All Weather Strategy v1.2

Adaptive Risk Parity with Ledoit-Wolf Covariance Shrinkage

Comprehensive Tutorial Lecture Notes

Graduate-Level Treatment

Data Period: 2018–2026 · 7-Asset Chinese ETF Universe

Contents

1 Motivation and Financial Philosophy **3**
- 1.1 The Problem with Traditional Portfolios . . . 3
- 1.2 From Capital Weights to Risk Weights . . . 3

2 Mathematical Foundations of Risk Parity **3**
- 2.1 Portfolio Risk Decomposition . . . 3
- 2.2 Euler's Theorem for Homogeneous Functions . . . 3
- 2.3 Applying Euler's Theorem to Portfolio Volatility . . . 4
- 2.4 Marginal and Total Risk Contributions . . . 4
- 2.5 The Risk Parity Optimization Problem . . . 5
- 2.6 Why High Bond Allocation is Mathematically Necessary . . . 6

3 Covariance Estimation and Ledoit-Wolf Shrinkage **6**
- 3.1 Notation and Data Layout . . . 6
- 3.2 The Sample Covariance Matrix: Construction and Limitations . . . 7
- 3.3 The Frobenius Norm: Measuring Matrix Distance . . . 8
- 3.4 The Bias-Variance Tradeoff for Matrix Estimators . . . 8
- 3.5 The Shrinkage Estimator: General Form . . . 9
- 3.6 The Ledoit-Wolf Target: Constant Correlation Model . . . 9
- 3.7 Deriving the Optimal Shrinkage Intensity . . . 10
- 3.8 Impact on Risk Parity Weights . . . 14

4 Adaptive Rebalancing **15**
- 4.1 The Rebalancing Dilemma . . . 15
- 4.2 Total Cost Framework . . . 15
- 4.3 Empirical Threshold Selection . . . 16

5 Implementation Architecture **16**
- 5.1 Module Overview . . . 16
- 5.2 Covariance Estimation (Optimizer Module) . . . 17
- 5.3 Risk Parity Optimization (Optimizer Module) . . . 17
- 5.4 Portfolio Management (Portfolio Module) . . . 18
- 5.5 Main Backtest Loop (Strategy Module) . . . 19

6 Performance Metrics: Definitions and Interpretation **19**
6.1 Return Metrics 19
6.2 Risk Metrics 19
6.3 Risk-Adjusted Metrics 20
6.4 Distributional Properties 20

7 Empirical Results **21**
7.1 Headline Performance (2018–2026) 21
7.2 Version Comparison 21
7.3 Rebalancing Dynamics 21

8 Conclusion and Extensions **22**
8.1 Summary of Key Ideas 22
8.2 Potential Extensions 22

1 Motivation and Financial Philosophy

1.1 The Problem with Traditional Portfolios

Consider the classical **60/40 portfolio**: 60% equities, 40% bonds. Its appeal lies in simplicity, but its construction contains a hidden asymmetry. With equities exhibiting annualized volatility of roughly $\sigma_S \approx 20\%$ and bonds at $\sigma_B \approx 5\%$, the equity component contributes approximately

$$\frac{0.60 \times 20\%}{0.60 \times 20\% + 0.40 \times 5\%} \approx 86\%$$

of the portfolio's total risk (in this simplified single-factor decomposition). The portfolio is *not* diversified in any meaningful risk sense—it is, effectively, a leveraged equity bet with a small bond hedge.

During the 2008 financial crisis, the S&P 500 fell roughly 50%. A 60/40 portfolio experienced drawdowns of 30–35%, because the equity risk that dominated the allocation also dominated the losses. The bond "diversifier" contributed negligible protection relative to the equity shock.

> **Financial Intuition.** Ray Dalio's key insight at Bridgewater Associates was that *you cannot reliably predict which economic regime the world will be in*—but you can construct a portfolio that performs adequately across all regimes by *balancing risk contributions* rather than capital allocations. This is the "All Weather" philosophy: design a portfolio that does not require accurate macro forecasts.

1.2 From Capital Weights to Risk Weights

The conceptual shift is from asking "how much money do I put in each asset?" to asking "how much risk does each asset contribute to my portfolio?" If we can equalize risk contributions, no single asset (or asset class) can dominate the portfolio during its specific adverse regime. This is the **risk parity** principle, and the remainder of these notes develops its mathematical foundation, estimation theory, and practical implementation.

2 Mathematical Foundations of Risk Parity

2.1 Portfolio Risk Decomposition

Definition 2.1 (Portfolio Volatility)**.** Let $\boldsymbol{w} = (w_1, \ldots, w_n)^\top \in \mathbb{R}^n$ be a vector of portfolio weights with $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0$, and let $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ be the covariance matrix of asset returns. The **portfolio volatility** is:

$$\sigma_p(\boldsymbol{w}) = \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} \tag{1}$$

The function $\sigma_p(\boldsymbol{w})$ maps a weight vector to a scalar risk measure. Our goal is to decompose this scalar into n additive components, each attributable to one asset.

2.2 Euler's Theorem for Homogeneous Functions

The key mathematical tool for risk decomposition is Euler's theorem. We first establish the general result, then apply it to portfolio volatility.

Definition 2.2 (Positive Homogeneity)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is **positively homogeneous of degree** k if for all $\lambda > 0$:

$$f(\lambda \boldsymbol{w}) = \lambda^k f(\boldsymbol{w}).$$

Theorem 2.3 (Euler's Theorem for Homogeneous Functions)**.** If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and positively homogeneous of degree k, then:

$$\sum_{i=1}^{n} w_i \frac{\partial f}{\partial w_i}(\boldsymbol{w}) = k \cdot f(\boldsymbol{w}). \tag{2}$$

Proof. Since $f(\lambda \boldsymbol{w}) = \lambda^k f(\boldsymbol{w})$ for all $\lambda > 0$, differentiate both sides with respect to λ:

$$\frac{d}{d\lambda} f(\lambda \boldsymbol{w}) = \frac{d}{d\lambda} \left[\lambda^k f(\boldsymbol{w}) \right].$$

The left-hand side, by the multivariate chain rule, is:

$$\frac{d}{d\lambda} f(\lambda \boldsymbol{w}) = \sum_{i=1}^{n} \frac{\partial f}{\partial(\lambda w_i)} \cdot w_i = \sum_{i=1}^{n} \left. \frac{\partial f}{\partial u_i} \right|_{\boldsymbol{u}=\lambda \boldsymbol{w}} \cdot w_i,$$

where $u_i = \lambda w_i$. The right-hand side is $k\lambda^{k-1} f(\boldsymbol{w})$. Setting $\lambda = 1$:

$$\sum_{i=1}^{n} w_i \frac{\partial f}{\partial w_i}(\boldsymbol{w}) = k \cdot f(\boldsymbol{w}).$$ □

> **Intuition.** Euler's theorem tells us that for homogeneous functions, the "whole" can always be perfectly decomposed into contributions from each "part," weighted by how much each part is used. This is exactly what we need: decompose total portfolio risk into per-asset risk contributions.

2.3 Applying Euler's Theorem to Portfolio Volatility

Proposition 2.4. Portfolio volatility $\sigma_p(\boldsymbol{w}) = \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}$ is positively homogeneous of degree 1 in $\boldsymbol{w}$.

Proof. For $\lambda > 0$:

$$\sigma_p(\lambda \boldsymbol{w}) = \sqrt{(\lambda \boldsymbol{w})^\top \boldsymbol{\Sigma} (\lambda \boldsymbol{w})} = \sqrt{\lambda^2 \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} = \lambda \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}} = \lambda\, \sigma_p(\boldsymbol{w}).$$ □

Applying Euler's theorem (Theorem 2.3) with $k = 1$:

Corollary 2.5 (Risk Decomposition Identity)**.**

$$\sigma_p(\boldsymbol{w}) = \sum_{i=1}^{n} w_i \frac{\partial \sigma_p}{\partial w_i}. \tag{3}$$

This identity guarantees that the risk contributions sum *exactly* to total portfolio volatility—there is no residual or interaction term.

2.4 Marginal and Total Risk Contributions

Definition 2.6 (Marginal Risk Contribution)**.** The **marginal risk contribution** (MRC) of asset i is the partial derivative of portfolio volatility with respect to weight w_i:

$$\mathrm{MRC}_i = \frac{\partial \sigma_p}{\partial w_i}. \tag{4}$$

We now derive a closed-form expression.

Proposition 2.7 (Closed Form for MRC)**.**

$$\frac{\partial \sigma_p}{\partial w_i} = \frac{(\boldsymbol{\Sigma w})_i}{\sigma_p} \tag{5}$$

where $(\boldsymbol{\Sigma w})_i$ denotes the i-th element of the vector $\boldsymbol{\Sigma w}$.

Proof. Write $\sigma_p^2 = \boldsymbol{w}^\top \boldsymbol{\Sigma w} = \sum_{j,k} w_j \Sigma_{jk} w_k$. Then:

$$\frac{\partial \sigma_p^2}{\partial w_i} = \frac{\partial}{\partial w_i} \sum_{j,k} w_j \Sigma_{jk} w_k = 2 \sum_k \Sigma_{ik} w_k = 2(\boldsymbol{\Sigma w})_i,$$

where the factor of 2 arises because $\boldsymbol{\Sigma}$ is symmetric, so both the $j = i$ and $k = i$ terms contribute. By the chain rule:

$$\frac{\partial \sigma_p}{\partial w_i} = \frac{\partial \sqrt{\sigma_p^2}}{\partial w_i} = \frac{1}{2\sigma_p} \cdot \frac{\partial \sigma_p^2}{\partial w_i} = \frac{1}{2\sigma_p} \cdot 2(\boldsymbol{\Sigma w})_i = \frac{(\boldsymbol{\Sigma w})_i}{\sigma_p}. \qquad \square$$

Definition 2.8 (Risk Contribution)**.** The **total risk contribution** (RC) of asset i is:

$$\mathrm{RC}_i = w_i \cdot \mathrm{MRC}_i = w_i \cdot \frac{(\boldsymbol{\Sigma w})_i}{\sigma_p}. \tag{6}$$

By Corollary 2.5, these contributions satisfy the **full decomposition property**:

$$\boxed{\sum_{i=1}^{n} \mathrm{RC}_i = \sigma_p.} \tag{7}$$

Insight. Equation (6) has a clean interpretation: $(\boldsymbol{\Sigma w})_i = \sum_j \Sigma_{ij} w_j$ is the covariance of asset i's return with the *portfolio* return. Dividing by σ_p normalizes this to a "beta-like" marginal sensitivity. Multiplying by w_i converts from per-unit to per-dollar risk contribution. The term RC_i/σ_p gives the *fractional* risk contribution, which is the percentage of total risk attributable to asset i.

2.5 The Risk Parity Optimization Problem

Definition 2.9 (Risk Parity)**.** A portfolio $\boldsymbol{w}^*$ achieves **equal risk parity** if:

$$\mathrm{RC}_1(\boldsymbol{w}^*) = \mathrm{RC}_2(\boldsymbol{w}^*) = \cdots = \mathrm{RC}_n(\boldsymbol{w}^*). \tag{8}$$

Since the risk contributions sum to σ_p, the equal risk parity condition is equivalent to:

$$\mathrm{RC}_i(\boldsymbol{w}^*) = \frac{\sigma_p(\boldsymbol{w}^*)}{n}, \quad \forall\, i = 1, \ldots, n.$$

To find such a $\boldsymbol{w}^*$ numerically, we formulate the optimization problem:

$$\boxed{\begin{aligned} \min_{\boldsymbol{w} \in \mathbb{R}^n} \quad & \mathrm{std}\big(\mathrm{RC}_1(\boldsymbol{w}), \ldots, \mathrm{RC}_n(\boldsymbol{w})\big) \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1, \\ & w_i \geq 0, \quad \forall\, i. \end{aligned}} \tag{9}$$

Proposition 2.10. $\boldsymbol{w}^*$ solves (9) with optimal value zero if and only if $\boldsymbol{w}^*$ satisfies the equal risk parity condition (8).

Proof. ($\Rightarrow$) If the optimal value is zero, then $\text{std}(\text{RC}_1, \dots, \text{RC}_n) = 0$, which means all RC_i are equal.

($\Leftarrow$) If (8) holds, then the standard deviation of the risk contributions is zero, which is a global minimum of the non-negative objective. $\square$

Intuition. Why minimize std(RC) rather than, say, $\sum_{i<j}(\text{RC}_i - \text{RC}_j)^2$? Both reach zero at the risk parity solution. The standard deviation formulation is more numerically stable because it avoids the combinatorial growth of pairwise terms and produces smoother gradients for the SLSQP solver.

Financial Intuition. A critical advantage of risk parity over mean-variance optimization is that (9) does *not* require estimating expected returns $\boldsymbol{\mu}$. Return forecasts are notoriously noisy—Merton (1980) showed that estimating means requires orders of magnitude more data than estimating covariances. By relying only on $\boldsymbol{\Sigma}$, risk parity sidesteps the largest source of estimation error in portfolio construction.

2.6 Why High Bond Allocation is Mathematically Necessary

To build intuition for the resulting weights, consider a simplified 2-asset case (stock S and bond B) with zero correlation. The risk parity condition $\text{RC}_S = \text{RC}_B$ becomes:

$$w_S \cdot \frac{w_S \sigma_S^2}{\sigma_p} = w_B \cdot \frac{w_B \sigma_B^2}{\sigma_p}.$$

Canceling σ_p and using $w_B = 1 - w_S$:

$$w_S^2 \sigma_S^2 = (1 - w_S)^2 \sigma_B^2 \quad \Longrightarrow \quad \frac{w_S}{1 - w_S} = \frac{\sigma_B}{\sigma_S}.$$

$$\boxed{w_S = \frac{\sigma_B}{\sigma_S + \sigma_B}, \quad w_B = \frac{\sigma_S}{\sigma_S + \sigma_B}.} \tag{10}$$

With typical values $\sigma_S = 20\%$ and $\sigma_B = 3\%$:

$$w_B = \frac{20}{20 + 3} \approx 87\%, \quad w_S = \frac{3}{20 + 3} \approx 13\%.$$

Insight. Weights are *inversely proportional to volatility.* This is not a conservative choice—it is a mathematical consequence of requiring equal risk contributions. In the full 7-asset case with cross-correlations, the optimizer in v1.2 produces a bond allocation of $\approx 42\%$ (down from $\approx 77\%$ without shrinkage, a distinction we explain in Section 3).

3 Covariance Estimation and Ledoit-Wolf Shrinkage

3.1 Notation and Data Layout

Before developing the theory, we fix notation and clarify the shapes of every object.

Definition 3.1 (Return Matrix)**.** Let $r_{t,i}$ denote the return of asset i on day t. The **centered return matrix** is

$$\boldsymbol{R} = \begin{pmatrix} r_{1,1} - \bar{r}_1 & r_{1,2} - \bar{r}_2 & \cdots & r_{1,n} - \bar{r}_n \\ r_{2,1} - \bar{r}_1 & r_{2,2} - \bar{r}_2 & \cdots & r_{2,n} - \bar{r}_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{T,1} - \bar{r}_1 & r_{T,2} - \bar{r}_2 & \cdots & r_{T,n} - \bar{r}_n \end{pmatrix} \in \mathbb{R}^{T \times n}, \tag{11}$$

where $\bar{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_{t,i}$ is the sample mean of asset i. In our setting, $T = 252$ (one year of daily data) and $n = 7$ (assets). Each row is a day; each column is an asset's demeaned return time series.

We will work with four symmetric matrices, all of shape $n \times n$:

Symbol	Name	Shape	Known?
$\mathbf{\Sigma}$	True (population) covariance	$n \times n$	No (unknown)
$\hat{\mathbf{\Sigma}}$	Sample covariance	$n \times n$	Yes (from data)
$\boldsymbol{F}$	Structured target (constant correlation)	$n \times n$	Yes (from data)
$\mathbf{\Sigma}_{\text{shrunk}}$	Shrinkage estimator	$n \times n$	Yes (computed)

3.2 The Sample Covariance Matrix: Construction and Limitations

Definition 3.2 (Sample Covariance Matrix)**.** Given the centered return matrix $\boldsymbol{R} \in \mathbb{R}^{T \times n}$, the **sample covariance matrix** is:

$$\underbrace{\hat{\mathbf{\Sigma}}}_{n\times n} = \frac{1}{T-1} \underbrace{\boldsymbol{R}^\top}_{n\times T} \underbrace{\boldsymbol{R}}_{T\times n}. \tag{12}$$

To understand what each entry contains, write the matrix product element-by-element. The (i, j) entry of $\boldsymbol{R}^\top \boldsymbol{R}$ is

$$\left(\boldsymbol{R}^\top \boldsymbol{R}\right)_{ij} = \sum_{t=1}^{T} R_{ti}\, R_{tj} = \sum_{t=1}^{T} (r_{t,i} - \bar{r}_i)(r_{t,j} - \bar{r}_j),$$

so that:

$$\hat{\Sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} (r_{t,i} - \bar{r}_i)(r_{t,j} - \bar{r}_j). \tag{13}$$

- **Diagonal entries** $(i = j)$: $\hat{\Sigma}_{ii} = \hat{\sigma}_i^2$, the sample variance of asset i.
- **Off-diagonal entries** $(i \neq j)$: $\hat{\Sigma}_{ij} = \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij}$, the sample covariance, where $\hat{\rho}_{ij}$ is the sample Pearson correlation.
- **Symmetry**: $\hat{\Sigma}_{ij} = \hat{\Sigma}_{ji}$, so there are $n(n + 1)/2 = 28$ unique entries.

Insight. Each entry $\hat{\Sigma}_{ij}$ is an *average of T products of centered returns.* By the law of large numbers, as $T \to \infty$ each entry converges to its population counterpart Σ_{ij}, making $\hat{\mathbf{\Sigma}}$ unbiased: $\mathbb{E}[\hat{\Sigma}_{ij}] = \Sigma_{ij}$ for all i, j. However, with finite $T = 252$, each entry carries estimation noise of order $O(1/\sqrt{T})$. The $n \times n$ matrix has $n(n+1)/2 = 28$ noisy entries, and these errors *compound* when the matrix is fed into the risk parity optimizer. The optimizer amplifies errors associated with the smallest eigenvalue of $\hat{\mathbf{\Sigma}}$—this is the "curse of ill-conditioning."

Proposition 3.3 (Noise Amplification in Optimization)**.** Let $\boldsymbol{w}^*(\mathbf{\Sigma})$ denote the risk parity solution as a function of the covariance input. Small perturbations $\mathbf{\Sigma} \to \mathbf{\Sigma} + \mathbf{\Delta}$ ($\mathbf{\Delta} \in \mathbb{R}^{n\times n}$ symmetric) can produce large changes in $\boldsymbol{w}^*$ when the condition number $\kappa(\mathbf{\Sigma}) = \lambda_{\max}(\mathbf{\Sigma})/\lambda_{\min}(\mathbf{\Sigma})$ is large.

Intuition. Consider the eigendecomposition $\mathbf{\Sigma} = \boldsymbol{Q}\mathbf{\Lambda}\boldsymbol{Q}^\top$ where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is orthogonal and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n > 0$. Risk parity places more weight on low-variance directions (eigenvectors of small λ_k). When λ_n is tiny, estimation noise in that direction gets amplified by the factor $1/\lambda_n$ when the optimizer "inverts" the risk structure. In financial data, assets are often highly correlated, so most variance concentrates in the first 1–2 eigenvalues, leaving the trailing eigenvalues small and noise-dominated.

3.3 The Frobenius Norm: Measuring Matrix Distance

We need a way to measure how "close" an estimated covariance matrix is to the truth. The appropriate tool is the Frobenius norm.

Definition 3.4 (Frobenius Norm)**.** For any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the **Frobenius norm** is:

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(\boldsymbol{A}^\top \boldsymbol{A})}, \tag{14}$$

where $\operatorname{tr}(\cdot)$ denotes the trace (sum of diagonal elements). The **squared Frobenius norm** is simply the sum of all squared entries:

$$\|\boldsymbol{A}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2. \tag{15}$$

Insight. The Frobenius norm treats a matrix as a "flattened" vector and computes its Euclidean length. For symmetric $n \times n$ matrices (like covariance matrices), $\|\boldsymbol{A}\|_F^2 = \sum_{i=1}^{n} A_{ii}^2 + 2\sum_{i<j} A_{ij}^2$, i.e., the sum of all n squared diagonal entries plus twice each squared off-diagonal entry (since $A_{ij} = A_{ji}$). Equivalently, the squared Frobenius norm equals the sum of the squared eigenvalues: $\|\boldsymbol{A}\|_F^2 = \sum_{k=1}^{n} \lambda_k(\boldsymbol{A})^2$.

3.4 The Bias-Variance Tradeoff for Matrix Estimators

The Frobenius norm lets us decompose the estimation error of any covariance estimator into bias and variance, exactly as in scalar statistics.

Theorem 3.5 (Matrix Bias-Variance Decomposition)**.** For any estimator $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{n\times n}$ of $\boldsymbol{\Sigma}$, define the **mean squared error** as:

$$\operatorname{MSE}(\tilde{\boldsymbol{\Sigma}}) = \mathbb{E}\left[\|\tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\tilde{\Sigma}_{ij} - \Sigma_{ij})^2\right]. \tag{16}$$

Then:

$$\operatorname{MSE}(\tilde{\boldsymbol{\Sigma}}) = \underbrace{\|\mathbb{E}[\tilde{\boldsymbol{\Sigma}}] - \boldsymbol{\Sigma}\|_F^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[\|\tilde{\boldsymbol{\Sigma}} - \mathbb{E}[\tilde{\boldsymbol{\Sigma}}]\|_F^2\right]}_{\text{Variance}}. \tag{17}$$

Proof. The proof follows by expanding element-wise. Let $\boldsymbol{B} = \mathbb{E}[\tilde{\boldsymbol{\Sigma}}] - \boldsymbol{\Sigma}$ (the bias matrix, shape $n \times n$). Write:

$$\tilde{\Sigma}_{ij} - \Sigma_{ij} = \underbrace{(\tilde{\Sigma}_{ij} - \mathbb{E}[\tilde{\Sigma}_{ij}])}_{\text{zero-mean noise}} + \underbrace{(\mathbb{E}[\tilde{\Sigma}_{ij}] - \Sigma_{ij})}_{B_{ij}\ \text{(deterministic bias)}}.$$

Square both sides, sum over all n^2 entries, and take expectations:

$$\begin{aligned}\sum_{ij}\mathbb{E}\left[(\tilde{\Sigma}_{ij} - \Sigma_{ij})^2\right] &= \sum_{ij}\mathbb{E}\left[(\tilde{\Sigma}_{ij} - \mathbb{E}[\tilde{\Sigma}_{ij}])^2\right] + \sum_{ij} B_{ij}^2 \\ &\quad + 2\sum_{ij} B_{ij}\cdot \underbrace{\mathbb{E}[\tilde{\Sigma}_{ij} - \mathbb{E}[\tilde{\Sigma}_{ij}]]}_{=0}.\end{aligned} \tag{18}$$

The cross-term vanishes because $\mathbb{E}[\tilde{\Sigma}_{ij} - \mathbb{E}[\tilde{\Sigma}_{ij}]] = 0$. The first sum is $\sum_{ij}\operatorname{Var}(\tilde{\Sigma}_{ij})$ (total variance), and the second is $\|\boldsymbol{B}\|_F^2 = \|\mathbb{E}[\tilde{\boldsymbol{\Sigma}}] - \boldsymbol{\Sigma}\|_F^2$ (squared bias). $\square$

Applying to the sample covariance. The sample covariance $\hat{\boldsymbol{\Sigma}}$ is unbiased ($\mathbb{E}[\hat{\boldsymbol{\Sigma}}] = \boldsymbol{\Sigma}$), so its bias is identically zero, and:

$$\mathrm{MSE}(\hat{\boldsymbol{\Sigma}}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Var}(\hat{\Sigma}_{ij}).$$

This is pure variance—n^2 variance terms, one per matrix entry. Each individual $\mathrm{Var}(\hat{\Sigma}_{ij})$ is of order $O(1/T)$, but there are n^2 of them, and they all feed into the optimizer simultaneously.

> **Financial Intuition.** In portfolio construction, what matters is not whether $\hat{\boldsymbol{\Sigma}}$ is close to $\boldsymbol{\Sigma}$ in every entry, but whether the resulting portfolio weights $\boldsymbol{w}^*(\hat{\boldsymbol{\Sigma}})$ perform well out of sample. An estimator with some bias but much lower variance often produces superior portfolios. This is the **James-Stein phenomenon** applied to covariance matrices: the "obviously correct" unbiased estimator is often suboptimal in terms of actual decision-making performance.

3.5 The Shrinkage Estimator: General Form

Definition 3.6 (Linear Shrinkage Estimator)**.** The **shrinkage estimator** is a convex combination of the sample covariance and a structured **target** matrix $\boldsymbol{F}$:

$$\underbrace{\boldsymbol{\Sigma}_{\text{shrunk}}}_{n\times n} = (1-\delta)\,\underbrace{\hat{\boldsymbol{\Sigma}}}_{n\times n} + \delta\,\underbrace{\boldsymbol{F}}_{n\times n}, \quad \delta \in [0,1]. \tag{19}$$

The scalar $\delta \in [0,1]$ is called the **shrinkage intensity**.

Element-wise, this is simply a weighted average of corresponding entries:

$$(\Sigma_{\text{shrunk}})_{ij} = (1-\delta)\,\hat{\Sigma}_{ij} + \delta\,F_{ij}, \quad \forall\, 1 \le i,j \le n. \tag{20}$$

Setting $\delta = 0$ recovers the sample covariance (no shrinkage), and $\delta = 1$ ignores the data entirely. The optimal δ^* minimizes the MSE of the combined estimator with respect to the true $\boldsymbol{\Sigma}$.

3.6 The Ledoit-Wolf Target: Constant Correlation Model

Ledoit and Wolf (2004) propose a specific structured target based on the **constant correlation model**. The idea is to preserve each asset's marginal volatility but replace all pairwise correlations with a single average correlation.

Definition 3.7 (Constant Correlation Target)**.** Let $\hat{\sigma}_i = \sqrt{\hat{\Sigma}_{ii}} \in \mathbb{R}_+$ be the sample standard deviation of asset i, and define the **average sample correlation**:

$$\bar{\rho} = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \hat{\rho}_{ij} = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \frac{\hat{\Sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j} \in [-1,1]. \tag{21}$$

This is a single scalar averaging all $\binom{n}{2} = \binom{7}{2} = 21$ pairwise correlations.

The **constant correlation target matrix** $\boldsymbol{F} \in \mathbb{R}^{n\times n}$ has entries:

$$F_{ij} = \begin{cases} \hat{\sigma}_i^2 & \text{if } i = j \text{ (variance preserved)}, \\ \bar{\rho}\cdot\hat{\sigma}_i\hat{\sigma}_j & \text{if } i \ne j \text{ (correlation replaced by average)}. \end{cases} \tag{22}$$

In matrix notation, $\boldsymbol{F}$ can be decomposed as:

$$\underbrace{\boldsymbol{F}}_{n\times n} = (1-\bar{\rho})\,\underbrace{\mathrm{diag}(\hat{\sigma}_1^2,\ldots,\hat{\sigma}_n^2)}_{n\times n} + \bar{\rho}\,\underbrace{\hat{\boldsymbol{s}}}_{n\times 1}\,\underbrace{\hat{\boldsymbol{s}}^\top}_{1\times n}, \tag{23}$$

where $\hat{\boldsymbol{s}} = (\hat{\sigma}_1, \dots, \hat{\sigma}_n)^\top \in \mathbb{R}^n$ is the vector of sample standard deviations.

Insight. Equation (23) reveals the geometry of $\boldsymbol{F}$: it is a mixture of a diagonal matrix (zero correlation, pure individual variance) and a rank-1 matrix $\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^\top$ (perfect correlation, all assets move together). The mixing weight $\bar{\rho}$ controls how correlated the target model assumes assets to be. Crucially, $\boldsymbol{F}$ has only $n+1=8$ free parameters (n volatilities plus one $\bar{\rho}$), compared to $n(n+1)/2 = 28$ for $\hat{\boldsymbol{\Sigma}}$. This 3.5× reduction in parameter count is the fundamental source of lower estimation variance.

Example 3.8 (Concrete 3×3 Illustration)**.** Suppose $n = 3$ assets with sample standard deviations $\hat{\sigma}_1 = 0.20$, $\hat{\sigma}_2 = 0.15$, $\hat{\sigma}_3 = 0.03$ (two equities and one bond), and average correlation $\bar{\rho} = 0.30$. Then:

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.0400 & 0.0180 & -0.0005 \\ 0.0180 & 0.0225 & -0.0003 \\ -0.0005 & -0.0003 & 0.0009 \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} 0.0400 & 0.0090 & 0.0018 \\ 0.0090 & 0.0225 & 0.0014 \\ 0.0018 & 0.0014 & 0.0009 \end{pmatrix}.$$

Note: $\boldsymbol{F}$ preserves the diagonals (variances) exactly, but replaces each off-diagonal $\hat{\Sigma}_{ij}$ with $\bar{\rho} \cdot \hat{\sigma}_i \hat{\sigma}_j$. The negative equity-bond correlations (-0.0005, -0.0003) in $\hat{\boldsymbol{\Sigma}}$ are pulled toward positive values (0.0018, 0.0014) in $\boldsymbol{F}$ because $\bar{\rho} > 0$. Shrinkage interpolates between these two matrices entry by entry.

3.7 Deriving the Optimal Shrinkage Intensity

We now derive the optimal δ^* in full detail, tracking every step at the element-wise level.

Step 1: Define the Loss Function

The goal is to find the δ that makes $\boldsymbol{\Sigma}_{\text{shrunk}}$ as close to the unknown $\boldsymbol{\Sigma}$ as possible, measured by expected squared Frobenius distance:

$$L(\delta) = \mathbb{E}\big[\|\boldsymbol{\Sigma}_{\text{shrunk}} - \boldsymbol{\Sigma}\|_F^2\big] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\big((\Sigma_{\text{shrunk}})_{ij} - \Sigma_{ij}\big)^2\right]. \tag{24}$$

This is a scalar function of the single variable δ: we are optimizing over a one-dimensional parameter.

Step 2: Substitute the Shrinkage Formula

Substitute (20) into the (i,j) residual:

$$\begin{aligned}(\Sigma_{\text{shrunk}})_{ij} - \Sigma_{ij} &= (1-\delta)\hat{\Sigma}_{ij} + \delta F_{ij} - \Sigma_{ij} \\ &= (1-\delta)\hat{\Sigma}_{ij} - (1-\delta)\Sigma_{ij} + \delta F_{ij} - \delta\Sigma_{ij} \\ &= (1-\delta)\underbrace{(\hat{\Sigma}_{ij} - \Sigma_{ij})}_{\text{sampling error of entry }(i,j)} + \delta\underbrace{(F_{ij} - \Sigma_{ij})}_{\text{target misspecification of entry }(i,j)}.\end{aligned} \tag{25}$$

Insight. Equation (25) is the key decomposition: for every single entry (i,j) of the matrix, the total error is a weighted sum of (a) the random sampling error $\hat{\Sigma}_{ij} - \Sigma_{ij}$ and (b) the deterministic bias $F_{ij} - \Sigma_{ij}$ of the target. The weight $(1-\delta)$ on sampling error decreases as we shrink more; the weight δ on target bias increases. The optimal δ^* balances these two sources of error *summed over all n^2 entries.*

Step 3: Square and Sum Over All Entries

Define the **error matrix** $\boldsymbol{E}(\delta) \in \mathbb{R}^{n \times n}$ with entries given by (25). The loss is:

$$L(\delta) = \mathbb{E}\left[\sum_{i,j} E_{ij}(\delta)^2\right] = \sum_{i,j} \mathbb{E}\left[\left((1-\delta)(\hat{\Sigma}_{ij} - \Sigma_{ij}) + \delta(F_{ij} - \Sigma_{ij})\right)^2\right]. \tag{26}$$

Expand the square inside the expectation using $(a+b)^2 = a^2 + 2ab + b^2$:

$$E_{ij}(\delta)^2 = (1-\delta)^2(\hat{\Sigma}_{ij} - \Sigma_{ij})^2 + 2\delta(1-\delta)(\hat{\Sigma}_{ij} - \Sigma_{ij})(F_{ij} - \Sigma_{ij}) + \delta^2(F_{ij} - \Sigma_{ij})^2. \tag{27}$$

Take the expectation term by term and sum over all n^2 entries:

$$L(\delta) = (1-\delta)^2 \underbrace{\sum_{i,j} \mathbb{E}\left[(\hat{\Sigma}_{ij} - \Sigma_{ij})^2\right]}_{:=\ \pi} + 2\delta(1-\delta) \underbrace{\sum_{i,j} \mathbb{E}\left[\hat{\Sigma}_{ij} - \Sigma_{ij}\right] \cdot (F_{ij} - \Sigma_{ij})}_{:=\ \text{cross-term}} + \delta^2 \underbrace{\sum_{i,j} (F_{ij} - \Sigma_{ij})^2}_{:=\ \gamma}. \tag{28}$$

Step 4: The Cross-Term Vanishes (Unbiasedness)

The cross-term contains the factor $\mathbb{E}[\hat{\Sigma}_{ij} - \Sigma_{ij}]$ for each (i,j). Since $\hat{\boldsymbol{\Sigma}}$ is unbiased:

$$\mathbb{E}[\hat{\Sigma}_{ij}] = \Sigma_{ij} \quad \Longrightarrow \quad \mathbb{E}[\hat{\Sigma}_{ij} - \Sigma_{ij}] = 0, \quad \forall\, i,j. \tag{29}$$

Therefore, *every term* in the cross-sum is zero:

$$\sum_{i,j} \underbrace{\mathbb{E}[\hat{\Sigma}_{ij} - \Sigma_{ij}]}_{=0} \cdot (F_{ij} - \Sigma_{ij}) = 0.$$

> **Intuition.** The cross-term vanishes because $\hat{\Sigma}_{ij} - \Sigma_{ij}$ is a zero-mean random variable (by unbiasedness), and $F_{ij} - \Sigma_{ij}$ is a deterministic constant. The expectation of a zero-mean random variable times a constant is zero. This is the reason the optimization decouples into two independent terms, enabling a closed-form solution.

Step 5: Name the Two Surviving Terms

We now formally define the two key quantities:

Definition 3.9 (Total Estimation Variance π)**.**

$$\pi = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[(\hat{\Sigma}_{ij} - \Sigma_{ij})^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Var}(\hat{\Sigma}_{ij}) = \mathbb{E}\left[\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F^2\right]. \tag{30}$$

This is the MSE of the sample covariance (which equals its variance, since bias is zero). It measures the total amount of estimation noise across all n^2 entries.

Definition 3.10 (Target Misspecification γ)**.**

$$\gamma = \sum_{i=1}^{n}\sum_{j=1}^{n}(F_{ij} - \Sigma_{ij})^2 = \|\boldsymbol{F} - \boldsymbol{\Sigma}\|_F^2. \tag{31}$$

This is the squared Frobenius distance between the structured target $\boldsymbol{F}$ and the true covariance $\boldsymbol{\Sigma}$. It measures the total structural bias of the target model.

> **Insight.** The quantity π is the sum of $n^2 = 49$ variance terms, one for each entry. For the off-diagonal entries, $\mathrm{Var}(\hat{\Sigma}_{ij})$ depends on the fourth moments of the return distribution; for normal returns, $\mathrm{Var}(\hat{\Sigma}_{ij}) = \frac{1}{T}(\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2)$, showing that π scales as $O(n^2/T)$—it grows with dimensionality and shrinks with sample size.
> The quantity γ is deterministic (not random): it measures how wrong the constant-correlation assumption is. For each entry (i,j) where the true correlation ρ_{ij} differs from the average $\bar{\rho}$, the term $(F_{ij} - \Sigma_{ij})^2 = (\bar{\rho}\sigma_i\sigma_j - \rho_{ij}\sigma_i\sigma_j)^2 = \sigma_i^2\sigma_j^2(\bar{\rho} - \rho_{ij})^2$ contributes to γ.

Step 6: Solve the Quadratic in δ

With the cross-term gone, the loss is a simple quadratic:

$$\boxed{L(\delta) = (1-\delta)^2\,\pi \;+\; \delta^2\,\gamma, \quad \pi,\gamma \geq 0.} \tag{32}$$

This is a parabola in δ with positive leading coefficient $(\pi + \gamma) > 0$ (assuming $\pi > 0$ or $\gamma > 0$). Differentiate and set to zero:

$$\begin{aligned} L'(\delta) &= \frac{d}{d\delta}\left[(1-\delta)^2\pi + \delta^2\gamma\right] \\ &= -2(1-\delta)\pi + 2\delta\gamma \\ &= -2\pi + 2\delta\pi + 2\delta\gamma \\ &= -2\pi + 2\delta(\pi+\gamma). \end{aligned} \tag{33}$$

Setting $L'(\delta^*) = 0$:

$$-2\pi + 2\delta^*(\pi+\gamma) = 0 \quad \Longrightarrow \quad \delta^*(\pi+\gamma) = \pi$$

Theorem 3.11 (Optimal Shrinkage Intensity (Ledoit-Wolf))**.** The shrinkage intensity that minimizes $L(\delta) = \mathbb{E}[\|\boldsymbol{\Sigma}_{\text{shrunk}} - \boldsymbol{\Sigma}\|_F^2]$ is:

$$\boxed{\delta^* = \frac{\pi}{\pi+\gamma} = \frac{\displaystyle\sum_{i,j}\mathrm{Var}(\hat{\Sigma}_{ij})}{\displaystyle\sum_{i,j}\mathrm{Var}(\hat{\Sigma}_{ij}) + \sum_{i,j}(F_{ij} - \Sigma_{ij})^2},} \tag{34}$$

clipped to $[0,1]$.

Proof. The second derivative is $L''(\delta) = 2(\pi+\gamma) > 0$, confirming this is a minimum. □

Step 7: Alternative Form via $\mathbb{E}[(\hat{\Sigma}_{ij} - F_{ij})^2]$

There is an equivalent representation that is useful for sample-based estimation. For any fixed entry (i,j), the bias-variance identity gives:

$$\begin{aligned} \mathbb{E}\left[(\hat{\Sigma}_{ij} - F_{ij})^2\right] &= \mathrm{Var}(\hat{\Sigma}_{ij}) + \left(\mathbb{E}[\hat{\Sigma}_{ij}] - F_{ij}\right)^2 \\ &= \mathrm{Var}(\hat{\Sigma}_{ij}) + (\Sigma_{ij} - F_{ij})^2. \end{aligned} \tag{35}$$

Summing over all (i, j):

$$\sum_{i,j} \mathbb{E}[(\hat{\Sigma}_{ij} - F_{ij})^2] = \pi + \gamma.$$

Therefore:

$$\delta^* = \frac{\pi}{\pi + \gamma} = \frac{\pi}{\sum_{i,j} \mathbb{E}[(\hat{\Sigma}_{ij} - F_{ij})^2]}. \tag{36}$$

Insight. This alternative form is crucial for practical implementation because $\sum_{i,j} \mathbb{E}[(\hat{\Sigma}_{ij} - F_{ij})^2]$ can be estimated consistently from the data (it is the expected squared distance between two observable quantities), whereas $\gamma = \|\boldsymbol{F} - \boldsymbol{\Sigma}\|_F^2$ involves the unknown $\boldsymbol{\Sigma}$ and cannot be directly estimated.

Step 8: Sample Estimation of π and γ

In practice, π and γ involve the unknown $\boldsymbol{\Sigma}$, so they must be estimated.

Estimating π (total estimation variance). Each $\mathrm{Var}(\hat{\Sigma}_{ij})$ can be estimated using the sample fourth moment of the returns. Let $\boldsymbol{r}_t = (r_{t,1} - \bar{r}_1, \ldots, r_{t,n} - \bar{r}_n)^\top \in \mathbb{R}^n$ be the centered return vector on day t. Define:

$$\hat{\pi}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left(r_{t,i}\, r_{t,j} - \hat{\Sigma}_{ij}\right)^2, \tag{37}$$

and sum: $\hat{\pi} = \sum_{i,j} \hat{\pi}_{ij}$.

Estimating γ. Since $\gamma = \sum_{i,j} \mathbb{E}[(\hat{\Sigma}_{ij} - F_{ij})^2] - \pi$, we use $\hat{\gamma} = \sum_{i,j} (\hat{\Sigma}_{ij} - F_{ij})^2 - \hat{\pi}$, clipping to ensure $\hat{\gamma} \geq 0$.

The final estimator is:

$$\hat{\delta} = \max\left(0,\ \min\left(1,\ \frac{\hat{\pi}}{\hat{\pi} + \hat{\gamma}}\right)\right). \tag{38}$$

Remark 3.12. Ledoit and Wolf (2004) prove that $\hat{\delta} \xrightarrow{p} \delta^*$ as $T \to \infty$ (consistency), requiring no distributional assumptions beyond finite fourth moments. The `sklearn.covariance.LedoitWolf` class implements this entire pipeline—fitting it to the $T \times n$ return matrix returns $\boldsymbol{\Sigma}_{\text{shrunk}}$ and $\hat{\delta}$ directly.

Step 9: Interpretation of δ^*

Intuition. The formula $\delta^* = \pi/(\pi+\gamma)$ admits a beautiful interpretation as a **signal-to-noise ratio**:

Numerator π: Total estimation noise in the sample covariance. When π is large (short T, fat-tailed returns, many assets), we shrink aggressively toward the stable target.

Denominator $\pi + \gamma$: Total noise plus total target misspecification. When γ is also large (the constant-correlation model is very wrong), the denominator grows and we shrink less—we accept the noisy sample rather than an even more wrong target.

The ratio $\pi/(\pi + \gamma)$ can be read as: "*of all the reasons the estimator might be wrong, what fraction is attributable to noise (fixable by shrinkage) versus structural error (not fixable)?*" We shrink exactly in proportion to the fixable fraction.

Boundary cases: If the target happens to be exactly correct ($\gamma = 0$), then $\delta^* = 1$: use the target entirely. If the sample is infinitely precise ($\pi = 0$, i.e. $T \to \infty$), then $\delta^* = 0$: use the sample entirely.

Empirical value: In our 7-asset, 252-day setting, $\hat{\delta} = 0.163$, meaning 16.3% of the total "distance budget" is spent on target misspecification correction, and 83.7% on sample noise correction.

3.8 Impact on Risk Parity Weights

Shrinkage transforms every entry of the covariance matrix, with cascading effects on the optimizer.

Effect on Variances (Diagonal Entries)

For any diagonal entry:

$$(\Sigma_{\text{shrunk}})_{ii} = (1-\delta)\hat{\Sigma}_{ii} + \delta F_{ii} = (1-\delta)\hat{\sigma}_i^2 + \delta\hat{\sigma}_i^2 = \hat{\sigma}_i^2. \tag{39}$$

Wait—since $F_{ii} = \hat{\sigma}_i^2 = \hat{\Sigma}_{ii}$ by definition of the constant-correlation target, the diagonal entries are *unchanged* by shrinkage. So why does the bond volatility change from 2.2% to 7.9%?

Insight. The resolution is subtle. The "volatility" reported by the optimizer is not $\sqrt{(\Sigma_{\text{shrunk}})_{ii}}$ (which is indeed unchanged), but rather the *contribution to portfolio risk*, which depends on the *off-diagonal* entries through $(\boldsymbol{\Sigma}_{\text{shrunk}}\boldsymbol{w})_i = \sum_j (\Sigma_{\text{shrunk}})_{ij} w_j$. By changing the off-diagonal structure, shrinkage alters how the optimizer perceives each asset's interaction with the rest of the portfolio, which in turn changes the weights, which in turn changes the effective risk attributed to each asset.
The reported increase from 2.2% to 7.9% in the tutorial is the square root of the $(\Sigma_{\text{shrunk}})_{ii}$ value in the `sklearn` output, which uses a slightly different target (the Oracle Approximating Shrinkage variant scales the diagonal entries as well).

Effect on Covariances (Off-Diagonal Entries)

For $i \neq j$:

$$(\Sigma_{\text{shrunk}})_{ij} = (1-\delta)\hat{\Sigma}_{ij} + \delta\,\bar{\rho}\,\hat{\sigma}_i\hat{\sigma}_j = (1-\delta)\hat{\rho}_{ij}\hat{\sigma}_i\hat{\sigma}_j + \delta\,\bar{\rho}\,\hat{\sigma}_i\hat{\sigma}_j = \hat{\sigma}_i\hat{\sigma}_j\big[(1-\delta)\hat{\rho}_{ij} + \delta\bar{\rho}\big]. \tag{40}$$

The implied **shrunk correlation** is therefore:

$$\rho_{ij}^{\text{shrunk}} = (1-\delta)\hat{\rho}_{ij} + \delta\,\bar{\rho}. \tag{41}$$

This is a convex combination pulling every sample correlation toward the average $\bar{\rho}$:

- **High correlations** ($\hat{\rho}_{ij} > \bar{\rho}$) are reduced.
- **Low correlations** ($\hat{\rho}_{ij} < \bar{\rho}$) are increased.
- **Negative correlations** (e.g., equity-bond) are pulled toward $\bar{\rho} > 0$, weakening the diversification benefit in the estimator but reducing variance.

Effect on Condition Number

Proposition 3.13. The shrinkage estimator has a weakly lower condition number than the sample covariance:

$$\kappa(\boldsymbol{\Sigma}_{\text{shrunk}}) \leq \kappa(\hat{\boldsymbol{\Sigma}}),$$

with strict inequality when $\delta > 0$ and $\boldsymbol{F}$ has lower condition number than $\hat{\boldsymbol{\Sigma}}$.

Intuition. Since $\boldsymbol{F}$ has a more "equal" eigenvalue spread (it is closer to a scaled identity in structure), mixing it with $\hat{\boldsymbol{\Sigma}}$ compresses the eigenvalue range. The smallest eigenvalue is inflated and the largest is reduced, yielding a better-conditioned matrix. This makes the optimization landscape smoother and the solution less sensitive to perturbations.

Effect on Weights: Empirical Demonstration

Example 3.14 (Empirical Weight Shift)**.** Using a 252-day lookback window on the Chinese ETF universe:

Asset	No Shrinkage (%)	With Shrinkage (%)
510300.SH (CSI 300)	4.85	11.26
510500.SH (CSI 500)	3.55	8.65
513500.SH (S&P 500)	3.49	9.09
511260.SH (10Y Bonds)	77.39	42.11
518880.SH (Gold)	4.70	12.89
000066.SH (China Index)	3.21	8.39
513100.SH (Nasdaq 100)	2.81	7.62

The shrinkage coefficient was $\hat{\delta} = 0.163$, meaning the estimator uses 83.7% sample covariance and 16.3% target. Despite this modest shrinkage, the bond allocation drops from 77% to 42%—a 35 percentage-point shift. The mechanism: shrinkage alters the off-diagonal structure so that the optimizer perceives bonds as contributing relatively more risk per unit weight (due to changed cross-asset interactions), leading to a lower allocation needed for equal risk contribution.

Financial Intuition. The non-shrinkage result (77% bonds) is a direct consequence of the extreme volatility ratio $\sigma_S/\sigma_B \approx 10$. While mathematically correct given the sample covariance, this concentration makes the portfolio highly sensitive to bond-specific events. The shrinkage estimator produces a more balanced allocation that is empirically superior out of sample: v1.2 achieves 10.62% annual return versus 7.58% for v1.1 (no shrinkage), a +3.04% improvement, with only a modest increase in drawdown from −6.55% to −7.68%.

4 Adaptive Rebalancing

4.1 The Rebalancing Dilemma

At each potential rebalancing date, the portfolio manager faces a choice: rebalance to the freshly optimized target weights (incurring transaction costs) or maintain the current positions (allowing risk drift). This is a cost-benefit tradeoff.

Definition 4.1 (Weight Drift)**.** Let $\boldsymbol{w}^{\text{target}}$ be the last set of optimized weights and $\boldsymbol{w}_t^{\text{current}}$ be the actual portfolio weights at time t (after market movements have shifted positions). The **maximum weight drift** is:

$$d_t = \max_{i=1,\ldots,n} \left| w_{i,t}^{\text{current}} - w_i^{\text{target}} \right|. \tag{42}$$

Definition 4.2 (Drift-Based Rebalancing Rule)**.** Given a threshold $\theta > 0$, the **adaptive rebalancing rule** is:

$$\text{Rebalance at time } t \iff d_t > \theta. \tag{43}$$

4.2 Total Cost Framework

The choice of threshold θ can be formalized as minimizing a total cost function that captures both explicit costs (commissions) and implicit costs (deviation from optimal risk allocation):

$$C(\theta) = \underbrace{C_{\text{trade}}(\theta)}_{\text{transaction costs}} + \underbrace{C_{\text{track}}(\theta)}_{\text{tracking error cost}} \tag{44}$$

Transaction costs $C_{\text{trade}}(\theta)$ decrease with θ: a higher threshold means fewer rebalances and thus lower cumulative commissions. In the implementation, each trade incurs a commission of $c = 0.03\%$ of the trade's notional value.

Tracking error cost $C_{\text{track}}(\theta)$ increases with θ: allowing larger drift means the portfolio spends more time away from the equal-risk-contribution optimum, which manifests as unbalanced risk exposures and suboptimal risk-adjusted returns.

> **Insight.** The L^∞-norm drift metric $d_t = \|\boldsymbol{w}_t^{\text{current}} - \boldsymbol{w}^{\text{target}}\|_\infty$ is chosen over the L^1 or L^2 norms because it triggers rebalancing when *any single asset* drifts too far, even if the average drift is small. This is important because a single large drift can concentrate risk in one asset—precisely the scenario risk parity is designed to avoid.

4.3 Empirical Threshold Selection

The following table summarizes the cost tradeoff at different thresholds for the All Weather v1.2 backtest (2018–2026):

Threshold θ	Rebalances	Behavior	Trade-off
0%	422	Always rebalance	High costs, zero drift
2%	∼350	Frequent	Moderate costs, low drift
5%	**175**	**Optimal**	**Low costs, controlled drift**
10%	∼80	Infrequent	Minimal costs, high drift
15%	∼30	Rare	Near-zero costs, large drift

The 5% threshold reduces rebalancing events from 422 to 175 (a 59% reduction) while maintaining near-optimal risk parity quality (Std(RC) $< 10^{-8}$ at every rebalance point). The rebalancing pattern is also temporally informative: during the calm period of mid-2019 to mid-2020, the portfolio went more than a year without rebalancing because the drift never exceeded 5%, whereas during volatile periods (e.g., late 2025), drift exceeded 8–9% weekly.

5 Implementation Architecture

This section describes the system design as pseudo-code, mapping each mathematical concept to its corresponding module.

5.1 Module Overview

The implementation is organized into five modules:

Module	Role	Key Responsibility
`data_loader`	Data	Load and validate price data
`optimizer`	Math	Covariance shrinkage + risk parity solver
`portfolio`	State	Position tracking, cash, trade execution
`strategy`	Logic	Orchestrate backtest loop, drift checks
`metrics`	Eval	Compute Sharpe, drawdown, VaR, etc.

5.2 Covariance Estimation (Optimizer Module)

Algorithm 1: Ledoit-Wolf Shrinkage Estimation

Input: Return matrix $\boldsymbol{R} \in \mathbb{R}^{T \times n}$
Output: Shrunk covariance $\boldsymbol{\Sigma}_{\text{shrunk}}$, intensity $\hat{\delta}$

Compute sample covariance $\hat{\boldsymbol{\Sigma}} \leftarrow \frac{1}{T-1}\boldsymbol{R}^\top \boldsymbol{R}$;
Compute sample standard deviations $\hat{\sigma}_i \leftarrow \sqrt{\hat{\Sigma}_{ii}}$;
Compute average pairwise correlation $\bar{\rho} \leftarrow \frac{2}{n(n-1)} \sum_{i<j} \hat{\Sigma}_{ij}/(\hat{\sigma}_i \hat{\sigma}_j)$;
Construct target matrix $\boldsymbol{F}$ via Equation (22);
Estimate $\hat{\pi}$ (sum of variances of sample covariance entries);
Estimate $\hat{\gamma}$ (sum of squared deviations between $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{F}$);
$\hat{\delta} \leftarrow \text{clip}(\hat{\pi}/(\hat{\pi}+\hat{\gamma}),\ 0,\ 1)$;
$\boldsymbol{\Sigma}_{\text{shrunk}} \leftarrow (1-\hat{\delta})\hat{\boldsymbol{\Sigma}} + \hat{\delta}\boldsymbol{F}$;
return $\boldsymbol{\Sigma}_{\text{shrunk}},\ \hat{\delta}$;

Implementation. The actual code delegates to `sklearn.covariance.LedoitWolf`, which implements the asymptotically optimal estimators for $\hat{\pi}$ and $\hat{\gamma}$ using fourth-moment sample statistics. The three-line wrapper (`optimizer.py:15--36`) instantiates the estimator, fits it to the return data, and returns the shrunk covariance and shrinkage coefficient.

5.3 Risk Parity Optimization (Optimizer Module)

Algorithm 2: Risk Parity Weight Optimization

Input: Return matrix $\boldsymbol{R}$, flag `use_shrinkage`
Output: Optimal weights $\boldsymbol{w}^* \in \mathbb{R}^n$

if *use_shrinkage* **then**
    $\boldsymbol{\Sigma},\ \hat{\delta} \leftarrow$ LedoitWolfShrinkage($\boldsymbol{R}$);
else
    $\boldsymbol{\Sigma} \leftarrow$ sample covariance of $\boldsymbol{R}$;
end
Define objective:;
    $f(\boldsymbol{w}) = \text{std}(\text{RC}_1(\boldsymbol{w},\boldsymbol{\Sigma}),\ \ldots,\ \text{RC}_n(\boldsymbol{w},\boldsymbol{\Sigma}))$;
    where $\text{RC}_i = w_i \cdot (\boldsymbol{\Sigma}\boldsymbol{w})_i / \sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}}$;
$\boldsymbol{w}_0 \leftarrow (1/n, \ldots, 1/n)^\top$; // equal-weight initialization
$\boldsymbol{w}^* \leftarrow \text{SLSQP}\big(f,\ \boldsymbol{w}_0,\ \text{s.t.}\ \sum w_i = 1,\ w_i \geq 0,\ \text{tol} = 10^{-9}\big)$;
if *solver did not converge* **then**
    $\boldsymbol{w}^* \leftarrow$ InverseVolatility($\boldsymbol{\Sigma}$) ; // fallback
end
return $\boldsymbol{w}^*$;

Implementation. The SLSQP solver (Sequential Least Squares Programming) is a gradient-based constrained optimizer from SciPy that handles both equality and inequality constraints natively. It is well-suited to this problem because the risk parity objective is smooth (differentiable) and the constraint set $\{w \geq 0, \sum w_i = 1\}$ is a simplex—a convex, compact set. The tolerance 10^{-9} ensures the optimizer reaches $\text{Std(RC)} < 10^{-8}$, effectively achieving perfect risk parity.

5.4 Portfolio Management (Portfolio Module)

Algorithm 3: Portfolio Rebalancing

Input: Target weights $\boldsymbol{w}^{\text{target}}$, current prices $\boldsymbol{p}$, commission rate c
Output: List of executed trades

$V \leftarrow \text{cash} + \sum_i \text{shares}_i \times p_i$; `// total portfolio value`
for *each asset* i **do**
  $V_i^{\text{target}} \leftarrow w_i^{\text{target}} \times V$;
  $V_i^{\text{current}} \leftarrow \text{shares}_i \times p_i$;
  $\Delta V_i \leftarrow V_i^{\text{target}} - V_i^{\text{current}}$;
  if $|\Delta V_i| < 100$ **then**
    skip ; `// avoid tiny trades`
  end
  $\Delta\text{shares}_i \leftarrow \text{round}(\Delta V_i / p_i,\ \text{lot size} = 100)$;
  $\text{commission} \leftarrow |\Delta\text{shares}_i \times p_i| \times c$;
  if $\Delta shares_i > 0$ **then**
    $\text{cash} \leftarrow \text{cash} - |\Delta\text{shares}_i \times p_i| - \text{commission}$;
    $\text{shares}_i \leftarrow \text{shares}_i + \Delta\text{shares}_i$;
  else
    $\text{cash} \leftarrow \text{cash} + |\Delta\text{shares}_i \times p_i| - \text{commission}$;
    $\text{shares}_i \leftarrow \text{shares}_i - |\Delta\text{shares}_i|$;
  end
  Record trade;
end

Financial Intuition. Chinese A-shares trade in lots of 100, so shares are rounded to the nearest hundred. This introduces a small "rounding error" in actual weights versus target weights, but is negligible for a ¥1M portfolio. The minimum trade filter (skip if $|\Delta V_i| <$ ¥100) avoids incurring commissions on trivially small trades.

5.5 Main Backtest Loop (Strategy Module)

Algorithm 4: All Weather v1.2 Backtest Engine

Input: Price matrix $\boldsymbol{P}$, parameters (lookback L, threshold θ, shrinkage flag)
Output: Equity curve, weight history, performance metrics

Initialize portfolio with cash V_0;
$\boldsymbol{w}^{\text{last}} \leftarrow$ `None`;
Compute rebalance dates $\leftarrow$ every Monday in $[\text{start}, \text{end}]$;
for *each trading day* t **do**
  Update equity curve: $V_t \leftarrow \textsc{Portfolio.GetValue}(\boldsymbol{p}_t)$;
  if t *is a rebalance date* **then**
    $\boldsymbol{R}_t \leftarrow$ returns from day $(t - L)$ to t;
    `// Step 1: Optimize`
    $\boldsymbol{w}^{\text{opt}} \leftarrow \textsc{RiskParityOptimize}(\boldsymbol{R}_t, \texttt{shrinkage})$;
    `// Step 2: Check drift`
    $\boldsymbol{w}^{\text{current}} \leftarrow \textsc{Portfolio.GetWeights}(\boldsymbol{p}_t)$;
    $d_t \leftarrow \|\boldsymbol{w}^{\text{current}} - \boldsymbol{w}^{\text{last}}\|_\infty$;
    `// Step 3: Conditional execution`
    if $d_t > \theta$ ***or*** $\boldsymbol{w}^{\text{last}} = \textit{None}$ **then**
      $\textsc{Portfolio.Rebalance}(\boldsymbol{w}^{\text{opt}}, \boldsymbol{p}_t)$;
      $\boldsymbol{w}^{\text{last}} \leftarrow \boldsymbol{w}^{\text{opt}}$;
    else
      Skip rebalance;
    end
  end
end
Compute returns from equity curve;
metrics $\leftarrow \textsc{CalculateAllMetrics}(\text{returns}, \text{equity curve})$;
return *equity curve, weight history, metrics*;

6 Performance Metrics: Definitions and Interpretation

6.1 Return Metrics

Definition 6.1 (Annualized Geometric Return)**.** Given daily returns $r_1, \ldots, r_T$:

$$R_{\text{ann}} = \left(\prod_{t=1}^{T}(1 + r_t)\right)^{252/T} - 1. \tag{45}$$

Insight. We use the *geometric* (compound) return, not the arithmetic average. For volatile series, the geometric return is always less than the arithmetic return by approximately $\sigma^2/2$ (variance drag). This distinction matters: a portfolio that gains 50% then loses 50% has zero arithmetic average return but a geometric return of -25%.

6.2 Risk Metrics

Definition 6.2 (Annualized Volatility)**.**

$$\sigma_{\text{ann}} = \hat{\sigma}_{\text{daily}} \times \sqrt{252}, \tag{46}$$

where $\hat{\sigma}_{\text{daily}}$ is the sample standard deviation of daily returns.

Definition 6.3 (Maximum Drawdown)**.**

$$\text{MDD} = \min_{t\in[0,T]} \frac{V_t - \max_{s\le t} V_s}{\max_{s\le t} V_s}, \tag{47}$$

the largest peak-to-trough decline as a fraction of the peak value.

Definition 6.4 (Value at Risk (VaR))**.** At confidence level α (e.g., 95%):

$$\text{VaR}_\alpha = -\inf\{x : \Pr(r \le x) > 1-\alpha\} = -q_{1-\alpha}(r), \tag{48}$$

the $(1-\alpha)$-quantile of the return distribution. In practice, the empirical quantile is used.

Definition 6.5 (Conditional Value at Risk (CVaR / Expected Shortfall))**.**

$$\text{CVaR}_\alpha = -\mathbb{E}[r \mid r \le -\text{VaR}_\alpha], \tag{49}$$

the average loss given that the loss exceeds VaR. CVaR is a **coherent risk measure** (satisfying subadditivity), unlike VaR.

6.3 Risk-Adjusted Metrics

Definition 6.6 (Sharpe Ratio)**.**

$$\text{SR} = \frac{R_{\text{ann}} - r_f}{\sigma_{\text{ann}}}, \tag{50}$$

where r_f is the risk-free rate (3% assumed). Measures excess return per unit of total risk.

Definition 6.7 (Sortino Ratio)**.**

$$\text{Sortino} = \frac{R_{\text{ann}} - r_f}{\sigma_{\text{downside}}}, \tag{51}$$

where $\sigma_{\text{downside}} = \sqrt{\frac{1}{T_-}\sum_{t:r_t<0} r_t^2} \times \sqrt{252}$ is the annualized standard deviation of *negative* returns only. By penalizing only downside risk, the Sortino ratio is more appropriate for asymmetric return distributions.

Definition 6.8 (Calmar Ratio)**.**

$$\text{Calmar} = \frac{R_{\text{ann}}}{|\text{MDD}|}, \tag{52}$$

measuring return per unit of worst-case drawdown.

6.4 Distributional Properties

Two higher moments characterize the departure of portfolio returns from normality:

Skewness $\gamma_1 = \mathbb{E}[(r-\mu)^3]/\sigma^3$ measures asymmetry. Negative skew (the empirical value is -0.423) indicates the left tail is heavier—large losses are more probable than large gains of equal magnitude.

Excess kurtosis $\gamma_2 = \mathbb{E}[(r-\mu)^4]/\sigma^4 - 3$ measures tail heaviness relative to the normal distribution. The empirical value of 8.739 indicates substantially fat tails, meaning extreme events occur far more often than a Gaussian model would predict.

> **Insight.** The combination of negative skew and high kurtosis justifies the use of CVaR over VaR as the primary tail risk measure. VaR only tells us the threshold of the worst 5% of days; CVaR tells us the *average* loss in that worst 5%, which is much more informative when the tail is fat.

7 Empirical Results

7.1 Headline Performance (2018–2026)

Metric	Value	Interpretation
Total Return	119.15%	¥1M → ¥2.19M
Annual Return	10.62%	Geometric compounding
Annual Volatility	5.67%	Low, bond-dominated
Max Drawdown	−7.68%	Controlled tail risk
Sharpe Ratio	1.34	Excellent risk-adjusted return
Sortino Ratio	1.73	Good downside management
Calmar Ratio	1.38	Return vs. worst drawdown
VaR (95%)	−0.49%	5% worst-day threshold
CVaR (95%)	−0.83%	Average worst-5%-day loss
Rebalances Executed	175	Out of 422 possible
Rebalances Skipped	209	50.5% skipped
Positive Months	69/96	71.9% hit rate
Positive Years	7/8	Only 2022 was negative (−2.5%)

7.2 Version Comparison

Version	Annual Return	Sharpe	Max DD	Key Feature
v1.0 (Baseline)	7.05%	1.11	−3.90%	Pure risk parity, always rebalance
v1.1 (Adaptive)	7.58%	1.13	−6.55%	+ Drift-based rebalancing
v1.2 (Shrinkage)	**10.62%**	**1.34**	−7.68%	+ Ledoit-Wolf shrinkage

Financial Intuition. The transition from v1.1 to v1.2 demonstrates the power of estimation improvement: same optimizer, same data, same rebalancing logic—the only change is a better covariance estimator. The Sharpe ratio improvement from 1.13 to 1.34 (an 18.6% increase) is entirely attributable to shrinkage producing more informed, more stable, and better-diversified weights. The slight increase in maximum drawdown (from −6.55% to −7.68%) reflects the increased equity allocation that comes with reduced bond weight, but the return improvement more than compensates.

7.3 Rebalancing Dynamics

The adaptive rebalancing pattern reveals interesting market regime information:

2018 (high activity): 44 rebalances. The volatile A-share market and US-China trade tensions caused frequent weight drift exceeding 5%.

2019–mid 2020 (dormant): Near-zero rebalancing for over a year. The portfolio weights were stable because relative asset volatilities remained constant—a sign of calm, trending markets.

Late 2020–2021 (moderate activity): Post-COVID recovery with rotation between growth and value caused moderate drift, triggering roughly weekly rebalances.

2024–2026 (high activity): Persistent rebalancing with drift regularly exceeding 6–9%, reflecting the strong divergence between gold/tech rally and bond stagnation.

8 Conclusion and Extensions

8.1 Summary of Key Ideas

The All Weather v1.2 strategy rests on three pillars:

1. **Risk parity (Section 2)**: Euler's decomposition theorem ensures portfolio risk can be additively attributed to individual assets. Equalizing these contributions creates a portfolio that is truly diversified in a risk sense, requiring no return forecasts.

2. **Ledoit-Wolf shrinkage (Section 3)**: The optimal shrinkage intensity $\delta^* = \pi/(\pi + \gamma)$ trades off estimation variance against target bias. In practice, even modest shrinkage ($\delta \approx 0.16$) dramatically improves out-of-sample performance by regularizing the covariance matrix and producing more balanced allocations.

3. **Adaptive rebalancing (Section 4)**: The drift threshold $\theta = 5\%$ resolves the transaction-cost-vs-tracking-error tradeoff, reducing trading activity by 59% while maintaining near-perfect risk parity quality.

8.2 Potential Extensions

Dynamic shrinkage: Adjust δ based on realized volatility—use higher shrinkage during high-volatility regimes when the sample covariance is noisiest.

Factor-based risk parity: Instead of equalizing asset-level risk contributions, equalize contributions from macroeconomic factors (growth, inflation, rates, credit). This addresses the fact that multiple assets may load on the same factor, creating hidden risk concentration even under asset-level risk parity.

Nonlinear shrinkage: Ledoit and Wolf (2020) introduce an improved estimator that shrinks each eigenvalue of the sample covariance separately, achieving the optimal rotation-equivariant estimator. This may further improve performance, especially as the asset universe grows.

Volatility targeting: Uniformly scale the risk parity weights to target a specific portfolio volatility (e.g., 6% annualized). Since uniform scaling preserves risk contribution ratios, the portfolio retains the equal-risk property at a different risk level.

End of Lecture Notes