motivating approximation theories that justify dimension reduction without abandoning feasibility [19].

A broadly useful class of dynamics on the simplex that delivers (i) well-posed updates, (ii) an optimization interpretation, and (iii) a continuous-time ordinary differential equation (ODE) limit is given by mirror-descent-type constructions and closely related simplex geometries that connect discrete updates, Lyapunov structure, and limiting flows. In particular, the time-scale dynamics on the simplex built from the incentive, adaptive, and time-scale components admit a unified Lyapunov analysis using

information-theoretic measures of distance and Riemannian geometries on the simplex. This analysis supports well-posed evolution and stability guarantees across a wide family of simplex dynamics, including replicator-type dynamics [20]. On the algorithmic side, mirror descent (MD), defined as an optimization-driven update rule based on a distance-generating function (DGF), has an explicit optimization interpretation and can be studied directly in continuous time in a stochastic networked setting, where a noisy continuous-time MD model is introduced, and the convergence behavior is characterized, to highlight how noise variance affects limiting performance [21]. Discrete-time simplex-feasible updates that align with this viewpoint also appear through exponential-weights-style methods on the simplex, where exponential weights and projected gradient descent are analyzed via a zeroth-order gradient estimator for smooth objectives on the simplex, emphasizing well-defined, feasible updates even when only function evaluations are available [22]. From a continuous-time perspective with a geometric interpretation, stochastic replicator dynamics provide an ODE-based evolution on the simplex that can be viewed as intrinsic Brownian motion under Aitchison geometry as well as approximation results which including a Jordan–Kinderlehrer–Otto (JKO) scheme to connect the stochastic dynamics to principled limiting and approximation interpretations [23]. Finally, a complementary dynamical systems viewpoint links convex optimization and continuous-time dynamics more generally by framing optimization algorithms as control-inspired dynamical systems that converge to optimal solutions, reinforcing the optimization interpretation of the flow that underlies discrete updates [24].

Algorithmic design strategies can empirically demonstrate bounded transaction costs (TC) control across computational reduction techniques by coupling reduced decision rules with evaluation protocols and explicit bounds. In portfolio optimization with TC, a new algorithm derives optimality conditions using only n-dimensional quantities, which reduces computational complexity and is shown in numerical experiments to outperform a barrier method (solving instances up to 79 times faster), with effectiveness increasing as TC rises, while also noting sensitivity to problem size, small-cost regimes, and the use of a random target vector in experiments [25]. In a dynamic setting, easy-to-compute heuristic trading strategies based on simpler models are complemented with upper bounds on performance to assess near-optimality. Furthermore, Monte Carlo (MC) simulations show heuristics performing close to these bounds, with type 2 gradient penalties producing better upper bounds than type 1 penalties, although the approach is limited by the growing state-space dimension and the possibility that heuristics and MC estimates are not fully optimal [26]. A related approach using equilibrium computation for financial economies with incomplete markets and TC uses differentiable homotopy and an equilibrium selection concept to manage multiple equilibria. However, it also highlights limitations in terms of mathematical complexity, handling of nondifferentiabilities and locally nonunique equilibria, and subjectivity in selection, which can constrain the strength of bounded TC control in practice [27]. More generally, polynomial-time black-box reductions that provide exact incentive compatibility for multidimensional and continuous-type spaces and yield an approximation scheme with inverse polynomial dependence on additive loss illustrate how formal guarantees can persist under reduction while emphasizing limits such as restriction to expected social welfare objectives, exponential dependence on dimension and type space, and the lack of empirical evidence in the study [28]. Empirical demonstration across such heterogeneous reductions depends on careful evaluation design: performance comparisons can vary with captime, instance set size, and scenario hardness, and predictive models can approximate empirical analysis to reduce computational cost without materially degrading accuracy, which supports more reliable cross-technique assessments of bounded TC control [29]. Complementary reduction at the

evaluation stage comes from simulation-based analysis using variance reduction (VR) techniques and simulation speedups, which improve the precision and efficiency of the empirical estimates used to compare algorithmic designs [30]. In addition, computation-intensive algorithm selection introduces low-bound evaluation and relaxation-based heuristics and reports empirical improvements in the critical path, throughput, and area, indicating how selecting among competing reduced designs can materially change realized cost-performance tradeoffs while also noting nondeterministic polynomial time (NP) completeness or NP-hardness and early stage maturity limitations [31]. Finally, the empirical evaluation of algorithmic debugging strategies using benchmarks shows how debugging can help identify which design components drive performance and cost while recognizing that some strategies can be costly and may require users to answer long series of connected questions, which can complicate systematic empirical validation [32].

Robustness and tracking in nonstationary markets, where both the objective and feasible sets change over time, can be formulated by combining (i) forward-looking tracking objectives that explicitly account for model deviations and noise, (ii) dynamic feasibility constraints that evolve with prices and admissible hedges, and (iii) time-consistent robust control formulations that remain meaningful when preferences or constraints shift. For index tracking, a robust optimization formulation based on quadratic and linear models was used to minimize the tracking error while controlling conservatism. Empirical results in [33] show reduced tracking error, improved performance measures, stronger behavior in noisy conditions, and more uniform and diversified weights. However, the study itself remains focused on static models and explicitly notes the need for multiperiod planning and additional constraints to address truly time-varying objectives and feasible sets [33]. A complementary tracking formulation is a forward-looking robust index tracking approach that uses Bregman divergence (a divergence used to measure deviations between nominal and actual distributions) to define robustness, derive a semianalytical robust portfolio, and report better out-of-sample and downturn performance than a nonrobust alternative. It simultaneously highlights that practical robustness depends on tuning the interplay of parameters such as λ and the contamination radius η (a parameter governing how far the actual distribution can deviate), and notes missing comparisons and unresolved joint requirements, such as sparsity with robustness [34]. When the feasible set is driven by hedging and no-arbitrage-style constraints that move with the market, robustness is formulated through a dynamic programming principle for superhedging prices, where superhedging prices evolve as a concave envelope of future prices evaluated at the current prices. Robust investment is posed as maximizing the expected utility of intertemporal consumption subject to a robust superhedging constraint, with the existence (and possible uniqueness) of optimal strategies under suitable assumptions but with an explicit limitation to discrete time and an open extension challenge to continuous time [35]. To keep robustness meaningful as objectives and constraints evolve, robust control formulations transform a penalty problem into a constraint problem and introduce an additional endogenous state variable to ensure time consistency of the constraint specification, using a recursive entropy structure for date-0 relative entropy and focusing constraint preferences on future probability distribution changes. This approach emphasizes practical and conceptual difficulties such as specifying the alternative-model set Q (the set of candidate probability models), potential temporal inconsistency under date-0 relative entropy constraints, and restrictions on how probability distortions are applied over time [36]. When objectives themselves move because preferences are time-inconsistent, robustness is formulated through a game-theoretic framework based on a subgame-perfect Nash equilibrium, yielding extended dynamic programming and Hamilton–Jacobi–Bellman–Isaacs equations (HJBI; the

robust-control dynamic programming equations) with a verification theorem and an application to robust dynamic mean-variance portfolio selection under constant and state-dependent risk aversion. The authors notes complexity and reliance on dominated model sets, and the given example does not directly resolve the full moving-objective and moving-feasible-set case [37]. For robustness in portfolio selection under changing uncertainty, a min–max mean-variance formulation with ambiguous covariance was developed using a McKean–Vlasov dynamic programming approach, providing explicit robust strategies, a robust efficient frontier, and a lower bound for the Sharpe ratio while acknowledging unresolved complexity when drift uncertainty is present due to feedback dependence on both wealth and expected wealth [38]. The design choice of how conservative robustness should be can also be addressed by explicitly managing the "price of robustness", where probabilistic bounds on constraint violations are used to adjust conservatism in a linear robust formulation, and numerical results (including for portfolio optimization) illustrate feasibility and near-optimality under data changes. This approach, however, can be overly conservative and does not directly address time-varying objectives and feasible sets [39]. A broader synthesis of robust portfolio formulations emphasized that robustness can be posed under alternative objectives such as mean-variance, mean-value-at-risk (VaR; a downside risk criterion), and mean-conditional-value-at-risk (CVaR; a tail-focused downside risk criterion), and it highlighted the sensitivity of mean-variance strategies to input and estimation errors alongside the role of robust estimation and Bayesian robust approaches while also stressing limited empirical validation and a gap in multiperiod robust studies with many assets, which is central when both objectives and feasible sets evolve over time [40]. Finally, an online learning perspective addressed moving objectives by robustly evolving a convex coverage set (a set of policies spanning trade-offs across objectives) of policies in nonstationary environments via continuous evaluation and enhancement, with strong empirical performance in nonstationary settings but with stated needs for improved preference exploration, potential nonlinear scalarization, and better generalization across environments [41].

Beyond the financial mathematics literature, several of the methodological themes used here, including multiscale time–frequency representations and constrained dynamic optimization, also appear in neighboring application areas. For instance, time-reassigned multisynchrosqueezing variants have been used for robust feature extraction in nonstationary and noisy mechanical signals [42]; multiobjective evolutionary designs have been developed for resource-constrained optimization problems such as unmanned aerial vehicle (UAV) scheduling [43]; and adaptive multitask optimization with anomaly-transfer mechanisms has been proposed for heterogeneous, time-varying sources [44]. Although these settings differ from portfolio selection, they underscore the generality of multiscale operators and dynamic optimization tools, and they motivate our emphasis on operator choice, stability, and robustness.

Research gaps addressed in this paper

The present work is motivated by the following gaps in the mathematical portfolio optimization literature:

Gap 1: Single-scale risk geometry vs. multiscale market structure. Most portfolio constructions adopt a single covariance geometry, implicitly assuming that risk is captured at an effective horizon. This leaves open the question of how to formulate, in a mathematically transparent way, a *multiscale* quadratic risk functional whose components correspond to scale-localized fluctuations and whose estimators remain well-posed under online updating.

Gap 2: High-dimensional optimization vs. reduced-order decision variables with approximation control. In large universes, estimating inputs and solving full-dimensional optimization at every time can amplify noise. Although factor models and principal component analysis (PCA) reductions are widely used, there is limited theory for portfolio maps that simultaneously: (i) preserve long-only simplex feasibility by construction, (ii) reduce the number of dynamic state variables to a small m, and (iii) admit an approximation interpretation in terms of the spectral tails of the underlying geometry.

Gap 3: Static reoptimization vs. simplex-preserving dynamics with an ODE interpretation. Many discrete-time methods repeatedly solve a static optimization problem, which obscures the dynamic structure of the allocation process. A mathematically natural alternative is to define the policy through a simplex-preserving mirror-flow (multiplicative update) that admits a Kullback–Leibler (KL) proximal characterization and a continuous-time limit of replicator type. However, the integration of such dynamics with multiscale risk and reduced-order parameterizations is still underdeveloped.

Gap 4: Transaction costs handled as a penalty vs. explicit turnover control (including basis refresh). Practical deployment requires controlling the turnover and understanding how it propagates through the model components. Existing turnover regularizations often do not yield explicit bounds that transfer from reduced to asset allocations. Moreover, when the investable basis is time-varying (e.g., via periodic PCA refresh), the turnover impact of basis changes must be quantified in terms of perturbations of the underlying risk geometry and spectral separation.

Gap 5: Nonstationarity and robustness vs. operator-sensitive models with principled diagnostics. Models that incorporate transforms (such as wavelets) and dimension reduction introduce operator choices (wavelet family, number of components, refresh period) that can materially affect system behavior. Therefore, a mathematical finance treatment benefits from a framework in which robustness can be discussed through stability bounds, sensitivity with respect to model "design parameters", and tracking-style diagnostics when the environment and comparator portfolios evolve over time.

Contributions. The main contributions of this work can be summarized as follows:

- We introduce a *structural* wavelet multiscale layer that defines both a multiscale quadratic risk geometry and a risk-normalized multiscale return signal, enabling online estimation of scale-specific covariances with stability guarantees.
- We propose an investable *reduced-order* parameterization that maps reduced weights on a low-dimensional simplex into N-asset long-only portfolios via a time-varying PCA basis and an entropy-regularized softmax map, along with perturbation and approximation-control interpretations.
- We derive a simplex-preserving *mirror-flow* update for the reduced weights with a KL-proximal characterization and an ODE-consistent replicator-type limit, providing a principled dynamical-system formulation of allocation.
- We establish explicit *turnover control* that transfers from reduced allocations to traded allocations and quantify additional turnover induced by periodic basis refresh.
- We validate the theoretical design parameters empirically on real equity data under transaction costs, including sensitivity analyses, regime robustness, and ablations that isolate the effect of the wavelet layer, PCA reduction, mirror dynamics, and inertia.

Outline. This study proposes a reduced-order multiscale portfolio optimization methodology in which long-only allocations evolve through a simplex-preserving mirror-flow driven by a wavelet-based multiscale signal and multiscale covariance geometry. We first formalize the multiscale construction and its induced quadratic risk functional and clarify how the same multiscale layer generates a risk-normalized return signal. We then introduce an investable reduced-order representation obtained by mapping a small set of latent allocation weights through a time-varying PCA-softmax basis and establish the stability properties of this basis under perturbations of the underlying geometry. Next, we present the resulting mirror-flow update rule in reduced coordinates, including its KL-proximal interpretation and its consistency with a replicator-type ODE that explains the qualitative behavior of the portfolio paths. Building on these components, we derive explicit turnover control statements that transfer from reduced allocations to asset allocations and quantify the additional turnover induced by periodic refresh. We conclude with an empirical study on real data under transaction costs, complemented by sensitivity analyses of the key tuning parameters and operator choices, and ablations designed to isolate the contribution of each mathematical component of the framework.

Paper structure. The remainder of the paper is organized as follows. Section 2 introduces the market setting, portfolio constraints, and the reduced parameterization used throughout the study. Section 3 develops the wavelet-based multiscale risk geometry and the corresponding risk-normalized signal construction. Section 4 presents the PCA-based reduction and the entropy-regularized softmax mapping that produces investable long-only basis portfolios as well as with stability results for the resulting basis map. Sections 5–8 develop the reduced objective and mirror-flow dynamics and establish the main theoretical properties of the proposed framework, including turnover control, a regret-type control result, and a closing discussion of the wavelet–PCA coupling. Section 9 reports the empirical evaluation, including baseline comparisons, sensitivity analyses, regime robustness, computational scaling, and ablation studies. Section 10 concludes and outlines directions for future work.

2. Market model, constraints, and reduced parameterization

Notation

To avoid ambiguity and to keep the exposition self-contained, Tables 1–8 collect and define the mathematical notation used throughout the study, including the theoretical development and Algorithm 1. All quantities not explicitly defined in the main text should be interpreted in accordance with these tables. Unless stated otherwise, vectors are treated as column vectors, vector inequalities are understood componentwise, and for vector arguments, the functions $\exp(\cdot)$ and $\log(\cdot)$ are applied componentwise; the symbol $\odot$ denotes the componentwise (Hadamard) product. Long-only full investment constraints are enforced by restricting the portfolio weights to the simplex Δ_n defined in (2.1).

$$\Delta_n = \{\boldsymbol{x} \in \mathbb{R}^n : \ \boldsymbol{x} \geq 0, \ \mathbb{1}^\top \boldsymbol{x} = 1\}. \tag{2.1}$$

Table 1. Notation: Indices, dimensions, and sets.

Symbol	Meaning / role
$t = 1, \ldots, T$	Discrete time index; T is the final observation time.
$i = 1, \ldots, N$	Asset index; N is the number of assets in the portfolio.
$j = 1, \ldots, J$	Wavelet scale (level) index; J is the number of scales.
$k = 1, \ldots, m$	Reduced component index; $m \in \{3, 4, 5\}$ is the reduced dimension.
ℓ, t_ℓ, K	Basis refresh index, refresh times, and refresh period (basis held constant between refreshes).
$\mathbb{R}^n$, $\Delta_n = \Delta_n$	Euclidean space and probability simplex.

Table 2. Notation: Market data and portfolio variables.

Symbol	Meaning / role
$P_{i,t} > 0$	Price of asset i at time t.
$r_{i,t}$, $\boldsymbol{r}_t \in \mathbb{R}^N$	Log-return and return vector: $r_{i,t} = \log P_{i,t} - \log P_{i,t-1}$.
$\boldsymbol{r}_{i,t}^{(L)} \in \mathbb{R}^L$	Rolling length-L return window for asset i ending at t.
$\boldsymbol{w}_t \in \Delta_N$	Traded (asset-level) portfolio weights at time t.
$\boldsymbol{u}_t \in \Delta_m$	Reduced (latent/basis) weights at time t; only m dynamic state variables.
$A_t \in \mathbb{R}^{N \times m}$	Basis matrix with columns $\boldsymbol{a}_k(t) \in \Delta_N$; portfolio map $\boldsymbol{w}_t = A_t \boldsymbol{u}_t$.

Table 3. Notation: Wavelet operators and multiscale features.

Symbol	Meaning / role
ψ	Chosen wavelet family (e.g., `db4`, `sym4`, `coif1`) for the stationary wavelet transform (SWT) instantiation.
$\mathcal{D}_j$, $\mathcal{S}$	Linear detail and smooth operators (abstract); instantiated via SWT in Remark 3.
$\boldsymbol{c}_{i,j,t}^A$, $\boldsymbol{c}_{i,j,t}^D$	SWT approximation/detail coefficient vectors (asset i, level j, window ending at t).
$W_{i,j,t}$, $\boldsymbol{W}_{j,t} \in \mathbb{R}^N$	Scalar detail feature (asset i) and stacked detail vector at level j.
$s_{i,t}$, $\boldsymbol{s}_t \in \mathbb{R}^N$	Smooth/trend feature (asset i) and stacked smooth vector.

Table 4. Notation: Multiscale covariance geometry.

Symbol	Meaning / role
$\Sigma_{j,t} \in \mathbb{R}^{N \times N}$	Scale-specific covariance estimator updated by an exponentially weighted moving average (EWMA). $\Sigma_{j,t} = (1 - \alpha)\Sigma_{j,t-1} + \alpha \boldsymbol{W}_{j,t} \boldsymbol{W}_{j,t}^\top + \varepsilon I_N$.
$\boldsymbol{\omega} \in \Delta_J$	Non-negative multiscale aggregation weights.
Σ_t, $\bar{\Sigma}_t$	Multiscale covariance $\Sigma_t = \sum_{j=1}^J \omega_j \Sigma_{j,t}$ and its stabilized version $\bar{\Sigma}_t = \Sigma_t + \varepsilon_0 I_N$.
$\mathcal{R}_t(\boldsymbol{w})$	Multiscale quadratic risk functional $\mathcal{R}_t(\boldsymbol{w}) = \boldsymbol{w}^\top \Sigma_t \boldsymbol{w}$.
α, ε, ε_0	EWMA rate and small ridge terms ensuring numerical stability/positive definiteness.

Table 5. Notation: Multiscale signal.

Symbol	Meaning / role
z_t	Raw multiscale score: $z_t = s_t + \lambda_d \sum_{j=1}^{J} d_j W_{j,t}$.
$d \in \Delta_J$, λ_d	Detail weights and detail mixing coefficients.
$\tilde{\mu}_t$	Risk-normalized score $\tilde{\mu}_t = \mathrm{diag}(\Sigma_t)^{-1/2} z_t$.
μ_t	Cross-sectionally demeaned and scaled signal $\mu_t = \kappa_\mu(\tilde{\mu}_t - \frac{1}{N}(\mathbb{1}^\top \tilde{\mu}_t)\mathbb{1})$.
$\widehat{\mu}_t$	EWMA-filtered effective signal $\widehat{\mu}_t = (1-\beta)\widehat{\mu}_{t-1} + \beta \mu_t$.
κ_μ, β	Signal scale and signal-smoothing rate.

Table 6. Notation: PCA basis construction.

Symbol	Meaning / role
S_t	Symmetric geometry matrix used for PCA (in experiments, $S_t := \bar{\Sigma}_t$).
$\lambda_k(t)$, $v_k(t)$	Eigenvalues and orthonormal eigenvectors of S_t (ordered $\lambda_1(t) \geq \cdots \geq \lambda_N(t) \geq 0$).
τ, softmax$(\cdot)$	Softmax temperature and simplex map $a_k(t) = \mathrm{softmax}(\tau v_k(t))$ (with sign convention in Remark 7).
$\delta_k(t)$, $\delta_*(t)$	Spectral gap at mode k and $\delta_*(t) = \min_{1 \leq k \leq m} \delta_k(t)$ (controls basis stability under perturbations).

Table 7. Notation: Reduced objective and mirror-flow dynamics.

Symbol	Meaning / role
$\mu_t^{(u)}$	Reduced signal $\mu_t^{(u)} = A_t^\top \widehat{\mu}_t \in \mathbb{R}^m$.
$\Sigma_t^{(u)}$, $\bar{\Sigma}_t^{(u)}$	Reduced covariance $\Sigma_t^{(u)} = A_t^\top \Sigma_t A_t$ and stabilized version $\bar{\Sigma}_t^{(u)} = A_t^\top \bar{\Sigma}_t A_t + \varepsilon_u I_m$.
$J_t(u)$	Reduced objective $J_t(u) = (\mu_t^{(u)})^\top u - \frac{\gamma}{2} u^\top \bar{\Sigma}_t^{(u)} u + \rho H(u)$.
$H(u)$	Shannon entropy $H(u) = -\sum_{k=1}^{m} u_k \log u_k$.
g_t	Gradient $g_t = \nabla J_t(u_t) = \mu_t^{(u)} - \gamma \bar{\Sigma}_t^{(u)} u_t - \rho(\log u_t + \mathbb{1})$.
u_{t+1}^{tar}	Mirror target iterate: $u_{t+1}^{\mathrm{tar}} \propto u_t \odot \exp(\eta\, g_t)$.
η, λ	Mirror step size and inertia (mixing) parameter: $u_{t+1} = \lambda u_t + (1-\lambda) u_{t+1}^{\mathrm{tar}}$.
γ, ρ	Risk aversion (quadratic term) and entropy regularization weight.
ε_u	Reduced-space ridge ensuring numerical stability / positive definiteness.

Table 8. Notation: Trading frictions, norms, and auxiliary operators.

Symbol	Meaning / role
TO_{t+1}	One-way turnover $\mathrm{TO}_{t+1} = \frac{1}{2} \lVert w_{t+1} - w_t \rVert_1$.
c, R_{t+1}^{gross}, R_{t+1}^{net}	Transaction-cost rate and (optional) net-return model: $R_{t+1}^{\mathrm{net}} = R_{t+1}^{\mathrm{gross}} - c\,\mathrm{TO}_{t+1}$.
$\lVert\cdot\rVert_1$, $\lVert\cdot\rVert_2$, $\lVert\cdot\rVert_\infty$	ℓ_1, ℓ_2, and ℓ_∞ norms; $\lVert\cdot\rVert_{1\to 1}$ denotes the induced operator norm.
diag$(\cdot)$, Tr$(\cdot)$, $\langle\cdot,\cdot\rangle$	Diagonal operator, trace, and Euclidean inner product.
$D_{\mathrm{KL}}(\cdot\Vert\cdot)$	Kullback–Leibler divergence (mirror map/entropy geometry).
I_N, I_m, $\mathbb{1}$	Identity matrices and all-ones vector.
B, M	Boundedness constants used in the EWMA drift bound corollary (for example $\lVert W_{j,t} \rVert_2 \leq B$ and $\lVert \Sigma_{j,t} \rVert_2 \leq M$).

We consider a market of $N \geq 2$ assets observed over discrete times $t = 1, 2, \ldots, T$. Let $P_{i,t} > 0$ denote the price of asset i at time t, and define log-returns

$$r_{i,t} = \log P_{i,t} - \log P_{i,t-1}, \qquad \boldsymbol{r}_t = (r_{1,t}, \ldots, r_{N,t})^\top \in \mathbb{R}^N. \tag{2.2}$$

A long-only, fully invested portfolio is a weight vector $\boldsymbol{w} \in \Delta_N$ where

$$\Delta_N := \{\boldsymbol{w} \in \mathbb{R}^N : \boldsymbol{w} \geq 0,\ \mathbb{1}^\top \boldsymbol{w} = 1\}. \tag{2.3}$$

The central modeling choice is to restrict the portfolio process to a time-varying, low-dimensional convex set inside Δ_N. Fix an integer $m \in \{3, 4, 5\}$ and define a basis matrix $A_t \in \mathbb{R}^{N\times m}$ whose columns are basis portfolios $\boldsymbol{a}_k(t) \in \Delta_N$:

$$A_t = [\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_m(t)], \qquad \boldsymbol{a}_k(t) \in \Delta_N. \tag{2.4}$$

The reduced weights are $\boldsymbol{u}_t \in \Delta_m$, and the investable portfolio is

$$\boldsymbol{w}_t = A_t \boldsymbol{u}_t. \tag{2.5}$$

Remark 1 (On the reduced dimension m). The integer m controls the number of *dynamic decision variables* and the size of the investable reduced set $\{\boldsymbol{w} = A_t\boldsymbol{u} : \boldsymbol{u} \in \Delta_m\} \subset \Delta_N$. Smaller m improves interpretability and reduces estimation and trading variability, but it also restricts the expressiveness of the feasible set. Section 4.4 makes this trade-off explicit by linking approximation quality to the spectral tail of the multiscale covariance geometry and stability to eigengap dependent perturbation bounds.

Remark 2 (Discrete basis refresh schedule). In the empirical section, the basis is refreshed on a discrete grid to reduce estimation noise and make turnover effects interpretable. Fix an integer refresh period $K \geq 1$ and define refresh times $\{t_\ell\}_{\ell\geq 0}$ with $t_{\ell+1} - t_\ell = K$. We compute A_{t_ℓ} from the current geometry (via PCA and (4.1)) and hold it fixed between refresh times as follows:

$$A_t := A_{t_\ell} \quad \text{for } t \in [t_\ell, t_{\ell+1}).$$

All feasibility and contraction statements are valid. In particular, explicit turnover bounds such as Theorem 5 apply to any interval where $A_{t+1} = A_t$.

Because each column of A_t lies in Δ_N, and $\boldsymbol{u}_t$ is a convex combination, $\boldsymbol{w}_t \in \Delta_N$ automatically follows. The remainder of the formulation is devoted to constructing (i) a multiscale geometry and signal through wavelets and (ii) a basis A_t through the PCA of the multiscale geometry. The portfolio decision $\boldsymbol{u}_t$ is then updated by a mirror-flow that is stable on Δ_m.

3. Wavelet multiscale geometry and multiscale signal

The wavelet layer is introduced into the model as a multiscale function. To avoid dependence on a particular implementation, we describe the transformation in the operator form.

3.1. Wavelet feature operators

Fix a window length $L \in \mathbb{N}$ and number of scales $J \in \mathbb{N}$. For each asset i and time $t \geq L$, define the return window vector

$$\boldsymbol{r}_{i,t}^{(L)} := (r_{i,t-L+1}, \ldots, r_{i,t})^{\top} \in \mathbb{R}^L. \tag{3.1}$$

For each scale $j \in \{1, \ldots, J\}$, let $\mathcal{D}_j : \mathbb{R}^L \to \mathbb{R}$ denote a linear detail coefficient functional and $\mathcal{S} : \mathbb{R}^L \to \mathbb{R}$ denote a linear trend functional. The scale-j detail feature and the trend feature for asset i at time t are defined by

$$W_{i,j,t} := \mathcal{D}_j(\boldsymbol{r}_{i,t}^{(L)}), \qquad s_{i,t} := \mathcal{S}(\boldsymbol{r}_{i,t}^{(L)}). \tag{3.2}$$

Remark 3 (SWT instantiation used in the empirical section). Although the analysis only assumes linear maps $\mathcal{D}_j$ and $\mathcal{S}$, the numerical experiments instantiate them via the *stationary wavelet transform* (SWT) with a chosen wavelet family ψ. Given the return window $\boldsymbol{r}_{i,t}^{(L)}$, let $\{(\boldsymbol{c}_{i,j,t}^{A}, \boldsymbol{c}_{i,j,t}^{D})\}_{j=1}^{J}$ denote the level-J SWT approximation/detail coefficient vectors, where $\boldsymbol{c}_{i,j,t}^{A}, \boldsymbol{c}_{i,j,t}^{D} \in \mathbb{R}^L$. We set

$$W_{i,j,t} := (\boldsymbol{c}_{i,j,t}^{D})_L, \qquad s_{i,t} := (\widehat{\boldsymbol{r}}_{i,t}^{A})_L,$$

where $\widehat{\boldsymbol{r}}_{i,t}^{A} \in \mathbb{R}^L$ is the SWT reconstruction using only the approximation coefficients (all details set to zero). This choice preserves the linearity required by Eq (3.2). For the stationary wavelet transform (SWT), we require L to be a multiple of 2^J.

Collect features across assets as follows:

$$\boldsymbol{W}_{j,t} := (W_{1,j,t}, \ldots, W_{N,j,t})^{\top} \in \mathbb{R}^N, \qquad \boldsymbol{s}_t := (s_{1,t}, \ldots, s_{N,t})^{\top} \in \mathbb{R}^N. \tag{3.3}$$

In applications, $\mathcal{D}_j$ and $\mathcal{S}$ may correspond to SWT detail coefficient and a smooth reconstruction; however, the analysis below only uses linearity and boundedness.

3.2. Multiscale covariance recursion and multiscale risk

For each scale j, we estimate a scale-specific covariance using an exponentially weighted recursion. Fix $\alpha \in (0, 1)$ and $\varepsilon > 0$, and define

$$\Sigma_{j,t} = (1 - \alpha)\Sigma_{j,t-1} + \alpha\, \boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^{\top} + \varepsilon I_N, \qquad j = 1, \ldots, J. \tag{3.4}$$

Let $\boldsymbol{\omega} \in \Delta_J$ be nonnegative multiscale weights, and define the multiscale covariance matrix

$$\Sigma_t := \sum_{j=1}^{J} \omega_j \Sigma_{j,t}. \tag{3.5}$$

Remark 4 (Stabilized multiscale covariance used in computation). In finite precision arithmetic, it is convenient to work with a strictly positive definite surrogate. Accordingly, in the empirical section, we use

$$\bar{\Sigma}_t := \Sigma_t + \varepsilon_0 I_N, \tag{3.6}$$

where $\varepsilon_0 > 0$ is small. All statements below remain valid with $\bar{\Sigma}_t$ in place of Σ_t; setting $\varepsilon_0 = 0$ recovers (3.5).

This induces a multiscale quadratic risk functional on Δ_N:

$$\mathcal{R}_t(\boldsymbol{w}) := \boldsymbol{w}^\top \Sigma_t \boldsymbol{w} = \sum_{j=1}^{J} \omega_j \, \boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w}. \tag{3.7}$$

The decomposition in (3.7) is a modeling statement: risk is aggregated across wavelet scales rather than extracted from a single unconditional covariance.

Provenance (standard well-posedness of the multiscale covariance recursion). The next statement is an elementary consequence of the exponentially weighted covariance recursion (3.4) with a positive ridge term. We include it to guarantee that the multiscale quadratic form (3.7) is strictly positive definite at all times, and therefore well-defined on the simplex.

Theorem 1 (Strict positive definiteness)**.** *Assume $\Sigma_{j,0} \succeq 0$ for $j = 1, \dots, J$, and let $\alpha \in (0, 1)$ and $\varepsilon > 0$ in* (3.4)*. Then, $\Sigma_{j,t} \succ 0$ for all j and all $t \geq 1$. Consequently, $\Sigma_t \succ 0$ for all $t \geq 1$ and all $\boldsymbol{\omega} \in \Delta_J$.*

Proof. Fix j and $t \geq 1$. Because $\boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^\top \succeq 0$, and $\Sigma_{j,t-1} \succeq 0$, we have $(1-\alpha)\Sigma_{j,t-1} + \alpha \boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^\top \succeq 0$. Adding $\varepsilon I_N \succ 0$ yields $\Sigma_{j,t} \succ 0$. For Σ_t, it is a convex combination of positive definite matrices and is therefore positive definite. □

Proposition 1 (Multiscale risk as exponentially weighted wavelet energy)**.** *Fix a scale $j \in \{1, \dots, J\}$, and let $(\Sigma_{j,t})_{t \geq 0}$ evolve according to* (3.4)*. For any portfolio vector $\boldsymbol{w} \in \mathbb{R}^N$, define the (scalar) wavelet coefficient of the portfolio at scale j and time t by*

$$\xi_{j,t}(\boldsymbol{w}) := \boldsymbol{w}^\top \boldsymbol{W}_{j,t}.$$

Then, for all $t \geq 1$,

$$\boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w} = (1-\alpha)^t \boldsymbol{w}^\top \Sigma_{j,0} \boldsymbol{w} + \alpha \sum_{s=1}^{t} (1-\alpha)^{t-s} \, \xi_{j,s}(\boldsymbol{w})^2 + \varepsilon \, c_t \, \|\boldsymbol{w}\|_2^2, \tag{3.8}$$

where $c_t := \sum_{\ell=0}^{t-1} (1-\alpha)^\ell = (1 - (1-\alpha)^t)/\alpha$. Consequently, the multiscale risk (3.7) *admits the explicit representation*

$$\begin{aligned} \mathcal{R}_t(\boldsymbol{w}) &= \sum_{j=1}^{J} \omega_j \, \boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w} \\ &= \sum_{j=1}^{J} \omega_j \Big[(1-\alpha)^t \boldsymbol{w}^\top \Sigma_{j,0} \boldsymbol{w} + \alpha \sum_{s=1}^{t} (1-\alpha)^{t-s} \xi_{j,s}(\boldsymbol{w})^2 + \varepsilon \, c_t \, \|\boldsymbol{w}\|_2^2 \Big]. \end{aligned}$$

In particular, if $\boldsymbol{w} \in \Delta_N$, then $\|\boldsymbol{w}\|_2^2 \leq 1$, and the ridge contribution is at most εc_t.

Proof. Unrolling the recursion (3.4) yields

$$\Sigma_{j,t} = (1-\alpha)^t \Sigma_{j,0} + \alpha \sum_{s=1}^{t} (1-\alpha)^{t-s} \boldsymbol{W}_{j,s}\boldsymbol{W}_{j,s}^\top + \varepsilon \sum_{\ell=0}^{t-1} (1-\alpha)^\ell I_N.$$

Left- and right-multiplying by $\boldsymbol{w}$ gives

$$\boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w} = (1-\alpha)^t \boldsymbol{w}^\top \Sigma_{j,0} \boldsymbol{w} + \alpha \sum_{s=1}^{t} (1-\alpha)^{t-s} \boldsymbol{w}^\top (\boldsymbol{W}_{j,s} \boldsymbol{W}_{j,s}^\top) \boldsymbol{w} + \varepsilon c_t \|\boldsymbol{w}\|_2^2 .$$

Finally, $\boldsymbol{w}^\top (\boldsymbol{W}_{j,s} \boldsymbol{W}_{j,s}^\top) \boldsymbol{w} = (\boldsymbol{w}^\top \boldsymbol{W}_{j,s})^2 = \xi_{j,s}(\boldsymbol{w})^2$, which gives (3.8). Summing across j with weights ω_j yields the stated multiscale identity. □

Remark 5 (Economic interpretation of multiscale quadratic risk). The risk functional $\mathcal{R}_t(\boldsymbol{w})$ in (3.7) can be interpreted as a horizon-decomposed variance. For each scale j, the scalar $\xi_{j,t}(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{W}_{j,t}$ is the portfolio's wavelet-detail coefficient at that scale. Proposition 1 shows that $\boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w}$ is an exponentially weighted average of $\xi_{j,s}(\boldsymbol{w})^2$, that is, a smoothed estimate of realized variability at the corresponding horizon. Hence, $\mathcal{R}_t(\boldsymbol{w}) = \sum_{j=1}^{J} \omega_j \boldsymbol{w}^\top \Sigma_{j,t} \boldsymbol{w}$ penalizes *simultaneously* the portfolio's short-horizon fluctuations (small j) and its medium/long-horizon fluctuations (large j), with the weights $\boldsymbol{\omega}$ encoding the investor's horizon emphasis. This offers a simple economic reading: rather than committing to a single sampling horizon, the geometry discourages portfolios whose risk is concentrated at one scale and supports allocations that are stable across time scales.

3.3. Risk-normalized multiscale signal

The same wavelet features also define a time-varying cross-sectional signal. Let $\boldsymbol{d} \in \Delta_J$ be the detail weights, and let $\lambda_d \geq 0$ control the contribution of details. Define a raw multiscale score

$$\boldsymbol{z}_t := \boldsymbol{s}_t + \lambda_d \sum_{j=1}^{J} d_j \boldsymbol{W}_{j,t}. \tag{3.9}$$

To prevent purely volatility-driven allocations, we normalize by the multiscale diagonal volatility and remove the cross-sectional mean. Fix $\kappa_\mu > 0$, and define

$$\tilde{\boldsymbol{\mu}}_t := \operatorname{diag}(\Sigma_t)^{-1/2} \boldsymbol{z}_t, \qquad \boldsymbol{\mu}_t := \kappa_\mu \Big(\tilde{\boldsymbol{\mu}}_t - \tfrac{1}{N} \mathbb{1}\mathbb{1}^\top \tilde{\boldsymbol{\mu}}_t \Big). \tag{3.10}$$

Remark 6 (Filtered signal used by the trading rule). For additional temporal stability, the empirical implementation uses exponentially weighted smoothing of the instantaneous score. Fix $\beta \in (0, 1]$, and define the filtered signal

$$\widehat{\boldsymbol{\mu}}_t := (1-\beta) \widehat{\boldsymbol{\mu}}_{t-1} + \beta \boldsymbol{\mu}_t, \tag{3.11}$$

initialized at the first tradable time t_0 by $\widehat{\boldsymbol{\mu}}_{t_0} := \boldsymbol{\mu}_{t_0}$. Because the analysis treats the sequence $(\boldsymbol{\mu}_t, \Sigma_t)$ as exogenous inputs, all results apply verbatim with $\widehat{\boldsymbol{\mu}}_t$ in place of $\boldsymbol{\mu}_t$. For notational simplicity, we continue to write $\boldsymbol{\mu}_t$ for the effective (possibly filtered) signal.

Therefore, $\mathbb{1}^\top \boldsymbol{\mu}_t = 0$. The signal $\boldsymbol{\mu}_t$ is interpreted as a score rather than a literal conditional expectation; it is designed to be time-varying and scale-aware.

4. PCA reduction and entropy-regularized basis portfolios

The multiscale covariance Σ_t defines the time-dependent geometry. We use this geometry to build a low-dimensional convex set of admissible portfolios, and we do so in a way that remains compatible with simplex constraints.

4.1. PCA directions of the multiscale geometry

Let $\bar{\Sigma}_t$ denote the stabilized multiscale covariance from Remark 4 (and, if desired, a smoothed version). Let $(\lambda_1(t), \boldsymbol{v}_1(t)), \ldots, (\lambda_N(t), \boldsymbol{v}_N(t))$ be eigenpairs of $\bar{\Sigma}_t$ with $\lambda_1(t) \geq \cdots \geq \lambda_N(t) \geq 0$ and orthonormal eigenvectors. We retain the first m directions $\boldsymbol{v}_1(t), \ldots, \boldsymbol{v}_m(t)$.

4.2. Softmax basis and its variational characterization

The PCA eigenvectors are sign-indefinite and do not satisfy the simplex constraints. We map each eigenvector to a long-only basis portfolio through a softmax transformation. Fix $\tau > 0$, and define

$$\boldsymbol{a}_k(t) := \mathrm{softmax}(\tau \boldsymbol{v}_k(t)) \in \Delta_N, \qquad k = 1, \ldots, m. \tag{4.1}$$

Remark 7 (Eigenvector orientation and temporal alignment). Eigenvectors are defined only up to a sign (and, when eigenvalues are nearly repeated, the associated directions are only identifiable up to rotations of the corresponding eigenspace). Accordingly, the stability statements below are understood under the eigengap condition in Lemma 2; when eigenvalues are clustered, the identifiable object is the eigenspace itself, and one may instead align the whole subspace (e.g., via an orthogonal Procrustes step), which we leave for future work. Because the softmax map $\boldsymbol{v} \mapsto \mathrm{softmax}(\tau \boldsymbol{v})$ is not sign-invariant, we impose a deterministic orientation rule.

At the first time a basis is computed (for instance, at the initial refresh time t_0), we choose the sign of each $\boldsymbol{v}_k(t_0)$ by requiring that its component with the largest absolute value is positive. At later refresh times, we *align* the new eigenvector with its previously oriented counterpart by maximal correlation:

$$\boldsymbol{v}_k^{\mathrm{al}}(t) := \sigma_k(t)\, \boldsymbol{v}_k(t), \qquad \sigma_k(t) := \mathrm{sign}\big(\big\langle \boldsymbol{v}_k^{\mathrm{al}}(t^-), \boldsymbol{v}_k(t) \big\rangle \big), \tag{4.2}$$

where t^- denotes the previous refresh time so that $\boldsymbol{v}_k^{\mathrm{al}}(t^-)$ is the last oriented eigenvector carried forward between refreshes and we adopt the convention $\mathrm{sign}(0) = +1$. This correlation-based rule eliminates artificial sign flips and is stable under perturbations whenever the eigengap condition ensures that consecutive eigenvectors remain acute; see Lemma 3.

We then obtain

$$(\mathrm{softmax}(\boldsymbol{x}))_i := \frac{e^{x_i}}{\sum_{\ell=1}^{N} e^{x_\ell}}. \tag{4.3}$$

This mapping ensures that every entry of $\boldsymbol{a}_k(t)$ is strictly positive, so the basis portfolios remain in the interior of the simplex.

To emphasize that (4.1) is not an arbitrary heuristic, we state its variational characterization.

Proposition 2 (Variational meaning of the softmax basis)**.** *Let $\boldsymbol{v} \in \mathbb{R}^N$ and $\tau > 0$. Define Shannon entropy $H(\boldsymbol{a}) := -\sum_{i=1}^{N} a_i \log a_i$ on Δ_N.Then,*

$$\mathrm{softmax}(\tau \boldsymbol{v}) = \arg\max_{\boldsymbol{a} \in \Delta_N} \big\{ \tau\, \boldsymbol{v}^\top \boldsymbol{a} + H(\boldsymbol{a}) \big\}. \tag{4.4}$$

Proof. The function $\boldsymbol{a} \mapsto \tau \boldsymbol{v}^\top \boldsymbol{a} + H(\boldsymbol{a})$ is strictly concave on the compact convex set Δ_N; therefore, a unique maximizer exists. Using Lagrange multipliers for the constraint $\sum_i a_i = 1$ and interior optimality (guaranteed by strict concavity and the barrier-like behavior of entropy near the boundary), we obtain

$$\tau v_i - (\log a_i + 1) + \lambda = 0 \quad \Rightarrow \quad a_i = C e^{\tau v_i}.$$

Normalization gives $C^{-1} = \sum_\ell e^{\tau v_\ell}$; hence, $a_i = e^{\tau v_i} / \sum_\ell e^{\tau v_\ell}$. □

4.3. Stability of the basis map

Because the basis is rebuilt from an estimated covariance, it is important to quantify how perturbations in $\bar{\Sigma}_t$ propagate to perturbations in $\boldsymbol{a}_k(t)$. We begin with a Lipschitz bound for the softmax function.

Lemma 1 (Softmax is globally Lipschitz)**.** *For any $\tau > 0$ and any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$,*

$$\|\mathrm{softmax}(\tau \boldsymbol{x}) - \mathrm{softmax}(\tau \boldsymbol{y})\|_2 \leq \frac{\tau}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2 . \tag{4.5}$$

Consequently,

$$\|\mathrm{softmax}(\tau \boldsymbol{x}) - \mathrm{softmax}(\tau \boldsymbol{y})\|_1 \leq \frac{\tau \sqrt{N}}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2 . \tag{4.6}$$

Proof. Let $\boldsymbol{p}(\boldsymbol{x}) := \mathrm{softmax}(\boldsymbol{x})$. A direct calculation shows that the Jacobian is

$$\nabla \boldsymbol{p}(\boldsymbol{x}) = \mathrm{diag}(\boldsymbol{p}(\boldsymbol{x})) - \boldsymbol{p}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x})^\top,$$

which is symmetric and positive semidefinite. For any $\boldsymbol{z}$ with $\|\boldsymbol{z}\|_2 = 1$,

$$\boldsymbol{z}^\top \nabla \boldsymbol{p}(\boldsymbol{x}) \boldsymbol{z} = \sum_{i=1}^N p_i z_i^2 - \Big(\sum_{i=1}^N p_i z_i \Big)^2 = \mathrm{Var}(Z),$$

where Z is a discrete random variable taking values $\{z_i\}$ with probabilities $\{p_i\}$. By Popoviciu's inequality, $\mathrm{Var}(Z) \leq (M - m)^2/4$, where $M = \max_i z_i$, and $m = \min_i z_i$. Moreover, $M - m = M + |m| \leq \sqrt{2}\sqrt{M^2 + m^2} \leq \sqrt{2}\|\boldsymbol{z}\|_2 = \sqrt{2}$. Hence, $\mathrm{Var}(Z) \leq (\sqrt{2})^2/4 = 1/2$, which implies $\|\nabla \boldsymbol{p}(\boldsymbol{x})\|_{2\to 2} \leq 1/2$. By the mean value theorem applied to the map $\boldsymbol{x} \mapsto \boldsymbol{p}(\tau \boldsymbol{x})$,

$$\|\boldsymbol{p}(\tau \boldsymbol{x}) - \boldsymbol{p}(\tau \boldsymbol{y})\|_2 \leq \sup_{\xi} \|\tau \nabla \boldsymbol{p}(\tau \xi)\|_{2\to 2} \, \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \frac{\tau}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2 .$$

The ℓ_1 bound follows from $\|\boldsymbol{q}\|_1 \leq \sqrt{N}\|\boldsymbol{q}\|_2$. □

We now link perturbations of $\bar{\Sigma}_t$ to perturbations of eigenvectors, and thus to perturbations of the softmax basis portfolios. To keep the proof self-contained, we state a simple eigenvector perturbation bound under the spectral gap condition.

Lemma 2 (A Davis–Kahan-type eigenvector perturbation bound)**.** *Let S and $\widetilde{S}$ be symmetric matrices in $\mathbb{R}^{N\times N}$. Let $(\lambda, \boldsymbol{v})$ be a unit eigenpair of S with a simple eigenvalue λ, and let $(\widetilde{\lambda}, \widetilde{\boldsymbol{v}})$ be a unit eigenpair of $\widetilde{S}$. Define $\Delta := \widetilde{S} - S$, and let the spectral gap of λ be*

$$\delta := \min_{\mu \in \mathrm{spec}(S)\setminus\{\lambda\}} |\lambda - \mu|.$$

Assume $\|\Delta\|_2 \leq \delta/2$. Then, there exists a sign $\sigma \in \{+1, -1\}$ such that

$$\big\| \boldsymbol{v} - \sigma \widetilde{\boldsymbol{v}} \big\|_2 \leq \frac{4}{\delta} \|\Delta\|_2 . \tag{4.7}$$

Proof. Let $\boldsymbol{v}$ be the eigenvector of S associated with λ, and decompose $\widetilde{\boldsymbol{v}}$ as

$$\widetilde{\boldsymbol{v}} = c\,\boldsymbol{v} + \boldsymbol{s}, \qquad c := \boldsymbol{v}^\top \widetilde{\boldsymbol{v}}, \qquad \boldsymbol{s} \perp \boldsymbol{v}.$$

Then, $\|\boldsymbol{s}\|_2 = \sqrt{1-c^2}$ is the sine of the angle between $\boldsymbol{v}$ and $\widetilde{\boldsymbol{v}}$. From $(S+\Delta)\widetilde{\boldsymbol{v}} = \widetilde{\lambda}\widetilde{\boldsymbol{v}}$, we obtain

$$(S - \widetilde{\lambda} I)\widetilde{\boldsymbol{v}} = -\Delta\widetilde{\boldsymbol{v}}.$$

Projecting onto the orthogonal complement of $\boldsymbol{v}$ yields

$$(I - \boldsymbol{v}\boldsymbol{v}^\top)(S - \widetilde{\lambda} I)\boldsymbol{s} = -(I - \boldsymbol{v}\boldsymbol{v}^\top)\Delta\widetilde{\boldsymbol{v}}.$$

On the subspace $\boldsymbol{v}^\perp$, the operator $S - \widetilde{\lambda} I$ has eigenvalues $\mu - \widetilde{\lambda}$, where $\mu \in \mathrm{spec}(S) \setminus \{\lambda\}$. By Weyl's inequality, $|\widetilde{\lambda} - \lambda| \le \|\Delta\|_2$. Hence, for $\mu \neq \lambda$,

$$|\mu - \widetilde{\lambda}| \ge |\mu - \lambda| - |\widetilde{\lambda} - \lambda| \ge \delta - \|\Delta\|_2 \ge \delta/2.$$

Therefore, $(S - \widetilde{\lambda} I)$ is restricted to $\boldsymbol{v}^\perp$ is invertible, and its inverse has an operator norm of at most $2/\delta$. Taking norms,

$$\|\boldsymbol{s}\|_2 \le \left\|(S - \widetilde{\lambda} I)^{-1}\big|_{\boldsymbol{v}^\perp}\right\|_{2\to 2} \|\Delta\|_2 \left\|\widetilde{\boldsymbol{v}}\right\|_2 \le \frac{2}{\delta}\,\|\Delta\|_2 .$$

Finally,

$$\min_{\sigma\in\{\pm 1\}} \left\|\boldsymbol{v} - \sigma\widetilde{\boldsymbol{v}}\right\|_2 = \sqrt{2-2|c|} = 2\sin(\tfrac{1}{2}\arccos|c|) \le 2\sqrt{1-c^2} = 2\,\|\boldsymbol{s}\|_2 \le \frac{4}{\delta}\,\|\Delta\|_2 .$$

□

By combining Lemmas 1 and 2, we obtain a stability bound for the basis portfolios constructed from the perturbed covariance matrices.

Provenance (new stability statement for our basis map). The eigenvector perturbation estimate used below is classical; the substantive point here is to propagate it through the *simplex-valued* softmax map that defines our investable PCA basis. The resulting Lipschitz-type bound is tailored to the composed map

$$S \;\mapsto\; \boldsymbol{v}_k(S) \;\mapsto\; \mathrm{softmax}(\tau\,\boldsymbol{v}_k(S)),$$

together with the sign-alignment rule in Remark 7.

Lemma 3 (Sign alignment by maximal correlation)**.** *Let $\boldsymbol{v}, \widetilde{\boldsymbol{v}} \in \mathbb{R}^N$ be unit vectors, and set*

$$\sigma := \mathrm{sign}(\boldsymbol{v}^\top\widetilde{\boldsymbol{v}}), \qquad \mathrm{sign}(0) := +1.$$

Then, $\sigma\widetilde{\boldsymbol{v}}$ is the closest sign choice to $\boldsymbol{v}$ in Euclidean norm:

$$\left\|\boldsymbol{v} - \sigma\widetilde{\boldsymbol{v}}\right\|_2 = \min_{s\in\{\pm 1\}} \left\|\boldsymbol{v} - s\widetilde{\boldsymbol{v}}\right\|_2 .$$

Moreover, $\sigma = +1$ whenever $\boldsymbol{v}^\top\widetilde{\boldsymbol{v}} > 0$ (equivalently, $\left\|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\right\|_2 < \sqrt{2}$).

Proof. For unit vectors,

$$\left\|\boldsymbol{v} - s\widetilde{\boldsymbol{v}}\right\|_2^2 = 2 - 2s\,\boldsymbol{v}^\top\widetilde{\boldsymbol{v}}, \qquad s \in \{\pm 1\}.$$

The minimizer therefore chooses s with the same sign as $\boldsymbol{v}^\top\widetilde{\boldsymbol{v}}$, and the second claim is immediate. □

Theorem 2 (Stability of the softmax–PCA basis under perturbations)**.** *Let S and $\widetilde{S}$ be symmetric matrices, and let $\boldsymbol{v}_k, \widetilde{\boldsymbol{v}}_k$ be unit eigenvectors associated with simple eigenvalues of S and $\widetilde{S}$, respectively, with gaps δ_k as in Lemma 2. Assume $\left\|\widetilde{S} - S\right\|_2 \le \delta_k/2$. Define the sign-aligned eigenvector*

$$\widetilde{\boldsymbol{v}}_k^{\mathrm{al}} := \sigma_k\,\widetilde{\boldsymbol{v}}_k, \qquad \sigma_k := \operatorname{sign}(\boldsymbol{v}_k^\top\widetilde{\boldsymbol{v}}_k), \quad \operatorname{sign}(0) := +1,$$

and define basis portfolios $\boldsymbol{a}_k = \operatorname{softmax}(\tau\boldsymbol{v}_k)$ and $\widetilde{\boldsymbol{a}}_k = \operatorname{softmax}(\tau\widetilde{\boldsymbol{v}}_k^{\mathrm{al}})$. Then,

$$\left\|\boldsymbol{a}_k - \widetilde{\boldsymbol{a}}_k\right\|_1 \le \frac{2\tau\sqrt{N}}{\delta_k}\left\|\widetilde{S} - S\right\|_2. \tag{4.8}$$

In particular, if eigenvectors are oriented over time via the correlation rule in Remark 7, the bound applies directly to consecutive basis refreshes, thereby excluding spurious "sign-flip" turnover.

Proof. Let σ be the sign that minimizes $\left\|\boldsymbol{v}_k - \sigma\widetilde{\boldsymbol{v}}_k\right\|_2$. By Lemma 3, we may take $\sigma = \operatorname{sign}(\boldsymbol{v}_k^\top\widetilde{\boldsymbol{v}}_k)$; hence, $\sigma\widetilde{\boldsymbol{v}}_k = \widetilde{\boldsymbol{v}}_k^{\mathrm{al}}$. Lemma 1 yields

$$\left\|\operatorname{softmax}(\tau\boldsymbol{v}_k) - \operatorname{softmax}(\tau\widetilde{\boldsymbol{v}}_k^{\mathrm{al}})\right\|_1 \le \frac{\tau\sqrt{N}}{2}\left\|\boldsymbol{v}_k - \widetilde{\boldsymbol{v}}_k^{\mathrm{al}}\right\|_2.$$

Apply Lemma 2 (with $\delta = \delta_k$) to obtain $\left\|\boldsymbol{v}_k - \widetilde{\boldsymbol{v}}_k^{\mathrm{al}}\right\|_2 \le \frac{4}{\delta_k}\left\|\widetilde{S} - S\right\|_2$, which gives (4.8). □

4.4. Choosing the reduced dimension m: an approximation–stability trade-off

The parameter m affects the framework through two distinct mathematical mechanisms. First, it determines how much of the covariance geometry is retained by PCA truncation. Second, it controls the sensitivity of the PCA basis to perturbations, because eigenvectors associated with smaller eigen-gaps are less stable. The following lemma provides a transparent "spectral tail" viewpoint for quadratic risk, which motivates why small values such as $m \in \{3, 4, 5\}$ can be adequate when the tail energy is small.

Lemma 4 (Spectral tail control of quadratic risk)**.** *Let $S \in \mathbb{R}^{N\times N}$ be symmetric, positive, and semidefinite with eigenpairs $(\lambda_i, \boldsymbol{v}_i)_{i=1}^N$ ordered so that $\lambda_1 \ge \cdots \ge \lambda_N \ge 0$. Let $\Pi_m := \sum_{i=1}^m \boldsymbol{v}_i\boldsymbol{v}_i^\top$ be the orthogonal projector onto the top-m eigenspace. Then, for any $\boldsymbol{w} \in \mathbb{R}^N$,*

$$\boldsymbol{w}^\top S\,\boldsymbol{w} = (\Pi_m\boldsymbol{w})^\top S\,(\Pi_m\boldsymbol{w}) + \big((I - \Pi_m)\boldsymbol{w}\big)^\top S\,\big((I - \Pi_m)\boldsymbol{w}\big), \tag{4.9}$$

and the residual term satisfies the bound

$$\big((I - \Pi_m)\boldsymbol{w}\big)^\top S\,\big((I - \Pi_m)\boldsymbol{w}\big) \le \lambda_{m+1}\,\|\boldsymbol{w}\|_2^2. \tag{4.10}$$

In particular, if $\boldsymbol{w} \in \Delta_N$, then $\|\boldsymbol{w}\|_2^2 \le 1$, and the risk contribution outside the top-m directions is at most λ_{m+1}.

Proof. Expand $\boldsymbol{w} = \sum_{i=1}^{N} c_i \boldsymbol{v}_i$ with $c_i = \boldsymbol{v}_i^\top \boldsymbol{w}$. Then, $\boldsymbol{w}^\top S \boldsymbol{w} = \sum_{i=1}^{N} \lambda_i c_i^2$, and $\Pi_m \boldsymbol{w} = \sum_{i=1}^{m} c_i \boldsymbol{v}_i$. Hence, (4.9) holds by separating the sum into $i \le m$ and $i > m$. For (4.10), use $\lambda_i \le \lambda_{m+1}$ for $i > m$ to obtain $\sum_{i>m} \lambda_i c_i^2 \le \lambda_{m+1} \sum_{i>m} c_i^2 \le \lambda_{m+1} \sum_{i=1}^{N} c_i^2 = \lambda_{m+1} \|\boldsymbol{w}\|_2^2$. □

A convenient scalar summary of "how much geometry is discarded" is the *spectral tail ratio*

$$\mathrm{Tail}_m(S) := \frac{\sum_{i=m+1}^{N} \lambda_i}{\sum_{i=1}^{N} \lambda_i} = 1 - \frac{\sum_{i=1}^{m} \lambda_i}{\mathrm{Tr}(S)} \in [0, 1]. \tag{4.11}$$

When $\mathrm{Tail}_m(S)$ is small, most variance energy of S is captured by the leading m directions, and the reduced geometry can be expected to approximate the full geometry well. This motivates working with small m once the spectral tail exhibits diminishing returns.

On the other hand, stability bounds for the PCA–softmax basis and for basis-refresh turnover depend on eigengaps. In particular, Theorem 6 involves $\delta_*(t) = \min_{1 \le k \le m} \delta_k(t)$, and the constant $\tau \sqrt{N}/\delta_*(t)$ typically *increases* with m because higher-order eigenvalues are often more tightly clustered. Thus, increasing m improves approximation (smaller tail) but can worsen stability and increase turnover through a smaller eigengaps. The restricted range $m \in \{3, 4, 5\}$ used in the empirical section reflects this mathematical trade-off: it stays in a low-dimensional regime where the tail energy is already small for many equity universes while avoiding fragile eigendirections with near-degenerate eigenvalues.

4.5. Implementation considerations for repeated PCA updates

Concerns naturally include the numerical stability of repeatedly computing eigendecompositions of estimated covariances. The framework already incorporates two stabilizing mechanisms: (i) ridge regularization of the multiscale covariance (Remark 4) and of the reduced covariance (Remark 10), and (ii) a discrete basis refresh schedule (Remark 2) that limits how often eigenvectors are recomputed.

Lemma 5 (Ridge regularization controls conditioning)**.** *Let $S \succeq 0$ with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_N \ge 0$, and let $\varepsilon_0 > 0$. Define $\bar{S} := S + \varepsilon_0 I_N$. Then, $\bar{S} > 0$, and $\lambda_{\min}(\bar{S}) \ge \varepsilon_0$. Moreover, the spectral condition number satisfies*

$$\kappa(\bar{S}) := \frac{\lambda_{\max}(\bar{S})}{\lambda_{\min}(\bar{S})} \le \frac{\lambda_1 + \varepsilon_0}{\varepsilon_0}.$$

Proof. Adding $\varepsilon_0 I_N$ shifts every eigenvalue of S by ε_0. Hence, $\lambda_i(\bar{S}) = \lambda_i(S) + \varepsilon_0 \ge \varepsilon_0$, and $\lambda_{\max}(\bar{S}) = \lambda_1 + \varepsilon_0$, which implies the claimed bound. □

Remark 8 (Practical stability and eigengap diagnostics). In finite precision arithmetic, two issues dominate: ill-conditioning and small eigengaps. Lemma 5 shows that the ridge parameter ε_0 prevents ill-conditioning by enforcing a uniform eigenvalue floor. Small eigengaps are more subtle: when $\delta_*(t)$ is small, eigenvectors may rotate substantially even under small perturbations of the covariance, which is exactly the mechanism quantified by Lemma 2 and Theorem 2. In implementation, we therefore treat $\delta_*(t)$ as a diagnostic: if it becomes too small, one can either reduce m, increase the refresh period K (to reduce estimation noise), or increase regularization (e.g. larger ε_0 or additional covariance smoothing). The sign convention in Remark 7 further removes artificial sign flips that would otherwise mimic "instability" and inflate turnover without changing the underlying subspace.

5. Reduced objective and mirror-flow dynamics

At each time t, the multiscale signal $\boldsymbol{\mu}_t$ and multiscale covariance Σ_t define a mean–variance geometry in $\mathbb{R}^N$. Because portfolios are restricted to $\boldsymbol{w} = A_t\boldsymbol{u}$, it is natural to work with the induced reduced quantities.

Remark 9 (Relation to neurodynamic portfolio optimization). Dynamical-system formulations have also been used as computational devices for portfolio selection, including collaborative neurodynamic optimization methods for minimax and biobjective problems [45]. The present construction differs in emphasis: the mirror-flow (5.5) and (5.6) is not merely a solver for a static optimization instance, but is the portfolio policy itself, driven online by the time-varying multiscale inputs $(\boldsymbol{\mu}_t, \Sigma_t)$ and constrained by the reduced representation. Moreover, the KL-prox characterization, simplex invariance, and the explicit turnover bounds derived below are tailored to trading implementability under transaction costs.

5.1. Reduced mean and reduced covariance

Let A_t be built using (4.1). The reduced signal and reduced covariance are defined as

$$\boldsymbol{\mu}_t^{(u)} := A_t^\top \boldsymbol{\mu}_t \in \mathbb{R}^m, \qquad \Sigma_t^{(u)} := A_t^\top \Sigma_t A_t \in \mathbb{R}^{m\times m}. \tag{5.1}$$

Remark 10 (Regularized reduced covariance). In the computations, we form the reduced geometry from the stabilized covariance $\bar{\Sigma}_t$ in (3.6) and add a small ridge on the reduced space:

$$\bar{\Sigma}_t^{(u)} := A_t^\top \bar{\Sigma}_t A_t + \varepsilon_u I_m, \tag{5.2}$$

with $\varepsilon_u > 0$ being very small. This guarantees the strict positive definiteness of the reduced quadratic form under finite precision and does not affect the conceptual structure of the model.

Because $\Sigma_t \succ 0$, and A_t has full column rank for generic data, $\Sigma_t^{(u)}$ is positive semidefinite and is typically positive definite. The reduced mean–variance term satisfies

$$(\boldsymbol{\mu}_t^{(u)})^\top \boldsymbol{u} = \boldsymbol{\mu}_t^\top (A_t\boldsymbol{u}), \qquad \boldsymbol{u}^\top \Sigma_t^{(u)} \boldsymbol{u} = (A_t\boldsymbol{u})^\top \Sigma_t (A_t\boldsymbol{u}).$$

Therefore, the reduced objective is consistent with the asset-space objective restricted to the reduced feasible set.

5.2. Concave reduced objective

We introduce a concave objective on Δ_m:

$$J_t(\boldsymbol{u}) = (\boldsymbol{\mu}_t^{(u)})^\top \boldsymbol{u} - \frac{\gamma}{2}\boldsymbol{u}^\top \Sigma_t^{(u)} \boldsymbol{u} + \rho\, H(\boldsymbol{u}), \qquad \boldsymbol{u} \in \Delta_m, \tag{5.3}$$

where $\gamma > 0$ controls risk aversion, and $\rho \geq 0$ controls diversification in the reduced space. Whenever $\rho > 0$, the entropy term discourages degenerate allocations and supports the stable dynamics.

For $\boldsymbol{u}$ in the interior of the simplex, the gradient is

$$\nabla J_t(\boldsymbol{u}) = \boldsymbol{\mu}_t^{(u)} - \gamma \Sigma_t^{(u)} \boldsymbol{u} - \rho(\log(\boldsymbol{u}) + \mathbb{1}), \tag{5.4}$$

with elementwise log.

5.3. Replicator-type ODE and a stable mirror discretization

A continuous-time simplex-preserving dynamic associated with (5.4) is

$$\dot{\boldsymbol{u}}(t) = \boldsymbol{u}(t) \odot \Big(\boldsymbol{g}(t) - (\boldsymbol{u}(t)^\top \boldsymbol{g}(t))\mathbb{1}\Big), \qquad \boldsymbol{g}(t) = \nabla J_t(\boldsymbol{u}(t)). \tag{5.5}$$

Rather than discretizing (5.5) by a naive Euler step (which may create boundary artifacts), we use a mirror step with the KL geometry. Given a step size $\eta > 0$, define

$$\boldsymbol{u}_{t+1}^{\mathrm{tar}} = \frac{\boldsymbol{u}_t \odot \exp(\eta \boldsymbol{g}_t)}{\mathbb{1}^\top(\boldsymbol{u}_t \odot \exp(\eta \boldsymbol{g}_t))}, \qquad \boldsymbol{g}_t := \nabla J_t(\boldsymbol{u}_t). \tag{5.6}$$

In applications, it is often desirable to smooth allocations; we include an inertia parameter $\lambda \in [0, 1)$ and set

$$\boldsymbol{u}_{t+1} = \lambda \boldsymbol{u}_t + (1 - \lambda)\boldsymbol{u}_{t+1}^{\mathrm{tar}}. \tag{5.7}$$

The investable weights are $\boldsymbol{w}_{t+1} = A_t \boldsymbol{u}_{t+1}$.

Provenance (standard property of entropic mirror descent). The simplex preservation and strict positivity in the next theorem are standard for exponentiated-gradient / entropic mirror-descent updates; see, for eample, [46,47]. We restate the result in our notation because feasibility is a central requirement of the reduced mirror flow.

Theorem 3 (Simplex invariance and strict positivity)**.** *Let $\boldsymbol{u}_t \in \Delta_m$ with strictly positive entries, and let $\eta > 0$. Then, $\boldsymbol{u}_{t+1}^{\mathrm{tar}} \in \Delta_m$ and has strictly positive entries. If $\lambda \in [0, 1)$, then $\boldsymbol{u}_{t+1} \in \Delta_m$ and has strictly positive entries.*

Proof. All entries of $\boldsymbol{u}_t \odot \exp(\eta \boldsymbol{g}_t)$ are positive. Normalization in (5.6) yields a positive vector that sums to one. The inertia update (5.7) is a convex combination of two points in Δ_m; strict positivity follows from $\lambda < 1$. □

Provenance (standard KL-prox representation). The following characterization is classical: the exponentiated-gradient update is equivalently a KL-proximal step; see, for example, [46,47]. We include it to make explicit the variational structure underlying (5.6).

Theorem 4 (KL-prox characterization of the mirror step)**.** *Fix t, and define $\boldsymbol{g}_t = \nabla J_t(\boldsymbol{u}_t)$. Then, $\boldsymbol{u}_{t+1}^{\mathrm{tar}}$ in* (5.6) *is the unique maximizer of*

$$\max_{\boldsymbol{u} \in \Delta_m} \Big\{\eta\, \boldsymbol{g}_t^\top \boldsymbol{u} - D_{\mathrm{KL}}(\boldsymbol{u} \| \boldsymbol{u}_t)\Big\}, \qquad D_{\mathrm{KL}}(\boldsymbol{u} \| \boldsymbol{u}_t) := \sum_{i=1}^{m} u_i \log\Big(\frac{u_i}{u_{t,i}}\Big). \tag{5.8}$$

Proof. The objective in (5.8) is strictly concave in Δ_m. A Lagrange multiplier calculation on the interior gives

$$\eta g_{t,i} - \left(\log\Big(\frac{u_i}{u_{t,i}}\Big) + 1\right) + \lambda = 0 \quad \Rightarrow \quad u_i = C\, u_{t,i} e^{\eta g_{t,i}}.$$

Normalizing the sum to one yields (5.6). □

Lemma 6 (Consistency with the replicator vector field)**.** *Assume* $\|\boldsymbol{g}_t\|_\infty \le G$ *for some* $G > 0$*. Let* $\boldsymbol{u}^+ = \boldsymbol{u}_{t+1}^{\mathrm{tar}}$ *be given by* (5.6)*. Then, for* η *small,*

$$\boldsymbol{u}^+ - \boldsymbol{u}_t = \eta\, \boldsymbol{u}_t \odot \Big(\boldsymbol{g}_t - (\boldsymbol{u}_t^\top \boldsymbol{g}_t)\mathbb{1}\Big) + \boldsymbol{\varepsilon}_t, \tag{5.9}$$

where the remainder satisfies $\|\boldsymbol{\varepsilon}_t\|_2 \le C\,\eta^2$ *with a constant* C *that depends only on* G *and* m*.*

Proof. Write $Z := \mathbb{1}^\top(\boldsymbol{u}_t \odot e^{\eta \boldsymbol{g}_t})$, and note $Z > 0$. Expanding the exponential elementwise,

$$e^{\eta g_{t,i}} = 1 + \eta g_{t,i} + \tfrac{1}{2}\eta^2 g_{t,i}^2 + O(\eta^3).$$

Hence,

$$Z = \sum_i u_{t,i}\Big(1 + \eta g_{t,i} + \tfrac{1}{2}\eta^2 g_{t,i}^2\Big) + O(\eta^3) = 1 + \eta(\boldsymbol{u}_t^\top \boldsymbol{g}_t) + \tfrac{1}{2}\eta^2 \sum_i u_{t,i} g_{t,i}^2 + O(\eta^3).$$

A standard expansion of $1/Z$ around 1 yields

$$\frac{1}{Z} = 1 - \eta(\boldsymbol{u}_t^\top \boldsymbol{g}_t) + O(\eta^2),$$

with the $O(\eta^2)$ coefficient bounded by a constant that depends on G and m. Now,

$$u_i^+ = u_{t,i} e^{\eta g_{t,i}} \cdot \frac{1}{Z} = u_{t,i}\Big(1 + \eta g_{t,i} + O(\eta^2)\Big)\Big(1 - \eta(\boldsymbol{u}_t^\top \boldsymbol{g}_t) + O(\eta^2)\Big),$$

so

$$u_i^+ - u_{t,i} = \eta u_{t,i}(g_{t,i} - (\boldsymbol{u}_t^\top \boldsymbol{g}_t)) + O(\eta^2).$$

Collecting terms yields (5.9) and the stated bounds. □

6. Turnover control and stability of the induced asset weights

The reduced parameterization separates the question of *decision dynamics* (how $\boldsymbol{u}_t$ evolves) from the question of *asset exposure* (how changes in $\boldsymbol{u}_t$ and A_t translate into changes in $\boldsymbol{w}_t$). This separation allows for a transparent control of turnover.

6.1. A contraction property of convex-hull mappings

Because each $\boldsymbol{a}_k(t) \in \Delta_N$, the matrix A_t has nonnegative entries, and each column sums to one:

$$A_t \ge 0, \qquad \mathbb{1}^\top A_t = \mathbb{1}^\top. \tag{6.1}$$

Proposition 3 (ℓ_1 contraction under a fixed basis)**.** *If* A *satisfies* (6.1)*, then for all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^m$*,*

$$\|A\boldsymbol{u} - A\boldsymbol{v}\|_1 \le \|\boldsymbol{u} - \boldsymbol{v}\|_1 .$$

In particular, if $\boldsymbol{w} = A\boldsymbol{u}$*, and* $\boldsymbol{w}' = A\boldsymbol{v}$*, then* $\|\boldsymbol{w} - \boldsymbol{w}'\|_1 \le \|\boldsymbol{u} - \boldsymbol{v}\|_1$.

Proof. Let $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}$. Then,

$$\|A\boldsymbol{x}\|_1 = \sum_{i=1}^N \Big|\sum_{k=1}^m A_{ik}x_k\Big| \le \sum_{i=1}^N \sum_{k=1}^m A_{ik}|x_k| = \sum_{k=1}^m |x_k| \sum_{i=1}^N A_{ik} = \sum_{k=1}^m |x_k| = \|\boldsymbol{x}\|_1 .$$

□

6.2. Deterministic turnover bound induced by inertia

Define the one-step ℓ_1 turnover of reduced allocations by $\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1$ and similarly for asset weights $\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1$.

Remark 11 (One-way turnover and proportional costs). Empirically, turnover is reported in the *one-way* convention

$$\mathrm{TO}_{t+1} := \frac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1, \qquad \mathrm{TO}^{(u)}_{t+1} := \frac{1}{2}\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1.$$

Under a fixed basis, Theorem 5 implies $\mathrm{TO}_{t+1} \le 1 - \lambda$. With a proportional transaction cost rate $c > 0$ (in basis points (bp) per unit one-way turnover), the net portfolio return is modeled as

$$R^{\mathrm{net}}_{t+1} = R^{\mathrm{gross}}_{t+1} - c\,\mathrm{TO}_{t+1}.$$

Provenance (control statement for the inertial mirror update). The bound below follows from a simple convex-combination argument, but it is a key mechanism of our framework: it yields an explicit *knob* for turnover control through the inertia parameter λ, and it is the starting point for the transfer bounds at the asset level.

Theorem 5 (Inertia yields explicit turnover control)**.** *Let $\boldsymbol{u}_{t+1} = \lambda\boldsymbol{u}_t + (1-\lambda)\boldsymbol{u}^{\mathrm{tar}}_{t+1}$ with $\lambda \in [0,1)$. Then,*

$$\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 = (1-\lambda)\left\|\boldsymbol{u}^{\mathrm{tar}}_{t+1} - \boldsymbol{u}_t\right\|_1 \le 2(1-\lambda).$$

If the basis is held fixed between t and $t+1$ (that is, $A_{t+1} = A_t$), then

$$\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1 \le 2(1-\lambda).$$

Proof. By definition, $\boldsymbol{u}_{t+1} - \boldsymbol{u}_t = (1-\lambda)(\boldsymbol{u}^{\mathrm{tar}}_{t+1} - \boldsymbol{u}_t)$; hence, $\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 = (1-\lambda)\|\boldsymbol{u}^{\mathrm{tar}}_{t+1} - \boldsymbol{u}_t\|_1$. Because $\boldsymbol{u}_t, \boldsymbol{u}^{\mathrm{tar}}_{t+1} \in \Delta_m$, their ℓ_1 distance is at most 2, and therefore, $\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 \le 2(1-\lambda)$.

If $A_{t+1} = A_t =: A$, then Proposition 3 yields $\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1 = \|A(\boldsymbol{u}_{t+1} - \boldsymbol{u}_t)\|_1 \le \|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1$. □

6.3. Effect of basis refresh on turnover

When A_t is updated, additional turnover arises, even if $\boldsymbol{u}_t$ changes slowly. A general decomposition follows from the triangle inequality as follows:

$$\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1 = \|A_{t+1}\boldsymbol{u}_{t+1} - A_t\boldsymbol{u}_t\|_1 \le \|A_{t+1}(\boldsymbol{u}_{t+1} - \boldsymbol{u}_t)\|_1 + \|(A_{t+1} - A_t)\boldsymbol{u}_t\|_1. \tag{6.2}$$

The first term is controlled by Proposition 3 whenever A_{t+1} is column-stochastic. The second term is controlled by the stability of A_t, which is precisely where the covariance-to-basis stability bounds become useful.

To express a clean bound, define the induced matrix norm

$$\|M\|_{1\to 1} := \max_{\|\boldsymbol{x}\|_1 = 1}\|M\boldsymbol{x}\|_1,$$

which is equal to the maximum absolute column sum. Because $\boldsymbol{u}_t \in \Delta_m$, we have $\|\boldsymbol{u}_t\|_1 = 1$, and thus

$$\|(A_{t+1} - A_t)\boldsymbol{u}_t\|_1 \le \|A_{t+1} - A_t\|_{1\to 1}.$$

Moreover, $\|A_{t+1} - A_t\|_{1\to 1} \le \max_{k\le m}\|\boldsymbol{a}_k(t+1) - \boldsymbol{a}_k(t)\|_1$. Theorem 2 then translates a bound on covariance perturbations into a bound on the basis turnover.

Provenance (novel turnover transfer with basis refresh). The next theorem is specific to our time-varying PCA–softmax basis and reduced-order parameterization. It separates the turnover into an "endogenous" component controlled by the reduced dynamics and an "exogenous" component induced by refreshing the basis matrix A_t.

Theorem 6 (Turnover bound with basis refresh)**.** *Let $\lambda \in [0, 1)$ be the inertia parameter in the reduced update* (5.7)*, and let A_t be column-stochastic for all t. Define the one-way turnover by* $\mathrm{TO}_{t+1} := \frac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1$*. Then, for every t,*

$$\mathrm{TO}_{t+1} \le (1 - \lambda) + \frac{1}{2}\|A_{t+1} - A_t\|_{1\to 1}. \tag{6.3}$$

In particular, if $A_{t+1} = A_t$ (no basis refresh), then $\mathrm{TO}_{t+1} \le 1 - \lambda$*.*

Moreover, suppose that A_t is constructed from the top-m eigenvectors of a symmetric matrix S_t via the softmax map $\boldsymbol{a}_k(t) = \mathrm{softmax}(\tau \boldsymbol{v}_k(t))$, with eigengaps $\delta_k(t) > 0$ as in Lemma 2. Assume that the eigenvectors are oriented using the sign convention in Remark 7. If

$$\|S_{t+1} - S_t\|_2 \le \frac{1}{2}\,\delta_*(t), \qquad \delta_*(t) := \min_{1\le k\le m} \delta_k(t),$$

then the turnover obeys the covariance-driven bound

$$\mathrm{TO}_{t+1} \le (1 - \lambda) + \frac{\tau\sqrt{N}}{\delta_*(t)}\,\|S_{t+1} - S_t\|_2\,. \tag{6.4}$$

Proof. Starting from the decomposition (6.2) and assuming that A_{t+1} is column-stochastic, Proposition 3 gives

$$\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1 \le \|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 + \|(A_{t+1} - A_t)\boldsymbol{u}_t\|_1\,.$$

By Theorem 5, $\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 \le 2(1 - \lambda)$. Because $\boldsymbol{u}_t \in \Delta_m$ implies $\|\boldsymbol{u}_t\|_1 = 1$, we also have $\|(A_{t+1} - A_t)\boldsymbol{u}_t\|_1 \le \|A_{t+1} - A_t\|_{1\to 1}$. Dividing by 2 yields (6.3).

For the covariance-driven bound, note that $\|A_{t+1} - A_t\|_{1\to 1}$ equals the maximum absolute column sum, and hence,

$$\|A_{t+1} - A_t\|_{1\to 1} \le \max_{1\le k\le m} \|\boldsymbol{a}_k(t+1) - \boldsymbol{a}_k(t)\|_1\,.$$

Under the stated perturbation condition and the consistent orientation, Theorem 2 applied with $S = S_t$ and $\widetilde{S} = S_{t+1}$ yields, for each $k \le m$,

$$\|\boldsymbol{a}_k(t+1) - \boldsymbol{a}_k(t)\|_1 \le \frac{2\tau\sqrt{N}}{\delta_k(t)}\,\|S_{t+1} - S_t\|_2\,.$$

Taking the maximum over $k \le m$ and using $\delta_*(t) = \min_{k\le m} \delta_k(t)$ gives

$$\|A_{t+1} - A_t\|_{1\to 1} \le \frac{2\tau\sqrt{N}}{\delta_*(t)}\,\|S_{t+1} - S_t\|_2\,.$$

Substituting this into (6.3) yields (6.4). □

Proposition 4 (From turnover bounds to transaction cost bounds)**.** *Let $\phi : [0,1] \to \mathbb{R}_+$ be a nondecreasing transaction cost function applied to one-way turnover, and define the per-step cost as*

$$\mathrm{TC}_{t+1} := \phi(\mathrm{TO}_{t+1}), \qquad \mathrm{TO}_{t+1} := \tfrac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1.$$

Assume ϕ is Lipschitz with constant L_ϕ on $[0,1]$ and $\phi(0) = 0$. If $\mathrm{TO}_{t+1} \le B_{t+1}$ for some deterministic bound $B_{t+1} \in [0,1]$, then

$$\mathrm{TC}_{t+1} \le L_\phi\, B_{t+1}.$$

In particular, for the proportional model $\phi(x) = cx$ used in the empirical section, we obtain $\mathrm{TC}_{t+1} \le c\, B_{t+1}$, and the turnover bounds in Theorems 5 and 6 translate directly into explicit cost bounds.

Proof. Because ϕ is nondecreasing, and $\mathrm{TO}_{t+1} \le B_{t+1}$, we have $\mathrm{TC}_{t+1} = \phi(\mathrm{TO}_{t+1}) \le \phi(B_{t+1})$. By Lipschitz continuity and $\phi(0) = 0$, $\phi(B_{t+1}) \le L_\phi B_{t+1}$. □

Corollary 1 (EWMA drift bound for the multiscale covariance)**.** *Assume the multiscale covariances are updated by*

$$\Sigma_{j,t} = (1-\alpha)\Sigma_{j,t-1} + \alpha\, \boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^\top + \varepsilon I_N, \qquad \Sigma_t = \sum_{j=1}^{J} \omega_j \Sigma_{j,t},$$

with $\alpha \in (0,1]$ and weights $\omega_j \ge 0$ satisfying $\sum_{j=1}^J \omega_j = 1$. Suppose there exists $B > 0$ such that $\left\|\boldsymbol{W}_{j,t}\right\|_2 \le B$ for all $j \in \{1,\dots,J\}$ and all t. Define

$$M := \max_{1\le j\le J} \max\Big\{ \left\|\Sigma_{j,0}\right\|_2,\ B^2 + \frac{\varepsilon}{\alpha}\Big\}.$$

Then, for all $t \ge 0$,

$$\left\|\bar{\Sigma}_{t+1} - \bar{\Sigma}_t\right\|_2 = \|\Sigma_{t+1} - \Sigma_t\|_2 \le \alpha(M + B^2) + \varepsilon. \tag{6.5}$$

Consequently, if the PCA basis is refreshed from $S_t = \bar{\Sigma}_t$, and the eigengap condition in Theorem 6 holds, then

$$\mathrm{TO}_{t+1} \le (1-\lambda) + \frac{\tau\sqrt{N}}{\delta_*(t)}(\alpha(M + B^2) + \varepsilon). \tag{6.6}$$

Proof. Because $\|I_N\|_2 = 1$, and $\|\boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^\top\|_2 = \|\boldsymbol{W}_{j,t}\|_2^2$, taking the operator norm in the update gives

$$\left\|\Sigma_{j,t}\right\|_2 \le (1-\alpha)\left\|\Sigma_{j,t-1}\right\|_2 + \alpha\left\|\boldsymbol{W}_{j,t}\right\|_2^2 + \varepsilon \le (1-\alpha)\left\|\Sigma_{j,t-1}\right\|_2 + \alpha B^2 + \varepsilon.$$

Iterating this scalar recursion yields

$$\left\|\Sigma_{j,t}\right\|_2 \le (1-\alpha)^t \left\|\Sigma_{j,0}\right\|_2 + (1-(1-\alpha)^t)B^2 + \varepsilon \sum_{s=0}^{t-1}(1-\alpha)^s \le \max\Big\{ \left\|\Sigma_{j,0}\right\|_2,\ B^2 + \frac{\varepsilon}{\alpha}\Big\} \le M.$$

Moreover,

$$\Sigma_{j,t+1} - \Sigma_{j,t} = \alpha(\boldsymbol{W}_{j,t+1}\boldsymbol{W}_{j,t+1}^\top - \Sigma_{j,t}) + \varepsilon I_N;$$

hence,

$$\left\|\Sigma_{j,t+1} - \Sigma_{j,t}\right\|_2 \le \alpha(\left\|\boldsymbol{W}_{j,t+1}\right\|_2^2 + \left\|\Sigma_{j,t}\right\|_2) + \varepsilon \le \alpha(B^2 + M) + \varepsilon.$$

Because Σ_t is a convex combination of the $\Sigma_{j,t}$, we obtain

$$\|\Sigma_{t+1} - \Sigma_t\|_2 = \left\| \sum_{j=1}^{J} \omega_j (\Sigma_{j,t+1} - \Sigma_{j,t}) \right\|_2 \leq \sum_{j=1}^{J} \omega_j \left\| \Sigma_{j,t+1} - \Sigma_{j,t} \right\|_2 \leq \alpha(M + B^2) + \varepsilon.$$

Finally, $\bar{\Sigma}_t = \Sigma_t + \varepsilon_0 I_N$ differs from Σ_t by a constant ridge term, so $\bar{\Sigma}_{t+1} - \bar{\Sigma}_t = \Sigma_{t+1} - \Sigma_t$, giving (6.5). Substituting (6.5) into (6.4) with $S_t = \bar{\Sigma}_t$ yields (6.6). □

7. A regret-type control for the reduced mirror-flow

Because the signal and covariance are time-varying, it is natural to quantify the cumulative suboptimality of the reduced decisions relative to a fixed comparator in the hindsight. A convenient route is to use KL-prox characterization and a classical argument for exponentiated-gradient mirror ascent.

Assumption 1 (Bounded gradient field)**.** *There exists $G > 0$ such that for all t and all iterates $\boldsymbol{u}_t$ generated by* (5.6)*, we have $\|\boldsymbol{g}_t\|_\infty \leq G$, where $\boldsymbol{g}_t = \nabla J_t(\boldsymbol{u}_t)$.*

Lemma 7 (A one-step KL inequality)**.** *Let $\boldsymbol{u}_{t+1}^{\mathrm{tar}}$ be defined by* (5.6) *and, assume $\|\boldsymbol{g}_t\|_\infty \leq G$. Then, for any comparator $\boldsymbol{u} \in \Delta_m$,*

$$\boldsymbol{g}_t^\top (\boldsymbol{u} - \boldsymbol{u}_t) \leq \frac{1}{\eta}\Big(D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_t) - D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_{t+1}^{\mathrm{tar}}) \Big) + \frac{\eta G^2}{2}. \tag{7.1}$$

Proof. Write the update as $u_{t+1,i}^{\mathrm{tar}} = u_{t,i} e^{\eta g_{t,i}} / Z_t$, where $Z_t = \sum_k u_{t,k} e^{\eta g_{t,k}}$. A direct computation yields

$$D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_{t+1}^{\mathrm{tar}}) - D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_t) = \sum_i u_i \log\Big(\frac{u_{t,i}}{u_{t+1,i}^{\mathrm{tar}}}\Big) = \log Z_t - \eta\, \boldsymbol{g}_t^\top \boldsymbol{u}.$$

Rearranging gives

$$\boldsymbol{g}_t^\top \boldsymbol{u} = \frac{1}{\eta} \log Z_t - \frac{1}{\eta}\Big(D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_{t+1}^{\mathrm{tar}}) - D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_t) \Big).$$

Therefore,

$$\boldsymbol{g}_t^\top (\boldsymbol{u} - \boldsymbol{u}_t) = \frac{1}{\eta} \log Z_t - \boldsymbol{g}_t^\top \boldsymbol{u}_t - \frac{1}{\eta}\Big(D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_{t+1}^{\mathrm{tar}}) - D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_t) \Big).$$

It remains to bound $(1/\eta) \log Z_t - \boldsymbol{g}_t^\top \boldsymbol{u}_t$. Note that $Z_t = \mathbb{E}_{i\sim \boldsymbol{u}_t}[e^{\eta g_{t,i}}]$ and that $g_{t,i} \in [-G, G]$. By Hoeffding's lemma, $\log \mathbb{E}[e^{\eta X}] \leq \eta \mathbb{E}[X] + \eta^2 (b-a)^2/8$ for $X \in [a, b]$. Here, $b - a = 2G$; hence,

$$\log Z_t \leq \eta\, \boldsymbol{g}_t^\top \boldsymbol{u}_t + \frac{\eta^2 G^2}{2}.$$

Substituting yields the proof. □

Provenance (standard static regret bound for exponentiated gradient). Static regret bounds of this form are classic results in online learning for exponentiated-gradient / mirror-descent methods; see, for example, [47, 48]. We state and prove the bound in our reduced notation to connect the WPROD-R dynamics to the standard regret-control perspective.

Theorem 7 (Static regret bound for the reduced mirror flow)**.** *Assume J_t is concave on Δ_m for each t and, Assumption 1 holds. Let $\boldsymbol{u}_{t+1}^{\mathrm{tar}}$ be generated by* (5.6) *(without inertia) from an initial $\boldsymbol{u}_1 \in \Delta_m$. Then, for any fixed comparator $\boldsymbol{u} \in \Delta_m$,*

$$\sum_{t=1}^{T} \left(J_t(\boldsymbol{u}) - J_t(\boldsymbol{u}_t)\right) \leq \frac{D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_1)}{\eta} + \frac{\eta G^2}{2} T. \tag{7.2}$$

Proof. Concavity implies $J_t(\boldsymbol{u}) - J_t(\boldsymbol{u}_t) \leq \langle \nabla J_t(\boldsymbol{u}_t), \boldsymbol{u} - \boldsymbol{u}_t \rangle = \boldsymbol{g}_t^{\top}(\boldsymbol{u} - \boldsymbol{u}_t)$. Apply Lemma 7, and sum over t. The KL terms telescope, yielding

$$\sum_{t=1}^{T} \boldsymbol{g}_t^{\top}(\boldsymbol{u} - \boldsymbol{u}_t) \leq \frac{D_{\mathrm{KL}}(\boldsymbol{u}\|\boldsymbol{u}_1)}{\eta} + \frac{\eta G^2}{2} T,$$

and the claim follows. □

Remark 12. The bound (7.2) is a rigorous quantitative statement that the reduced process does not drift arbitrarily far from the best fixed reduced allocation chosen in hindsight, even though the multiscale signal and covariance change over time. Inertia (5.7) can be viewed as an additional regularization that improves turnover behavior; its effect on regret can be controlled by standard averaging arguments when needed, but the essential stability mechanism already appears in the mirror step.

Remark 13 (Economic interpretation under a time-varying basis). The regret bound in Theorem 7 is formulated for the reduced variables $\boldsymbol{u}_t$ because $\boldsymbol{u}_t$ is the only dynamic decision state. Nevertheless, it has a direct portfolio-management interpretation in the traded space. If the basis is fixed (no refresh), each comparator $\boldsymbol{u} \in \Delta_m$ corresponds to a fixed traded portfolio $\boldsymbol{w} = A\boldsymbol{u}$, and the regret bound compares the strategy to the best fixed convex combination of the basis portfolios in hindsight. If the basis is refreshed, the same fixed reduced comparator $\boldsymbol{u}$ induces an *implementable time-varying* benchmark $\boldsymbol{w}_t(\boldsymbol{u}) := A_t\boldsymbol{u}$ whose constituents rotate with the estimated multiscale geometry while the exposure vector $\boldsymbol{u}$ is held constant. In this sense, static regret on Δ_m corresponds to a tracking guarantee relative to the best *constant exposure* to a time-varying family of basis portfolios. The additional economic "price" of allowing the benchmark class to vary through A_t is reflected in the basis-variation term of the turnover bound (Theorem 6), which links implementability costs to covariance drift and eigengap separation.

Remark 14 (Relation to inertial iterates). Theorem 7 is stated for the *target* mirror sequence $\{\boldsymbol{u}_{t+1}^{\mathrm{tar}}\}$ (i.e., without inertia). When inertia is used, the realized iterate is the convex combination $\boldsymbol{u}_{t+1} = \lambda\boldsymbol{u}_t + (1-\lambda)\boldsymbol{u}_{t+1}^{\mathrm{tar}}$. Thus, $\|\boldsymbol{u}_{t+1} - \boldsymbol{u}_t\|_1 = (1-\lambda)\left\|\boldsymbol{u}_{t+1}^{\mathrm{tar}} - \boldsymbol{u}_t\right\|_1$, so inertia reduces variation of the realized sequence. A regret bound for $\{\boldsymbol{u}_t\}$ can be obtained by standard stability arguments (viewing inertia as an averaging step); in the empirical section, we additionally report a regret proxy evaluated on the realized iterates.

8. Closing remark on the wavelet–PCA coupling

The mathematical role of the wavelet layer is explicit: it induces the multiscale geometry Σ_t and multiscale score $\boldsymbol{\mu}_t$. The PCA layer is then applied to this geometry, so the reduced feasible set adapts to the multiscale structure rather than to a single unconditional covariance. The mirror flow acts on the reduced simplex, which is essential for keeping the model dimension fixed while still allowing time variation through $(\boldsymbol{\mu}_t, \Sigma_t, A_t)$. The stability results above show how each layer affects the next: bounded wavelet features yield controlled covariance updates, controlled covariance perturbations yield stable PCA directions and stable basis portfolios, and stable basis and KL-prox dynamics yield feasible, positive, and tunable portfolio paths.

8.1. Algorithmic summary of the WPROD-R (core procedure)

For completeness, we summarize the core online procedure used in the empirical section. The method maintains: (i) a multiscale risk geometry $(\Sigma_{j,t})_{j=1}^{J}$ updated by an EWMA recursion on wavelet detail coefficients; (ii) a time-varying investable basis $A_t = [\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_m(t)]$ obtained from the top-m eigenvectors of a stabilized geometry matrix through the softmax mapping; and (iii) reduced mixing weights $\boldsymbol{u}_t \in \Delta_m$ evolved by an entropy-regularized mirror-flow with inertia. The traded portfolio is always $\boldsymbol{w}_t = A_t \boldsymbol{u}_t \in \Delta_N$; hence, the long-only full investment is enforced by construction. The basis refresh schedule (Remark 2) separates turnover induced by portfolio dynamics from that induced by changes in the estimated basis.

Algorithm 1 WPROD-R: Wavelet–PCA reduced mirror-flow portfolio dynamics

Require: Prices $\{P_{i,t}\}_{t=0}^{T}$ $(i = 1, \ldots, N)$, log-returns $\boldsymbol{r}_t$ as in (2.2). Parameters: L (multiple of 2^J), J, α, β, $\boldsymbol{\omega} \in \Delta_J$, $\boldsymbol{d} \in \Delta_J$, $\lambda_d \geq 0$, $\kappa_\mu > 0$, $\tau > 0$, m, refresh period K, $\gamma > 0$, $\rho \geq 0$, $\eta > 0$, $\lambda \in [0, 1)$, $\varepsilon, \varepsilon_0, \varepsilon_u > 0$, cost rate $c \geq 0$.

Ensure: $\{\boldsymbol{w}_t\}$, $\{\boldsymbol{u}_t\}$ (and optionally $\{R_t^{\text{net}}\}$).

Set $t_0 := L$. Initialize $(\Sigma_{j,t_0})_{j=1}^{J}$, set $\widehat{\boldsymbol{\mu}}_{t_0}$, and form $\Sigma_{t_0} := \sum_{j=1}^{J} \omega_j \Sigma_{j,t_0}$, $\bar{\Sigma}_{t_0} := \Sigma_{t_0} + \varepsilon_0 I_N$.
Construct A_{t_0} from the top-m eigenvectors of S_{t_0} (in experiments $S_t := \bar{\Sigma}_t$), via $\boldsymbol{a}_k(t_0) = \mathrm{softmax}(\tau \boldsymbol{v}_k(t_0))$ (Remark 7).
Set $\boldsymbol{u}_{t_0} := \frac{1}{m}\mathbb{1}$ and $\boldsymbol{w}_{t_0} := A_{t_0}\boldsymbol{u}_{t_0}$.
for $t = t_0, \ldots, T-1$ **do**
    Compute SWT features $(\boldsymbol{W}_{j,t})_{j=1}^{J}$ and smooth/trend vector $\boldsymbol{s}_t$ on the rolling window (Remark 3).
    for $j = 1, \ldots, J$ **do**
        $\Sigma_{j,t} \leftarrow (1-\alpha)\Sigma_{j,t-1} + \alpha\, \boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^{\top} + \varepsilon I_N$.
    end for
    $\Sigma_t \leftarrow \sum_{j=1}^{J} \omega_j \Sigma_{j,t}$, $\quad \bar{\Sigma}_t \leftarrow \Sigma_t + \varepsilon_0 I_N$.
    $\boldsymbol{z}_t \leftarrow \boldsymbol{s}_t + \lambda_d \sum_{j=1}^{J} d_j \boldsymbol{W}_{j,t}$, $\quad \tilde{\boldsymbol{\mu}}_t \leftarrow \mathrm{diag}(\Sigma_t)^{-1/2}\boldsymbol{z}_t$.
    $\boldsymbol{\mu}_t \leftarrow \kappa_\mu(\tilde{\boldsymbol{\mu}}_t - \frac{1}{N}(\mathbb{1}^{\top}\tilde{\boldsymbol{\mu}}_t)\mathbb{1})$, $\quad \widehat{\boldsymbol{\mu}}_t \leftarrow (1-\beta)\widehat{\boldsymbol{\mu}}_{t-1} + \beta\boldsymbol{\mu}_t$.
    if t is a refresh time (multiple of K) **then**
        Recompute A_t from the top-m eigenvectors of $S_t = \bar{\Sigma}_t$.
    else
        $A_t \leftarrow A_{t-1}$.
    end if

$\boldsymbol{\mu}_t^{(u)} \leftarrow A_t^\top \widehat{\boldsymbol{\mu}}_t, \quad \bar{\Sigma}_t^{(u)} \leftarrow A_t^\top \bar{\Sigma}_t A_t + \varepsilon_u I_m.$

$\boldsymbol{g}_t \leftarrow \boldsymbol{\mu}_t^{(u)} - \gamma \bar{\Sigma}_t^{(u)} \boldsymbol{u}_t - \rho(\log \boldsymbol{u}_t + \mathbb{1}).$

$\boldsymbol{u}_{t+1}^{\mathrm{tar}} \leftarrow \dfrac{\boldsymbol{u}_t \odot \exp(\eta \boldsymbol{g}_t)}{\mathbb{1}^\top(\boldsymbol{u}_t \odot \exp(\eta \boldsymbol{g}_t))}.$

$\boldsymbol{u}_{t+1} \leftarrow \lambda \boldsymbol{u}_t + (1-\lambda)\boldsymbol{u}_{t+1}^{\mathrm{tar}}.$

$\boldsymbol{w}_{t+1} \leftarrow A_t \boldsymbol{u}_{t+1}.$

$\mathrm{TO}_{t+1} \leftarrow \frac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1, \quad R_{t+1}^{\mathrm{net}} \leftarrow R_{t+1}^{\mathrm{gross}} - c\,\mathrm{TO}_{t+1}.$

end for

9. Empirical results

This section reports an empirical evaluation of the wavelet–PCA reduced-order dynamics (WPROD-R) framework using daily equity data. Our goal is to assess whether the theoretical design features–multiscale risk geometry, reduced-order representation, and simplex-preserving mirror-flow updates–translate into stable performance under realistic trading frictions. We present baseline comparisons, regime-split results, and sensitivity and ablation analyses. Performance is summarized using standard risk-adjusted and drawdown metrics together with turnover-based trading diagnostics.

9.1. Empirical setup and baseline configuration

Our empirical analysis is based on a reproducible experimental setup. The dataset consists of daily closing prices for a universe of 10 large-cap United States (US) equities spanning 1 January 2019 to 1 January 2025. The selected assets provide a diversified mix of technology-heavy growth stocks and constituents from the energy, health care, and consumer staples sectors. The data were sourced from the Stooq database. Transaction costs are modeled using a proportional one-way turnover charge of $c = 10$ basis points (bp).

The baseline configuration for WPROD-R, used throughout unless otherwise stated, is reported in Table 9. The parameters correspond to the core model components: the wavelet multiscale geometry (e.g., wavelet family ψ, number of scales J, and window length L), the PCA-based basis construction (e.g., number of principal components m and softmax temperature τ), and the mirror-flow dynamics (e.g., step size η and inertia parameter λ).

9.2. Comparative performance analysis

We begin by evaluating the performance of the full WPROD-R model against two classical benchmarks: A mean-variance (MV) optimization strategy and a passive equal-weight (EW) portfolio. The MV strategy is implemented using a rolling window of length L_{MV} (taken equal to the WPROD-R lookback length L) to estimate the sample mean and sample covariance of returns, and it solves a long-only, fully invested Markowitz problem on Δ_N at each rebalance, with the same risk aversion parameter γ as in the WPROD-R model. To avoid numerical singularity, we add a small ridge term $\varepsilon_{\mathrm{MV}} I_N$ to the sample covariance, but we do not apply shrinkage or Bayesian priors in this baseline. This yields a transparent classical comparator. To guard against an unstable sample-covariance baseline, the appendix reports strengthened comparators based on Ledoit–Wolf covariance shrinkage and risk-parity allocations [49, 50]. The EW portfolio represents a naive diversification approach with zero turnover. Table 10 summarizes the key performance metrics for these strategies, both before (gross) and after

(net) transaction costs. We report the compound annual growth rate (CAGR), annualized volatility, Sharpe ratio, and maximum drawdown (Max DD), together with average and maximum daily one-way turnover.

Table 9. Baseline model configuration.

Parameter	Value	Description
m	4	Number of principal components (basis portfolios)
ψ	db4	Wavelet family (Daubechies 4)
J	3	Number of wavelet scales
L	256	Lookback window for wavelet transform
α	0.08	EWMA parameter for wavelet covariance
τ	3.5	Softmax temperature for basis portfolio construction
K	21	Basis refresh period (trading days) for the PCA–softmax basis A_t
γ	4.0	Risk aversion parameter in the objective function
ρ	0.02	Entropy regularization parameter
η	0.35	Learning rate for the mirror descent update
λ	0.8	Inertia parameter for smoothing portfolio weights
c	10 bp	One-way transaction cost

Table 10. Performance comparison of WPROD-R against baselines.

	Net-of-Cost Performance				Turnover	
Strategy	CAGR (%)	Volatility (%)	Sharpe Ratio	Max DD (%)	Avg Daily (%)	Max Daily (%)
WPROD-R (Full)	33.90	28.12	1.18	−34.17	2.39	74.96
MV (Markowitz)	36.17	38.61	0.99	−44.20	0.42	2.32
Equal-Weight (EW)	27.65	24.43	1.12	−31.58	0.00	0.00

The results in Table 10 reveal that the WPROD-R model achieves a net Sharpe ratio of 1.18, outperforming the traditional MV benchmark (0.99) and being competitive with the passive EW strategy (1.12). In this sample, the MV strategy attains a higher gross CAGR, but its net performance is more sensitive to trading frictions and to estimation error in rolling mean/covariance inputs, and it exhibits a larger maximum drawdown. By contrast, WPROD-R exposes an explicit turnover knob through the inertia parameter λ in the reduced mirror flow (Theorem 5), which yields a principled adaptivity–turnover trade-off and supports cost-aware deployment.

Figure 1 provides a visual representation of the cumulative wealth generated by each strategy in the study. The WPROD-R model demonstrates a more stable growth trajectory than the volatile path of the MV portfolio. The figure also highlights the model's resilience during market downturns, such as the coronavirus disease 2019 (COVID-19) crash in March 2020 and the 2022 drawdown, where it effectively mitigates losses compared to benchmarks.

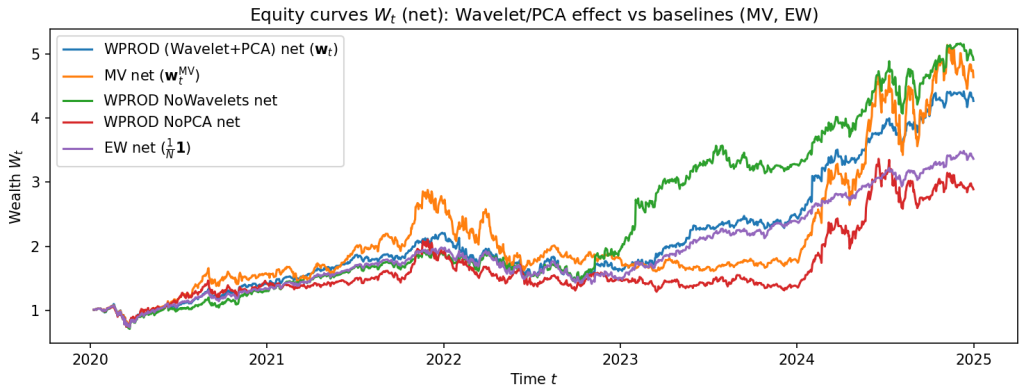


Figure 1. Net wealth evolution for the WPROD-R model and benchmark strategies. The plot reports cumulative net returns over the backtesting period and facilitates comparison of risk-adjusted performance and drawdown behavior across strategies.

9.3. Robustness to alternative transaction cost models

The baseline backtests adopt the proportional transaction cost model in Remark 11, namely $R_{t+1}^{\text{net}} = R_{t+1}^{\text{gross}} - c\,\text{TO}_{t+1}$ with $c = 10$ bp per unit one-way turnover. Although this specification is standard in academic backtests, practical trading costs are often *convex* in turnover due to market-impact effects and liquidity frictions. From a mathematical perspective, Proposition 4 shows that the turnover bounds in Theorems 5–6 immediately imply explicit transaction cost bounds for any nondecreasing Lipschitz cost map $\phi : [0, 1] \to \mathbb{R}_+$ applied to TO_{t+1}. The empirical role of this subsection is therefore to stress-test the *same realized weight paths* under several representative convexifications of ϕ and to verify that the resulting net performance conclusions are qualitatively stable.

We consider the following three cost mappings (all calibrated to share the same linear slope c at the origin):

$$\phi_{\text{lin}}(x) = c\,x, \qquad \phi_{\text{pw}}(x) = c\,\min\{x, \theta\} + \xi c\,\max\{x - \theta, 0\}, \qquad \phi_{\text{LQ}}(x) = c\,x + \kappa x^2,\ \ \kappa = \frac{c}{\theta}, \tag{9.1}$$

with threshold $\theta = 0.05$ (5% one-way turnover) and high-turnover multiplier $\xi = 3$. For each strategy, we keep the trading weights $(\boldsymbol{w}_t)_t$ *unchanged* and recompute net returns by

$$R_{t+1}^{\text{net},\phi} = R_{t+1}^{\text{gross}} - \phi(\text{TO}_{t+1}), \qquad \text{TO}_{t+1} = \tfrac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1, \tag{9.2}$$

so that the experiment isolates sensitivity to the cost curvature rather than to a change of the trading rule.

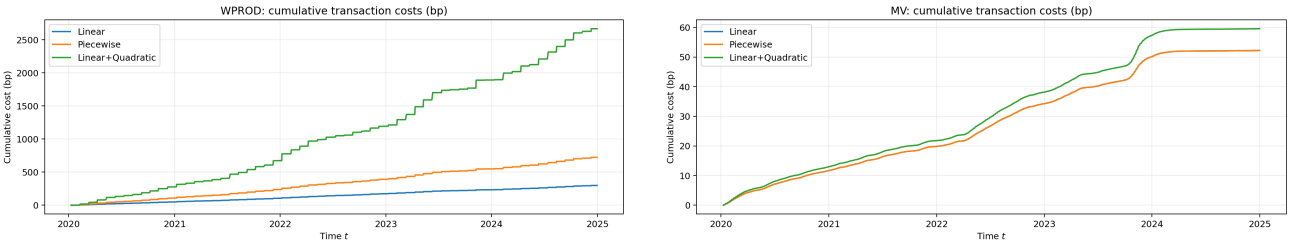
Table 11 summarizes the resulting net performance, and Figures 2–4 provide a visual diagnosis. Two patterns are immediate. First, the MV benchmark exhibits uniformly small turnover (max daily turnover below θ); hence, the piecewise model coincides with the linear model, and the linear+quadratic model produces only a marginal additional cost (on the order of a few basis points over the full sample). Second, WPROD-R exhibits a low typical daily turnover (average $\approx 2.4\%$) but a pronounced turnover tail (max daily turnover $\approx 75\%$), which is precisely the regime in which convex cost penalties become economically meaningful. Under the linear+quadratic mapping, cumulative costs increase substantially, and the net wealth curve is visibly compressed relative to the proportional-cost baseline.

Table 11. Sensitivity of net performance to transaction-cost specification. Net returns are recomputed by applying alternative cost mappings $\phi(\mathrm{TO}_t)$ to the realized one-way turnover path; weights are unchanged. The threshold is $\theta = 0.05$, and the linear slope is $c = 10$ bp.

Strategy	Cost model	CAGR (%)	Sharpe	Max DD (%)	Avg TO (%)	Max TO (%)	Cum. cost (bp)	$\Pr(\mathrm{TO}_t > \theta)$ (%)
WPROD-R	Linear	33.90	1.18	−34.18	2.39	74.96	299.3	4.71
WPROD-R	Piecewise	32.76	1.15	−34.63	2.39	74.96	724.3	4.71
WPROD-R	Lin.+Quad.	27.66	1.01	−36.62	2.39	74.96	2667.8	4.71
MV	Linear	36.17	0.99	−44.20	0.42	2.32	52.2	0.00
MV	Piecewise	36.17	0.99	−44.20	0.42	2.32	52.2	0.00
MV	Lin.+Quad.	36.15	0.99	−44.21	0.42	2.32	59.6	0.00
EW	Linear	27.65	1.12	−31.58	0.00	0.00	0.0	0.00
EW	Piecewise	27.65	1.12	−31.58	0.00	0.00	0.0	0.00
EW	Lin.+Quad.	27.65	1.12	−31.58	0.00	0.00	0.0	0.00

(a) WPROD-R. **(b)** MV benchmark.

Figure 2. Net wealth W_t under alternative transaction-cost mappings (9.1). Each panel reports the net wealth trajectory (top) and the relative deviation $100\,(W_t/W_t^{\mathrm{lin}} - 1)$ from the proportional-cost baseline (bottom). Cost models barely affect MV because turnover remains small, whereas convex penalties materially affect WPROD-R because of refresh-induced turnover spikes.



(a) WPROD-R. **(b)** MV benchmark.

Figure 3. Cumulative transaction costs (in basis points) implied by the alternative cost mappings (9.1). The convex models amplify the contribution of the turnover tail for WPROD-R, whereas MV remains largely insensitive because its turnover distribution stays in the low-turnover regime.

The turnover tail is structurally linked to basis refresh. Because A_t is column-stochastic by condition (6.1) (by construction), the reduced-to-asset map $\boldsymbol{w}_t = A_t \boldsymbol{u}_t$ is ℓ_1-nonexpansive (Proposition 3). In our baseline configuration the PCA–softmax basis, A_t is refreshed every $K = 21$ trading days and held constant in between. Consequently, the event $\{\mathrm{TO}_{t+1} > \theta\}$ occurs on 59 out of 1253 trading days (4.71%), which is close to the refresh frequency $1/K$. This is consistent with the decomposition in Theorem 6: most trading days are governed by the inertia-controlled term (Theorem 5), whereas rare large-turnover events arise from the refresh term $\|A_{t+1} - A_t\|_{1\to 1}$. Figure 4 makes this mechanism explicit by contrasting the bulk of the turnover distribution with its tail on a log-survival scale.

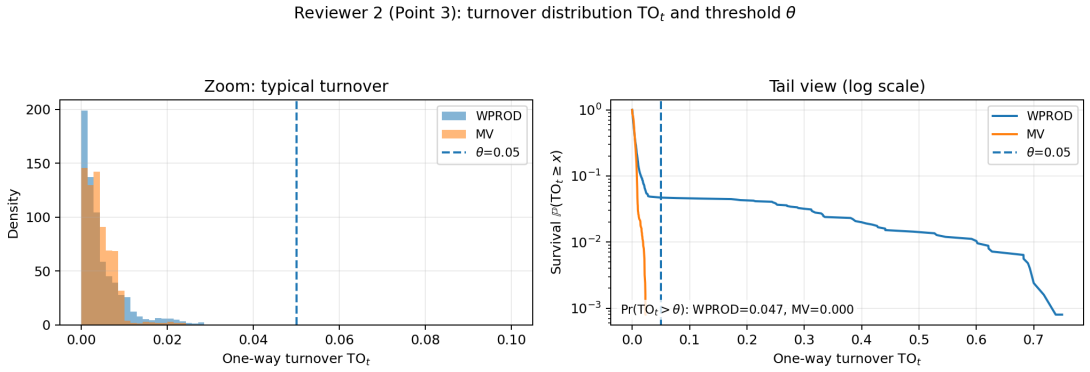
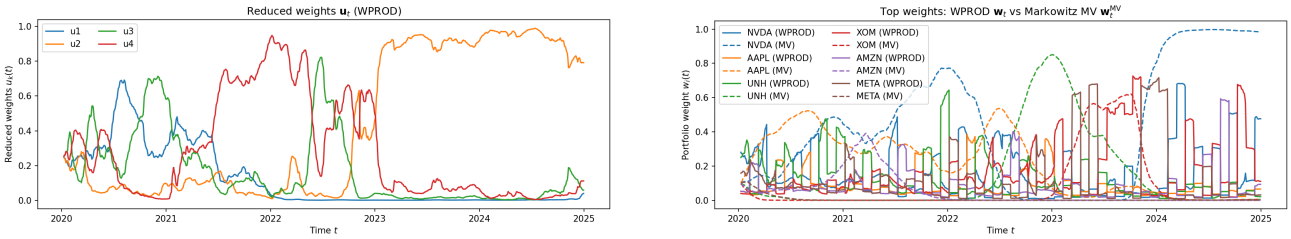


Figure 4. Distribution of one-way turnover $\mathrm{TO}_{t+1} = (1/2)\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1$ and threshold $\theta = 0.05$. Left: bulk turnover histogram. Right: log-survival view $\Pr(\mathrm{TO}_{t+1} \geq x)$ emphasizing tail behavior. The exceedance probability $\Pr(\mathrm{TO}_{t+1} > \theta)$ is nonzero for WPROD-R (refresh days) and effectively zero for MV in this dataset, consistent with the refresh decomposition in Theorem 6.

9.4. Portfolio dynamics and diagnostics

To better understand the source of the performance differential, we examine the internal dynamics of WPROD-R. The reduced weights $\mathbf{u}_t \in \Delta_m$ represent the allocation among the m time-varying basis portfolios and evolve under the mirror discretization of a replicator-type ODE. Figure 5a displays $\mathbf{u}_t$ over the backtest. The inertia parameter $\lambda = 0.8$ induces gradual changes by averaging the target update with the previous iterate, which limits abrupt shifts in the reduced allocation.



(a) Evolution of reduced weights $\mathbf{u}_t \in \Delta_4$.

(b) Asset-level weights $\mathbf{w}_t$ for WPROD-R vs. MV.

Figure 5. Internal dynamics of the WPROD-R model. (a) The reduced weights $\mathbf{u}_t$ evolve smoothly on the simplex Δ_4, representing allocations to latent basis portfolios. (b) The resulting asset weights $\mathbf{w}_t$ are compared to the MV portfolio, showing more stable and diversified positions.

These reduced weights are then mapped to the investable asset weights $\mathbf{w}_t = A_t\mathbf{u}_t$. Figure 5b compares the resulting asset-level allocations for WPROD-R with those of the classical MV benchmark for the top five assets by final weight. Relative to MV, WPROD-R yields more stable and diversified allocations and tends to avoid extreme corner solutions. This behavior is consistent with the softmax construction of the basis portfolios A_t and the smoothing effect of the mirror-flow dynamics.

To quantify the difference between the two strategies, we plot the ℓ_1 distance between the WPROD-R and MV portfolio weights, $(1/2)\|\mathbf{w}_t^{\mathrm{WPROD}} - \mathbf{w}_t^{\mathrm{MV}}\|_1$, over time in Figure 6. This distance, which corresponds to the one-way turnover required to switch from one portfolio to the other, serves as a proxy for disagreement between the two models. The plot reveals significant spikes during periods of market stress, such as in 2020 and 2022. These spikes indicate that during turbulent regimes, the multiscale covariance geometry Σ_t of the WPROD-R model induces a substantially different view of risk and opportunity compared to the single-scale covariance matrix used by the MV optimizer, leading to divergent allocations and, consistent with the performance results reported above, improved risk-adjusted outcomes.

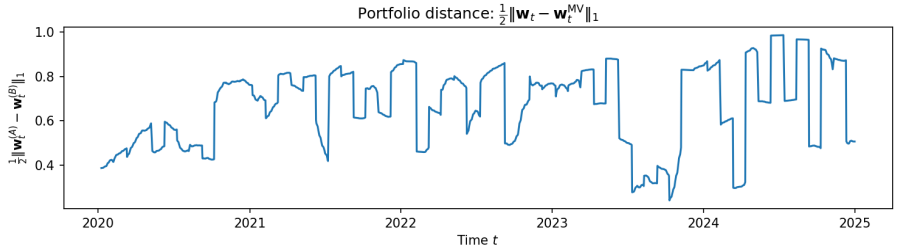


Figure 6. Time series of the ℓ_1 distance between WPROD-R and MV portfolio weights. Spikes in the distance highlight periods of market stress when the two models' optimal allocations diverge significantly, reflecting the different underlying geometric structures.

9.5. Validation of theoretical bounds

A key contribution of the WPROD-R framework is the establishment of theoretical guarantees for the stability and approximation quality of portfolio dynamics. We empirically validate these theoretical results by examining the impact of the key model parameters, or knobs, that control these bounds.

9.5.1. Turnover control via inertia (λ)

Theorem 5 establishes a deterministic upper bound on the turnover of the reduced weights, which is controlled by the inertia parameter λ. Specifically, under a fixed basis A_t, the one-way turnover of asset weights is bounded: $\|\mathbf{w}_{t+1} - \mathbf{w}_t\|_1 \leq 2(1-\lambda)$. To validate this, we conducted an experiment in which the PCA basis was held fixed, and we swept through different values of the learning rate η and the inertia parameter λ.

Figure 7 plots the maximum observed one-way turnover against the theoretical bound, $1-\lambda$. The empirical results confirm the theoretical prediction: the maximum turnover remains well below the bound and decreases monotonically as λ approaches 1. This provides strong evidence for the effectiveness of the inertia mechanism as a tool for explicit turnover control, which is crucial for managing transaction costs in practical settings. Table 12 further details this relationship, showing how the average turnover

responds to changes in both η and λ. Although a higher learning rate η increases turnover as expected, the dominant factor remains inertia λ.

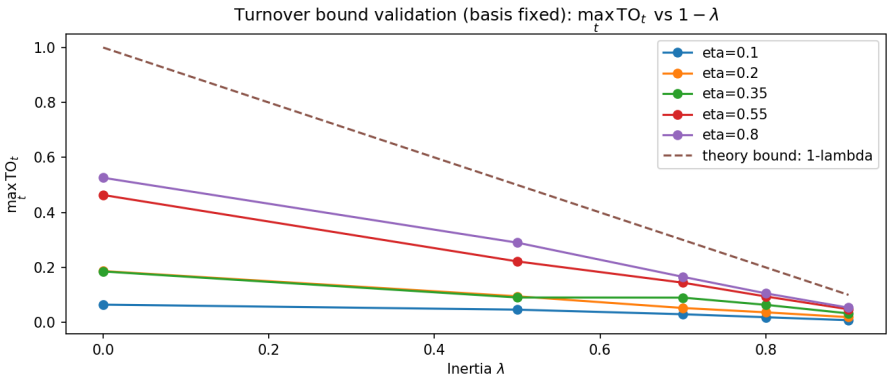


Figure 7. Empirical validation of the turnover bound. The maximum observed one-way turnover is plotted against the inertia parameter λ under a fixed basis, confirming the theoretical bound $\mathrm{TO}_t \leq 1 - \lambda$.

Table 12. Average daily turnover (%) as a function of η and λ.

	Inertia (λ)				
Learning Rate (η)	0.0	0.5	0.7	0.8	0.9
0.10	0.48	0.38	0.29	0.24	0.15
0.35	0.79	0.64	0.58	0.53	0.42
0.80	1.40	1.56	1.48	1.29	0.90

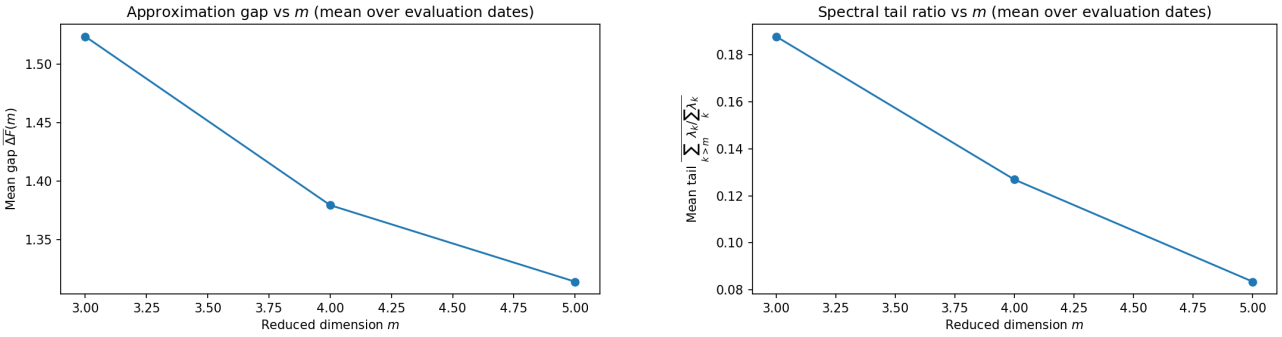
9.5.2. Reduced-order approximation quality (m)

The second key theoretical aspect is the quality of the reduced-order approximation, which is controlled by the number of principal components, m. The framework posits that a small number of basis portfolios can effectively span the space of desirable allocation. We test this by analyzing the approximation gap, defined as the difference in the objective function value between the optimal portfolio in the full asset space and the best portfolio within the reduced subspace spanned by A_t.

Table 13 shows that as m increases, both the approximation gap and the spectral tail ratio (the proportion of variance not captured by the top m eigenvectors) decrease. This confirms that adding more basis portfolios improves the reduced model's expressive power. However, the improvement exhibits diminishing returns, as shown in Figure 8. The approximation gap begins to saturate for $m > 4$, suggesting that a small number of principal components is sufficient to capture the most significant risk-return trade-offs. This finding provides practical guidance for model selection, indicating that a parsimonious choice of m can achieve near-optimal performance while minimizing the model complexity and risk of overfitting.

Table 13. Approximation gap and spectral tail vs. reduction dimension m.

Dimension (m)	Approx. Gap	Spectral Tail Ratio
3	1.524	0.188
4	1.379	0.127
5	1.314	0.083



(a) Mean approximation gap vs. m.

(b) Mean spectral tail ratio vs. m.

Figure 8. Validation of the reduced-order approximation. (a) The approximation gap shrinks as m increases, indicating that the reduced model better approximates the full optimum. (b) The spectral tail ratio, which represents the uncaptured variance, also diminishes. Both plots show saturating benefits for $m > 4$.

9.6. Sensitivity to model hyperparameters

The performance of the WPROD-R model depends on the structural choices regarding the multiscale geometry and dimensionality of the reduced space. We now investigate the sensitivity of the model to two key hyperparameters: the wavelet family, ψ, which determines the nature of the multiscale decomposition, and the number of principal components, m, which controls the degree of dimensionality reduction.

Table 14 reports a simple grid over five wavelet families and three values of m. Because the wavelet layer enters *structurally*—it determines both the multiscale risk geometry and the risk-normalized signal—changing ψ can materially alter the resulting policy. In our full-sample grid, the `coif1` (Coiflet) wavelet attains the highest net Sharpe ratio (1.52 at $m = 3$) and performs strongly across all tested m. At the same time, the spread across wavelets illustrates an important methodological point: ψ should be treated as an operator-level *design parameter*, not as a universal recommendation. To guard against reporting a peak of an idiosyncratic U-shaped curve, the appropriate protocol is walk-forward (time-series) validation across multiple subperiods; we adopt this protocol in the computational workflow and report a time-split robustness check in the appendix. Figure 9 is therefore used here primarily as a sensitivity diagnostic, highlighting the region of hyperparameters that warrants robust out-of-sample checking.

Table 14. Net annualized Sharpe ratio across wavelet families (ψ) and reduction dimensions (m).

	Reduction Dimension (m)		
Wavelet Family (ψ)	3	4	5
db2	1.33	1.28	1.24
db4 (Baseline)	1.34	1.18	1.18
sym4	1.35	1.38	1.35
coif1	**1.52**	1.50	1.49
haar	1.22	1.09	1.11

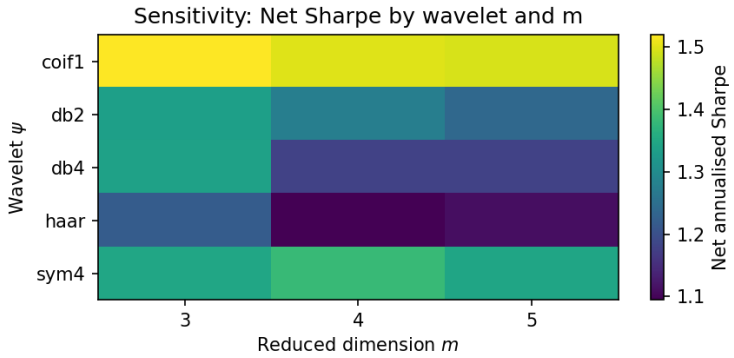
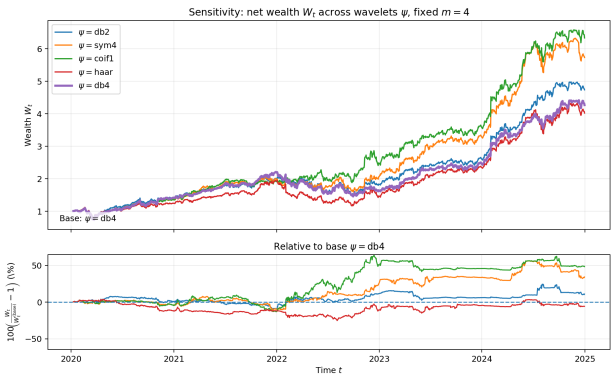
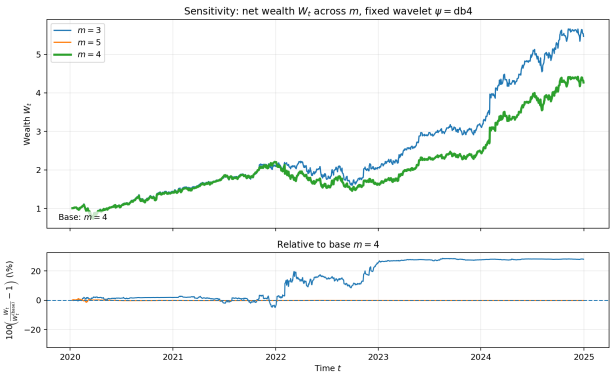


Figure 9. Heatmap of net annualized Sharpe ratio over the hyperparameter grid of wavelet family (ψ) and reduction dimension (m). In this sample, the `coif1` wavelet lies in the highest-performing region; the figure is intended as a sensitivity diagnostic rather than a claim of universality.



(a) Net wealth curves across wavelet families at fixed $m = 4$.



(b) Net wealth curves across reduction dimensions m at fixed wavelet family `db4`.

Figure 10. Sensitivity of wealth trajectories to hyperparameter choices. The plots show that both the wavelet family and the reduction dimension materially alter the portfolio's long-term performance.

To further illustrate the economic significance of these choices, Figure 10 presents net wealth curves for different hyperparameter settings. Figure 10a shows the performance divergence when varying the wavelet family at a fixed $m = 4$. In this sample, the `coif1` and `sym4` wavelets generate higher terminal wealth than the baseline `db4`. Similarly, Figure 10b shows that although increasing m from 3 to 5 with the `db4` wavelet does not yield a monotonic improvement, the differences in terminal wealth are still notable. Finally, Figure 11 directly compares the equity curve of the best-performing configuration (i.e., `coif1`, $m = 3$) against the baseline, illustrating the performance differences associated with hyperparameter selection in this sample.

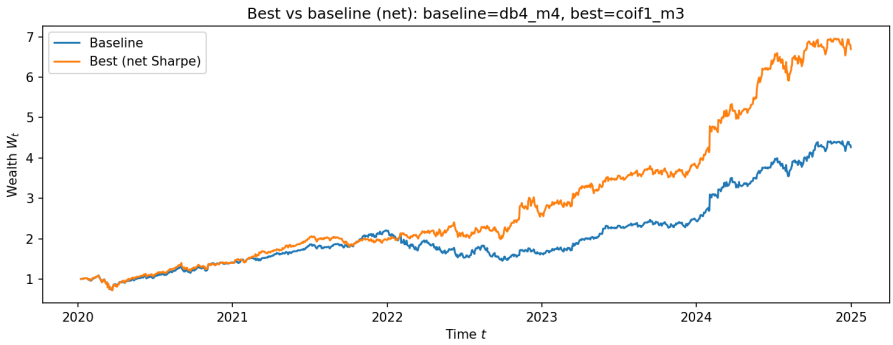


Figure 11. Comparison of the best configuration (`coif1`, $m = 3$) against the baseline (`db4`, $m = 4$). In this sample, the selected configuration yields higher net wealth; operator choices should be validated out-of-sample to assess stability.

9.7. Robustness across market regimes and tracking performance

A desirable feature of any adaptive strategy is its ability to perform robustly across different market conditions. To assess WPROD-R's performance, we classify the backtesting period into three distinct regimes—Low-, Mid-, and High-volatility—based on the rolling volatility of the equal-weight benchmark. Table 15 reports the performance of the WPROD-R model in each of these regimes.

Table 15. Net performance metrics by market regime.

Regime	Days	Sharpe Ratio	CAGR (%)	Avg. Turnover (%)
High-volatility	405	0.72	18.44	2.45
Mid-volatility	393	1.98	47.16	2.73
Low-volatility	393	2.23	57.07	2.06

The model exhibits stronger performance in low- and mid-volatility environments, with Sharpe ratios of 2.23 and 1.98, respectively. Performance deteriorates in the high-volatility regime but remains positive. Average turnover does not increase markedly during high-volatility periods, which is consistent with the stabilizing role of inertia and entropy regularization in the update rule.

From a theoretical standpoint, the WPROD-R model is an online learning algorithm designed to track time-varying optimal portfolios. A key measure of its effectiveness is dynamic regret, which quantifies the suboptimality of the algorithm's decisions relative to a clairvoyant benchmark. We construct an empirical proxy for the instantaneous regret, $J_t(u_t^*) - J_t(u_t)$, where u_t^* is the (in-sample) optimal reduced allocation at time t. Figure 12 shows the regret proxy over time.

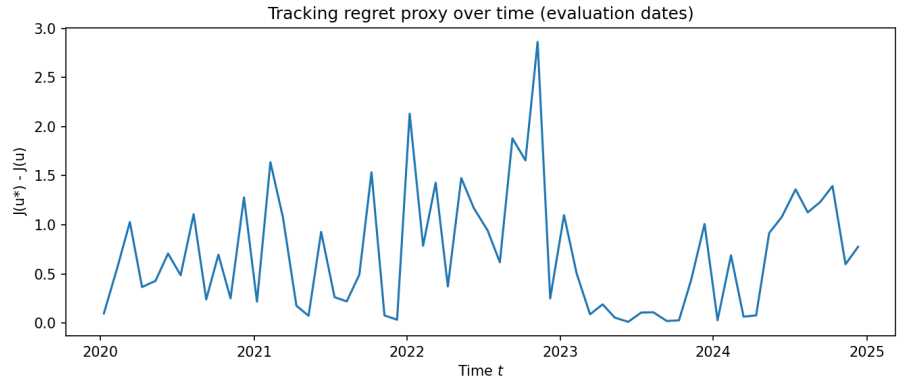


Figure 12. Time series of the instantaneous regret proxy. The plot shows the difference between the objective function value of the optimal in-sample allocation and the one chosen by the model, providing a measure of tracking efficiency.

Regret is generally low and stable but exhibits significant spikes that coincide with major market dislocations, such as the onset of the COVID-19 pandemic. These spikes represent periods when the market structure changes so rapidly that the adaptive model cannot keep pace with the ex-post optimal strategy. Table 16 confirms this intuition by showing that the mean and maximum regret are substantially higher in the high-volatility regime. This analysis provides an empirical analogue to the theoretical regret bounds discussed in the literature, demonstrating that the tracking performance of the model is directly linked to the stability of the underlying market environment.

Table 16. Distribution of instantaneous regret by market regime.

Regime	Mean	Median	Max	Count
High-volatility	0.956	0.824	2.862	20
Mid-volatility	0.637	0.327	2.131	18
Low-volatility	0.540	0.492	1.535	19

9.8. Ablation study: deconstructing the model

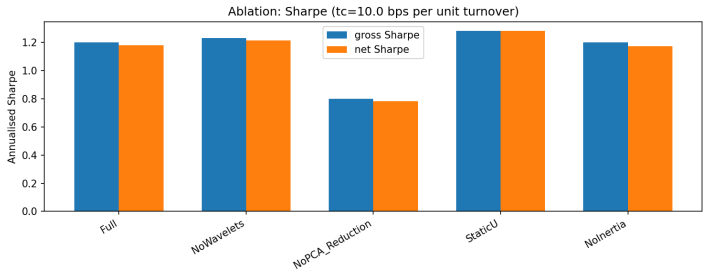
To isolate the contribution of each core component of the WPROD-R framework, we conducted an ablation study. We systematically removed or disabled each key mathematical layer of the model and evaluated its impact on performance. The variants considered are as follows:

- **NoWavelets**: The multiscale wavelet covariance Σ_t is replaced with a standard single-scale rolling covariance matrix. This tests the value of multiscale geometry.
- **NoPCA_Reduction**: The PCA-based dimensionality reduction is disabled. Mirror descent dynamics operate directly in the full N-dimensional asset space. This tests the role of the reduced-order manifold in the flow field.
- **StaticU**: The reduced weights $\mathbf{u}_t$ are held constant at a uniform allocation, disabling the mirror-flow dynamics. This tests the contribution of adaptive allocation mechanisms.
- **NoInertia**: The inertia parameter λ is set to 0, removing the smoothing component of the weight updates.

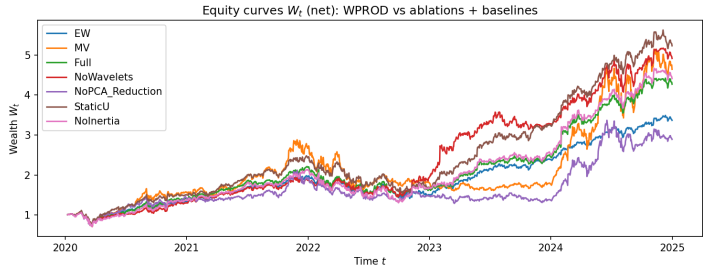
Table 17 reports the net performance metrics for each ablation. Removing the wavelet layer (NoWavelets) yields a modest improvement in this particular sample, suggesting that a simpler covariance estimate can suffice for the assets and period considered. In contrast, removing the dimensionality-reduction layer (NoPCA_Reduction) substantially degrades risk-adjusted performance: the net Sharpe ratio drops from 1.18 to 0.78. This pattern is consistent with the reduced-order representation acting as a regularizer that mitigates estimation noise in a high-dimensional simplex. The StaticU variant isolates the contribution of dynamic allocation; although it performs well in this backtest, its zero turnover makes it primarily a diagnostic comparison rather than a feasible trading strategy.

Table 17. Ablation study: Net performance metrics.

Variant	Sharpe Ratio	CAGR (%)	Volatility (%)	Max DD (%)	Avg. Turnover (%)
Full Model	1.18	33.90	28.12	−34.17	2.39
NoWavelets	1.21	37.74	30.14	−33.61	2.08
NoPCA_Reduction	0.78	23.80	35.13	−38.74	2.17
StaticU	1.28	39.48	29.32	−37.59	0.00
NoInertia	1.17	34.78	29.04	−38.95	2.95



(a) Net Sharpe ratios for ablation variants.



(b) Net wealth curves for ablation variants.

Figure 13. Visual summary of the ablation study. (a) The bar plot illustrates the contribution of the PCA reduction to the model's risk-adjusted performance. (b) The equity curves show the divergent wealth paths resulting from the removal of key model components.

Figure 13 provides a compact visual summary of the ablation study. The bar plot of Sharpe ratios (Figure 13a) highlights the performance loss when key components are removed, most notably the PCA reduction. The equity curves (Figure 13b) show the corresponding changes in wealth trajectories, with

the NoPCA_Reduction variant exhibiting higher volatility and deeper drawdowns than the other variants. Overall, the ablation results indicate that the reduced-order representation and the mirror-flow dynamics are important contributors to the observed risk-adjusted performance, and the incremental benefit of the wavelet layer is more sample dependent.

10. Conclusion and future work

This paper introduces the wavelet–PCA reduced-order dynamics (WPROD-R) framework, a mathematically grounded approach to online portfolio optimization that combines a wavelet-induced multiscale covariance geometry, dimensionality reduction, and entropy-regularized mirror flow dynamics on the simplex. The theoretical development provides a coherent chain of stability and control results: bounded multiscale features lead to well-posed covariance updates; stable principal directions induce a simplex-feasible reduced-order parameterization through the PCA–softmax basis map; and the resulting mirror flow admits explicit turnover control through inertia and basis refresh, together with regret-style diagnostic guarantees.

The empirical study reported in the main text shows that, within the tested sample and protocol, WPROD-R can deliver improved risk-adjusted performance relative to standard baselines while maintaining stable trading behavior under turnover control. The ablation evidence further indicates that PCA-based reduction plays a central role in the framework: restricting the decision process to a low-dimensional manifold is consistent with a regularization effect that mitigates noise amplification in allocation decisions. Beyond the core experiments, the appendix reports an additional evaluation in a substantially higher-dimensional setting (106 assets), which serves as a stress test of the wavelet–PCA coupling and illustrates how the reduced-order design behaves when the ambient dimension of the simplex is materially larger, see Table C3. Readers are referred to the appendix for the full high-dimensional setup and results.

A limitation intrinsic to operator-driven multiscale constructions is sensitivity to the choice of wavelet family. Because the wavelet operator enters structurally—affecting both the multiscale risk geometry and the signal—variation in realized performance across wavelet families should be interpreted as a model-selection issue rather than evidence of universal dominance of any single operator. Accordingly, selecting ψ and related design parameters such as J, L, and the scale weights is best guided by walk-forward validation across multiple out-of-sample subperiods and, when possible, by economically meaningful horizon considerations. In this manuscript, wavelet-grid results are reported as diagnostics, and operator sensitivity is treated as an explicit caveat.

For transparency and reproducibility, the appendix consolidates the model setting and experimental design in a single place. In particular, the appendix summarizes the core hyperparameters and their mathematical roles in the multiscale reduced-order mirror-flow model, and it records the sensitivity and ablation protocols used to validate turnover control, reduced-order approximation behavior, and operator dependence. This organization is intended to make the design space explicit and to facilitate replication without interrupting the main narrative.

Several directions for future research emerge naturally from the present formulation and diagnostics. First, although the appendix provides evidence in a higher-dimensional universe, further scaling to substantially larger universes would sharpen understanding of the computational and statistical trade-offs in the wavelet–PCA coupling and may motivate hierarchical, sparse, or structured PCA variants. Second,

although the paper considers proportional transaction costs (and related robustness checks), extending the decision rule to incorporate market impact and liquidity effects directly would improve practical relevance, particularly in high-dimensional settings where basis refresh and rebalancing can interact with trading frictions. Third, developing principled adaptive mechanisms for hyperparameters such as η, λ, τ, and the refresh period K could improve responsiveness under nonstationarity while preserving the stability guarantees established in the paper. Finally, the theoretical analysis could be strengthened by establishing high-probability guarantees for regret and turnover under stochastic assumptions, providing sharper control statements in risk-sensitive environments.

Use of Generative-AI tools declaration

The author declare he has used Large Language Models (LLMs) in the creation of this article.

The AI tools were employed solely for language polishing, including grammar correction, improving sentence structure, enhancing clarity, and organizing the presentation of content.

AI-assisted language refinements were applied across all sections of the manuscript. The author reviewed and approved all modifications.

Acknowledgments

The author would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work.

Conflict of interest

The author declares that there is no conflict of interest in this paper.

References

1. X. Wu, Y. Xiong, A fractal market perspective on improving futures pricing and optimizing cash-and-carry arbitrage strategies, *Quant. Financ. Econ.*, **9** (2025), 713–744. https://doi.org/10.3934/QFE.2025025
2. Z. Li, B. Chen, S. Lu, G. Liao, The impact of financial institutions' cross-shareholdings on risk-taking, *Int. Rev. Econ. Financ.*, **92** (2024), 1526–1544. https://doi.org/10.1016/j.iref.2024.02.080
3. P. Joshi, Regime-specific interdependencies in cryptocurrency markets: A high-frequency gmm-var approach, *Data Sci.n Financ. Econ.*, **5** (2025), 419–439 https://doi.org/10.3934/DSFE.2025017
4. J. Pekár, I. Brezina, M. Reiff, Green investments: Portfolio selection based on risk measure and esg indicators. impact of environmental indicators on portfolio selection, *Green Financ.*, **7** (2025), 223–246. https://doi.org/10.3934/GF.2025009
5. M. Murgui-García, J. Ruiz-Tamarit, Economic resilience in the short-run: A dynamic macroeconomic approach, *Nat. Account. Rev.*, **7** (2025), 290–308. https://doi.org/10.3934/NAR.2025013

6. H. Zhao, Y. Chen, X. Wang, D. Wang, H. Xu, W. Deng, Joint optimization scheduling using AHMQDE-ACO for key resources in smart operations, *IEEE T. Consum. Electr.*, **71** (2025), 9261–9273. https://doi.org/10.1109/TCE.2025.3619833
7. W. Deng, X. Li, Y. Sun, H. Zhao, Privacy protection-enhanced vertical-horizontal federated learning secure sharing for multisource heterogeneous data, *IEEE T. Ind. Inform.*, 2026, 1–10. https://doi.org/10.1109/TII.2025.3649540
8. W. Deng, H. Li, H. Zhao, Anti-noise bearing fault diagnosis using time-reassigned multisynchrosqueezing transform and complex sparse learning dictionary, *IEEE T. Instrum. Meas.*, **71** (2025), 3557310. https://doi.org/10.1109/TIM.2025.3604987
9. T. Conlon, J. Cotter, An empirical analysis of dynamic multiscale hedging using wavelet decomposition, *J. Futures Markets*, **32** (2012), 272–299. https://doi.org/10.1002/fut.20519
10. M. U. Torun, A. N. Akansu, M. Avellaneda, Portfolio risk in multiple frequencies, *IEEE Signal Proc. Mag.*, **28** (2011), 61–71. https://doi.org/10.1109/MSP.2011.941552
11. C. Abad, G. Iyengar, Portfolio selection with multiple spectral risk constraints, *SIAM J. Financ. Math.*, **6** (2014), 467–486. https://doi.org/10.1137/140967635
12. F. Fernández-Navarro, L. Martínez-Nieto, M. Carbonero-Ruz, T. Montero-Romero, Mean squared variance portfolio: A mixed-integer linear programming formulation, *Mathematics*, **9** (2021), 223. https://doi.org/10.3390/math9030223
13. V. A. Nguyen, S. Shafiee, D. Filipovi'c, D. Kuhn, Mean-covariance robust risk measurement, *Social Science Research Network*, 2021, arXiv: 2112.09959. https://doi.org/10.48550/arXiv.2112.09959
14. Y. S. Kim, R. Giacometti, S. T. Rachev, F. J. Fabozzi, D. Mignacca, Measurin financial risk and portfolio optimization with a non-Gaussian multivariate model, *Ann. Oper. Res.*, **2012** (2012), 325–343. https://doi.org/10.1007/s10479-012-1229-8
15. P. A. Bilokon, Implementing portfolio risk management and hedging in practice, 2023, arXiv: 2309.15767v1. https://doi.org/10.48550/arXiv.2309.15767
16. K. Vu, P. L. Poirion, C. D'Ambrosio, L. Liberti, Random projections for quadratic programs over a Euclidean ball, In: *Integer Programming and Combinatorial Optimization: 20th International Conference*, 2019, 442–452. https://doi.org/10.1007/978-3-030-17953-3_33
17. A. Frangioni, F. Furini, C. Gentile, Improving the approximated projected perspective reformulation by dual information, *Oper. Res. Lett.*, **45** (2017), 519–524. https://doi.org/10.1016/j.orl.2017.08.001
18. W. Shen, J. Wang, Transaction costs-aware portfolio optimization via fast Löwner-John ellipsoid approximation, In: *AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, 1854–1860.
19. R. Mansini, W. Ogryczak, M. G. Speranza, Twenty years of linear programming based portfolio optimization, *Eur. J. Oper. Res.*, **234** (2014), 518–535. https://doi.org/10.1016/j.ejor.2013.08.035
20. M. Harper, D. Fryer, Lyapunov functions for time-scale dynamics on Riemannian geometries of the simplex, *Dyn. Games Appl.*, **5** (2015), 318–333. https://doi.org/10.1007/s13235-014-0124-0
21. M. Raginsky, J. Bouvrie, Continuous-time stochastic Mirror Descent on a network: Variance reduction, consensus, convergence, In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2012, 6793–6800. https://doi.org/10.1109/CDC.2012.6426639

22. T. Zrnic, E. Mazumdar, A note on Zeroth-order optimization on the simplex, 2022, arXiv: 2208.01185. https://doi.org/10.48550/arXiv.2208.01185

23. T. Lehmann, Darwinian evolution as Brownian motion on the simplex: A geometric perspective on stochastic replicator dynamics, *Ann. Appl. Probab.*, **33** (2023), 344–375. https://doi.org/10.1214/22-AAP1817

24. J. Wang, N. Elia, A control perspective for centralized and distributed convex optimization, In: *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 2011, 3800–3805. https://doi.org/10.1109/CDC.2011.6161503

25. M. J. Best, J. Hlouskova, An algorithm for portfolio optimization with transaction costs, *Manage. Sci.*, **51** (2005), 1676–1688. https://doi.org/10.1287/mnsc.1050.0418

26. D. B. Brown, J. E. Smith, Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds, *Manage. Sci.*, **57** (2011), 1752–1770. https://doi.org/10.1287/mnsc.1110.1377

27. P. J. J. Herings, K. Schmedders, Computing equilibria in finance economies with incomplete markets and transaction costs, **27** (2006), 493–512. https://doi.org/10.1007/s00199-004-0583-4

28. S. Dughmi, J. D. Hartline, R. D. Kleinberg, R. Niazadeh, Bernoulli factories and black-box reductions in mechanism design, *J. ACM*, **68** (2021), 10. https://doi.org/10.1145/3440988

29. F. Hutter, H. H. Hoos, K. Leyton-Brown, Tradeoffs in the empirical evaluation of competing algorithm designs, *Ann. Math. Artif. Intell.*, **60** (2010), 65–89. https://doi.org/10.1007/s10472-010-9191-0

30. C. McGeoch, Analyzing algorithms by simulation: Variance reduction techniques and simulation speedups, *ACM Comput. Surv.*, **24** (1992), 195–212. https://doi.org/10.1145/130844.130853

31. M. Potkonjak, J. Rabaey, Algorithm selection: A quantitative optimization-intensive approach, In: *ICCAD '94: Proceedings of the 1994 IEEE/ACM international conference on Computer-aided design*, 1994, 90–95.

32. J. Silva, A survey on algorithmic debugging strategies, *Adv. Eng. Softw.*, **42** (2011), 976–991. https://doi.org/10.1016/j.advengsoft.2011.05.024

33. M. Gharakhani, F. Z. Fazlelahi, S. J. Sadjadi, A robust optimization approach for index tracking problem, *J. Comput. Sci.*, **10** (2014), 2450–2463. https://doi.org/10.3844/jcssp.2014.2450.2463

34. S. Penev, P. V. Shevchenko, W. Wu, Myopic robust index tracking with Bregman divergence, *Quant. Financ.*, **22** (2021), 289–302. https://doi.org/10.1080/14697688.2021.1950918

35. L. Carassus, J. Obłój, J. Wiesel, The robust superreplication problem: A dynamic approach, *SIAM J. Financ. Math.*, **10** (2019), 907–941. https://doi.org/10.1137/18M1235934

36. L. P. Hansen, T. J. Sargent, Robust control and model uncertainty, *Am. Econ. Rev.*, **91** (2001), 60–66. https://doi.org/10.1257/aer.91.2.60

37. C. S. Pun, Robust time-inconsistent stochastic control problems, *Automatica*, **94** (2018), 249–257. https://doi.org/10.1016/j.automatica.2018.04.038

38. A. Ismail, H. Pham, Robust Markowitz mean-variance portfolio selection under ambiguous covariance matrix, *Math. Financ.*, **29** (2019), 174–207. https://doi.org/10.1111/mafi.12169

39. D. Bertsimas, M. Sim, The price of robustness, *Oper. Res.*, **52** (2004), 35–53. https://doi.org/10.1287/opre.1030.0065

40. F. J. Fabozzi, D. Huang, G. Zhou, Robust portfolios: Contributions from operations research and finance, *Ann. Oper. Res.*, **176** (2010), 191–220. https://doi.org/10.1007/s10479-009-0515-6

41. S. Abdelfattah, K. Kasmarik, J. Hu, A robust policy bootstrapping algorithm for multi-objective reinforcement learning in non-stationary environments, *Adapt. Behav.*, **28** (2019), 273–292.

42. W. Deng, H. Li, H. Zhao, Antinoise bearing fault diagnosis using time-reassigned multisynchrosqueezing transform and complex sparse learning dictionary, *IEEE T. Instrum. Meas.*, **74** (2025), 3557310. https://doi.org/10.1109/TIM.2025.3604987

43. C. Wang, Y. Peng, W. Deng, A dendrite net learning multi-objective artificial bee colony algorithm for UAV, *Appl. Soft Comput.*, **189** (2026), 114449. https://doi.org/10.1016/j.asoc.2025.114449

44. Y. Song, C. Song, Adaptive evolutionary multitask optimization based on anomaly detection transfer of multiple similar sources, *Expert Syst. Appl.*, **283** (2025), 127599. https://doi.org/10.1016/j.eswa.2025.127599

45. M. F. Leung, J. Wang, Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization, *IEEE T. Neur. Net. Lear.*, **32** (2021), 2825–2836. https://doi.org/10.1109/TNNLS.2019.2957105

46. A. Beck, M. Teboulle, Mirror descent and nonlinear projected subgradient methods for convex optimization, *Oper. Res. Lett.*, **31** (2003), 167–175. https://doi.org/10.1016/S0167-6377(02)00231-6

47. J. Kivinen, M. K. Warmuth, Exponentiated gradient versus gradient descent for linear predictors, *Inform. Comput.*, **132** (1997), 1–63. https://doi.org/10.1006/inco.1996.2612

48. N. Cesa-Bianchi, G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2009. https://doi.org/10.1017/CBO9780511546921

49. O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, *J. Multivariate Anal.*, **88** (2004), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

50. S. Maillard, T. Roncalli, J. Teiletche, The properties of equally weighted risk contributions portfolios, *J. Portfoli. Manag.*, **36** (2010), 60–70. https://doi.org/10.3905/jpm.2010.36.4.060

Appendix

A. Core hyperparameters and sensitivity protocols

The proposed framework is fully determined by a small set of *core mathematical hyperparameters* that control (i) the multiscale wavelet-induced covariance geometry, (ii) the associated risk-normalized signal, (iii) the PCA–softmax investable basis and its refresh schedule, (iv) the reduced mirror-flow dynamics on the simplex, and (v) the frictional accounting through turnover and transaction costs. For transparency and reproducibility, Table A1 records these parameters using the same notation as in the theoretical development and Algorithm 1. Table A2 summarizes the sensitivity and ablation protocols used to validate the theoretical mechanisms (turnover control, reduced-order approximation, operator dependence) and assess robustness under nonstationarity and trading frictions.

Table A1. Core mathematical hyperparameters of the multiscale reduced-order mirror-flow model (WPROD-R), expressed in the notation of the theory and Algorithm 1.

Symbol	Typical values	Mathematical role in the model
Reduced-order representation and investable basis		
m	$m \in \{3, 4, 5\}$	Reduced dimension: reduced weights $\boldsymbol{u}_t \in \Delta_m$ generate traded weights $\boldsymbol{w}_t = A_t \boldsymbol{u}_t \in \Delta_N$.
S_t	$S_t = \bar{\Sigma}_t$	Symmetric geometry used to compute PCA directions $\{\boldsymbol{v}_k(t)\}_{k=1}^m$.
τ	$\tau > 0$	Softmax temperature in the basis map $\boldsymbol{a}_k(t) = \mathrm{softmax}(\tau \boldsymbol{v}_k(t)) \in \Delta_N$ (controls the concentration of basis portfolios).
K	$K \in \mathbb{N}$ (e.g. $K = 21$)	Basis refresh period: A_t is recomputed at refresh times and held constant between refreshes (piecewise-constant A_t).
Wavelet multiscale layer (operators and geometry)		
ψ	for example db4, sym4, coif1, haar	Wavelet family defining the multiscale operators (SWT instantiation of smooth/detail components).
J	$J \in \mathbb{N}$ (e.g. $J = 3$)	Number of scales; defines detail features $\{\boldsymbol{W}_{j,t}\}_{j=1}^J$ and scale covariances $\{\Sigma_{j,t}\}_{j=1}^J$.
L	$L \in \mathbb{N}$ (e.g. $L = 256$)	Rolling window length for multiscale features; constrained by $L \equiv 0 \pmod{2^J}$ for SWT.
α	$\alpha \in (0, 1)$	EWMA rate in $\Sigma_{j,t} = (1-\alpha)\Sigma_{j,t-1} + \alpha \boldsymbol{W}_{j,t}\boldsymbol{W}_{j,t}^\top + \varepsilon I_N$.
$\boldsymbol{\omega}$	$\boldsymbol{\omega} \in \Delta_J$	Scale aggregation weights in the multiscale geometry $\Sigma_t = \sum_{j=1}^J \omega_j \Sigma_{j,t}$.
Multiscale signal construction and smoothing		
$\boldsymbol{d}$	$\boldsymbol{d} \in \Delta_J$	Detail mixture weights in $\sum_{j=1}^J d_j \boldsymbol{W}_{j,t}$.
λ_d	$\lambda_d \geq 0$	Detail coefficient in $\boldsymbol{z}_t = \boldsymbol{s}_t + \lambda_d \sum_{j=1}^J d_j \boldsymbol{W}_{j,t}$.
κ_μ	$\kappa_\mu > 0$	Signal scale in $\boldsymbol{\mu}_t = \kappa_\mu(\tilde{\boldsymbol{\mu}}_t - \frac{1}{N}(\mathbb{1}^\top \tilde{\boldsymbol{\mu}}_t)\mathbb{1})$, with $\tilde{\boldsymbol{\mu}}_t = \mathrm{diag}(\Sigma_t)^{-1/2}\boldsymbol{z}_t$.
β	$\beta \in (0, 1]$	Signal smoothing: $\widehat{\boldsymbol{\mu}}_t = (1-\beta)\widehat{\boldsymbol{\mu}}_{t-1} + \beta \boldsymbol{\mu}_t$.
Reduced mirror-flow dynamics on the simplex		
γ	$\gamma > 0$	Risk aversion in the reduced objective $J_t(\boldsymbol{u}) = \boldsymbol{\mu}_t^{(u)\top}\boldsymbol{u} - \frac{\gamma}{2}\boldsymbol{u}^\top \bar{\Sigma}_t^{(u)}\boldsymbol{u} + \rho H(\boldsymbol{u})$.
ρ	$\rho \geq 0$	Entropy regularization weight; $H(\boldsymbol{u}) = -\sum_k u_k \log u_k$.
η	$\eta > 0$	Mirror step size in $\boldsymbol{u}_{t+1}^{\mathrm{tar}} \propto \boldsymbol{u}_t \odot \exp(\eta \nabla J_t(\boldsymbol{u}_t))$.
λ	$\lambda \in [0, 1)$	Inertia parameter in $\boldsymbol{u}_{t+1} = \lambda \boldsymbol{u}_t + (1-\lambda)\boldsymbol{u}_{t+1}^{\mathrm{tar}}$ (turnover control).
Stabilization, feasibility floors, and frictions		
ε	$\varepsilon > 0$	Ridge term in each $\Sigma_{j,t}$ update (ensures strict positive definiteness).
ε_0	$\varepsilon_0 > 0$	Ridge term in $\bar{\Sigma}_t = \Sigma_t + \varepsilon_0 I_N$ (stabilized multiscale geometry).
ε_u	$\varepsilon_u > 0$	Ridge term in $\bar{\Sigma}_t^{(u)} = A_t^\top \bar{\Sigma}_t A_t + \varepsilon_u I_m$ (stabilized reduced geometry).
$\epsilon_{\min}$	$\epsilon_{\min} > 0$	Positivity floor to avoid $\log 0$ and ensure strict interior iterates on simplex.
c	$c \geq 0$	Proportional cost rate in $R_{t+1}^{\mathrm{net}} = R_{t+1}^{\mathrm{gross}} - c\,\mathrm{TO}_{t+1}$.
TO_{t+1}	$\mathrm{TO}_{t+1} = \frac{1}{2}\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\|_1$	One-way turnover used in frictional accounting and turnover-control statements.

Table A2. Sensitivity and robustness protocols used to connect empirical behavior to the core theoretical mechanisms: turnover control via inertia, reduced-order approximation via PCA truncation, and multiscale operator dependence via wavelet choice.

Sensitivity / robustness protocol	Varied quantity (grid)	Mathematical purpose and diagnostic interpretation
Turnover-control validation	(η, λ) on a grid; basis held fixed ($K \leq 0$)	Tests the theoretical turnover mechanism induced by inertia in the mirror-flow and isolates turnover generated purely by reduced dynamics (no basis-refresh contribution).
Reduced-order approximation study	$m \in \{3, 4, 5\}$	Evaluates how the reduced feasible set $\{\boldsymbol{w} = A_t \boldsymbol{u} : \boldsymbol{u} \in \Delta_m\}$ approaches the full simplex as m increases; supports the spectral-tail/approximation-gap narrative.
Operator dependence (multiscale sensitivity)	Wavelet family ψ (often combined with m)	Quantifies sensitivity to the multiscale operator choice; supports that the wavelet layer is structural (affects geometry and signal) rather than a cosmetic preprocessing step.
Ablation analysis (component isolation)	Remove one component at a time: wavelet layer, PCA reduction, mirror dynamics, inertia	Provides evidence that each mathematical ingredient is necessary for the observed behavior; clarifies which component drives stability, adaptivity, or turnover control.
Regime robustness	Conditioning on volatility regimes (low/mid/high)	Tests the stability of the dynamics under nonstationarity; connects to tracking-style interpretations when the environment changes over time.
Transaction-cost stress	Vary cost rate c	Verifies that improvements persist under frictions; links turnover-control bounds to realized net performance and implementability.

Model data organization. Figure A1 provides a schematic view of how the WPROD-R framework organizes the input data and intermediate objects used by the allocation rules. Starting from a panel of asset prices, we form log returns and construct rolling windows of length L. A stationary wavelet transform (with wavelet family ψ and J scales) is applied to each rolling window to produce a smooth component and scale-specific detail components. The detailed components induce a multiscale covariance geometry through exponentially weighted scale covariances, which are aggregated and stabilized before applying principal component analysis. The leading principal directions are mapped to investable basis portfolios via a softmax map, yielding a periodically refreshed time-varying basis

matrix. The traded portfolio is then generated by a reduced simplex state, which evolves dynamically under the mirror-flow and is mapped back to the asset simplex through the basis.

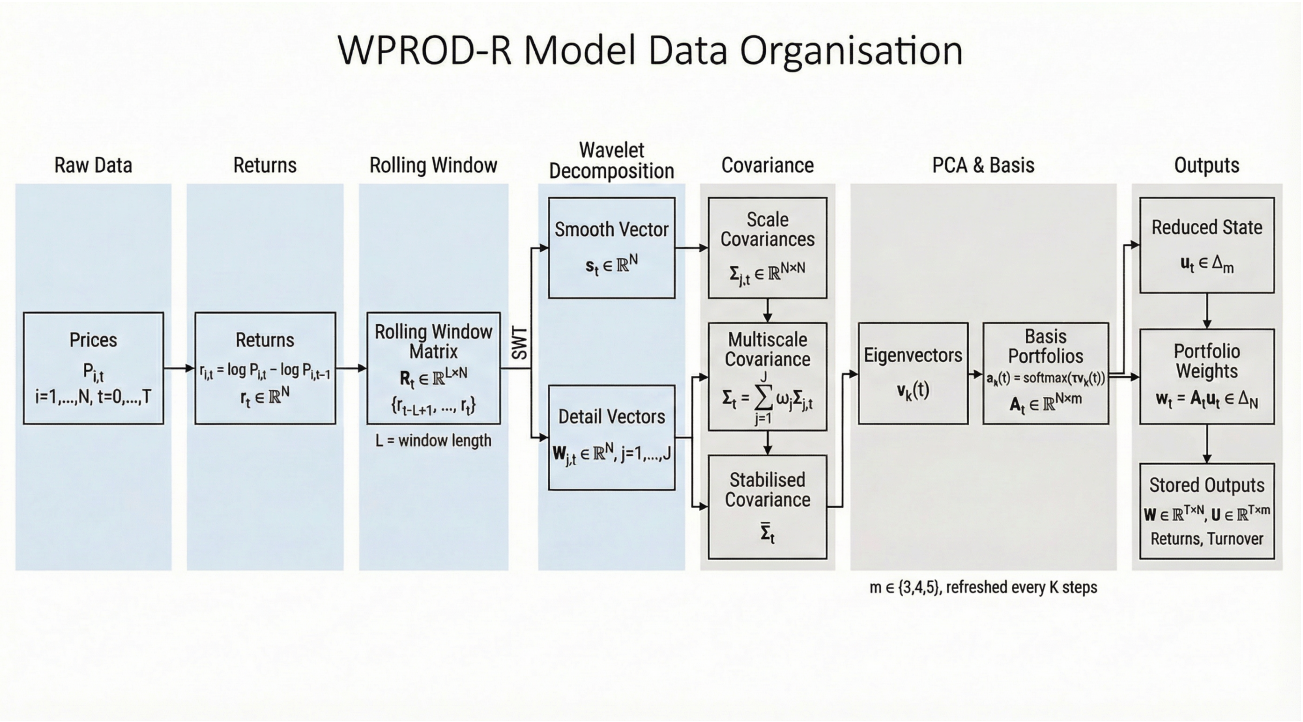


Figure A1. Data organization in WPROD-R. From prices $\{P_{i,t}\}_{i=1}^N$, we compute log-returns $\boldsymbol{r}_t \in \mathbb{R}^N$ and form the rolling window matrix $R_t \in \mathbb{R}^{L\times N}$ containing $\{\boldsymbol{r}_{t-L+1}, \ldots, \boldsymbol{r}_t\}$. Applying a stationary wavelet transform at J scales (wavelet family ψ) yields a smooth vector $\boldsymbol{s}_t \in \mathbb{R}^N$ and detail vector $\boldsymbol{W}_{j,t} \in \mathbb{R}^N$. The multiscale covariance geometry is constructed from exponentially weighted scale covariances $\Sigma_{j,t} \in \mathbb{R}^{N\times N}$ (driven by the detail vectors) and aggregated as $\Sigma_t = \sum_{j=1}^J \omega_j \Sigma_{j,t}$; a stabilized matrix $\bar{\Sigma}_t$ is then used for PCA. Let $\boldsymbol{v}_k(t)$ denote the leading eigenvectors of $\bar{\Sigma}_t$; the investable basis portfolios are defined by $\boldsymbol{a}_k(t) = \text{softmax}(\tau\, \boldsymbol{v}_k(t)) \in \Delta_N$ and collected into $A_t = [\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_m(t)] \in \mathbb{R}^{N\times m}$. The allocation is parameterized by reduced weights $\boldsymbol{u}_t \in \Delta_m$ (with $m \in \{3, 4, 5\}$), which evolve by the mirror-flow and generate traded weights $\boldsymbol{w}_t = A_t \boldsymbol{u}_t \in \Delta_N$. The basis is refreshed every K steps, and the stored outputs include $\{\boldsymbol{u}_t\}$, $\{\boldsymbol{w}_t\}$, realized returns, and turnover.

B. Additional empirical checks: strengthened baselines and time-split wavelet robustness

This appendix reports two focused empirical checks that address baseline strength and operator sensitivity. The aim is to separate these clarifications from the main empirical narrative and to link the evidence directly to the mathematical mechanisms established in the theory.

B.1. Strengthened baselines: shrinkage mean–variance and risk parity

A frequent concern in rolling mean–variance comparisons is that an unregularized sample covariance can be unstable (especially under short windows or near-collinearity), which may artificially weaken the baseline. To ensure that the conclusions do not rely on a fragile opponent, we report two strengthened

benchmarks in addition to equal-weight (EW) and the sample-covariance MV baseline: (i) a rolling MV optimizer using Ledoit–Wolf (LW) shrinkage covariance (MV–LW) [49], and (ii) a rolling risk-parity (RP) allocation computed from the same Ledoit–Wolf (LW) covariance (RP–LW) [50].

All baselines are implemented under the same long-only full-investment constraint $\boldsymbol{w}_t \in \Delta_N$, use the same rolling estimation window length $L = 256$ days for mean and covariance inputs, and are evaluated under proportional transaction costs $c = 10$ bp per unit one-way turnover. The MV baselines use the same risk-aversion level γ as the reduced objective in WPROD-R, and all reported quantities are computed from the realized *net* return series.

Table B1 shows that WPROD-R maintains a higher net Sharpe ratio than both MV baselines (sample and shrinkage) and the RP–LW benchmark, while remaining competitive with EW on risk-adjusted performance. The MV and MV–LW strategies deliver larger terminal wealth in this sample, but with materially higher volatility and deeper drawdowns; this is consistent with the fact that the Markowitz objective is sensitive to estimation noise and can concentrate weight aggressively, even under long-only constraints. The turnover statistics also highlight the dynamic nature of WPROD-R: average turnover is larger than for MV-type reoptimization, and the observed spikes in max TO are consistent with the additional refresh-induced term bounded in Theorem 6 on top of the inertia-controlled component in Theorem 5.

Table B1. Net performance comparison under transaction costs ($c = 10$ bp) for WPROD-R and strengthened baselines: rolling mean–variance with sample covariance (MV), mean–variance with Ledoit–Wolf shrinkage (MV–LW), risk parity on Ledoit–Wolf covariance (RP–LW), and equal-weight (EW). Metrics are computed from daily net returns over 2020–2024.

Strategy	$W_T - 1$	CAGR	Vol	Sharpe	MaxDD	$\overline{\text{TO}}$	max TO
WPROD	3.270	0.339	0.281	1.179	−0.342	0.024	0.750
MV (sample)	3.642	0.362	0.386	0.992	−0.442	0.004	0.023
MV (Ledoit–Wolf)	3.757	0.368	0.390	1.000	−0.450	0.004	0.022
Risk parity (LW)	1.771	0.228	0.217	1.053	−0.308	0.003	0.465
EW	2.366	0.276	0.244	1.122	−0.316	0.000	0.000

Figure B1 reports the net wealth trajectories W_t under proportional transaction costs ($c = 10$ bp) for WPROD-R and the strengthened benchmark set, namely EW, mean–variance (MV) with sample covariance, MV with Ledoit–Wolf shrinkage, and Ledoit–Wolf risk parity. The lower panel shows the relative deviation, defined as $100\,(W_t/W_t^{\mathrm{EW}} - 1)$, with respect to the EW benchmark.

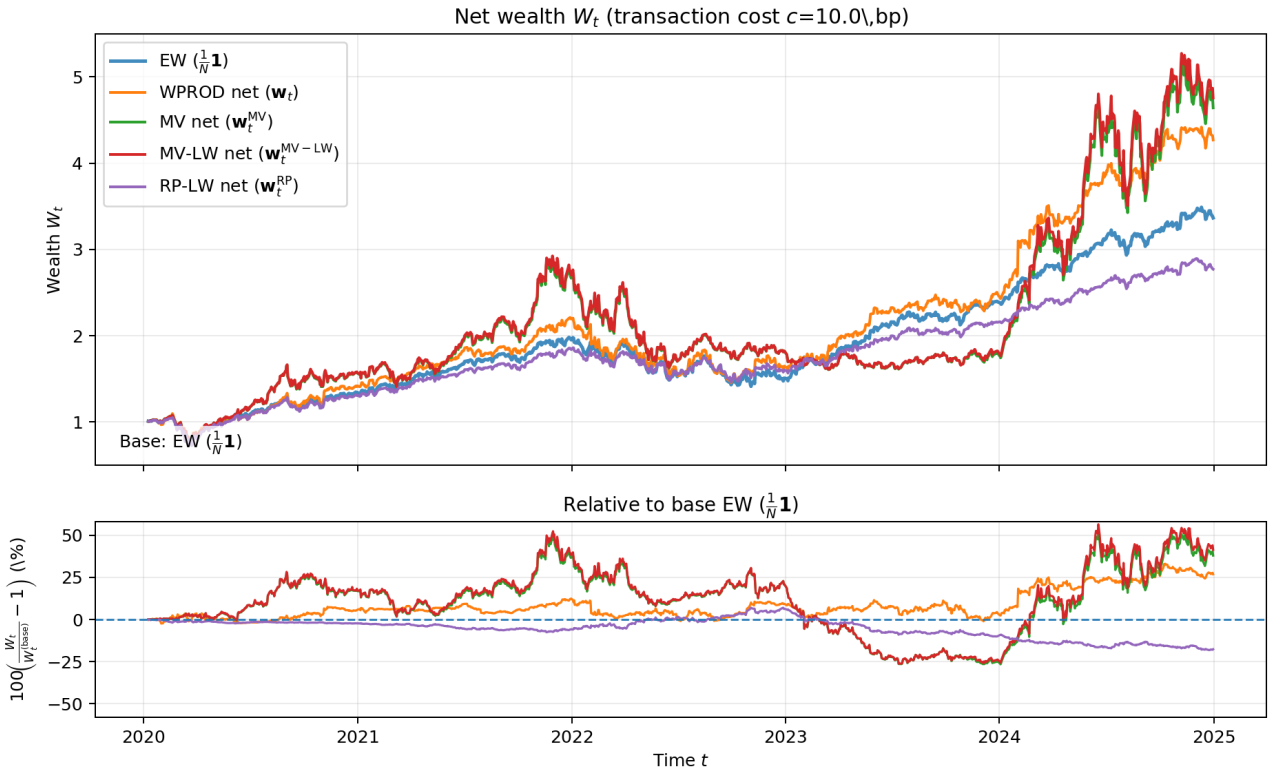


Figure B1. Net wealth comparison against strengthened baselines. Net wealth W_t under proportional transaction costs ($c = 10$ bp) for WPROD-R and the strengthened benchmark set: EW, MV with sample covariance, MV with Ledoit–Wolf shrinkage, and Ledoit–Wolf risk parity. The lower panel reports the relative deviation $100\,(W_t/W_t^{\mathrm{EW}} - 1)$ from the EW baseline.

B.2. Time-split wavelet robustness: cross-validation across subperiods

The wavelet family ψ is a structural operator in WPROD-R: It affects both the multiscale covariance geometry (through the scale covariances $\Sigma_{j,t}$) and the risk-normalized multiscale signal that drives the mirror flow. Therefore, sensitivity to ψ is not an anomaly but a reflection of the model design. To test whether the apparent advantage of a given wavelet is a sample-specific peak, we perform a time-split robustness check by evaluating each wavelet family on three consecutive, nonoverlapping subperiods: Period 1 (2020-01-09 to 2021-09-02), Period 2 (2021-09-03 to 2023-05-03), and Period 3 (2023-05-04 to 2024-12-31). All other hyperparameters are kept fixed (including $m = 4$, $J = 3$, $L = 256$, and the same transaction cost level), and the multiscale covariance update remains strictly positive-definite for each ψ by Theorem 1.

Table B2 and Figure B2 show that the ranking is not invariant across regimes: the best-performing wavelet can change as the return and dependence structure shifts. However, the results also demonstrate that the improvement is not a single-period artifact. For example, $\psi = \mathrm{coif1}$ dominates the first two subperiods and achieves the best full-sample Sharpe, while remaining close to the best performer in the third subperiod. This supports the interpretation of ψ as a genuine design parameter of the multiscale operator layer rather than a cosmetic preprocessing choice. At the same time, we regard operator sensitivity as a limitation that should be acknowledged explicitly: a practical implementation should

select ψ via an out-of-sample protocol (time-split cross-validation) rather than fixing it from in-sample performance alone.

Table B2. Net annualized Sharpe ratios (SR) of WPROD-R across wavelet families ψ and three nonoverlapping evaluation subperiods (time-split robustness check). The last column reports the full-sample Sharpe.

Wavelet	SR_1	SR_2	SR_3	SR_{Full}
$\psi = db2$	1.28	0.53	2.32	1.28
$\psi = db4$	1.30	0.31	2.21	1.18
$\psi = sym4$	1.38	0.74	2.17	1.38
$\psi = coif1$	1.44	1.07	2.17	1.50
$\psi = haar$	1.05	0.37	2.12	1.09

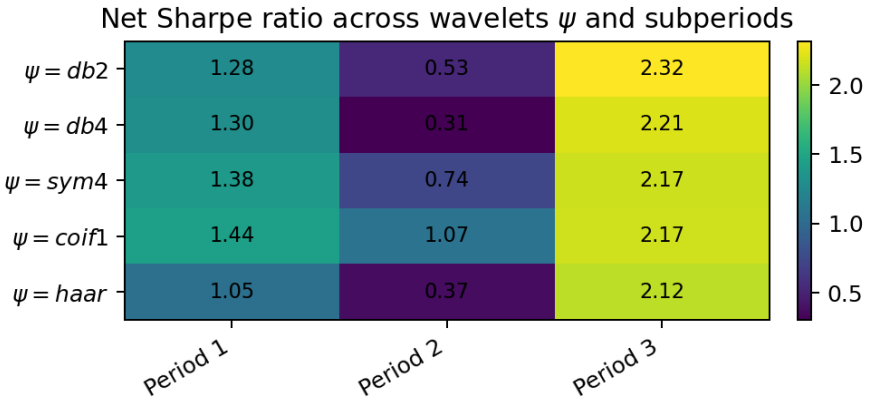


Figure B2. Time-split wavelet robustness. Net annualized Sharpe ratios across wavelet families ψ and three consecutive evaluation subperiods.

C. Additional empirical analyses: Parameter selection, high-dimensional stability, and computational feasibility

This subsection reports additional empirical evidence intended to clarify (i) the basis for key hyperparameter choices, (ii) the benefit of reduced-order dynamics in larger asset universes, and (iii) the practical computational cost of integrating multiscale wavelet features with periodic PCA basis updates. Unless stated otherwise, experiments use daily close prices from Stooq over the sample 2019–01–01 to 2025–01–01, and net portfolio returns include proportional transaction costs of 10 bp applied to one-way turnover (defined as half the ℓ_1 variation of successive portfolio weights). Annualized Sharpe ratios use a 252-trading-day convention.

C.1. Hyperparameter selection and sensitivity

The baseline configuration used throughout the empirical section is $L = 256$ (rolling window length), $J = 3$ (multiscale levels), $\alpha = 0.08$ (EWMA rate for multiscale covariance updates), and $\tau = 3.5$ (softmax temperature for constructing long-only PCA basis portfolios), with reduced dimension $m = 4$

and basis refresh period $K = 21$ trading days. These choices follow two simple principles. First, the stationary wavelet transform requires that the window length be compatible with the multiscale decomposition; we therefore choose L as a multiple of 2^J and restrict J to moderate values to avoid overfitting short-horizon noise. Second, α and τ control, respectively, the effective memory in the covariance recursion and the concentration of the investable basis; we interpret α through its effective half-life $h(\alpha) = \log(0.5)/\log(1-\alpha)$ and evaluate the stability of performance and turnover across a small grid of values.

Table C1 reports sensitivity across (L, J), and Table C2 varies (α, τ) and includes diversification diagnostics (average maximum weight and effective number of bets). The results show that performance is not driven by a knife-edge selection. In particular, Sharpe ratios remain strong across the tested (L, J) grid and across a range of EWMA half-lives, whereas turnover and concentration increase gradually as τ increases, consistent with the interpretation of τ as a basis "sharpness" parameter.

Table C1. Sensitivity to wavelet window length L and multiscale depth J (net performance). CAGR, volatility, maximum drawdown (MDD), and turnover are reported in percent. Turnover is the average daily one-way turnover.

L	J	Sharpe	CAGR (%)	Vol (%)	MDD (%)	TO (%/day)
128	2	1.391	37.3	25.1	31.8	1.97
128	3	1.156	31.8	27.0	36.2	2.17
256	2	1.275	37.0	27.7	34.5	2.22
256	3	1.179	33.9	28.1	34.2	2.39

Table C2. Sensitivity to EWMA rate α (reported with effective half-life) and softmax temperature τ. CAGR, turnover, and average maximum weight are reported in percent. Eff. N is the effective number of bets, $1/\sum_i w_{i,t}^2$, averaged over time. The baseline setting used in the main experiments corresponds to $(\alpha, \tau) = (0.08, 3.5)$.

α	Half-life (days)	τ	Sharpe	CAGR (%)	TO (%/day)	Max w (%)	Eff. N
0.04	17.0	2.5	1.274	34.7	1.14	28.0	6.9
0.04	17.0	3.5	1.334	39.4	1.75	38.7	5.2
0.04	17.0	4.5	1.361	42.9	2.28	47.1	4.1
0.08	8.3	2.5	1.185	31.6	1.57	26.7	7.1
0.08	8.3	3.5	1.179	33.9	2.39	38.1	5.1
0.08	8.3	4.5	1.162	35.8	3.01	47.4	3.9
0.12	5.4	2.5	1.119	28.7	1.47	26.0	7.2
0.12	5.4	3.5	1.126	30.4	2.22	36.5	5.4
0.12	5.4	4.5	1.128	32.1	2.82	44.8	4.3

Figure C1 presents heatmaps of net annualized Sharpe ratios across the hyperparameter grids (L, J) and (α, τ). The plots highlight broad regions of stable performance, providing a visual justification for parameter selection rather than relying on a single point estimate.

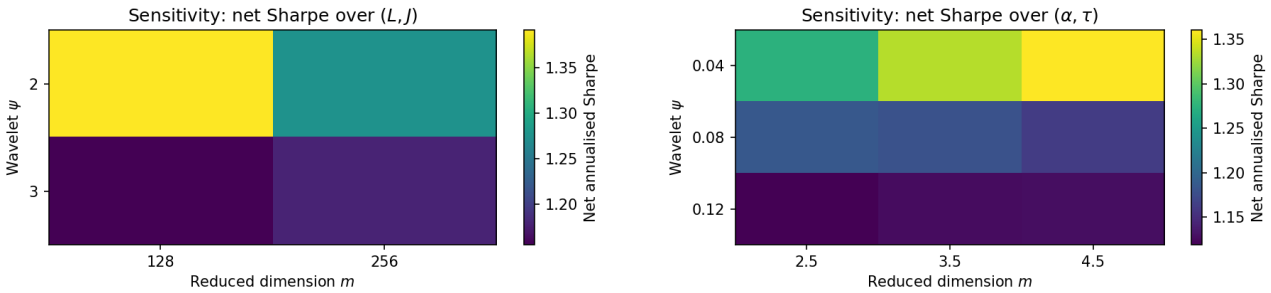


Figure C1. Heatmaps of net annualized Sharpe ratios over (L, J) (left) and (α, τ) (right). The heatmaps provide a visual parameter-selection justification by highlighting broad regions of stable performance rather than a single point estimate.

C.2. High-dimensional evaluation and stability of reduced-order dynamics

To test whether the reported results are specific to a small hand-picked universe, we repeat the backtests on larger asset spaces. We assemble a broad U.S. equity universe of 107 tickers (large-cap and sector-diverse); after data availability filtering on Stooq, 106 tickers remain. For each target dimension $N \in \{10, 30, 50\}$, we sample three independent random subsets and run identical model parameters (including $m = 4$). We compare: (i) **WPROD-R**, the proposed reduced-order dynamics with PCA basis; (ii) **NoPCA (full mirror)**, a full-dimensional mirror-flow baseline in which the allocation evolves directly in the N-simplex without PCA reduction (same multiscale signal and covariance geometry); and (iii) **EW (static)**, a constant equal-weight (buy-and-hold) allocation.

Table C3. High-dimensional evaluation on random subsets of a larger equity universe. Reported values are means across three independent subsets for each target N. CAGR, turnover (TO), and average maximum weight are in percent; Eff. N is the effective number of bets. Net returns include 10 bp transaction costs applied to one-way turnover.

N	Strategy	Sharpe	CAGR (%)	TO (%/day)	Max w (%)	Eff. N
10	WPROD-R	0.517	10.1	2.19	31.7	6.0
10	NoPCA (full mirror)	0.136	−0.0	2.13	68.6	2.1
10	EW (static)	0.522	9.5	0.00	10.0	10.0
30	WPROD-R	0.650	13.0	1.05	10.9	23.1
30	NoPCA (full mirror)	0.332	6.5	2.36	62.4	2.7
30	EW (static)	0.653	12.6	0.00	3.3	30.0
50	WPROD-R	0.724	14.3	0.69	5.5	43.7
50	NoPCA (full mirror)	−0.043	−5.0	2.66	60.5	3.2
50	EW (static)	0.659	12.5	0.00	2.0	50.0

Table C3 reports averages across the three subsets for each N. Two empirical patterns are noteworthy. First, the full-dimensional mirror baseline becomes increasingly concentrated as N grows, with the average maximum weight exceeding 60% and an effective number of bets near 3 at $N = 50$, consistent with noise amplification when directly optimizing in a high-dimensional simplex. Second, WPROD-R remains diversified as N increases (effective number of bets above 40 at $N = 50$) and achieves materially

lower turnover than the full-dimensional baseline, supporting the paper's motivation that reduced-order dynamics mitigate estimation noise and keeping the allocation genuinely time-varying.

C.3. Computational complexity and runtime evidence

The proposed implementation performs three dominant computations per trading step: multiscale feature extraction, multiscale covariance updates, and (periodic) basis refresh. In terms of leading-order operations, multiscale feature extraction on a rolling window of length L at J scales has cost on the order of $O(N\,L\,J)$. The multiscale covariance update consists of J rank-one matrix updates and a weighted aggregation, which is $O(J\,N^2)$ in a dense representation. Basis construction requires the top-m eigenvectors of an $N \times N$ symmetric matrix; a dense eigendecomposition is $O(N^3)$, but this operation is amortized by refreshing only every K days and by keeping m small (in practice, truncated eigensolvers reduce the cost toward $O(N^2 m)$). Finally, the reduced mirror update is $O(Nm + m^2)$ and is negligible for $m \in \{3, 4, 5\}$.

To complement the complexity discussion with practical evidence, we profile wall-clock runtimes by instrumenting the code with per-block timing over a representative 252-trading-day segment (mean and 95th percentile statistics). Table C4 reports the mean and 95th percentile per-step times (in milliseconds) at the baseline refresh period $K = 21$, across $N \in \{10, 30, 50\}$. The wavelet feature extraction is the dominant cost component, and the amortized PCA refresh time is two orders of magnitude smaller than the total step time. Table C5 further shows the expected decline in amortized PCA refresh cost as K increases at fixed $N = 50$.

Table C4. Runtime breakdown (milliseconds per trading step) at refresh period $K = 21$. Mean times are reported for major blocks, along with the 95th percentile of the total state-update time. These results indicate that wavelet feature computation dominates runtime and the amortized PCA refresh overhead remains small.

N	Wavelet	Cov. update	PCA refresh	Mirror step	Total (mean)	Total (p95)
10	4.10	0.117	0.015	0.074	4.37	5.45
30	7.71	0.125	0.021	0.068	7.98	9.62
50	17.52	0.201	0.030	0.052	17.91	24.92

Table C5. Effect of basis refresh period K on amortized PCA refresh time at $N = 50$ (milliseconds per step). As expected, increasing K decreases the per-step PCA overhead.

K	Wavelet (mean)	PCA refresh (mean)	Total state (mean)
5	12.85	0.125	13.19
21	17.52	0.030	17.91
63	13.00	0.008	13.34

As illustrated in Figure C2, the computational cost increases with the number of assets N, primarily driven by the wavelet feature extraction stage, while the relative contribution of PCA basis refresh becomes negligible as the refresh period K increases.

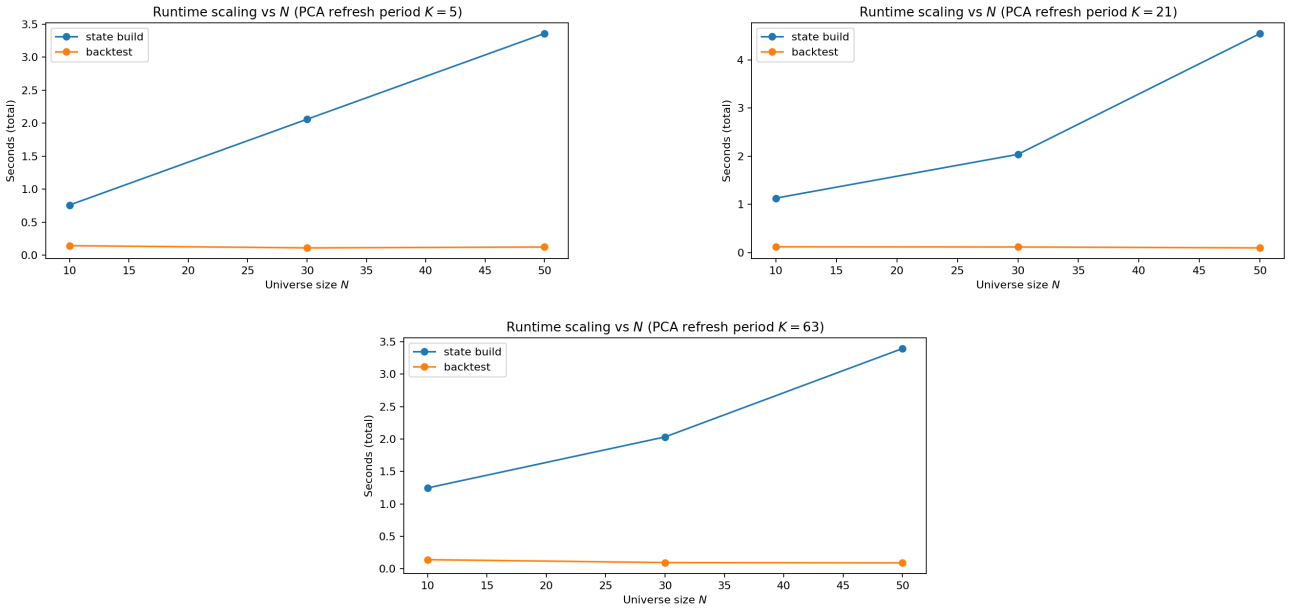


Figure C2. Runtime scaling plots for different basis refresh periods $K \in \{5, 21, 63\}$. The results are consistent with the theoretical decomposition: the wavelet feature block drives overall runtime growth with N, and PCA refresh becomes negligible as K increases.