UCLouvain

LOUVAIN
School of Management

Louvain School of Management

The Virtue of Complexity: Benign Overfitting in Portfolio Construction

Author: **Arnaud MERTENS**
Supervisor: **Prof. Nathan LASSANCE**
Academic year 2025–2026
Master [120] in Business Engineering, professional focus
Major in Financial Engineering
Daytime Schedule

Foreword

This Master's thesis was written under the supervision of Prof. Nathan Lassance, whom the author thanks for his guidance throughout the project.

Declaration of originality. The author of this work declares on his honour that it is the result of his own original work. Any use of the ideas or statements of a third party, whether through translation or information processing software, is clearly identified and referenced as such.

Declaration regarding the use of AI tools. In the preparation of this Master's thesis, the author made use of generative AI assistants (Anthropic's Claude) as a support tool for the development and debugging of the Python research code, for editorial assistance with the LaTeX manuscript, and for proofreading. All research questions, methodological choices, analyses and conclusions are the author's own; every empirical result reported in this thesis was produced by the author's code on the author's machine, and all AI-assisted material was reviewed, verified and validated by the author.

Statements on responsible use of generative AI.

- ☒ The content of my master's thesis is my original output, even if I used generative AI to help prepare it.
- ☒ I master the theories, tools and methods that I use for the thesis.
- ☒ I verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.
- ☒ I acknowledge that I master the final output and that I am able to explain it and answer questions about it.
- ☒ I made sure not to provide generative AI tools with access to confidential data or information.

Signed by Mertens, Arnaud Date: 1st August 2026

Executive Summary

Modern machine learning has shown that models complex enough to fit their training data perfectly can nonetheless generalize, a phenomenon known as *benign overfitting*. This stands in sharp contrast to the classical view in portfolio theory, where estimation error is expected to make optimized portfolios underperform the naive $1/N$ allocation. This thesis examines whether benign overfitting can also be observed in financial data and, if so, whether its statistical advantages translate into economically meaningful improvements in portfolio performance once factor exposures and transaction costs are taken into account.

To answer this question, a two-part empirical study is conducted using Fama–French portfolios covering U.S., European, and emerging equity markets. The first part investigates the existence of double descent by sweeping a return regression through the interpolation threshold and testing the spectral conditions proposed by Bartlett et al. (2020). The second part evaluates whether these statistical properties improve portfolio construction by backtesting six portfolio strategies across rolling estimation windows. Performance is assessed through Sharpe ratios, five-factor attribution, transaction costs, and effective-rank diagnostics computed on the covariance matrices effectively inverted by each strategy.

The results show that double descent is indeed present in financial data, and that introducing a very small amount of ridge regularization reduces the interpolation peak by two orders of magnitude. In highly over-parameterized settings, stabilized minimum-variance portfolios outperform the naive $1/N$ benchmark by roughly 30% in Sharpe ratio while remaining less risky, and this advantage survives a five-factor attribution, suggesting that it reflects genuine estimation alpha rather than systematic factor tilts. However, the premium is highly sensitive to implementation costs and disappears outside favourable market environments. The proposed diagnostics identify these environments before portfolio construction, while the refined Ridgelet 2 estimator is shown to fail on style-portfolio universes, exposing a limitation of the recent zero-variance portfolio framework of Chang et al. (2026).

Overall, the findings suggest that the benefits of model complexity in portfolio allocation are neither universal nor accidental. They arise under specific and measurable conditions, namely short estimation windows, high-dimensional settings, covariance-based objectives, and sufficiently dispersed covariance spectra. Rather than a general principle, benign overfitting therefore provides a practical framework for identifying when additional model complexity is likely to improve real-world portfolio construction.

Acknowledgements

I would like to thank my supervisor, Professor Nathan Lassance. He was available to answer my questions throughout the year, suggested directions worth exploring when the work called for them, and gave me feedback on what I produced. Our discussions helped me decide where to take the thesis, and I am grateful for his time and his input.

I am deeply grateful to my family for their constant support throughout my studies. They have always encouraged me to follow my own path and trusted the choices I made, even when they were not the most obvious ones. Their patience, encouragement, and confidence have meant more than I can express, and I would not have reached this point without them.

I would also like to thank my friends at university. Sharing the challenges, deadlines, and long days of studying with them made these past years far more enjoyable. Their friendship and support have been an important part of this journey.

Finally, I would like to thank the Portfolio Management team at Edmond de Rothschild in Brussels, which I joined as an intern. Working alongside them showed me how investment decisions are shaped by transaction costs, turnover limits, and client mandates. That experience informed the way I approached several of the questions studied in this thesis, in particular the net-of-cost evaluation of the strategies.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation

The foundational doctrines of classical financial econometrics and modern portfolio theory have long been anchored to the principle of parsimony. Since the work of Markowitz (1952), researchers have been warned against the perils of overfitting: estimating optimization models in which the number of parameters or investable assets approaches the number of historical observations has traditionally been viewed as a recipe for estimation failure (Kan and Zhou, 2007; DeMiguel et al., 2009b). In the under-parameterized regime, this trade-off between accessing investment opportunities and limiting estimation risk even admits an interior optimum: Lassance et al. (2025) show that the portfolio size maximizing expected out-of-sample utility is typically well below the number of available assets. In standard linear frameworks, when the number of assets or predictors grows relative to the sample size, the sample covariance matrix loses its full-rank properties, becoming unstable and prone to maximizing estimation errors (Michaud, 1989; Ledoit and Wolf, 2004). Textbooks accordingly proscribe data interpolation (a model fitting its training data perfectly, with zero residual variance) on the grounds that the fitted parameters merely capture sample-specific noise that fails to generalize out of sample (Hastie et al., 2009).

Yet recent asset pricing research increasingly argues in the opposite direction: capturing fleeting anomalies, time-varying risk premia, and complex cross-sectional interactions requires richly parameterized models (Kelly and Xiu, 2023; Kelly et al., 2024). Over the short look-back windows relevant to Tactical Asset Allocation (TAA), this prescription makes the estimation problem inherently sample-poor but information-rich, and mechanically pushes the empirical system into an overparameterized regime in which the number of assets or predictive features dominates the temporal sample size ($N \gg T$).

Recent advances in statistical learning theory have turned this historical prohibition into a diagnostic question (Belkin et al., 2019; Bartlett et al., 2020). Rather than uniformly degrading out-of-sample accuracy, letting a model expand far beyond the interpolation threshold can trigger a *second descent* in prediction risk, with test error falling again as model capacity keeps growing. This thesis is motivated by this intersection: asking whether the complexity that classical doctrine treats as a hazard can instead be deliberately exploited in systematic portfolio construction, while taking seriously the growing body of evidence that the specific noise environment of financial markets constrains how far this program can be pushed (Fallahgoul, 2025; Nagel, 2025).

1.2 From classical regularization to benign overfitting

The traditional resolution to high-dimensional instability in finance relies on explicit regularization. Within the classical bias–variance trade-off, adding explanatory variables or investable assets mechanically reduces in-sample approximation bias but exposes the system to a sharp increase in estimation variance (Hastie et al., 2009). To combat this curse of dimensionality, researchers have historically turned to penalization methods such as ridge regression (ℓ_2 penalty) and the Lasso (ℓ_1 penalty), which deliberately trade a small amount of structural bias for a large reduction in out-of-sample forecast variance (Hastie et al., 2009). Yet, as the number of parameters or assets approaches the available sample size ($N \approx T$), an acute statistical crisis emerges regardless: the sample covariance matrix becomes singular, estimation error spikes, and portfolio performance collapses (Kan and Zhou, 2007; Lu et al., 2024).

The modern literature on double descent upends this narrative by showing that the textbook U-shaped risk curve is only the first half of a larger generalization framework (Belkin et al., 2019; Hastie et al., 2022). Once model capacity extends past the interpolation threshold into the overparameterized regime, the model has more degrees of freedom than data points and fits the training set perfectly, yet the implicit regularization of the minimum-ℓ_2-norm solution distributes the signal, and dilutes the noise, across a large number of dimensions, allowing out-of-sample risk to decline in a second descent (Bartlett et al., 2020; Kelly et al., 2024).

To bring this shift into quantitative asset allocation, this thesis is built on a three-member estimator suite: **Ridge**, **Ridgeless**, and **Ridgelet**. The Ridge architecture is the classical benchmark, with an explicitly calibrated penalty smoothing the empirical covariance structure. The Ridgeless estimator pushes the penalty to its limit ($\lambda \to 0^+$), where the Moore–Penrose pseudo-inverse interpolates the training data exactly (Hastie et al., 2022). In practice, however, Ridgeless solutions are prone to severe weight instability in the low signal-to-noise environment of financial time series near the threshold (Hastie et al., 2022; Chang et al., 2026). This vulnerability motivates the recently proposed **Ridgelet** estimator of Chang et al. (2026): a tuning-free device that bypasses the numerical fragility of the pseudo-inverse by attaching a microscopic, scale-invariant stabilizer to the sample covariance matrix, achieving out-of-sample optimality in heavily overparameterized regimes without any cross-validation.

1.3 Research questions and main results

To evaluate the validity and practical viability of this body of theory inside quantitative portfolio selection, the thesis addresses three research questions:

1. **Spectral validity of benign overfitting in return prediction.** Does the double descent curve manifest when predicting equity returns from the Fama–French cross-section? Beyond the curve itself, do the empirical covariance matrices of financial predictors

formally satisfy the effective-rank conditions (a dominant low-dimensional eigenvalue block combined with a slowly decaying, high-effective-rank tail, formalized through the effective ranks introduced in Chapter 3) under which the theory of Bartlett et al. (2020) permits noise to become benign?

2. **Translation of statistical risk into portfolio utility.** Does the statistical reduction of out-of-sample forecast error in the overparameterized regime translate into an economic *double ascent* of the out-of-sample Sharpe ratio (Lu et al., 2024)? Specifically, can an interpolating, zero-in-sample-variance portfolio built with the Ridgelet operator outperform traditional allocations (the naive equal-weight $(1/N)$ baseline, classical mean-variance optimization, and explicit shrinkage (DeMiguel et al., 2009b; Ledoit and Wolf, 2004)) when evaluated on the metrics asset managers actually use: out-of-sample volatility, Jensen's alpha, maximum drawdown, Value-at-Risk, and portfolio turnover (Suhonen et al., 2017)?
3. **Horizon accessibility: tactical versus strategic allocation.** Does access to the overparameterized regime depend on the investment horizon? Short tactical look-back windows push N/T above one mechanically, whereas the long estimation windows of Strategic Asset Allocation (SAA) keep the problem in the classical underparameterized domain (Coqueret and Laguerre, 2025). The question is therefore whether the virtue of complexity is, in practice, confined to TAA.

Main results in brief. Anticipating the full analysis, the answers are as follows. The double descent curve does appear in financial return data, and a tiny tuning-free ridge divides its test-error peak at the interpolation threshold by a factor of roughly one hundred (from 20.4 to 0.19 in the U.S. sweep); the spectral conditions of Bartlett et al. (2020) are however only partially satisfied, and the second descent, while real, does not fall below the best under-parameterized risk level. The allocation experiment reverses this verdict: the Sharpe double ascent not only appears but *overtakes* the first regime: deep in the over-parameterized region, the stabilized minimum-variance funds (Sharpe ratios of 1.06–1.08) beat both the naive $1/N$ benchmark (0.81) and the best under-parameterized configuration of the same window (0.93), with an alpha that survives a five-factor attribution. The premium is consumed by transaction costs beyond 10–25 basis points per unit of turnover, is confined to short (tactical) estimation windows, and vanishes on the two international universes (including Europe, a *developed* market whose covariance spectrum is nearly as concentrated as the emerging-markets one, a feature the diagnostics flag in advance), making the virtue of complexity in allocation a tactical, conditional, and measurable phenomenon in which the spectrum, not the market label, determines the payoff.

1.4 Contributions

This thesis makes three contributions to asset pricing, financial machine learning, and quantitative asset management.

First, it delivers a direct empirical test of the effective-rank conditions of Bartlett et al. (2020) on real financial return data rather than synthetic benchmarks. The eigenstructure of empirical covariance matrices is mapped on three universes of different maturity and breadth (the U.S. 100-portfolio grid, whose size deciles provide a natural segmentation by idiosyncratic noise, and the pooled 24-series European and emerging-markets universes), so that the sensitivity of the diagnostics to the dataset is itself an empirical result and draws a boundary against data snooping (Bailey et al., 2017; Fabozzi, 2025).

Second, the thesis establishes a like-for-like mapping between statistical risk and asset-management utility: the same Ridge/Ridgeless/Ridgelet suite is run through a regression protocol (mean squared error) and a portfolio protocol (full performance dashboard) on identical data. The dynamics of turnover and of the ℓ_2 norm of the weights document that the Ridgelet operator stabilizes allocations in the over-parameterized regime, where the raw pseudo-inverse produces unstable weights and, in the tangency formulation, uninvestable leverage (Chang et al., 2026; Kan and Zhou, 2007).

Third, the thesis articulates an operational *horizon accessibility* framework distinguishing TAA from SAA. The virtue of complexity (Kelly et al., 2024) is shown to be governed by the ratio N/T: structurally out of reach for strategic allocations, but a risk-stabilizing and performance-enhancing engine for systematic tactical allocation. This grounds the practice of quantitative funds that ingest wide signal sets over short estimation windows (Kelly and Xiu, 2023).

1.5 Thesis structure

Chapter 2 reviews the literature, from the classical case for parsimony to benign overfitting and its financial applications. Chapter 3 develops the theoretical framework: double descent, the effective-rank conditions of Bartlett et al. (2020), the Ridgeless and Ridgelet estimators, and estimation risk in mean-variance portfolios. Chapter 4 presents the data and the two empirical protocols, run identically on all universes. Chapter 5 reports the results: double descent and spectral diagnostics in regression, then the double ascent of the Sharpe ratio in allocation. Chapter 6 discusses the evidence and its limitations, and concludes with recommendations for systematic funds and directions for future research.

2. Literature review

The literature behind this thesis can be organized as a narrative in five stages: a classical era in which statistical complexity was feared and deliberately avoided; the empirical explosion of dimensionality in asset pricing that made this avoidance untenable; the theoretical rehabilitation of complexity in statistical learning; the recent transposition of these insights into quantitative finance; and, finally, an emerging strand of caution arguing that the specific signal-to-noise environment of financial markets constrains how far these insights can be pushed.

2.1 Estimation risk and the classical case for parsimony

Modern portfolio theory begins with Markowitz (1952), whose mean-variance framework converts beliefs about expected returns and covariances into optimal weights. Its practical weakness has always been *estimation*: the optimizer amplifies whatever error its estimated inputs contain. Merton (1980) identified the deepest root: means are estimated far less precisely than covariances from the same sample, so mean-leaning rules inherit an order of magnitude more noise; Michaud (1989) branded the optimizer an "error maximizer" for this reason. Kan and Zhou (2007) characterize the loss analytically and DeMiguel et al. (2009b) demonstrate it empirically: none of the optimizing strategies they evaluate consistently beats the estimation-free $1/N$ rule. Constructive responses concentrated on the covariance: Ledoit and Wolf (2004) shrink it toward a structured target, and Jagannathan and Ma (2003) show that even *wrong* constraints help because they act as implicit shrinkage, early evidence that the cure is spectral regularization. Cassio (2024), in his master's thesis at the Louvain School of Management, provided a first exploration of these questions on Fama–French data; the present thesis extends that work from the regression setting to the portfolio-allocation problem itself: unconstrained funds backtested over a complete (N, T) grid, with factor attribution and transaction costs. The same distrust of heavily parameterized models pervades classical statistics through the bias–variance trade-off (Hastie et al., 2009): complexity beyond the minimum was understood to buy in-sample fit at the cost of generalization, and the era's prescriptions (parsimony, regularization, cross-validation) all flow from this premise.

2.2 High-dimensional finance and the factor zoo

While theory counselled parsimony, empirical asset pricing moved decisively in the opposite direction. Cochrane (2011), in his presidential address, coined the term "factor zoo" to describe the proliferation of characteristics claimed to explain the cross-section of expected returns. Harvey et al. (2016) catalogue hundreds of published factors and argue, on multiple-testing grounds, that

a large fraction are likely false discoveries, recommending substantially raised statistical hurdles for new factors; Chen (2021) reaches the opposite verdict, arguing on Bayesian grounds that most claimed findings in cross-sectional return predictability are likely true. Feng et al. (2020) develop a systematic (regularized two-pass) procedure for evaluating whether each new factor adds explanatory power relative to the existing zoo. The relevant point here is not which factors survive but the structural fact these papers document: the predictor space of modern empirical finance is *intrinsically* high-dimensional. Whether returns are modelled with hundreds of firm characteristics, thousands of engineered features, or the cross-section of portfolio returns itself, the candidate parameters now routinely rival or exceed the available time-series observations. High dimensionality in finance is not a modelling choice to be avoided; it is the operating environment.

2.3 Benign overfitting and double descent

The tension between the classical prescription (avoid complexity) and the practice of modern machine learning (embrace it) was resolved (in machine learning's favour) by a sequence of theoretical papers. Belkin et al. (2019) documented the *double descent* risk curve, showing that test risk, having peaked at the interpolation threshold, descends a second time in the over-parameterized regime. Hastie et al. (2022) characterized this behaviour exactly for *ridgeless* least squares (the minimum-norm interpolating solution obtained through the Moore–Penrose pseudo-inverse) using random matrix theory. Bartlett et al. (2020) isolated the precise spectral conditions (the effective ranks of the covariance matrix, developed in Section 3.2) under which such interpolators can nonetheless deliver near-optimal prediction risk: the phenomenon they named *benign overfitting*. Tsigler and Bartlett (2023) extended the characterization to ridge regression, showing that explicit regularization achieves benign generalization under strictly weaker spectral conditions. Together, these papers replaced a universal prohibition ("never interpolate") with a diagnostic question ("does the data's covariance spectrum permit interpolation?"), a question this thesis asks empirically of financial data.

2.4 Machine learning in return prediction and portfolio choice

A rapidly growing literature imports these results into finance. Kelly et al. (2024) prove, and document on U.S. equity data, that market timing strategies built on ridgeless random-feature regressions achieve Sharpe ratio improvements that increase in model complexity (the "virtue of complexity"). They build on Gu et al. (2020), who show that machine learning materially improves cross-sectional return prediction, with Kelly et al. (2022) extending the evidence across asset classes and Xiong (2026) tracing the regularization mechanics to the self-induced shrinkage of large factor models. Lu et al. (2024) transpose the double descent logic from prediction

to *allocation*, showing theoretically and empirically that the out-of-sample Sharpe ratio of mean-variance portfolios traces a double-ascent curve in the number of assets. Most recently, Chang et al. (2026) attack the over-parameterized minimum-variance problem directly: they prove that the *Ridgeless* estimator (pseudo-inverse) fails to generalize out of sample in the $N > T$ regime, and introduce the *Ridgelet* estimator (a tuning-free, vanishingly small ridge attached to the sample covariance matrix) which restores out-of-sample optimality and exhibits the double descent phenomenon in portfolio risk. The Ridge/Ridgeless/Ridgelet triad that structures the empirical work of this thesis (Sections 5.1 and 5.4) is drawn directly from this literature.

2.5 Critiques and open questions

A final strand of recent literature warns against over-extrapolating the virtue of complexity to financial data. Fallahgoul (2025) derives information-theoretic sample complexity bounds showing that, under the weak signal-to-noise ratios typical of return prediction, reliable learning of complex predictive relationships can be outright impossible at the training sample sizes used in the empirical virtue-of-complexity literature; his analysis of ridgeless regression further shows that its *effective* complexity is bounded by the sample size rather than the nominal feature dimension. Chapter 5 speaks directly to this warning: the second descent documented there never falls below the best under-parameterized risk level, and the dominant-block index k^* remains pinned at the sample-size boundary, exactly the bounded effective complexity his analysis predicts. Nagel (2025) argues, through a kernel-regression representation, that the empirical success of high-complexity market timing strategies may reflect mechanisms other than genuine complexity benefits. Finally, a longer-standing literature on backtest overfitting (Bailey et al., 2017; López de Prado, 2018) reminds us that apparent out-of-sample success in finance can itself be an artifact of multiple testing and non-stationarity, a caution that applies with full force to complexity sweeps like those performed here, where dozens of configurations are evaluated on a single historical sample.

2.6 Research gap

Within this literature, the present thesis occupies a specific niche: rather than assuming that financial data satisfy the conditions for benign overfitting (as the applied literature implicitly does) or assuming they do not (as the critiques above suggest), it *tests* the spectral conditions of Bartlett et al. (2020) directly on Fama–French data and maps the same estimator suite (Ridge/Ridgeless/Ridgelet) from statistical risk (Section 5.1) to portfolio performance (Section 5.4) on the same data, allowing the regression-level and allocation-level manifestations of the phenomenon to be compared like-for-like.

3. Theoretical framework

3.1 From the bias–variance trade-off to double descent

Classical learning theory ties model complexity to a bias–variance trade-off: as the parameter count N grows, training error falls monotonically while test risk traces a U-shape, and interpolating the training data ($N \geq T$) is treated as the paradigmatic symptom of overfitting, to be avoided through parsimony, regularization or cross-validation (Hastie et al., 2009). The empirical success of heavily over-parameterized models (which routinely interpolate their training data yet generalize well) contradicts this prescription. Belkin et al. (2019) named the reconciliation *double descent*: below $N = T$ the classical U-shape holds; at the *interpolation threshold* $N = T$ risk peaks sharply, because the unique interpolating solution must fit the realized noise exactly with no degrees of freedom left to average it out; beyond it, interpolating solutions form a continuum and a selection principle (for least squares, the minimum-ℓ_2-norm solution delivered by the Moore–Penrose pseudo-inverse) picks one that can generalize, so that risk *descends a second time*, sometimes below the best under-parameterized benchmark. The classical curve is thus not wrong but incomplete: it is the left branch of a larger picture, whose canonical illustration is reproduced in Figure I.1. Hastie et al. (2022) make the picture exact in the proportional regime $N/T \to \gamma$: random-matrix formulas reproduce the double descent shape, characterize the divergence at $\gamma = 1$ analytically, and show that an arbitrarily small ridge penalty removes the singularity entirely, a fact that motivates the tiny-ridge estimators at the heart of this thesis.

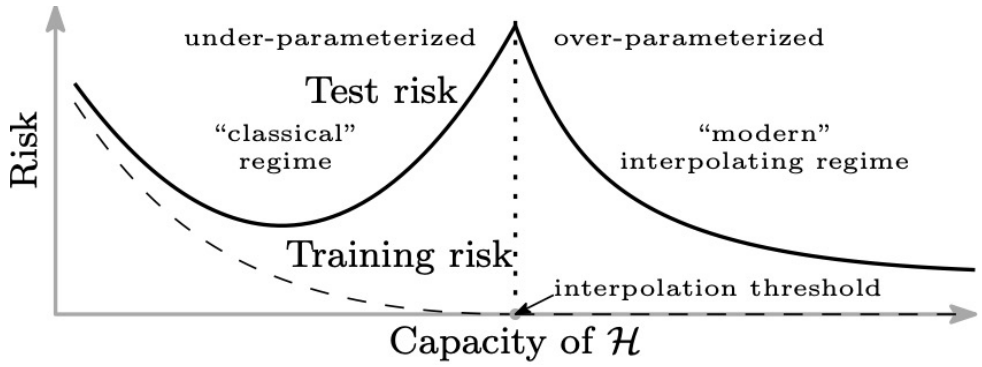

Figure I.1: The double descent risk curve: U-shaped test risk in the classical regime, a peak at the interpolation threshold, and a second descent in the over-parameterized regime.
Source: reproduced from Belkin et al. (2019), Figure 1, p. 15850.

3.2 Benign overfitting and its spectral conditions

The double descent phenomenon documents that interpolating estimators can generalize well in the over-parameterized regime; it does not by itself explain when and why this occurs. Bartlett et al. (2020) provide the key theoretical answer for linear regression, coining the term *benign overfitting* for the situation in which an estimator interpolates noisy training data exactly (zero training error) yet still achieves a test risk close to the lowest achievable risk.

3.2.1 Setup

Consider a linear model $y = x^\top \beta^* + \varepsilon$, where $\beta^* \in \mathbb{R}^N$ is the true (unknown) coefficient vector, $x \in \mathbb{R}^N$ has covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N > 0$, and ε is mean-zero noise with variance σ^2. When $N \geq T$, the ordinary least-squares problem is underdetermined; Bartlett et al. (2020) study the *minimum-norm interpolating estimator*

$$\hat{\beta} = X^+ y,$$

where X^+ denotes the Moore–Penrose pseudo-inverse of the design matrix $X \in \mathbb{R}^{T \times N}$: the unique matrix satisfying the four Penrose conditions $XX^+X = X$, $X^+XX^+ = X^+$, $(XX^+)^\top = XX^+$ and $(X^+X)^\top = X^+X$ (Penrose, 1955). When X has full row rank (the generic case for $N \geq T$), it admits the closed form $X^+ = X^\top (XX^\top)^{-1}$, and $\hat{\beta} = X^+ y$ solves $X\hat{\beta} = y$ exactly. This estimator achieves zero training error by construction whenever $N \geq T$ (generically), and among all interpolating solutions, it is the one of smallest Euclidean norm.

3.2.2 Effective ranks

The central theoretical device is a pair of *effective rank* functions of the covariance spectrum, defined for $k = 0, 1, 2, \ldots$ as

$$r_k(\Sigma) = \frac{\sum_{i>k} \lambda_i}{\lambda_{k+1}}, \tag{3.1}$$

$$R_k(\Sigma) = \frac{(\sum_{i>k} \lambda_i)^2}{\sum_{i>k} \lambda_i^2}. \tag{3.2}$$

Both quantities summarize the "size" of the tail of the spectrum beyond the k-th eigenvalue: $r_k(\Sigma)$ compares the tail's total mass to the next eigenvalue, while $R_k(\Sigma)$ is a squared-mass-to-squared-mass ratio that behaves like an effective dimension of the tail (it equals m exactly when the trailing m eigenvalues are all equal, and is smaller when the tail is itself unevenly spread).

Bartlett et al. (2020) define the effective rank threshold

$$k^* = \min\{k \geq 0 : r_k(\Sigma) \geq b\,T\}$$

for a universal constant $b > 0$, i.e., the smallest number of leading eigen-directions one must remove before the remaining tail mass, relative to the next eigenvalue, falls below (a constant multiple of) the sample size T.

3.2.3 The benign overfitting risk bound

The performance of an estimator $\hat{\beta}$ is measured by its excess risk, the expected out-of-sample prediction loss beyond the irreducible noise floor: $\text{Excess Risk}(\hat{\beta}) = \mathbb{E}\big[(y - x^\top\hat{\beta})^2\big] - \mathbb{E}\big[(y - x^\top\beta^*)^2\big] = \mathbb{E}_x\big[\big(x^\top(\hat{\beta} - \beta^*)\big)^2\big]$, where the expectation is over a new draw (x, y) independent of the training sample. Under the model above, Bartlett et al. (2020) give upper and lower bounds of the same order on this quantity for the minimum-norm interpolator, of the form

$$\mathbb{E}\,[\text{Excess Risk}] \;\asymp\; \underbrace{\|\beta^*\|^2\,\lambda_{k^*+1}}_{\text{bias}} \;+\; \underbrace{\sigma^2\left(\frac{k^*}{T} + \frac{T}{R_{k^*}(\Sigma)}\right)}_{\text{variance}}. \tag{3.3}$$

Benign overfitting occurs (i.e., the interpolator's risk converges to the noise floor as $T \to \infty$) if and only if, simultaneously,

$$\frac{k^*}{T} \longrightarrow 0 \qquad \text{and} \qquad \frac{T}{R_{k^*}(\Sigma)} \longrightarrow 0. \tag{3.4}$$

Interpretation. Condition (3.4) demands a specific spectral shape: a *few dominant eigenvalues* (small k^*/T, keeping the variance contribution negligible) followed by a *long, flat tail* (large $R_{k^*}(\Sigma)$ relative to T, so label noise is diluted across many nearly equal directions). Sharply decaying spectra (strong common factors with nothing beyond them) fail the condition: with no flat tail to absorb noise, the interpolator's variance explodes, which is exactly the mechanism behind the threshold divergence of Belkin et al. (2019) and Hastie et al. (2022). The two conditions are *separately necessary* (small k^*/T does not compensate a bounded $T/R_{k^*}(\Sigma)$, nor conversely) a distinction that matters directly in Section 5.1.

A practical checklist. For empirical work the theory can be restated as four points, with their mapping onto the rigorous statement kept explicit: (i) *overparameterization* $N \geq T$: the *regime* in which interpolation exists, rather than a condition; (ii) *a dominant low-dimensional signal block* ($k^*/T \to 0$); (iii) *a long, flat spectral tail* ($T/R_{k^*}(\Sigma) \to 0$); (iv) *signal alignment* of β^* with the dominant eigen-directions, which controls the bias term of Equation (3.3). Points (ii)–(iii) are

the two formal conditions of (3.4); (i) and (iv) are the regime and bias-side requirements that accompany them.

3.3 Extension to ridge regression

Tsigler and Bartlett (2023) extend the characterization to ridge estimators $\hat{\beta}_\lambda = (X^\top X + \lambda I)^{-1} X^\top y$, $\lambda \geq 0$, and show that a positive penalty achieves benign generalization under substantially weaker spectral conditions: shrinking the small eigen-directions smoothly substitutes for the raw spectral geometry that the ridgeless case must rely on. Ridge regression should therefore be much more robust than the minimum-norm interpolator to the divergence at $N = T$, the central object of comparison in Section 5.1.

3.4 Ridgeless and Ridgelet estimators

The estimator suite used throughout the empirical work of this thesis comprises three members, ordered by decreasing explicit regularization:

1. **Ridge**: $\hat{\beta}_\alpha = (X^\top X + \alpha I)^{-1} X^\top y$ with a macroscopic penalty α, trading bias for variance in the classical manner (Tsigler and Bartlett, 2023);
2. **Ridgeless**: the minimum-ℓ_2-norm interpolator $\hat{\beta} = X^+ y$, computed through the Moore–Penrose pseudo-inverse, the exact object studied by Bartlett et al. (2020) and Hastie et al. (2022);
3. **Ridgelet**: $\hat{\beta}_\tau = (X^\top X + \tau I)^{-1} X^\top y$ with a *vanishingly small, tuning-free* penalty τ, following Chang et al. (2026). As those authors emphasize, τ is not a tuning parameter chosen to balance bias and variance, but a fixed, negligible stabilizer that replaces the hard rank truncation of the pseudo-inverse with a smooth damping of the near-null spectral directions.

A theoretical subtlety governs when the Ridgeless and Ridgelet estimators differ from one another, and it maps onto the two halves of this thesis. In the *regression* setting of Section 5.1, $X^\top y$ lies by construction in the row space of X, so the Ridgelet converges to the minimum-norm (Ridgeless) solution as $\tau \to 0$: the two estimators coincide up to numerical differences near the interpolation threshold, where hard rank truncation and smooth damping treat the near-zero eigenvalues differently. In the *portfolio* setting of Section 5.4, the vector to which the generalized inverse is applied ($\mathbf{1}$ or $\hat{\mu}$) generically has null-space components whenever $N > T$, i.e., a nonzero projection onto the null space of $\hat{\Sigma}$, the subspace of directions v with $\hat{\Sigma} v = 0$ that appears as soon as the sample covariance is rank-deficient. The pseudo-inverse annihilates them; the Ridgelet amplifies them by $1/\tau$, so its portfolio converges to the minimum-norm *in-sample zero-variance* portfolio, a genuinely different object, and Chang et al. (2026) prove that it is this object that

generalizes out of sample and attains optimal risk when $N \gg T$. The regression comparison thus doubles as a validation check, while the substantive difference is expected in allocation.

The zero-variance portfolio: interpolation in allocation space. The title of Chang et al. (2026) names the allocation-side analogue of interpolation. When $N > T$, the sample covariance S is singular, so fully invested weight vectors with $w^\top S w = 0$ exist: portfolios that fit the training sample perfectly, as an over-parameterized regression fits its labels. Two results of that paper sharpen the reading. First, the pseudo-inverse weights *fail even to interpolate* (past the threshold their in-sample variance rises again), so only the Ridgelet path attains the zero-variance portfolio and its out-of-sample optimality; the Ridgeless–Ridgelet gap of Chapter 5 is the visible trace of this qualitative difference. Second, the same authors propose a refined "Ridgelet 2" in which the identity inside the tiny ridge is replaced by an estimator of the idiosyncratic covariance; both versions are backtested here, and Chapter 5 shows the refinement's premise fails on style-portfolio universes (finding 3′).

Shrinkage as adaptive ridge. Linear shrinkage and ridge regularization of the covariance are the same operation as far as minimum-variance weights are concerned, a connection formalized by DeMiguel et al. (2009a). The Ledoit and Wolf (2004) estimator is $\hat{\Sigma}_{\text{LW}} = (1 - \hat{\delta})\, S + \hat{\delta}\, \bar{\mu} I$; since GMV weights are invariant to a rescaling of Σ, $w_{\text{GMV}}(\hat{\Sigma}_{\text{LW}}) = w_{\text{GMV}}(S + \tau_{\text{eff}} I)$ with $\tau_{\text{eff}} = \hat{\delta}\, \bar{\mu}/(1 - \hat{\delta})$. The GMV-LW fund is thus a ridge-stabilized GMV fund with an adaptive penalty, and the GMV-Ridgelet a member of the same family with a vanishingly small fixed one. Two predictions follow. Near $N = T$, where the sample covariance is maximally ill-conditioned, the data-driven intensity $\hat{\delta}$ (the ratio of the estimated sampling error of S to the dispersion of its eigenvalues around their mean) should be substantial, so the LW fund should sail over the singularity that destroys the unstabilized funds. Deep in the over-parameterized regime, both funds solve the same ridge-regularized GMV problem and differ only in penalty intensity, so their performance should converge as long as neither penalty distorts the dominant eigen-directions that drive the weights. Both predictions are confirmed in Section 5.4: the LW fund is the only one undisturbed at the threshold, and the two funds' deep-regime Sharpe ratios end up within two points of one another.

3.5 The virtue of complexity in return prediction

Kelly et al. (2024) bring this body of theory directly to bear on return prediction in finance. Working with market-return forecasting models built from random-feature expansions of a small number of underlying predictive signals, they prove theoretically, and document empirically on U.S. equity market data, that increasing model complexity beyond the interpolation threshold ($N \gg T$) does not degrade (and can in fact monotonically improve) both out-of-sample predictive

R^2 and the realized Sharpe ratio of a market-timing strategy built on the resulting forecasts. This "virtue of complexity" is a direct economic manifestation of benign overfitting: heavily overparameterized, ridgeless or lightly-regularized return-prediction models are shown to *not* suffer the classical bias–variance penalty for complexity, provided the underlying predictor covariance exhibits a spectral structure broadly consistent with (3.4). Their results form the direct empirical template for the regression analysis carried out in this thesis: rather than predicting a single market index from a handful of macro-finance signals, we work with the cross-section of Fama–French portfolio returns (the N portfolio series observed each month), expanding the feature space via return lags instead of their random nonlinear feature expansions, but the underlying complexity-sweep methodology and diagnostic apparatus (effective ranks, double descent curves) are the same.

3.6 Estimation risk in mean-variance portfolios

3.6.1 The Markowitz tangency portfolio

Markowitz (1952) formalized portfolio choice as a mean–variance trade-off: for N assets with excess-return mean μ and covariance Σ, and defining the Sharpe ratio of a portfolio w as its expected excess return per unit of volatility, $\mathrm{SR}(w) = w^\top \mu / \sqrt{w^\top \Sigma w}$, the Sharpe-maximizing (*tangency*) portfolio has weights

$$w^* \;\propto\; \Sigma^{-1}\mu, \tag{3.5}$$

normalized to $\mathbf{1}^\top w^* = 1$. In practice both inputs are estimated from T past returns, yielding the *plug-in* tangency (MSR) portfolio $\hat{w} \propto \hat{\Sigma}^{-1}\hat{\mu}$.

3.6.2 The curse of dimensionality in portfolio estimation

Chapter 2 reviewed why the plug-in rule fails: Merton's estimation asymmetry between means and covariances (Merton, 1980), Michaud's error maximization (Michaud, 1989), and the verdicts of Kan and Zhou (2007) and DeMiguel et al. (2009b). Two technical facts complete the setup here. First, $\hat{\Sigma}$ is singular whenever $N \geq T$ (rank at most $T - 1$), leaving $\hat{\Sigma}^{-1}$ undefined or extremely sensitive to noise in near-null directions. Second, DeMiguel et al. (2009b) estimate that the plug-in strategy would need thousands of months of data to beat $1/N$ reliably for 25–50 assets, which is why the equally-weighted portfolio serves throughout this thesis as the no-estimation-risk benchmark.

3.6.3 The pseudo-inverse and the parallel with regression

The parallel with regression is exact. When $N \geq T$, $\hat{\Sigma}$ is singular and the Moore–Penrose pseudo-inverse $\hat{\Sigma}^{+}$ replaces $\hat{\Sigma}^{-1}$ in Equation (3.5), the same generalized-inverse operation that

defines the minimum-norm interpolator $\hat{\beta} = X^{+}y$ of Section 3.2. Under this correspondence the asset count N plays exactly the role of the feature count of the regression problem, the window T that of its sample size (which is why the thesis deliberately uses the single pair (N, T) in both settings), and $N = T$ is the portfolio-theoretic *interpolation threshold*. Lu et al. (2024) formalize the parallel: the out-of-sample Sharpe ratio of plug-in portfolios first rises with N, deteriorates as $N \to T$ (the Kan–Zhou overfitting region), then *rises again* for $N \gg T$, a *double ascent* curve mirroring double descent, driven by the same mechanism: deep in the over-parameterized regime the minimum-norm solution behaves like an implicitly regularized estimator rather than a noise-fitting one, now operating on the return covariance instead of a regression design matrix.

3.6.4 Regularized alternatives

A complementary strand regularizes $\hat{\Sigma}$ directly rather than relying on the pseudo-inverse's implicit regularization: robustified high-dimensional estimators (Petukhina et al., 2024) in the same spirit as the ridge extension of Tsigler and Bartlett (2023). This motivates the inclusion in the strategy suite of a shrinkage-based minimum-variance fund built on the Ledoit and Wolf (2004) estimator, the classical robust benchmark against which the Ridgeless and Ridgelet alternatives are compared, and, per Section 3.4, itself a member of the ridge family with adaptive intensity.

4. Data and methodology

4.1 The Fama–French data library

Data source. The analysis uses monthly data from Kenneth French's data library (Fama and French, 1993, 2015): the Fama–French five-factor set (`F-F_Research_Data_5_Factors_2x3.csv`), used as a coverage cross-check, and the 100 portfolios formed on size and book-to-market (`100_Portfolios_10x10.csv`), which supply the cross-section of returns used to build both the regression target and the feature set.

In both cases, the average value-weighted monthly returns are used (the raw files stack several tables under a shared layout, and only the first block, the value-weighted returns, is retained). The two files have different historical coverage by construction: the 100-portfolios file begins in July 1926, while the five-factor file begins in July 1963, when the accounting data required by the profitability and investment factors become reliably available. Both extend to the most recent month in the sample and are aligned on their common date index before any processing.

4.2 The U.S. asset universe

Data and asset universe. The analysis uses the same 100 Fama–French portfolios formed on size and book-to-market as in Section 5.1 (`100_Portfolios_10x10.csv`), parsed with the same corrected routine that retains only the first contiguous block of dated rows (Section 4.1). Unlike Section 5.1, where each of the 100 series was used to build *regression features*, here each of the 100 series is treated as an *investable asset*: the 10×10 sort on size (`ME`, market equity deciles) and book-to-market (`BM` deciles) provides a diversified, pre-cleaned cross-section of quasi-assets, in the tradition of DeMiguel et al. (2009b), who use the same class of Fama–French sorted portfolios as one of their seven empirical test universes. Only months with a complete 100-asset cross-section (no missing portfolio) are retained, since early decades of this file can have thinly populated extreme size/book-to-market deciles.

4.3 The emerging-markets universe

The market-maturity comparison relies on the emerging-markets section of the same data library. Five files are used: the emerging five-factor set (`Emerging_5_Factors.csv`, providing the emerging `Mkt-RF` and `RF` series used for the CAPM regressions), and the four 2×3 portfolio sorts on size crossed with book-to-market, operating profitability, investment, and prior 12–2 momentum. Since each sort yields only six portfolios, the four sorts are pooled (with column names prefixed by sort to preserve uniqueness) into a 24-series investable universe, the widest

cross-section the library offers for emerging markets. The files carry the same stacked-tables structure as their U.S. counterparts (handled by the same parser), and their coverage differs by sort: the size/book-to-market sort begins in July 1989, but the profitability and investment sorts (which require accounting data) only in July 1992, so the pooled sample (inner join on dates, complete cross-sections only) spans July 1992 to May 2026, or 407 months. Note the asymmetry with the U.S. universe: 24 series over 407 months versus 100 series over 749 months. It is deliberate: differences in cross-sectional breadth and noise level are precisely what the comparison is designed to measure.

4.4 The European universe

Following the same construction, a third universe is built from the European section of the library (developed markets): the European five-factor file and the four 2×3 sorts on size crossed with book-to-market, profitability, investment, and momentum, pooled into a 24-series universe. The complete-cross-section sample spans November 1990 to May 2026 (427 months). Europe thus matches emerging markets in breadth and sample length almost exactly, while carrying the *developed* label, which is precisely what makes it valuable: if outcomes are driven by the developed/emerging label, Europe should behave like the United States; if they are driven instead by the shape of the covariance spectrum (in particular, its degree of eigenvalue concentration), the descriptive statistics below predict it will behave like emerging markets.

4.5 Descriptive statistics

Table A.1 summarizes the three universes along the dimensions that matter for the analyses to come: cross-sectional averages of the per-series annualized moments, the average pairwise correlation, the shares of total correlation-matrix variance carried by the first and first five principal components, and the properties of each universe's equally-weighted portfolio and market factor.

Three aspects of the table anticipate results of Chapter 5. First, the *principal-component concentration*. All universes exhibit the "spiked" profile, but to very different degrees: the first principal component carries 74.6% of total correlation variance in the U.S. grid, against 92.3% in emerging markets, where the first five components carry 97.5% and leave barely 2.5% for the entire tail. The emerging-markets structure is a spike without a tail; since benign generalization requires a long, flat spectral tail over which noise can dissipate, this statistic already forecasts the outcome of the diagnostics of Sections 5.2 and 5.3. Second, the *signal-to-noise gap*: the emerging-markets cross-section is narrower (24 versus 100 series) and noisier, with a higher average pairwise correlation (0.92 versus 0.74), more negative skewness, and an equally-weighted Sharpe ratio of 0.53 against 0.80 for the U.S. universe. Third, the *size segmentation*: the U.S.

10×10 grid splits into a small-capitalization half (ME1–ME5) and a large-capitalization half (ME6–ME10), a within-U.S. noise axis reported in Table A.1. The two axes line up: from U.S. large caps to U.S. small caps to emerging markets, the first-eigenvalue share rises monotonically (74.4%, 79.5%, 92.3%) while the equally-weighted Sharpe ratio falls (0.87, 0.71, 0.53): noisier universes are also more one-dimensional, compressing exactly the spectral tail on which benign generalization relies.

4.6 Regression protocol

Throughout the two experiments, the central quantity is the *complexity ratio* N/T of parameters to observations: features over training months in the regression experiment below, assets over window length in the portfolio experiment of Section 4.7. The interpolation threshold sits at a ratio of one in both settings. The regression protocol described below for the U.S. universe is replicated identically on the European and emerging-markets universes; the design of these international variants is specified at the end of the section.

Target variable. Because the theory of Section 3.2 is formulated for scalar-response linear regression, the target variable y_{t+1} is defined as the equal-weighted average of the 100 portfolio returns at month $t + 1$, used as a scalar proxy for the broad cross-section (an alternative choice, e.g. a single size/value portfolio, is equally valid and is left as a robustness check).

Feature set. The predictor matrix X_t consists of 12 monthly lags of all 100 portfolio return series, yielding a maximum feature-space dimension of $N_{\max} = 100 \times 12 = 1{,}200$. To trace out a complexity sweep from very low to very high N without confounding complexity with the specific ordering of lags, the 1,200 candidate columns are randomly permuted once (fixed seed), and increasing-size prefixes of this permuted column list are used to build feature matrices of size $N = 2, \ldots, 1{,}200$, with a denser sampling grid placed around the anticipated interpolation threshold $N = T_{\text{train}}$.

Train/test split and models. The sample (after aligning features and target and dropping the resulting missing values) is split chronologically (no shuffling) into a training set of $T_{\text{train}} = 515$ months and a test set of $T_{\text{test}} = 221$ months (the sample of complete 100-portfolio cross-sections spans July 1945 to April 2026, and is thus now fully aligned with the sample used in the portfolio protocol below), giving a maximum aspect ratio of $N_{\max}/T_{\text{train}} = 1{,}200/515 \approx 2.33$. Following the estimator suite of Section 3.4, three estimators are compared at each complexity level N:

- the *Ridgeless* minimum-norm estimator, computed exactly through the Moore–Penrose pseudo-inverse (numerically, a rank-revealing least-squares solve), the interpolator of Section 3.2;

- the *Ridgelet* estimator of Chang et al. (2026): a tuning-free, vanishingly small ridge, implemented as $\tau = 10^{-6} \times \mathrm{tr}(X^\top X)/p$ (a fixed, negligible fraction of the average eigenvalue, never cross-validated), solved in primal ($N \times N$) or dual ($T \times T$) form depending on which dimension is smaller;
- the *Ridge* estimator with a fixed macroscopic penalty $\alpha = 1$, a deliberately heavy, uncalibrated benchmark: the penalty is never tuned to the data, so the estimator is strongly over-regularized rather than optimally chosen (a limitation discussed in Section 6.1).

As established in Section 3.4, the Ridgeless and Ridgelet estimators are theoretically coincident in the regression setting (up to numerical differences near the interpolation threshold); their joint inclusion here serves as an internal validation check, with the substantive Ridgeless–Ridgelet divergence expected only in the portfolio setting of Section 5.4. Out-of-sample performance is measured by mean squared error (MSE) on the held-out test set. For each complexity level, the effective ranks $r_k(\Sigma)$ and $R_k(\Sigma)$ of Equations (3.1)–(3.2) are also computed on the training-sample feature covariance, with the effective-rank threshold set at $b = 0.2$ (a value validated on a synthetic factor-plus-noise benchmark with known effective rank).

International variants. The identical protocol is run on the pooled 24-series emerging-markets universe (Section 4.3) with 24 monthly lags, giving $N_{\max} = 24 \times 24 = 576$ and, on the 407-month common sample, a chronological split of $T_{\text{train}} = 267$ / $T_{\text{test}} = 115$ and a maximum aspect ratio $N_{\max}/T_{\text{train}} = 2.16$, deliberately close to the U.S. sweep's 2.33; the European replication is identical in design (24 series, 24 lags, $T_{\text{train}} = 281$ on the 427-month sample, maximum ratio 2.05). The three universes are thus compared over equivalent complexity ranges.

4.7 Portfolio protocol

Portfolio construction: six virtual funds. Six strategies are compared at each complexity level N, spanning the estimator suite of Section 3.4 transposed to the allocation problem:

- **EW**: the equally-weighted portfolio, $w_i = 1/N$, requiring no estimation of μ or Σ: the "no-estimation-risk" control fund (DeMiguel et al., 2009b);
- **MSR**: the plug-in tangency portfolio of Equation (3.5), $\hat{w} \propto \hat{\Sigma}^+ \hat{\mu}$: classical Markowitz, exposed to estimation error in both $\hat{\mu}$ and $\hat{\Sigma}$;
- **GMV-Ridgeless**: the global-minimum-variance portfolio via the pseudo-inverse, $\hat{w} \propto \hat{\Sigma}^+ \mathbf{1}$: the "Ridgeless" estimator of Chang et al. (2026), free of $\hat{\mu}$-estimation risk but exposed to the null-space truncation of the pseudo-inverse once $N > T$;
- **GMV-Ridgelet**: the same GMV problem solved with the tuning-free tiny ridge of Chang et al. (2026), $\hat{w} \propto (\hat{\Sigma} + \tau I)^{-1} \mathbf{1}$ with $\tau = 10^{-6} \times \mathrm{tr}(\hat{\Sigma})/N$, converging, for $N > T$, to the minimum-norm in-sample zero-variance portfolio;

- **GMV-Ridgelet2**: the refined estimator of Chang et al. (2026), attaching the tiny ridge to a POET-type factor-plus-sparse estimator of the idiosyncratic covariance (Principal Orthogonal complEment Thresholding, Fan et al., 2013: the residual covariance beyond the leading K principal components, adaptively soft-thresholded), $w \propto (S + \tau\,\hat{\Omega})^{-1}\mathbf{1}$: the "Ω-aligned" ridge selects the zero-variance portfolio of minimal idiosyncratic variance. Implementation choices, documented for replicability: K by the eigenvalue-ratio criterion (capped at 8), thresholding constant 0.5;
- **GMV-LW**: the GMV portfolio built on the Ledoit and Wolf (2004) linear shrinkage covariance estimator: the classical robust high-dimensional alternative, standing in for the broader shrinkage family discussed by Petukhina et al. (2024).

Rolling out-of-sample backtest. For a given number of assets N and a fixed estimation window T, weights are re-estimated every month using only the trailing T months of data (no look-ahead), and realized against the actual return of the following month. This produces a time series of out-of-sample portfolio returns for each strategy, from which the annualized Sharpe ratio is computed as

$$\widehat{\text{Sharpe}} = \frac{\bar{r}}{\hat{\sigma}(r)}\sqrt{12},$$

where $\bar{r}$ and $\hat{\sigma}(r)$ denote the sample mean and sample standard deviation of the strategy's monthly out-of-sample returns. The Sharpe ratio is reported alongside a full performance dashboard: annualized return and volatility, maximum drawdown, monthly Value-at-Risk at the 95% level, average one-way turnover[1], and Jensen's alpha and beta from a CAPM regression of excess portfolio returns on the market excess return `Mkt-RF` (taken, together with the risk-free rate, from the Fama–French five-factor file). This rolling-window design follows the standard evaluation protocol of DeMiguel et al. (2009b) and Kan and Zhou (2007).

Multi-factor performance attribution. Because the asset universe is itself built from size and book-to-market sorts, a single-factor alpha is vulnerable to the objection that it merely proxies for passive exposures to standard factors. Alongside the CAPM regression, each fund's excess returns are therefore attributed against the full Fama–French five-factor model (Fama and French, 2015),

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_{p,M}(R_{m,t} - R_{f,t}) + \beta_{p,S}\,\text{SMB}_t + \beta_{p,H}\,\text{HML}_t + \beta_{p,R}\,\text{RMW}_t + \beta_{p,C}\,\text{CMA}_t + \varepsilon_{p,t},$$

[1] Computed as the mean ℓ_1-distance between each month's target weights and the previous month's weights *drifted* by realized returns, $w_{t-1}^{\text{drift}} = w_{t-1} \odot (1 + r_t)/(1 + w_{t-1}^{\top} r_t)$. The measure therefore captures both the re-optimization component and the rebalancing of price drift; in particular, the EW strategy exhibits its true rebalancing turnover (of the order of the cross-sectional dispersion of monthly returns, ≈ 0.02–0.04) rather than an artificial zero.

so that the reported five-factor alpha isolates the performance component that cannot be replicated by static loadings on market, size, value, profitability and investment, the relevant test of whether the over-parameterized funds generate genuine estimation alpha rather than implicit smart-beta tilts.

Net-of-cost evaluation. To price implementation frictions, each fund's Sharpe ratio is also computed net of transaction costs by deducting, ex post, a proportional cost $c \times \text{Turnover}_t$ from each month's realized return, for one-way costs of $c \in \{10, 25, 50\}$ basis points per unit of ℓ_1 turnover. The nature of this exercise should be stated precisely. The Fama–French portfolios are paper constructions (value-weighted aggregates of underlying stocks whose composition and liquidity are not observable here) so no actual trading cost can be *measured*; what is exactly measurable is each strategy's turnover across these quasi-assets, which is also the only dimension along which the six funds differ, since they hold the same universe. The net-of-cost figures are therefore a *stylized sensitivity analysis*: treating each portfolio as an investable style segment and sweeping the unknown per-unit cost c over a plausible institutional range, they ask whether the *relative* ranking of the strategies survives frictions, the same convention as DeMiguel et al. (2009b), who evaluate their strategies on comparable portfolio datasets net of a 50-basis-point proportional cost, itself tracing back to Balduzzi and Lynch (1999). Three frictions remain outside this simplified cost setting: the heterogeneity of true costs across segments (small-capitalization exposure is costlier to trade than the uniform c assumes), the internal annual reconstitution turnover of the Fama–French portfolios themselves (incurred roughly equally by all six funds, and thus largely neutral for their comparison), and nonlinear market impact. Costs are deducted ex post, leaving the optimization objectives unchanged: the question is whether the gross complexity gains survive plausible frictions, not how to design cost-aware weights.

Spectral diagnostics on the return covariance. Finally, to close the analogy between the regression and allocation experiments, the effective-rank diagnostics of Section 3.2 are computed directly on the objects the portfolio estimators actually invert: at each (N, T) node, k^*/T and $T/R_{k^*}(\hat{\Sigma})$ are evaluated on the rolling T-month return covariance matrices (sub-sampled across the backtest at a fixed stride and averaged), so that the spectral conditions for benign generalization are tested on the asset return covariance itself rather than only by analogy with the regression feature covariance.

Complexity sweep and SAA/TAA windows. As in Section 5.1, the 100 candidate assets are randomly permuted once (fixed seed), and increasing-size prefixes of this permuted list define asset sets of size $N = 2, \ldots, 100$. The entire sweep is run for three estimation windows, $T \in \{24, 48, 120\}$ months, corresponding respectively to a tactical (TAA), an intermediate, and a strategic (SAA) estimation horizon. This design choice has a mechanical consequence that is

central to the thesis: the maximum reachable aspect ratio is $N_{\max}/T$, i.e., 4.17 for $T = 24$, 2.08 for $T = 48$, but only 0.83 for $T = 120$. *Long-window strategic allocation therefore cannot reach the over-parameterized regime on this universe*: only short, tactical estimation windows push the allocation problem past the interpolation threshold, which is precisely where the theory of Sections 3.2 and 3.4 predicts the benefits of complexity to reside.

International variants. The same six funds and rolling protocol are run on the two 24-asset international universes (Europe and emerging markets) with $T \in \{12, 24, 48\}$ months. The narrower cross-sections shift the whole design one notch toward shorter windows: N/T reaches at most 2.0 (at $T = 12$), exactly 1.0 (at $T = 24$), and only 0.5 (at $T = 48$), so that the deep over-parameterized regime accessible in the United States at $T = 24$ is, internationally, only partially reachable even at the most tactical window. The Jensen regressions use each universe's own `Mkt-RF` and `RF` series. As a robustness check against composition noise (the dependence of the N-sweep on which assets enter the growing universe first) the emerging-markets sweep is repeated over ten random asset orderings (seeds 42–51) and reported as cross-permutation means with dispersion; by construction, at $N = N_{\max}$ all orderings coincide and the dispersion is exactly zero.

5. Empirical results

5.1 Double descent in the regression experiment

Table A.2 reports out-of-sample MSE at a representative subset of the complexity levels N tested (the full grid comprises 55 values of N; see the accompanying `partie1_double_descent_results.csv` for the complete series), for the three estimators of the suite. The Ridgeless column is computed exactly through the pseudo-inverse; the Ridgelet uses the tuning-free tiny ridge $\tau = 10^{-6}\,\mathrm{tr}(X^\top X)/p$.

The Ridgeless–Ridgelet comparison delivers the internal validation anticipated in Section 3.4: the two columns are identical to all reported digits except in the immediate vicinity of the interpolation threshold, where the Ridgelet's smooth damping of the near-zero eigenvalues tames the divergence that the hard rank truncation leaves exposed. At $N = T_{\text{train}}$ exactly, the Ridgeless MSE explodes to 20.40 while the Ridgelet contains it at 0.192, a factor of roughly 106; one step off the threshold, the two differ by less than 4%. Away from it, the coincidence confirms that in regression the two estimators are the same object, so the main comparison takes place in the portfolio setting of Section 5.4.

Table A.2 and Figure II.1 show the double descent pattern clearly. As N grows from 2 to $515 = T_{\text{train}}$, the minimum-norm estimator's test MSE rises monotonically and then explosively, peaking at the interpolation threshold ($N/T_{\text{train}} = 1$) at a value more than three orders of magnitude ($\approx$ 7,600 times) larger than at $N = 2$. This is precisely the divergence predicted by the asymptotic theory of Hastie et al. (2022) at aspect ratio $\gamma = 1$, and reflects the unique interpolating solution becoming extremely sensitive to the noise realized in the 515-month training sample. Beyond the threshold, MSE falls again as N increases further, reaching 0.00574 at $N = 1{,}200$ ($N/T_{\text{train}} = 2.33$), a genuine second descent, consistent with Belkin et al. (2019).

Two features matter for the discussion in Section 6.1. First, the recovery is incomplete: MSE at $N = 1{,}200$ (0.00574) remains about 2.1 times higher than at $N = 2$ (0.00269), so the second descent appears but does not fall below the best under-parameterized level. Second, the ridge estimator with $\alpha = 1$ is nearly flat across the entire sweep (MSE between 0.00269 and 0.00319), avoiding the interpolation-threshold singularity entirely, as Tsigler and Bartlett (2023) predict: a fixed positive penalty removes the divergence at $\gamma = 1$ at the price of a small bias. Part of this flatness, however, reflects the heaviness of the uncalibrated $\alpha = 1$ itself: the curve shows a strongly over-regularized benchmark, not an optimally tuned ridge (limitation in Section 6.1).

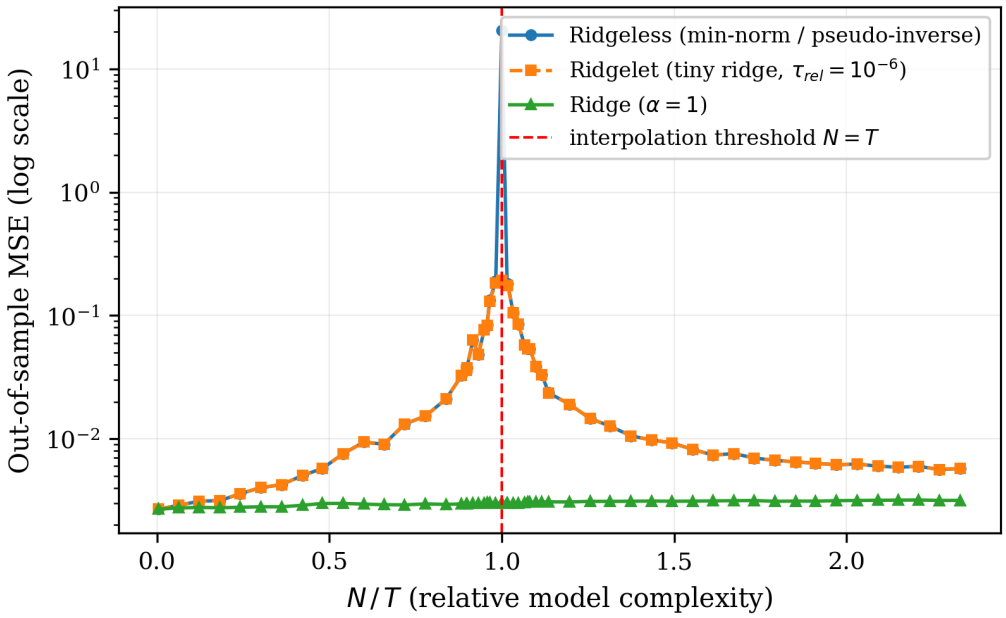

Figure II.1: Out-of-sample MSE versus model complexity N/T_{train}, U.S. universe. The dashed line marks the interpolation threshold $N = T_{\text{train}} = 515$.

Source: author's calculations.

5.2 Spectral diagnostics of the feature covariance

Table A.3 reports the effective-rank diagnostics k^*/T_{train} and $T_{\text{train}}/R_{k^*}(\Sigma)$ at the complexity levels of Table A.2, and Figure A.1 (Appendix A) plots them. Two regimes emerge. For $N \lesssim 455$, the threshold-crossing index sits at the very edge of the empirical spectrum (R_{k^*} numerically undefined[1]) and k^*/T grows steadily with N (from 0.002 to 0.882). Between $N = 455$ and $N = 462$ a marked transition occurs: k^*/T collapses to ≈ 0.17 and $R_{k^*}(\Sigma)$ becomes finite for the first time. Beyond it, both diagnostics improve monotonically as N grows (k^*/T down to 0.072 and $T/R_{k^*}(\Sigma)$ from 2.69 to 1.87 at $N = 1{,}200$) moving in the direction the theory predicts, yet with $T/R_{k^*}(\Sigma)$ never approaching the favorability threshold of 0.2. (The precise location of the abrupt transition should not be over-interpreted: it depends on where, along an evolving feature set, the tail-mass threshold happens first to be crossed; cf. the limitations in Section 6.1.)

The two requirements of condition (3.4) are not equally satisfied. The dominant-block condition $k^*/T \to 0$ approximately holds in the over-parameterized regime (0.072–0.18 for $N \gtrsim 462$), consistent with a spectrum dominated by a handful of common directions, plausibly the priced factors underlying the Fama–French portfolios. The tail-mass condition $T/R_{k^*}(\Sigma) \to 0$

[1]Whenever the crossing rule $r_k(\Sigma) \geq bT$ is not met at any k before the edge of the nonzero spectrum, k^* lands on the last (numerically) nonzero eigenvalue: the empirical feature covariance has rank at most $\min(N, T)$, so the tail beyond k^* then contains only zero eigenvalues and $R_{k^*}(\Sigma) = (\sum_{i>k^*} \lambda_i)^2 / \sum_{i>k^*} \lambda_i^2$ is a 0/0 ratio. In the tables, T/R_{k^*} is reported as ∞ at these nodes, the conservative reading: the tail-mass condition is maximally violated, as there is no spectral tail left over which noise could dissipate.

fails: the ratio stays between 1.9 and 2.7, an order of magnitude above the threshold. The eigenvalue tail is relatively flat, but neither flat nor long enough to dilute the noise in the sense of Bartlett et al. (2020). This is consistent with the *incomplete* second descent of Section 5.1 (test MSE at $N = 1{,}200$ remains $\approx 2.1\times$ its low-dimensional level): Equation (3.3) predicts partial recovery whenever the variance term $\sigma^2(k^*/T + T/R_{k^*}(\Sigma))$ does not vanish, and with $T/R_{k^*}(\Sigma)$ of order 2 it cannot. The theoretical diagnostic and the empirical risk curve corroborate one another.

5.3 The European and emerging-markets regressions

To test whether the nature of the dataset affects the strength of benign overfitting, the entire regression protocol is replicated on the emerging-markets universe (Section 4.6): 24 pooled series, 24 lags ($N_{\max} = 576$), $T_{\text{train}} = 267$, giving $N_{\max}/T_{\text{train}} = 2.16$, deliberately comparable to the U.S. sweep's 2.33; the European replication is identical in design ($T_{\text{train}} = 281$, $N_{\max}/T_{\text{train}} = 2.05$). Table A.4 reports the emerging-markets results; the corresponding curves and effective-rank diagnostics for both international universes are plotted in Figures A.2–A.5 in Appendix A.

Three comparative findings emerge.

First, **the double descent phenomenon itself is universal across the three universes**: the emerging-markets sweep reproduces the qualitative shape of the U.S. one, with a rise from a low-complexity baseline, an explosive peak at $N = T_{\text{train}}$ (Ridgeless MSE of 2.75, some 1,300 times the $N = 2$ baseline of 0.00204), and a second descent thereafter, and the Ridgelet again contains the threshold singularity that the pseudo-inverse leaves fully exposed (0.149 versus 2.75 at $N = T$, a factor of ≈ 18; compare the factor of ≈ 106 found at the U.S. threshold).

Second, **the recovery is much weaker in emerging markets**. At the deepest tested complexity, the emerging-markets test MSE settles at ≈ 3.9 times its low-complexity baseline (0.00787 versus 0.00204), against ≈ 2.1 times for the United States. The second descent exists, but reclaims less of the lost ground.

Third (and this is the finding that ties the comparison to the theory) **the effective-rank diagnostics deteriorate in the direction that explains the weaker recovery**. Whereas the U.S. feature covariance, past its transition, exhibits a small dominant block ($k^*/T \approx 0.07$–0.18), the emerging-markets diagnostics show k^*/T *pinned at* 0.996 (the very edge of the empirical spectrum) throughout almost the entire over-parameterized range tested ($1.0 \leq N/T \lesssim 2.0$), with $R_{k^*}(\Sigma)$ undefined there; only at the extreme end of the sweep ($N/T \gtrsim 2.1$) does a finite diagnostic emerge ($k^*/T \approx 0.40$, $T/R_{k^*} \approx 2.7$), and even then at values far outside the favorable region. In the language of Section 3.2: for the U.S. universe, the first Bartlett condition ($k^*/T \to 0$) is approximately satisfied and only the second fails; for the emerging-markets universe, *both* conditions fail, the first one outright. The spectral theory thus does not merely rationalize each universe's outcome ex post, it correctly *ranks* the universes: the universe with the worse

effective-rank profile (emerging markets, built from only 24 underlying series with a higher idiosyncratic noise load) is precisely the one with the weaker second descent. Nor is the ranking a surprise by the time it arrives: the descriptive statistics of Table A.1 had already shown the emerging-markets correlation structure to be a spike without a tail (92.3% of variance in the first principal component, 97.5% in the first five), leaving almost no spectral mass over which interpolation noise could dissipate. The nature of the dataset does affect the detection of benign overfitting, and the effective ranks (previewed by the raw eigenvalue concentration, formalized by the diagnostics) are the instrument that quantifies by how much.

Europe: a developed market with an emerging spectrum. The European replication seals the comparative argument. Its double descent curve is the mildest of the three (threshold peak 0.314 against a 0.0025 baseline, tamed to 0.174 by the Ridgelet), and its recovery ratio ($\approx 2.5\times$ the low-dimensional level) sits between the U.S. ($2.1\times$) and emerging-markets ($3.9\times$) values. Its spectral diagnostics, however, are the worst of the three: k^*/T remains pinned at 0.996 (the very edge of the spectrum) across the *entire* over-parameterized range, with R_{k^*} undefined throughout, whereas even the emerging-markets diagnostics open up at the far end of their sweep. In the binary classification, the European second descent appears but does not fall below the first regime's level. The pattern is exactly what the descriptive statistics predicted: with a first principal component carrying 90.6% of variance, Europe's covariance structure is nearly as concentrated as that of the emerging universe despite its developed-market label, and it is this structure that determines the outcome.

5.4 The double ascent of the Sharpe ratio

The rolling backtest uses the full available history of the 100 Fama–French portfolios with a complete cross-section (no missing asset): 749 usable months between July 1945 and April 2026, once the 449 months with an incomplete cross-section are dropped. These are concentrated in, but not confined to, the early decades, when extreme size/book-to-market deciles are thinly populated; the rolling windows are defined on this filtered sample. Figure II.2 plots the annualized out-of-sample Sharpe ratio of the six funds against N/T for each estimation window.

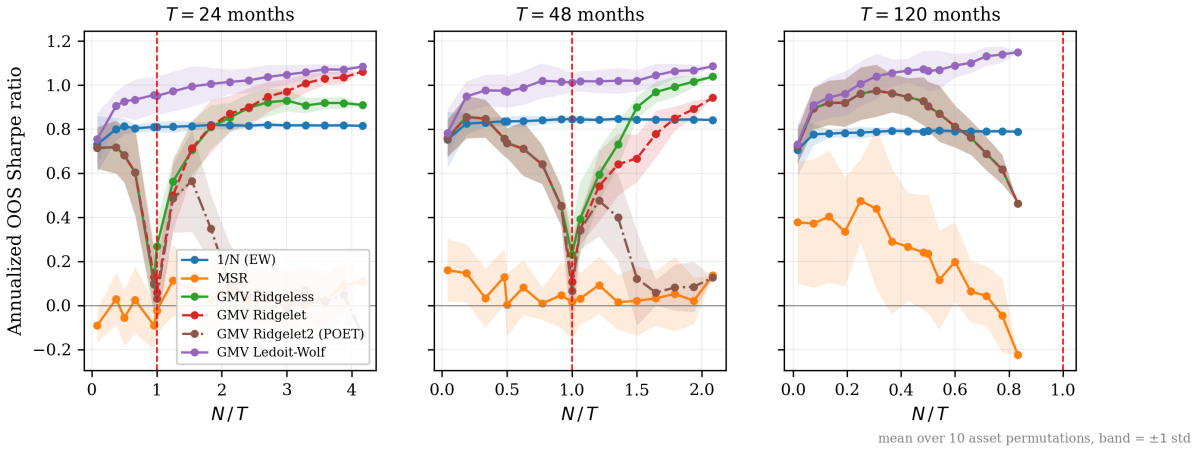

Figure II.2: Annualized out-of-sample Sharpe ratio of the six funds versus $N/T, T \in \{24, 48, 120\}$ months, U.S. universe. The dashed line marks the interpolation threshold $N = T$.
Source: author's calculations.

Table I.1 reports the Sharpe ratios at the three reference regimes (under-parameterized, $N/T \approx 0.5$; interpolation, $N/T \approx 1$; and the deepest reachable over-parameterization).

	$T = 24$ (TAA)			$T = 48$			$T = 120$ (SAA)	
N/T:	0.5	1.0	4.17	0.5	1.0	2.08	0.5	0.83
EW	0.812	0.810	0.814	0.835	0.845	0.841	0.791	0.788
MSR	−0.056	−0.024	0.110	0.003	0.016	0.136	0.235	−0.225
GMV-Ridgeless	0.681	0.267	0.909	0.737	0.231	**1.038**	0.903	0.462
GMV-Ridgelet	0.681	0.059	**1.061**	0.737	0.108	0.942	0.903	0.463
GMV-Ridgelet2	0.681	0.031	−0.134	0.737	0.066	0.127	0.903	0.462
GMV-LW	**0.926**	**0.950**	1.084	**0.969**	**1.011**	1.085	**1.063**	**1.149**

Table I.1: Annualized out-of-sample Sharpe ratios by complexity regime and estimation window, U.S. universe: means over ten asset orderings ($N = 100$ throughout; at $T = 120$ the last column is the maximum reachable $N/T = 0.83$).
Source: author's calculations on monthly data from the Kenneth R. French Data Library.

Table I.2 reports the full metrics dashboard in the flagship configuration ($T = 24$, $N = 100$, $N/T = 4.17$). The complete grid (3 windows × 15–17 values of N × 6 strategies × 8 metrics) is provided in the accompanying `partie3_full_metrics.csv`.

Strategy	SR	Net SR	Ret.	Vol.	Max DD	VaR_{95}	Turn.	α	α_5	R^2_5	β
EW	0.814	**0.811**	0.151	0.174	−0.485	−0.070	**0.022**	−0.001	−0.001	0.994	1.108
MSR	0.110	−0.112	2.317	11.47	−20.02	−0.264	85.1	0.636	0.555	0.010	−4.044
GMV-Ridgeless	0.909	0.487	0.135	0.140	−0.410	−0.055	1.975	0.017	−0.004	0.721	0.767
GMV-Ridgelet	**1.061**	0.679	0.151	**0.133**	**−0.343**	**−0.048**	1.700	**0.038**	**0.024**	0.553	0.654
GMV-Ridgelet2	−0.134	−0.428	−0.924	17.28	−32.57	−0.461	290.8	−3.929	−4.495	0.027	13.76
GMV-LW	1.084	0.808	0.148	0.128	−0.358	−0.047	1.175	0.033	0.017	0.664	0.689

Table I.2: Full metrics dashboard at $T = 24$, $N = 100$ ($N/T = 4.17$). Returns, volatility and alphas annualized; VaR monthly at 95%; "Net SR" deducts 25 bps per unit of drift-inclusive turnover.

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

Five findings emerge.

(1) The double ascent of the Sharpe ratio is clearly present, in the GMV family. For both windows that reach the over-parameterized regime, the minimum-variance strategies trace exactly the double-ascent shape predicted by Lu et al. (2024) and Chang et al. (2026). At $T = 24$, the GMV-Ridgelet Sharpe ratio declines from 0.68 at $N/T = 0.5$ to a trough of 0.06 at the interpolation threshold, then *ascends monotonically* (up to sampling noise) through the over-parameterized regime to 1.06 at $N/T = 4.17$, ending well above its low-dimensional level. The same pattern holds at $T = 48$ (trough of 0.11 at $N/T = 1$, recovery to 0.94 at $N/T = 2.08$). This is the portfolio-level counterpart of the regression double descent documented in Section 5.1, with the trough located exactly at $N = T$ in both windows.

(2) In the deep over-parameterized regime, optimized complexity beats naive diversification. At $T = 24$, $N/T = 4.17$, the GMV-Ridgelet (1.061) and GMV-LW (1.084) funds deliver Sharpe ratios roughly 30% above the EW benchmark (0.814), with *lower* volatility (0.133 and 0.128 versus 0.174), *shallower* maximum drawdowns (−0.34 and −0.36 versus −0.49), smaller tail risk, positive annualized Jensen's alphas (3.8% and 3.3% against a market beta of only ≈ 0.65–0.69), and moderate turnover. The DeMiguel et al. (2009b) verdict (that optimization never beats $1/N$) is thus confirmed for the classical tangency portfolio (finding 4 below) but *overturned*, and, in the binary terms of the double ascent, the second regime here both appears and *overtakes* the first: deep-regime Sharpe ratios (1.061–1.084) exceed the best under-parameterized configuration of the same window (0.926 for GMV-LW at $N/T = 0.5$), for regularized minimum-variance portfolios operating deep in the over-parameterized regime: on this universe and sample, complexity, properly stabilized, pays. Two further tests qualify this headline in opposite directions. The premium is *genuine estimation alpha, not disguised smart beta*: under the five-factor attribution, the Ridgelet's alpha falls from 3.8% to 2.4% annualized but remains material, with only 55% of the fund's variance explained by the five factors, while the Ridgeless fund's alpha is entirely absorbed (1.7% to −0.4%), one more respect in which the tiny-ridge stabilization dominates the

raw pseudo-inverse, and the EW benchmark's five-factor R^2 of 0.994 validates the attribution mechanically. However, *the capture of this alpha is cost-constrained*: at a 25 basis-point one-way cost per unit of turnover, the Ridgelet's net Sharpe ratio (0.679) falls below the nearly cost-free benchmark (0.811) and the shrinkage fund's (0.808) statistically ties it. The gross complexity gain is real and factor-robust, but its economic capture is bounded: at 10 basis points the premium survives nearly intact (net Sharpe ratios of 0.909 for the Ridgelet and 0.974 for the shrinkage fund, against 0.813 for the benchmark), so the break-even cost level lies between 10 and 25 basis points per unit of turnover, a demanding but realistic requirement for liquid, diversified exposures.

(3) The Ridgeless–Ridgelet divergence appears exactly where the theory places it. In the under-parameterized regime the two GMV variants are numerically identical at every tested configuration (e.g., both 0.681 at $T = 24$, $N/T = 0.5$), as they must be while $\hat{\Sigma}$ remains invertible. They separate only at and beyond the interpolation threshold (where $\mathbf{1}$ acquires null-space components) and in the deepest reachable regime ($T = 24$, $N/T = 4.17$) the Ridgelet dominates the Ridgeless (1.061 versus 0.909) with lower turnover (1.70 versus 1.98), consistent with the theoretical prediction of Chang et al. (2026) that the tiny-ridge solution, not the pseudo-inverse one, attains out-of-sample optimality as N/T grows. At the more moderate $T = 48$, $N/T = 2.08$, the ordering is reversed (0.942 versus 1.038), suggesting that the Ridgelet advantage requires genuinely deep over-parameterization to manifest reliably, an observation consistent with the asymptotic ($N \gg T$) character of the Chang et al. result. In the deep regime, moreover, the tuning-free Ridgelet performs within two points of Sharpe of the cross-sectionally calibrated Ledoit–Wolf fund (1.061 versus 1.084), echoing the claim of Chang et al. (2026) that careful tuning is unnecessary once the tiny ridge is in place.

(3′) The refined Ridgelet2 fails exactly where its theory predicts it should shine, and the failure is informative. The recommended estimator of Chang et al. (2026), which aligns the tiny ridge on a sparse estimate of the idiosyncratic covariance, behaves identically to the baseline funds below the interpolation threshold (where the tiny ridge is inert, an implementation validation) and then collapses exactly in the over-parameterized regime it was designed for: at $T = 24$, $N = 100$ its Sharpe ratio is -0.134 against the baseline Ridgelet's 1.061, with an annualized volatility of 17.3, a market beta of 13.8 and a monthly turnover near 291; the same collapse appears at $T = 48$ (0.127 versus 0.942) and, in weaker form, in the emerging-markets deep regime (0.45 versus 0.55); the Ridgelet2 fund is omitted from Table I.3 for space, its emerging-markets series being available in `partie3_full_metrics_EM.csv`. The mechanism is a violated premise rather than a flawed theorem: Ridgelet2's optimality (their Theorem 2) requires a *consistent* estimate of a substantial, sparse idiosyncratic covariance, plausible for hundreds of individual stocks, but not for diversified style portfolios in which most idiosyncratic risk is removed by construction (Table A.1: the first five principal components carry 97.5% of emerging-markets

variance) and $T = 24$ observations cannot estimate what little remains. Aligned on estimation noise, the Ω-weighted ridge selects, among the infinitely many zero-variance portfolios, one with explosive leverage. An implementation issue in the POET target estimator cannot be ruled out entirely as a contributing factor; two observations, however, bound its plausibility: the fund coincides exactly with the baseline strategies everywhere below the interpolation threshold, where the tiny ridge is inert (so the shared machinery is validated), and the collapse scales with the spectral concentration of the universe exactly as the violated-premise reading predicts. The practical lesson sharpens the thesis's central message: the choice *within* the Ridgelet family is itself spectrum-dependent, on factor-saturated aggregate universes, the identity-aligned Ridgelet1 is the robust member of the family, and the refinement's advantage should be expected only where a rich idiosyncratic cross-section exists to be exploited.

(4) The classical tangency portfolio is uninterpretable, not merely bad. The MSR fund's dashboard entries (annualized "returns" of 232%, volatility of 1,147%, drawdowns exceeding −2,000%, average turnover of 85) are not performance figures in any economic sense: they are artifacts of the normalization $\hat{w} \propto \hat{\Sigma}^{+}\hat{\mu}$ dividing by $\mathbf{1}^{\top}\hat{\Sigma}^{+}\hat{\mu}$, which passes arbitrarily close to zero in some estimation windows and produces unbounded leverage, the finite-sample pathology characterized analytically by Kan and Zhou (2007). The economically meaningful comparison in this analysis is therefore EW versus the GMV family; the MSR column is retained as a clear illustration of what interpolation-regime estimation risk does to an estimator that must also estimate $\hat{\mu}$. Its erratic behaviour across the whole sweep explains why the clean double-ascent shape only becomes visible once mean estimation is removed from the problem: an EW-versus-MSR comparison alone (the natural first cut of this analysis) shows only noise around the threshold, with no discernible double-ascent shape.

(5) The benign regime is a tactical, not strategic, phenomenon. At $T = 120$ (SAA), where N/T never exceeds 0.83, the GMV-Ridgeless and GMV-Ridgelet funds behave exactly as classical theory predicts: performance *deteriorates* monotonically as N/T rises toward the (unreachable) threshold, from 0.90 at $N/T = 0.5$ to 0.46 at $N/T = 0.83$, and the only strategies worth holding are the shrinkage fund (GMV-LW, 1.15) and the naive benchmark. The double ascent is not available to a strategic allocator on this universe, not because the theory fails, but because the regime in which it operates cannot be reached with $N_{\max} = 100$ assets and a ten-year window. Conversely, the tactical window ($T = 24$) is precisely what pushes the allocation problem into $N/T \gg 1$ territory where the complexity gains of finding (2) become available. This is the thesis's central finding in numbers: *benign overfitting in portfolio construction is a tactical asset allocation phenomenon*. The distinction, it should be stressed, is not about using recent (TAA) versus older (SAA) data per se: what governs the outcome is solely the size of the ratio N/T, and the TAA/SAA divide matters only because, by construction, a strategic window makes T

large and N/T small, while a tactical window does the opposite. The cost accounting adds a complementary nuance: the strategic shrinkage premium, built on low turnover, is the most cost-robust result in the study (GMV-LW at $T = 120$ still clears the benchmark by 0.12 of Sharpe net of 25 basis points), long-window allocators need no complexity, and the classical prescription they should follow also happens to be the cheapest to implement.

5.4.1 European and emerging-markets portfolios

The portfolio protocol is replicated on the two 24-asset international universes (Europe and emerging markets) with estimation windows $T \in \{12, 24, 48\}$ months. The narrow cross-section makes the exercise structurally constrained: even at the shortest tactical window ($T = 12$), the maximum reachable aspect ratio is $N/T = 2.0$, while $T = 24$ only just reaches the interpolation threshold and $T = 48$ never leaves the classical regime ($N/T \leq 0.5$). Figure II.3 plots the Sharpe ratio curves and Table I.3 reports the three reference regimes at the tactical window; four observations follow (full grid in `partie3_full_metrics_EM.csv`).

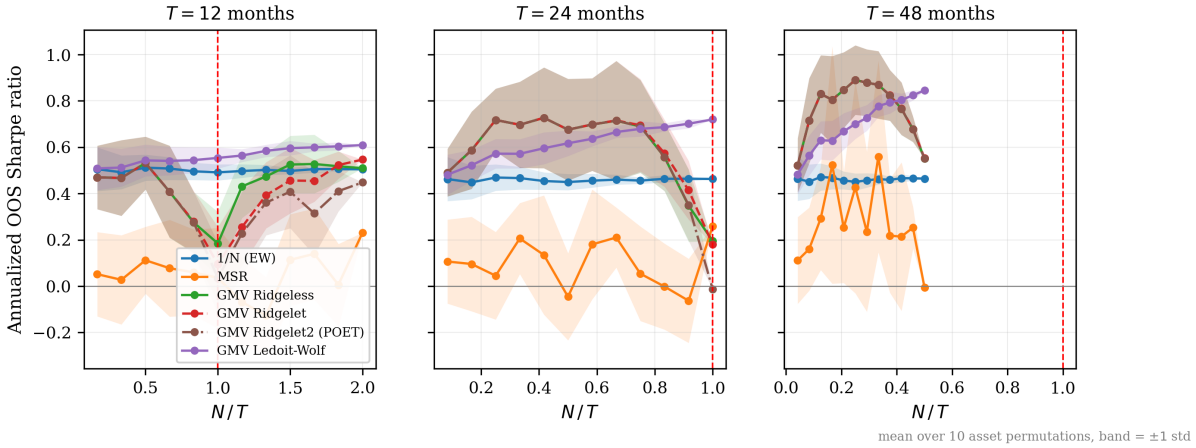

Figure II.3: Annualized out-of-sample Sharpe ratio of the six funds versus N/T, $T \in \{12, 24, 48\}$ months, emerging-markets universe. The dashed line marks the interpolation threshold $N = T$.
Source: author's calculations.

First, the interpolation trough replicates, and does so robustly. All emerging-markets figures are averaged over *ten* random asset permutations (see the robustness design below), so the collapse at $N = T$ cannot be an artifact of one particular universe composition: at $T = 12$, the GMV-Ridgelet Sharpe ratio falls from 0.53 ($N/T = 0.5$) to 0.10 at the threshold, a gap several times larger than the cross-permutation dispersion at either point (±0.11 and ±0.16 respectively), before recovering to 0.55 at $N/T = 2$; at $T = 24$, it falls from 0.68 (±0.22) at $N/T = 0.5$ to 0.18 at the threshold, which is also the end of the reachable range.

	Sharpe (std)			Turnover			α (ann.)		
N/T:	0.5	1.0	2.0	0.5	1.0	2.0	0.5	1.0	2.0
EW	0.512 (.06)	0.491 (.04)	0.504 (.00)	0.00	0.00	0.00	0.006	0.001	0.004
MSR	0.112 (.15)	0.036 (.22)	0.231 (.00)	444	571	203	(	)	—
GMV-Ridgeless	0.532 (.11)	0.184 (.14)	0.510 (.00)	4.73	37.8	11.0	0.055	0.032	0.040
GMV-Ridgelet	0.532 (.11)	0.095 (.16)	**0.547** (.00)	4.73	130.1	9.92	0.055	0.070	**0.061**
GMV-LW	**0.543** (.07)	**0.554** (.05)	**0.609** (.00)	**0.43**	**0.78**	**1.26**	0.012	0.015	0.027

Table I.3: Emerging markets, $T = 12$: Sharpe ratio (mean over ten asset permutations, standard deviation in parentheses), average monthly turnover and annualized Jensen's alpha by regime.
Source: author's calculations on monthly data from the Kenneth R. French Data Library.

Second, the deep-regime ranking matches the U.S. one. At the deepest reachable configuration ($T = 12$, $N/T = 2.0$), GMV-LW (0.609) > GMV-Ridgelet (0.547) > GMV-Ridgeless (0.510) > EW (0.504) $\gg$ MSR (0.231), but the gains are much smaller: the Ridgelet's edge over the naive benchmark is roughly +9% in Sharpe ratio, against +30% in the U.S. deep regime. Two non-exclusive explanations fit the evidence: the reachable depth is half the U.S. one ($N/T = 2$ versus 4.17), and the universe operates at a lower signal-to-noise level throughout (EW Sharpe ratio of ≈ 0.5 against ≈ 0.8).

Third, the comparison reinforces the conclusion of Section 5.3 at the allocation level: the same mechanism operates in all three universes (trough at $N = T$, recovery beyond, identical estimator ordering), but its economic payoff scales with the spectral quality of the universe: small gains where the effective-rank diagnostics are poor (Europe and emerging markets), sizeable gains where they are partially favorable (United States).

Fourth, implementability deteriorates alongside performance. At the deepest reachable regime, the GMV-Ridgelet turns over ≈ 9.9 monthly (one-way ℓ_1) against ≈ 1.7 in the U.S. deep regime, both pseudo-inverse funds spike above 30–100 at the threshold, and only the shrinkage fund stays implementable (0.4–1.3). At 25 basis points of one-way cost, the deep-regime GMV-Ridgelet falls from 0.55 gross to −0.80 net, the GMV-Ridgeless from 0.51 to −0.98, and even the shrinkage fund (0.40 net) slips below the naive benchmark (0.50 net). *Net of even modest costs, nothing beats* $1/N$ *in the deepest reachable regime of this universe*; away from it, only the classical shrinkage fund survives costs (0.56 net versus 0.46 at $T = 24$), and its premium owes nothing to over-parameterization. The verdict of DeMiguel et al. (2009b), overturned gross-of-cost in the U.S. deep regime, is restored here once implementation is priced. The five-factor attribution points the same way: the deep-regime Ridgelet's CAPM alpha of 6.1% shrinks to 1.0% once SMB, HML, RMW and CMA exposures are controlled for (five-factor $R^2 = 0.21$), so most of the apparent "complexity alpha" was factor exposure to begin with, while the benchmark's R^2 of 0.999 validates the attribution mechanically.

The European portfolios. The European backtests (Figure A.6, Appendix A) mirror the emerging-markets pattern, only more starkly. At the deepest reachable node ($T = 12$, $N/T = 2$), the stabilized funds capture nothing: the Ridgelet delivers a Sharpe ratio of 0.25 against 0.57 for the naive benchmark and 0.64 for Ledoit–Wolf, and only the raw pseudo-inverse edges above the benchmark gross (0.67), an advantage annihilated at 25 basis points (net −1.15, against 0.57 for the benchmark). At $T = 24$ the interpolation threshold delivers its usual collapse (Ridgelet −0.24 versus Ledoit–Wolf 0.86), and on longer windows the shrinkage fund dominates unchallenged (0.91 gross, 0.78 net at $T = 48$). In the binary classification, the European second ascent appears for the pseudo-inverse but never for the stabilized funds, and nothing complexity-driven survives costs, the same verdict as emerging markets, delivered by a developed market whose first principal component carries 90.6% of variance. Across the three universes, the ranking of gains (US $\gg$ Europe $\approx$ EM) follows the spectral ranking exactly, not the development label.

Robustness across universe compositions. A single-ordering sweep mixes estimation noise (the object of study) with *composition* noise (which assets enter first); to separate them, the sweeps of all three universes were repeated over ten random orderings (seeds 42–51), in the spirit of DeMiguel et al. (2009b) and Kelly et al. (2024), and every portfolio figure and table reports means across repetitions. The headline features (the threshold collapse of the pseudo-inverse funds, the stability of the shrinkage fund) hold with margins far exceeding the cross-permutation dispersion; fine-grained sub-threshold orderings move within one standard deviation (±0.02–0.27 in Sharpe) and should not be over-interpreted; and at $N = N_{\max}$ every permutation holds the identical universe, making the deepest-regime comparisons mechanically composition-invariant (see limitation (i), Section 6.2).

5.4.2 Spectral diagnostics of the return covariances

The diagnostics of Section 5.2 were computed on the regression feature covariance; the portfolio experiment, however, inverts a different object, the rolling T-month return covariance. Table I.4 therefore reports the same diagnostics on those matrices, averaged over sub-sampled rolling windows at each (N, T) node.

Universe	Regime	T	N/T	k^*/T	T/R_{k^*}
US	threshold	24	1.00	0.770	2.68
US	deep (TAA)	24	4.17	**0.107**	1.79
US	threshold	48	1.00	0.638	2.60
US	deep	48	2.08	**0.102**	1.98
US	max reachable	120	0.83	0.750	∞
Europe	threshold	12	1.00	0.917	∞
Europe	deep	12	2.00	0.819	2.83
Europe	max reachable	24	1.00	0.958	∞
EM	threshold	12	1.00	0.917	∞
EM	deep	12	2.00	0.848	2.78
EM	max reachable	24	1.00	0.958	∞

Table I.4: Effective-rank diagnostics of the rolling return covariance matrices at selected (T, N) nodes, averaged over rolling windows (12-month sub-sampling step). *Threshold* denotes the interpolation node $N/T = 1$; *deep* the deepest over-parameterized node reached by the sweep at that window; *max reachable* the largest attainable N/T when the window is too long for the threshold to be crossed ($N_{\max}/T < 1$) or the cross-section too narrow to go beyond it. Entries marked ∞ correspond to windows in which k^* falls at the edge of the nonzero spectrum of the rank-deficient rolling covariance, so the tail beyond k^* contains no nonzero eigenvalues and R_{k^*} is a $0/0$ ratio; T/R_{k^*} is then reported as ∞, the conservative reading (tail-mass condition maximally violated).

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

Three conclusions close the analogy. First, **the partial-conditions configuration of the regression experiment (dominant low-dimensional block present, tail mass insufficient) replicates on the matrices the portfolio estimators actually invert**: past the interpolation threshold the U.S. return-covariance spectrum opens up (k^*/T collapses from ≈ 0.77 at $N = T$ to 0.107 at $N/T = 4.17$) while the tail-mass ratio T/R_{k^*} declines but never approaches the favorable region. Notably, in the deep tactical regime the dominant-block condition is, for the first time in this thesis, *formally satisfied* ($k^*/T \approx 0.11 < 0.2$ at the deepest $T = 24$ node, ≈ 0.10 at $N/T = 2.08$ for $T = 48$): the Sharpe double ascent delivers its gains exactly where the spectrum comes closest to the benign configuration. Second, **the strategic window never gets there**: at $T = 120$ the spectrum stays pinned near the boundary at every reachable N ($k^*/T \approx 0.75$, R_{k^*} undefined in nearly every window), so finding (5) is visible in the eigenvalues themselves, not only in the Sharpe ratios. Third, **the cross-universe ranking replicates**: the European and emerging-markets return covariances stay at the spectral edge ($k^*/T \approx 0.82$–0.96) throughout, as in Section 5.3, and these are the universes with the smallest complexity gains. Measured on the actual objects of the allocation problem, the effective ranks thus order regimes within a universe, windows within a strategy, and universes against each other: they work as an operational pre-trade diagnostic, not only a theoretical device.

6. Discussion and conclusion

6.1 Discussion of the regression evidence

Three conclusions can be drawn from the combined evidence of Sections 5.1 and 5.2.

First, the qualitative double descent pattern predicted by Belkin et al. (2019) and characterized precisely by Hastie et al. (2022) is present in this Fama–French return-prediction setting: the sharp risk peak at the interpolation threshold $N = T_{\text{train}}$, followed by a genuine second descent as N grows further, is a robust empirical feature of the OLS/minimum-norm estimator applied to lagged portfolio returns, not an artifact of a particular modeling choice.

Second, strict benign overfitting in the sense of Bartlett et al. (2020) is *not* established by this analysis: while the first spectral condition ($k^*/T \to 0$) is reasonably well satisfied in the over-parameterized regime, the second ($T/R_{k^*}(\Sigma) \to 0$) is not, remaining an order of magnitude above the favorability threshold throughout the tested complexity range. In practice, this shortfall appears as the incomplete recovery of test MSE documented in Section 5.1: the theory and the empirics agree that some, but not all, of the variance penalty of interpolation is eliminated by over-parameterization here. For a finance application, this intermediate outcome is more informative than either extreme (clean benign overfitting, or no double descent at all).

Third, the comparison between the (near-)ridgeless estimator and the $\alpha = 1$ ridge estimator empirically confirms the theoretical prediction of Tsigler and Bartlett (2023) that a fixed positive regularization penalty removes the interpolation-threshold singularity almost entirely, and does so *without* requiring the favorable spectral structure that the ridgeless estimator would need: the ridge estimator's MSE stays flat and low throughout, independently of the fact that $T/R_{k^*}(\Sigma)$ never reaches the favorable range. This has a direct practical implication for the portfolio analysis of Section 5.4: whenever the strict spectral conditions for ridgeless benign overfitting cannot be verified or are not expected to hold, ridge regularization provides an effective safeguard against the interpolation singularity, at a negligible cost in the under-parameterized regime.

Limitations. The regression analysis rests on a single chronological train/test split, an equal-weighted prediction target whose sensitivity is untested, and a fixed rather than cross-validated ridge penalty; the latter deserves emphasis, since the heavy, uncalibrated $\alpha = 1$ leaves the ridge benchmark strongly over-regularized, and its near-flat risk curve partly reflects that heaviness. Re-running the sweep with a cross-validated penalty, and plotting the eigenvalues directly around the abrupt transition at $N \approx 455$–462, are the most direct extensions; none threatens the qualitative conclusions.

6.2 Discussion of the portfolio evidence

Three conclusions follow, each connecting back to Chapters 3–5.

First, the statistical double descent of Section 5.1 and the financial double ascent of Section 5.4 are now documented *on the same data, with the same estimator suite, and with the trough at the same threshold* ($N = T$ in both cases). The mechanism is the same (the minimum-norm resolution of an underdetermined system) and the correspondence between statistical risk and financial utility that Lu et al. (2024) develop theoretically is here exhibited empirically end-to-end. The one systematic difference is instructive: the regression experiment recovers only partially (test MSE at maximal complexity remains $\approx 2.1\times$ its low-dimensional level, Section 5.1), whereas the allocation experiment in the deep tactical regime recovers *beyond* its low-dimensional level. This asymmetry is consistent with the diagnostic finding of Section 5.2 that the Bartlett conditions hold only in part for the *lagged-return prediction* problem (whose signal-to-noise ratio is notoriously poor), while the GMV problem requires no return prediction at all: only covariance structure, which is far better identified in financial data. Where the signal is (covariance), complexity adds value; where it is weakest (mean prediction, cf. the MSR fund), complexity remains destructive, the boundary that the cautionary literature (Fallahgoul, 2025; Nagel, 2025) draws for the virtue-of-complexity programme.

Second, the estimator ordering found in the deep over-parameterized regime (Ridgelet $\approx$ Ledoit–Wolf > Ridgeless > EW $\gg$ MSR) matches the theoretical predictions of Chang et al. (2026) in both its placement of the Ridgelet above the Ridgeless and its suggestion that tuning-free stabilization suffices. For practice, this ordering carries a clear message: the gains from over-parameterized allocation do not require delicate calibration of tuning parameters, but they do require *some* stabilization of the covariance inverse: the raw pseudo-inverse leaves a material fraction of the available Sharpe ratio on the table, and raw tangency weights destroy it entirely. An even blunter reading of Figure II.2 deserves to be stated plainly: the Ledoit–Wolf fund is at or near the top of every panel, at every window and every N/T, so the two-decade-old practice of shrinkage estimation already delivers essentially all of the achievable gains. The contribution of the over-parameterization lens is not to displace that practice but to explain *why* it works (shrinkage as adaptive ridge, Section 3.4), to identify the regime in which a tuning-free alternative matches it, and to show where the singularity it sails over actually sits.

Third, the SAA/TAA asymmetry of finding (5), which is purely an asymmetry in the reachable ratio N/T, clarifies what benign overfitting is for in asset management. It is not a general improvement to portfolio construction: on long windows, the classical prescriptions (shrinkage, parsimony) remain correct and sufficient. It matters for short-window, wide-universe allocation, the regime into which any tactical process reacting to recent information across many assets is forced, and where the interpolation singularity has to be crossed rather than avoided.

Limitations and next steps. Five limitations qualify these results. (i) The composition-noise check (ten random asset orderings) covers all three universes, confirming the threshold collapse well beyond cross-permutation dispersion. It cannot, however, touch the deepest-regime comparisons, which are evaluated at $N = N_{\max}$ where every ordering holds the identical universe: their residual uncertainty is temporal, and would require a block bootstrap of the return history. (ii) The transaction-cost accounting is a simplified proportional sensitivity on turnover, not a measurement (Section 4.7), and its verdict is c-dependent: the U.S. gross premium is entirely consumed at 25 basis points, and every over-parameterization-driven fund falls below the naive benchmark in emerging markets at that same level; segment-heterogeneous costs and market impact, both outside the simplified cost setting, would sharpen these boundaries. (iii) No formal inference accompanies the point estimates: Sharpe-ratio differences and alphas are reported without heteroskedasticity-robust tests, so economically material gaps (e.g., 1.06 versus 0.81) are interpreted on magnitude and cross-regime consistency rather than statistical significance. (iv) The effective-rank diagnostics now cover both experiments (Section 5.4.2); what remains open is the calibration of the thresholding rule, $b = 0.2$ and the favorability cutoffs carry asymptotic conditions into finite samples and are, to that extent, heuristic. (v) All results assume enough stationarity for a trailing window to inform the following month; over-parameterized interpolators are exposed to *posterior drift* (Coqueret and Laguerre, 2025), and the minimum-norm and tiny-ridge solutions, leaning on the fine structure of the trailing covariance, have no built-in protection against abrupt breaks between monthly re-estimations. A sub-sample analysis around known break dates would be required to quantify, rather than flag, this vulnerability.

6.3 Synthesis of findings

Taken together, the two strands of evidence support a single operational reading: *benign overfitting in portfolio construction is real, but it is conditional*, and the conditions are now identifiable in advance rather than only in hindsight.

The first condition concerns *what is estimated*. Complexity added value only in the minimum-variance family, where the covariance matrix is the sole estimated input; wherever the mean vector entered the problem (the tangency portfolio), overfitting remained destructive at every complexity level, exactly as the estimation-risk literature predicts. For a practitioner, this is the sharpest line in the thesis: over-parameterized allocation is a covariance technology, not a forecasting technology.

The second condition concerns *how the inverse is taken*. The singularity at $N = T$ destroys any strategy built on an unstabilized inverse, but the two stabilizations tested here (the tuning-free tiny ridge of Chang et al. (2026) and the Ledoit and Wolf (2004) shrinkage) both convert the region beyond the threshold from a hazard into a source of gains, and end up within a few points

of Sharpe of one another in the deep regime, as expected from Section 3.4, which shows them to be one ridge family differing only in penalty intensity; the gains do not depend on delicate tuning.

The third condition concerns *the region of the* (N, T) *grid in which the strategy operates*. The deep over-parameterized regime is reachable only with short estimation windows on wide universes (the tactical corner of asset management), and only there do the gains materialize. On strategic horizons the classical prescriptions remain correct and sufficient.

The fourth condition concerns *the universe itself.* The effective-rank diagnostics, computed before any backtest, correctly anticipated which universe would reward complexity and which would not; and the cost accounting drew the economic boundary in all three: the thin European and emerging-markets margins are consumed outright, and even the U.S. premium (genuine, five-factor-robust alpha though it is) is entirely consumed at a 25 basis-point one-way cost, surviving only for low-cost implementations. Together, a pre-trade spectral diagnostic and cost-aware accounting turn the virtue of complexity from a slogan into a checklist: deploy stabilized minimum-variance complexity (GMV-Ridgelet or GMV-LW) on short-window, broad, liquid, high-SNR books; retain classical shrinkage on strategic mandates; never feed estimated means into an over-parameterized optimizer.

This thesis set out to determine whether the modern statistical theory of benign overfitting survives contact with financial data, and whether its statistical benefits translate into portfolio utility. Four findings summarize the answer.

First, *double descent is present in financial return prediction, but strict benign overfitting is not*. On the U.S. Fama–French cross-section, out-of-sample MSE traces the textbook curve (an explosive peak exactly at the interpolation threshold, followed by a genuine second descent) yet the effective-rank conditions of Bartlett et al. (2020) are only partially satisfied (the dominant-block condition approximately holds; the tail-mass condition fails), and, consistently, the second descent recovers only partially. The tuning-free Ridgelet estimator coincides with the Ridgeless one everywhere except at the threshold itself, where it tames a hundredfold MSE explosion, the predicted regression-side behaviour of the estimator suite.

Second, *the statistical phenomenon has a genuine financial counterpart*. In the minimum-variance family, the out-of-sample Sharpe ratio traces a double ascent in N/T with its trough exactly at $N = T$; in the deep tactical regime ($T = 24$ months, $N/T = 4.17$), the stabilized funds beat the $1/N$ benchmark by roughly 30% in Sharpe ratio with lower volatility, shallower drawdowns and positive Jensen's alpha, overturning in that regime the classical verdict of DeMiguel et al. (2009b). Two tests frame this result: the premium survives a Fama–French five-factor attribution (2.4% annualized residual alpha, against an entirely consumed alpha for the raw pseudo-inverse fund), but, while intact at a 10 basis-point one-way transaction cost, it is consumed at 25: genuine alpha, economically capturable only at low implementation cost.

Third, *the phenomenon is tactical, not strategic*. With 100 assets and a ten-year window, the over-parameterized regime is mechanically unreachable ($N/T \leq 0.83$); only short estimation windows push allocation past the threshold into the region where complexity pays. The virtue of complexity, in portfolio construction, is a property of short-window, high-breadth tactical allocation.

Fourth, *the nature of the dataset determines the size of the gains*. Replicated on the pooled European and emerging-markets universes, the entire mechanism reappears (same trough, same estimator ordering) but with weaker statistical recoveries, thinner Sharpe margins, and turnover levels that hand the residual gains to transaction costs. The most telling case is Europe: a developed market whose near-emerging covariance spectrum (first principal component at 90.6%) the diagnostics flag before any backtest, and whose gains indeed match the emerging-markets verdict rather than the U.S. one. The effective-rank diagnostics, computed ex ante, correctly rank all three universes (the spectrum, not the development label, determines the outcome) and serve as a practical pre-trade check.

6.4 Recommendations for systematic funds

Three recommendations follow for systematic managers.

First, *do not feed estimated means into an over-parameterized optimizer*. Across three universes, three windows and the full range of N/T, the plug-in tangency portfolio was not just dominated: its output was economically uninterpretable, with leverage pathologies of the kind characterized by Kan and Zhou (2007). Complexity should be spent on the covariance side of the problem, where the signal is.

Second, *stabilize the inverse first; only then increase breadth, shorten the window, and keep costs low*. On broad, liquid, high-SNR universes managed over short tactical windows, minimum-variance portfolios built on a tuning-free tiny ridge or on Ledoit–Wolf shrinkage captured a material gross Sharpe premium over $1/N$ that survives multi-factor attribution, but whose economic capture requires implementation costs toward the low end of the institutional range, the premium surviving intact at 10 but being entirely consumed at 25 basis points per unit of turnover; the same recipe on a narrow, noisy universe produced margins consumed outright by costs. A pre-trade check of the effective-rank diagnostics, together with honest net-of-cost accounting, separates these situations before capital is committed.

Third, *treat complexity sweeps with the scepticism that the multiple-testing literature calls for*. A design that evaluates dozens of configurations on one historical sample is exposed to backtest overfitting (Bailey et al., 2017; López de Prado, 2018), and the theoretical warnings of Fallahgoul (2025) and Nagel (2025) about weak-signal environments apply squarely to finance. The robustness practices used here (averaging over universe compositions, reporting dispersion,

and refusing to interpret differences within it) should be considered a minimum standard, not an optional extra.

6.5 Future research

Several extensions follow naturally. On the data side, the three-universe comparison could be extended to Asia-Pacific and Japanese sorts, or, more pointedly, to *individual-stock* universes, the habitat for which Ridgelet 2 was designed and where its idiosyncratic-covariance premise (Section 5, finding 3′) could finally be met; the size segmentation of the U.S. grid also remains available as a full experimental axis. On the methodology side, two robustness layers remain open: a moving-block bootstrap of the return history, to attach formal inference to the Sharpe-ratio gaps that this thesis interprets on magnitude and cross-regime consistency alone (limitation (iii)); and long-only constrained variants of the six funds, to measure how much of the gain survives the constraint that Jagannathan and Ma (2003) show acts as implicit shrinkage. The same diagnostic-first logic transfers to neighbouring fields. Derivatives hedging is a genuine gap: a mature over-parameterized literature exists there, from deep hedging (Buehler et al., 2019) and its reinforcement-learning implementations (Cao et al., 2021) to the option-pricing strand surveyed by Ruf and Wang (2020), yet no interpolation threshold has been located and no effective-rank diagnostics computed on the relevant covariances; the transfer is non-mechanical, since simulated training paths make the binding ratio network capacity against simulation budget. Credit scoring and volatility modelling, where high-dimensional features meet scarce observations, are further candidates, as are the generative-model approaches now reaching European asset pricing (Mattusch, 2025). On the theory side, the effective-rank diagnostics (now computed on both the regression features and the return covariances) would benefit from principled finite-sample calibration of their thresholds, replacing the heuristic constants used here; and the stability of the benign regime under structural change deserves attention, since over-parameterized models are known to be exposed to posterior drift when the data-generating process moves (Coqueret and Laguerre, 2025). Finally, an open direction is the feature space itself: the tactical settings that motivate this thesis increasingly ingest alternative data and NLP-extracted sentiment, generating feature counts that dwarf any sample size (Kelly and Xiu, 2023); whether the benign-overfitting checklist developed here (spectral diagnostics, stabilized inverses, cost-aware evaluation) carries over to those genuinely high-dimensional signal sets is the most consequential open question.

Bibliography

Bailey, D. H., Borwein, J. M., López de Prado, M., and Zhu, Q. J. (2017). The probability of backtest overfitting. *Journal of Computational Finance*, 20(4):39–70. https://doi.org/10.21314/JCF.2016.322.

Balduzzi, P. and Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78. https://doi.org/10.1016/S0304-405X(99)00004-5.

Bartlett, P. L., Long, P. M., Lugosi, G., and Tsigler, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48):30063–30070. https://doi.org/10.1073/pnas.1907378117.

Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854. https://doi.org/10.1073/pnas.1903070116.

Buehler, H., Gonon, L., Teichmann, J., and Wood, B. (2019). Deep hedging. *Quantitative Finance*, 19(8):1271–1291. https://doi.org/10.1080/14697688.2019.1571683.

Cao, J., Chen, J., Hull, J., and Poulos, Z. (2021). Deep hedging of derivatives using reinforcement learning. *Journal of Financial Data Science*, 3(1):10–27. https://doi.org/10.3905/jfds.2020.1.052.

Cassio, T. (2024). Benign overfitting and its applications in finance. Master's thesis, UCLouvain. https://thesis.dial.uclouvain.be/entities/masterthesis/f2b02dfe-ddaf-4506-a172-bff9a1e2fea7.

Chang, J., Ding, Y., Shi, Z., and Zhang, B. (2026). Zero variance portfolio. *arXiv preprint arXiv:2602.19462*. https://arxiv.org/abs/2602.19462.

Chen, A. Y. (2021). Most claimed statistical findings in cross-sectional return predictability are likely true. *Journal of Finance: Insights, forthcoming*. https://doi.org/10.2139/ssrn.3912915.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108. https://doi.org/10.1111/j.1540-6261.2011.01671.x.

Coqueret, G. and Laguerre, M. (2025). Overparametrized models with posterior drift. *arXiv preprint arXiv:2506.23619*. https://arxiv.org/abs/2506.23619.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812. https://doi.org/10.1287/mnsc.1080.0986.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953. https://doi.org/10.1093/rfs/hhm075.

Fabozzi, F. J. (2025). A conversation with don marcos (aka marcos lópez de prado). *Journal of Financial Data Science*, 7(1). https://www.pm-research.com/content/iijjfds/7/1/10.

Fallahgoul, H. (2025). High-dimensional learning in finance. *arXiv preprint arXiv:2506.03780*. https://arxiv.org/abs/2506.03780.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56. https://doi.org/10.1016/0304-405X(93)90023-5.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22. https://doi.org/10.1016/j.jfineco.2014.10.010.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B*, 75(4):603–680. https://doi.org/10.1111/rssb.12016.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370. https://doi.org/10.1111/jofi.12883.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273. https://doi.org/10.1093/rfs/hhaa009.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68. https://doi.org/10.1093/rfs/hhv059.

Hastie, T., Montanari, A., Rosset, S., and Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2):949–986. https://doi.org/10.1214/21-AOS2133.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition. https://doi.org/10.1007/978-0-387-84858-7.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4):1651–1683. https://doi.org/10.1111/1540-6261.00580.

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

Kelly, B. T., Malamud, S., and Zhou, K. (2022). The virtue of complexity everywhere. *Swiss Finance Institute Research Paper No. 22-57*. https://doi.org/10.2139/ssrn.4166368.

Kelly, B. T., Malamud, S., and Zhou, K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503. https://doi.org/10.1111/jofi.13298.

Kelly, B. T. and Xiu, D. (2023). Financial machine learning. *Foundations and Trends in Finance*, 13(3–4):205–363. https://doi.org/10.1561/0500000064.

Lassance, N., Vanderveken, R., and Vrins, F. (2025). Optimal portfolio size under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, pages 1–41. https://doi.org/10.1017/S0022109025102457.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411. https://doi.org/10.1016/S0047-259X(03)00096-4.

López de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley. https://www.wiley.com/en-us/Advances+in+Financial+Machine+Learning-p-9781119482086.

Lu, Y., Yang, Y., and Zhang, T. (2024). Double descent in portfolio optimization: Dance between theoretical sharpe ratio and estimation accuracy. *arXiv preprint arXiv:2411.18830*. https://arxiv.org/abs/2411.18830.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91. https://doi.org/10.2307/2975974.

Mattusch, M. (2025). Generative ai for european asset pricing: Alleviating the momentum anomaly. *The European Journal of Finance*, 31(7):850–888. https://doi.org/10.1080/1351847X.2024.2439979.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361. https://doi.org/10.1016/0304-405X(80)90007-0.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42. https://doi.org/10.2469/faj.v45.n1.31.

Nagel, S. (2025). Seemingly virtuous complexity in return prediction. *Chicago Booth Research Paper No. 25-10*. https://doi.org/10.2139/ssrn.5335012.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413. https://doi.org/10.1017/S0305004100030401.

Petukhina, A., Klochkov, Y., Härdle, W. K., and Zhivotovskiy, N. (2024). Robustifying markowitz. *Journal of Econometrics*, 239(2):105387. https://doi.org/10.1016/j.jeconom.2022.12.006.

Ruf, J. and Wang, W. (2020). Neural networks for option pricing and hedging: A literature review. *Journal of Computational Finance*, 24(1):1–46. https://arxiv.org/abs/1911.05620.

Suhonen, A., Lennkh, M., and Perez, F. (2017). Quantifying backtest overfitting in alternative beta strategies. *The Journal of Portfolio Management*, 43(2):90–104. https://doi.org/10.3905/jpm.2017.43.2.090.

Tsigler, A. and Bartlett, P. L. (2023). Benign overfitting in ridge regression. *Journal of Machine Learning Research*, 24(123):1–76. https://jmlr.org/papers/v24/22-1398.html.

Xiong, X. (2026). What do large factor models learn? self-induced regularization, cost of overfitting, and self-adaptivity. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.6039374.

A. Supplementary tables and figures

This appendix gathers the detailed numerical tables and the supplementary figures underlying results discussed in the main text. Each table is referenced at the relevant point of the argument; the corresponding figures and headline numbers appear in the body of the thesis.

	US (all)	US small	US large	Europe	EM
Number of series	100	50	50	24	24
Months	749	749	749	427	407
XS ann. mean return	13.9%	14.0%	13.9%	9.6%	10.6%
XS ann. volatility	20.4%	22.2%	18.6%	17.9%	21.0%
XS skewness	−0.15	−0.08	−0.22	−0.54	−0.49
XS excess kurtosis	2.86	3.43	2.29	2.36	2.27
Avg pairwise correlation	0.74	0.79	0.74	0.90	0.92
PC1 variance share	74.6%	79.5%	74.4%	90.6%	92.3%
PC1–PC5 variance share	82.9%	85.9%	83.2%	98.4%	97.5%
EW ann. return	13.9%	14.0%	13.9%	9.6%	10.6%
EW ann. volatility	17.5%	19.7%	16.0%	17.1%	20.2%
EW Sharpe ratio	0.80	0.71	0.87	0.56	0.53
Mkt-RF Sharpe ratio	0.57	0.57	0.57	0.40	0.38

Table A.1: Descriptive statistics of the investment universes: cross-sectional averages of annualized per-series moments, spectral concentration (PC shares) and the equally-weighted portfolio.

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

N	N/T_{train}	Ridgeless	Ridgelet	Ridge ($\alpha = 1$)
2	0.00	0.00269	0.00269	0.00269
124	0.24	0.00359	0.00359	0.00278
247	0.48	0.00575	0.00575	0.00299
370	0.72	0.01316	0.01316	0.00291
455	0.88	0.03259	0.03257	0.00297
506	0.98	0.18952	0.18259	0.00301
515	**1.00**	**20.398**	**0.19246**	**0.00302**
523	1.02	0.18029	0.17401	0.00301
549	1.07	0.05742	0.05730	0.00303
616	1.20	0.01905	0.01905	0.00308
800	1.55	0.00821	0.00821	0.00312
1015	1.97	0.00616	0.00616	0.00315
1200	2.33	0.00574	0.00574	0.00316

Table A.2: United States: out-of-sample MSE as a function of model complexity N, for the Ridgeless (minimum-norm), Ridgelet (tuning-free tiny ridge) and Ridge ($\alpha = 1$) estimators. $T_{\text{train}} = 515$.

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

N	N/T_{train}	k^*/T_{train}	$T_{\text{train}}/R_{k^*}(\Sigma)$	Favorable?
2	0.00	0.002	∞	No
124	0.24	0.239	∞	No
247	0.48	0.478	∞	No
370	0.72	0.717	∞	No
455	0.88	0.882	∞	No
462	0.90	0.175	2.692	No
506	0.98	0.142	2.513	No
515	**1.00**	**0.136**	**2.493**	**No**
616	1.20	0.101	2.316	No
800	1.55	0.080	2.123	No
1015	1.97	0.072	1.978	No
1200	2.33	0.072	1.874	No

Table A.3: United States: effective-rank diagnostics of Bartlett et al. (2020) ($b = 0.2, T_{\text{train}} = 515$); ∞ marks nodes where $R_{k^*}(\Sigma)$ is numerically undefined at the spectral edge, i.e., where k^* falls on the last nonzero eigenvalue of the rank-deficient empirical covariance and the residual tail sum is zero (a $0/0$ ratio, reported conservatively as $T/R_{k^*} = \infty$).

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

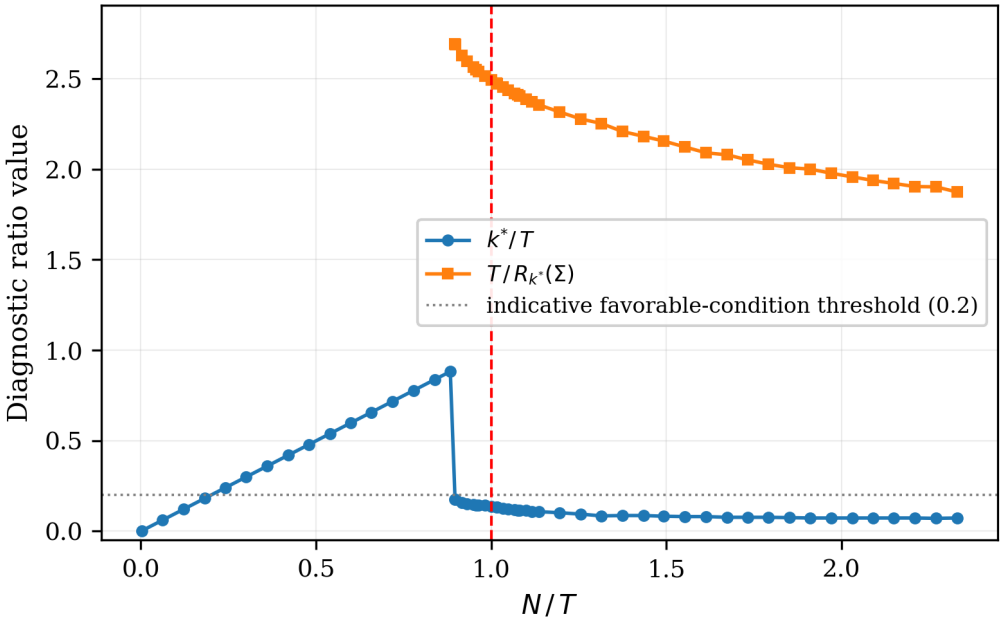

Figure A.1: Effective-rank diagnostics k^*/T and $T/R_{k^*}(\Sigma)$ versus N/T_{train} ($b = 0.2$). Dotted line: favorability threshold; dashed line: interpolation threshold.

Source: author's calculations.

N	N/T_{train}	Ridgeless	Ridgelet	Ridge	k^*/T	T/R_{k^*}
2	0.01	0.00204	0.00204	0.00207	0.004	∞
60	0.22	0.00309	0.00309	0.00216	0.221	∞
134	0.50	0.00497	0.00497	0.00224	0.498	∞
193	0.72	0.00657	0.00656	0.00227	0.719	∞
241	0.90	0.03601	0.03550	0.00230	0.899	∞
258	0.97	0.07334	0.06202	0.00231	0.963	∞
266	**1.00**	**2.7479**	**0.14919**	**0.00230**	**0.993**	∞
275	1.03	0.05668	0.05242	0.00230	0.996	∞
340	1.27	0.01979	0.01977	0.00233	0.996	∞
443	1.66	0.01145	0.01145	0.00235	0.996	∞
576	2.16	0.00787	0.00787	0.00235	0.404	2.699

Table A.4: Emerging markets: out-of-sample MSE and effective-rank diagnostics as a function of model complexity N. $T_{\text{train}} = 267$, threshold $b = 0.2$; ∞ entries as in Table A.3 (k^* at the edge of the nonzero spectrum, R_{k^*} a $0/0$ ratio).

Source: author's calculations on monthly data from the Kenneth R. French Data Library.

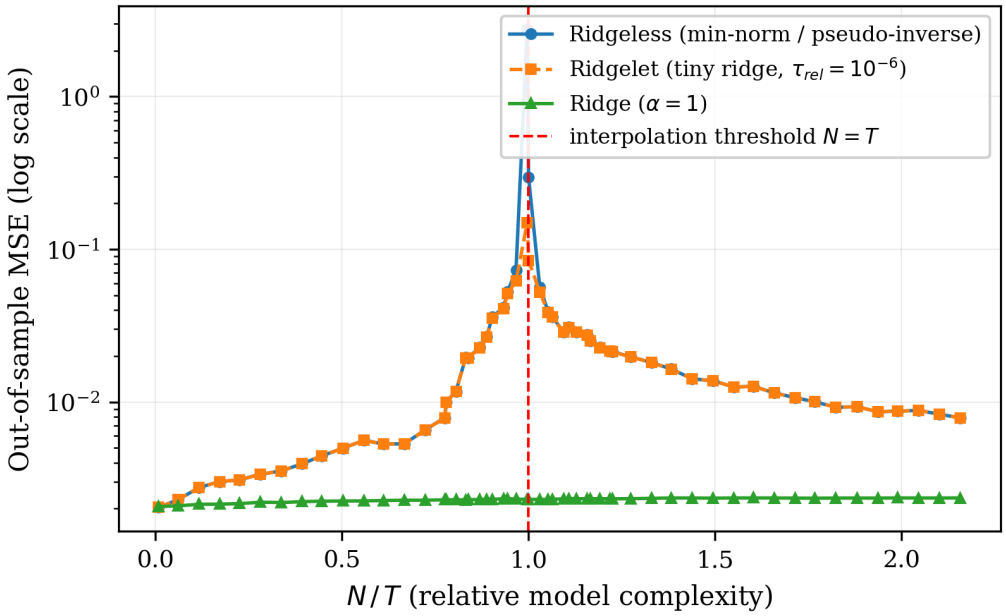

Figure A.2: Out-of-sample MSE of the Ridgeless and Ridgelet estimators versus N, emerging-markets universe (24 pooled series, 24 lags, $T_{\text{train}} = 267$).
Source: author's calculations.

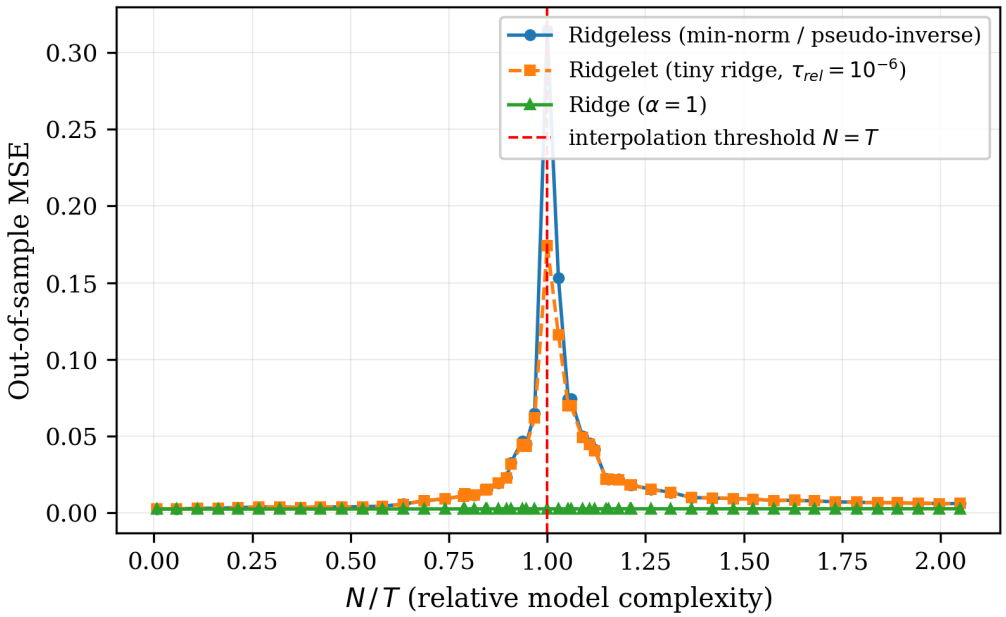

Figure A.3: Out-of-sample test MSE of the Ridgeless and Ridgelet estimators as a function of the number of features N, European universe (24 pooled series, 24 monthly lags, $T_{\text{train}} = 281$).
Source: author's calculations.

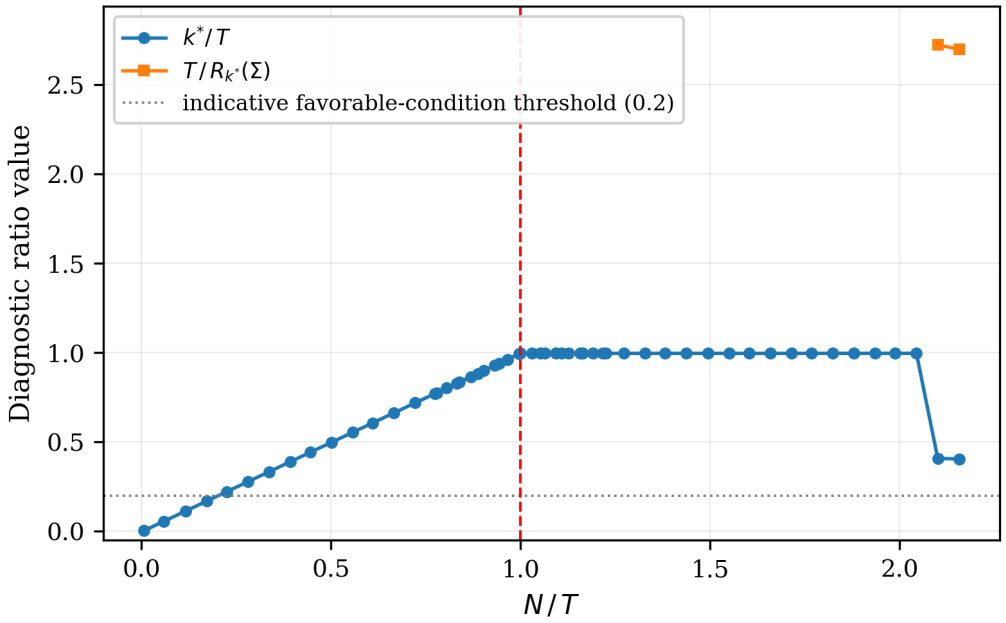

Figure A.4: Effective-rank diagnostics k^*/T and $T/R_{k^*}(\Sigma)$ along the feature sweep, emerging-markets universe. The dominant-block index opens up only at the far end of the sweep.
Source: author's calculations.

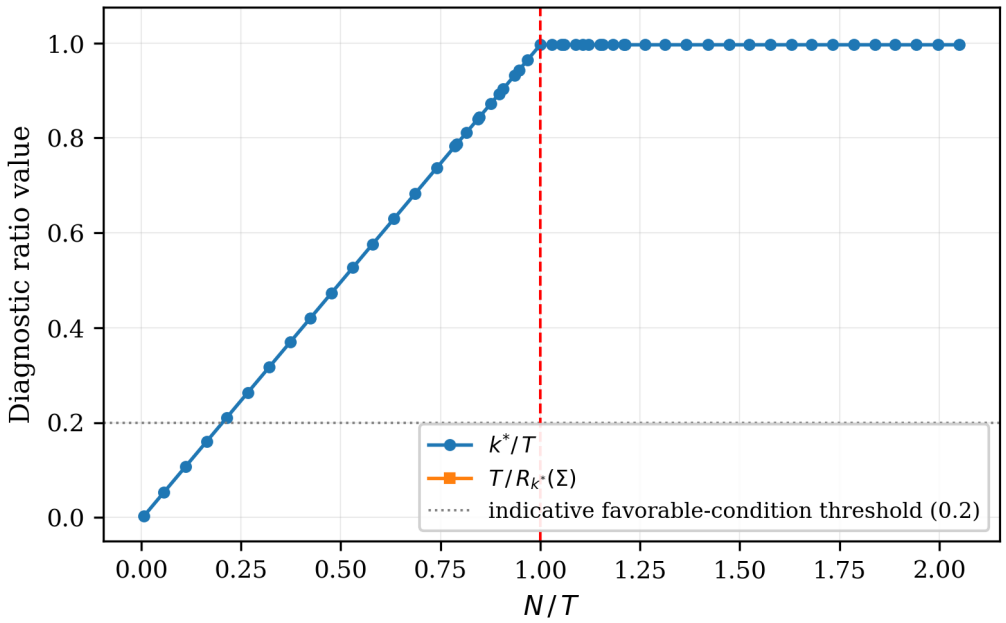

Figure A.5: Effective-rank diagnostics k^*/T and $T/R_{k^*}(\Sigma)$ along the feature sweep, European universe. The dominant-block index remains pinned at the spectral edge ($k^*/T \approx 0.996$) across the entire over-parameterized range.
Source: author's calculations.

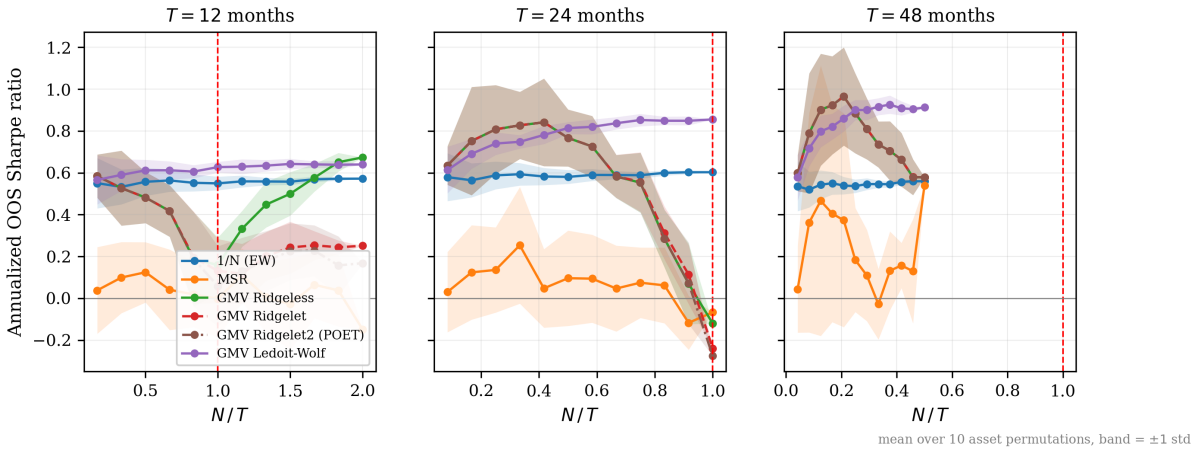

Figure A.6: Annualized out-of-sample Sharpe ratio of the six virtual funds as a function of N/T, European universe (24 pooled assets, $T \in \{12, 24, 48\}$ months, mean over ten asset orderings).
Source: author's calculations.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique

www.uclouvain.be/lsm