

# PCA on Implied Volatility Surfaces

**Author:** Trent Potter

**Graduation Year:** '25

**Institution:** Booth School of Business

**Class:** Data, Learning, and Algorithms

## Motivation and Industry Context

Implied volatility (IV) is a critical input in the pricing and hedging of options portfolios. Unlike observable variables such as underlying prices or interest rates, IV is derived indirectly from market prices, encapsulating expectations about future volatility and deviations from simpler pricing assumptions. Traders rely on IV extensively, as it represents collective market uncertainty, including deviations from log-normal returns, risk premiums, dividend uncertainties, and model misspecifications.

For practical portfolio management, understanding the covariance structure of IV across different strikes and maturities is essential. Portfolio risk stemming from IV changes frequently rivals or exceeds the direct underlying price risks (delta and gamma), especially for hedged positions where underlying price movements are neutralized.

However, constructing a reliable covariance matrix for IV surfaces poses significant challenges:

- **High dimensionality:** Individual equities can have thousands of traded options across numerous expirations and strikes, creating enormous covariance matrices that are computationally intensive and numerically unstable.
- **Noise and liquidity issues:** Bid-ask spreads, liquidity gaps, and data anomalies significantly impact IV estimates, further complicating reliable covariance estimation.

## The Dimensionality Challenge and Data Representation

A critical step in modeling IV surfaces is converting observed market data (options listed by discrete strikes and expirations) into a structured, less-sparse representation. First translating for strikes and expiration dates to  $\Delta$  and TTM makes observations comparable across different time periods. We still have a problem that  $(\text{delta}, \text{time-to-maturity}) \in \mathbb{R}^2$ . For empirical observations, we need a way to address this issue:

- **Functional Representation:** Model IV as a continuous function over  $(\text{delta}, \text{TTM})$ . Estimation of this  $\mathbb{R}^2$  valued function.
- **Grid-Based Sampling:** Generate a discrete sampling grid and populate points through interpolation or kernel-weighted averaging. Yields indexing over a finite subset of  $\Delta \times T$

Given practical constraints, this report adopts the latter method, constructing a standardized, interpolated grid of IV observations across delta and TTM. While this discretized representation simplifies covariance calculation, we recognize it involves potential interpolation errors and sacrifices structural information from the full surface.

In this analysis, we utilize options data provided by OptionMetrics, spanning from 2008 to 2022. OptionMetrics employs a kernel sampling function based on distances in time-to-maturity (TTM), delta, and an indicator variable distinguishing puts and calls separately. Specifically, the kernel is defined as:

The IV surface constructed from this data for AAPL options on 2017-03-03 is visualized below:

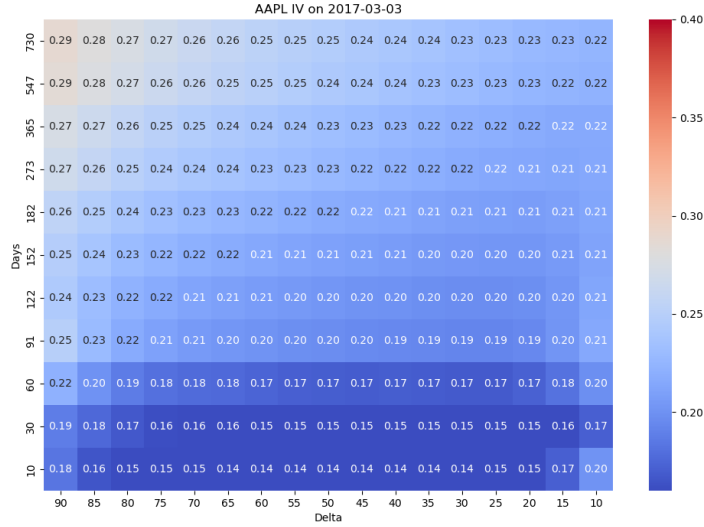


Figure 1: AAPL IV Surface on 2017-03-03

Data covering index options is also available and can be used similarly for further analyses.

## Dimension Reduction as a Foundation for Prediction

Dimension reduction techniques, such as PCA, also serve as foundational tools for traders who want to employ IV surfaces as covariates in more complex prediction models. Once IV surfaces have been distilled into a lower-dimensional set of factors, these factors become efficient inputs to predictive regressions, machine learning models, or time-series forecasting methodologies. By simplifying the IV representation, traders can significantly enhance the interpretability,

stability, and computational efficiency of predictive modeling efforts, enabling more sophisticated analyses of market dynamics, volatility forecasting, and portfolio optimization.

## Mathematical Foundations: From Black-Scholes to IV Surfaces

### Black-Scholes Option Pricing Basics

The classical framework for pricing European call options is given by the Black-Scholes formula:

$$C(S, K, t, r, d, \sigma) = Se^{-dt}N(d_1) - Ke^{-rt}N(d_2)$$

with

$$d_1 = \frac{\ln(\frac{S}{K}) + (r - d + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t}$$

where:

- $C$ : Call option price
- $S$ : Price of the underlying asset
- $K$ : Strike price
- $t$ : Time to maturity
- $r$ : Risk-free interest rate (or carry cost)
- $d$ : Dividend yield
- $\sigma$ : Volatility of the underlying asset (assumed constant)
- $N(\cdot)$ : Standard normal cumulative distribution function (CDF)

### Implied Volatility: Definition and Interpretation

Given a market-observed option price  $C_{market}$ , we numerically invert the Black-Scholes equation to solve for the volatility parameter  $\sigma$ , termed the *implied volatility (IV)*:

$$\sigma_{IV} = BS^{-1}(C_{market}; S, K, t, r, d)$$

This implied volatility encapsulates all market information not explicitly modeled by the basic Black-Scholes assumptions, including market expectations about volatility, risk premiums, dividend uncertainty, and deviations from log-normal return distributions.

## PCA and its Impact on Covariance and Precision Matrices

PCA decomposes the covariance matrix of IV changes into orthogonal principal components (PCs), capturing the primary dimensions of variation:

$$\Sigma = U\Lambda U^\top$$

where:

- $\Sigma$  is the covariance matrix of IV changes,
- $U$  contains eigenvectors (principal components), and
- $\Lambda$  is a diagonal matrix of eigenvalues representing variance explained by each factor.

For dimensionality reduction, we select the top  $K$  principal components, creating a reduced-rank covariance matrix:

$$\Sigma_K = U_K \Lambda_K U_K^\top$$

The corresponding precision matrix (inverse covariance matrix) is:

$$\Sigma_K^{-1} = U_K \Lambda_K^{-1} U_K^\top$$

Critically, the smallest eigenvalues in  $\Sigma$  produce the largest reciprocal eigenvalues in  $\Sigma^{-1}$ . Thus, the choice of  $K$  significantly impacts numerical stability, as retaining too many components can introduce instability from small eigenvalues, while retaining too few can omit critical information.

## Interpretation of Principal Components

### PC 1 (Level Factor)

- Represents a broad level shift in IV, consistently positive across the surface.
- Relatively lower variation observed in call options.

### PC 2 (Skew Factor)

- Captures skew in underlying volatility perceptions across delta space.
- Primarily delta-driven, highlighting shifts in distributional skewness.

### PC 3 (Term Structure Slope Factor)

- Reflects variation exclusively along maturity dimension.
- Analogous to yield curve slope in fixed income markets.

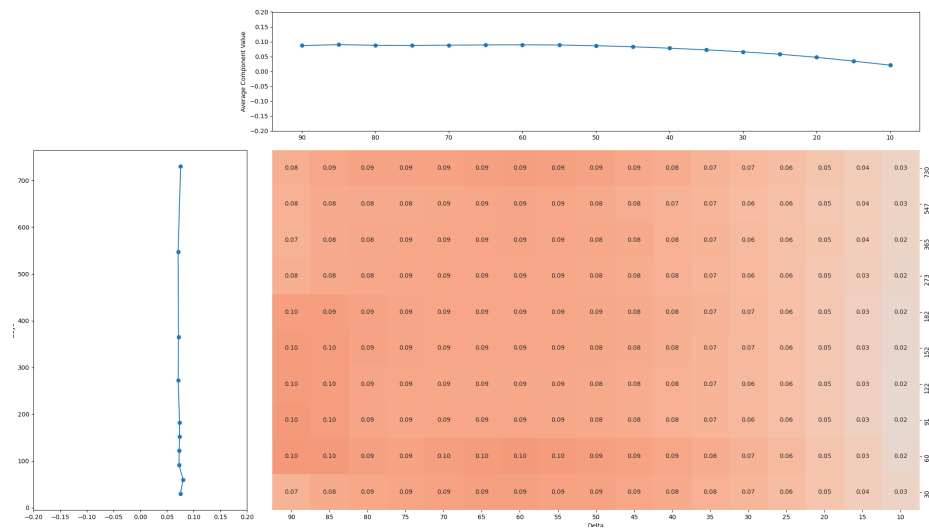


Figure 2: PC 1 - Level Factor

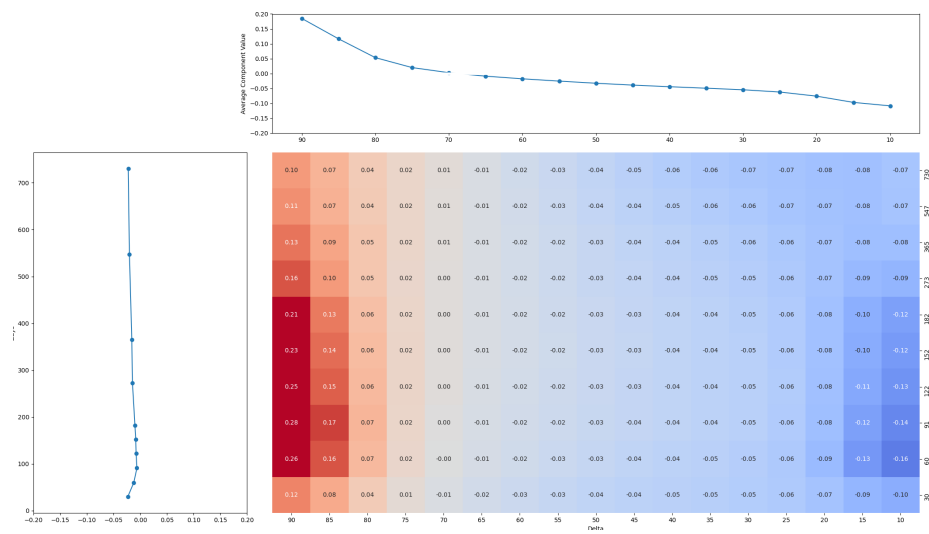


Figure 3: PC 2 - Skew Factor

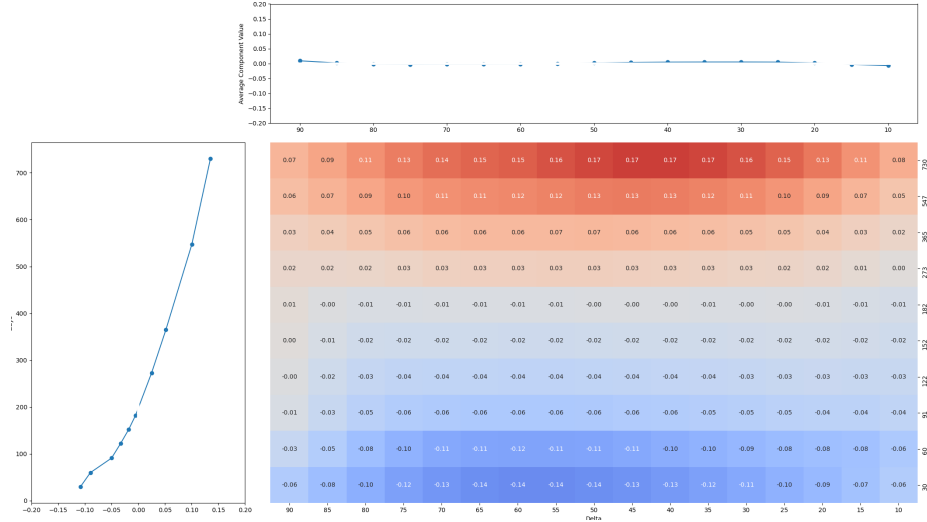


Figure 4: PC 3 - Term Structure Slope Factor

#### PC 4 (Decoupled Across Term Structure Factor)

- Captures variation that allows decoupled movements between shorter-dated and longer-dated expirations.
- Exhibits notable skewness in the far delta space (deep out-of-the-money puts).

#### PC 5 (Term Structure Curvature Factor)

- Represents curvature in the maturity dimension and kurtosis in IV surfaces.

### Trading Applications of PCA-Based Covariance Matrix Estimation

#### Marginal and Total Portfolio IV Risk

Principal Component Analysis (PCA) allows traders to assess portfolio risks by constructing either full-rank or reduced-rank covariance matrices. We define the Vega-weighted position vector as ( $w$ ). The marginal and total portfolio IV risk measures are given by:

$$\text{Marginal Risk} = w^\top \Sigma, \quad \text{Portfolio IV Risk} = w^\top \Sigma w$$

When a sufficient number of principal components ( $K > 5$ ) are retained, the reduced-rank covariance matrix  $\Sigma_K$  closely approximates the full-rank matrix. Importantly, PCA allows decomposition of the total portfolio IV risk into distinct sub-risks associated with each principal component:

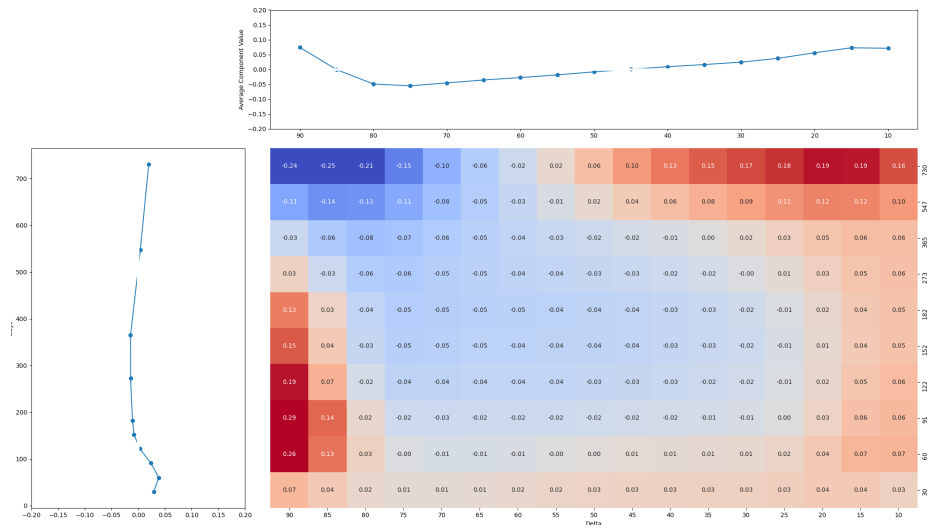


Figure 5: PC 4 - Decoupled Across Term Structure Factor

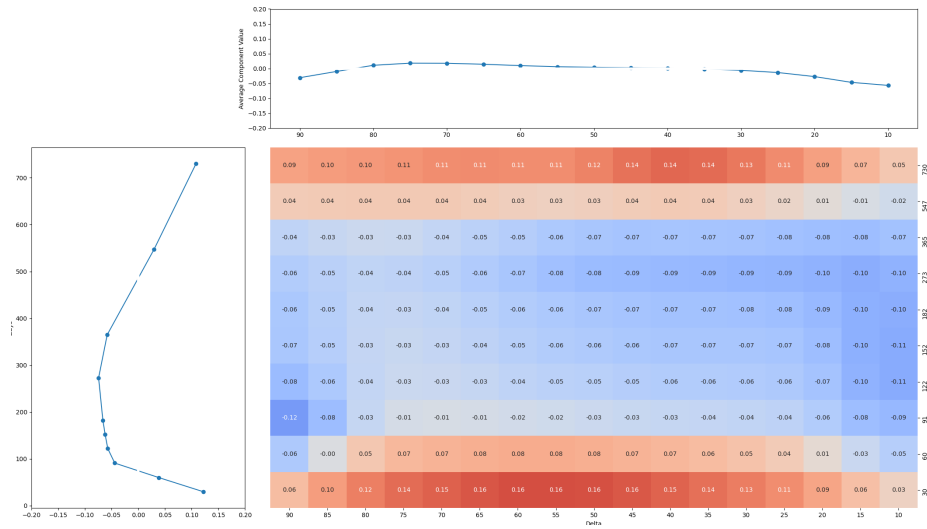


Figure 6: PC 5 - Term Structure Curvature Factor

$$\text{Risk Decomposition} = w^\top U_K \Lambda_K$$

This decomposition offers clear insights into individual sources of volatility risk.

### Mean-Variance Optimal Portfolio Construction

Given the expected returns vector  $\mu$  and the covariance matrix  $\Sigma$ , the optimal portfolio weights  $w$  can be obtained by solving the following optimization problem:

$$\begin{aligned} & \underset{w}{\text{maximize}} && \frac{w^\top \mu}{\sqrt{w^\top \Sigma w}} \\ & \text{subject to} && \sum_i w_i = 1 \end{aligned}$$

$$w_K = \frac{\Sigma_K^{-1} \mu}{\mathbf{1}^\top \Sigma_K^{-1} \mu}$$

Direct inversion of the full covariance matrix typically yields numerically unstable results due to high condition numbers, resulting in unrealistic and excessively large weights.

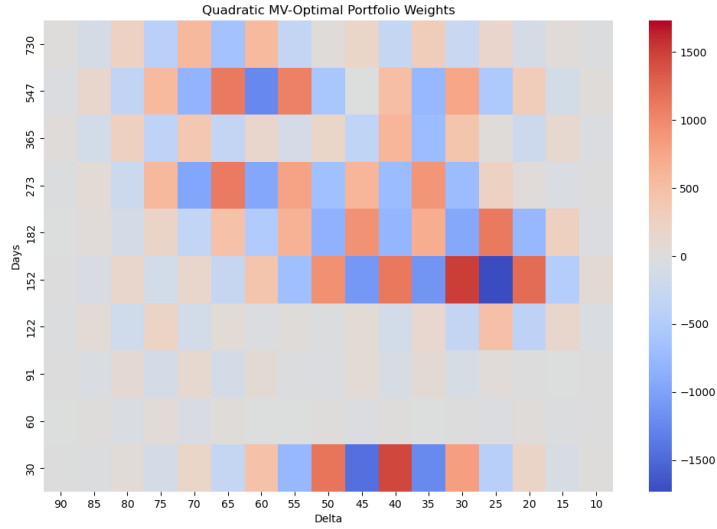


Figure 7: Unstable Position



Although MVO assumptions (normal distributions, linear payoffs) are violated by options portfolios, the practical ease of implementation makes it a common approximation in trading environments.

### Reduced-Rank PCA Covariance as a Stabilization Technique

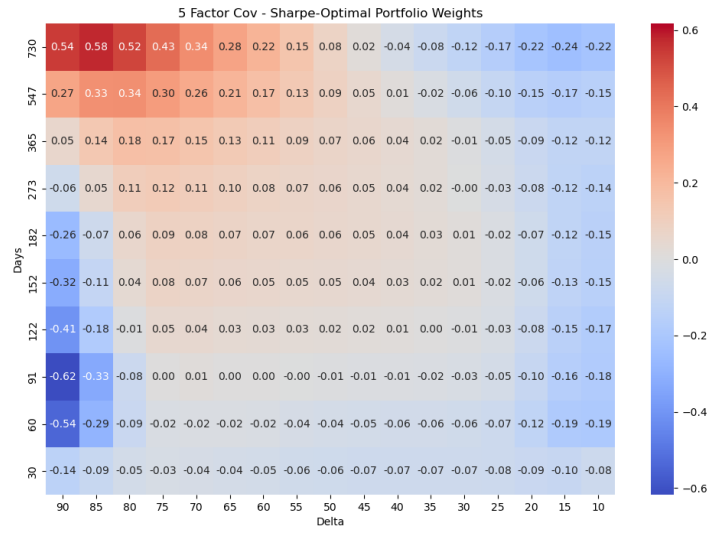
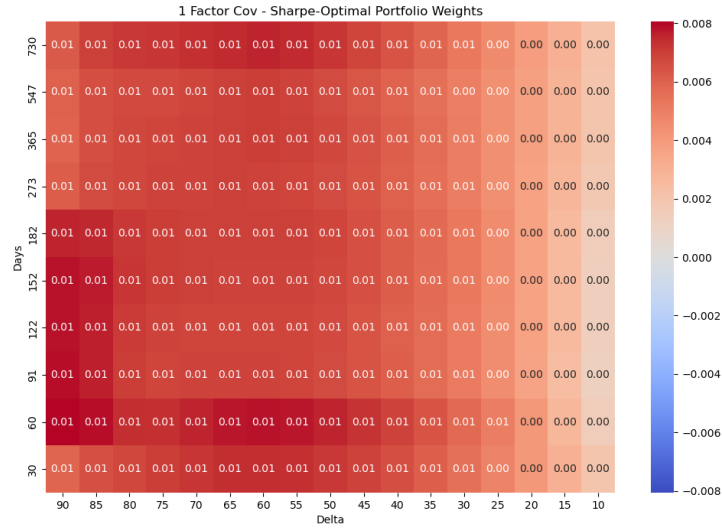
To address numerical instability, traders employ a reduced-rank PCA approach, removing smaller eigenvalues that destabilize the precision matrix. The reduced-rank precision matrix is defined as:

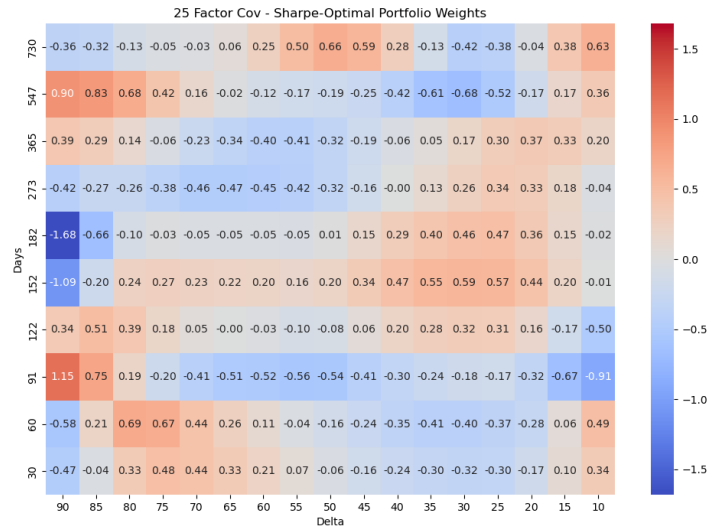
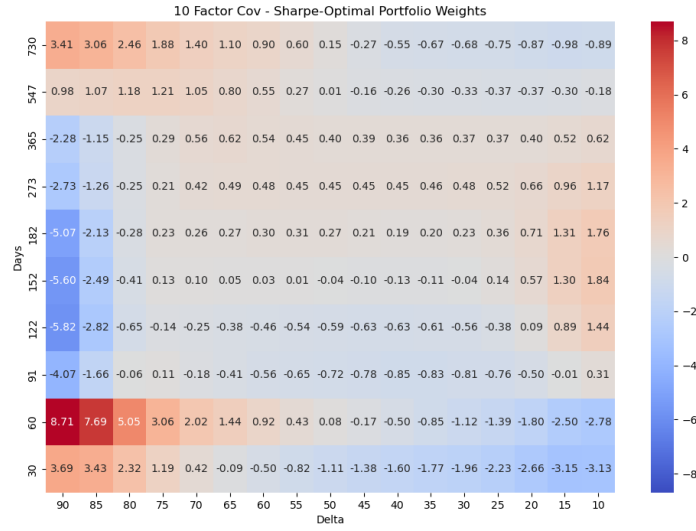
$$\Sigma_K^{-1} = U_K \Lambda_K^{-1} U_K^\top$$

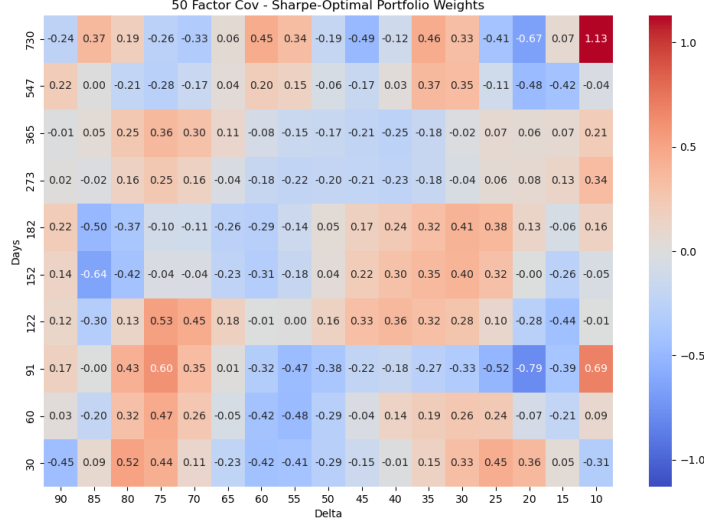
Choosing an appropriate number of principal components  $K$  is critical. Practically, traders employ several heuristics to guide this choice:

- **Explained Variance (Scree Plot):** Select  $K$  based on capturing a high percentage (e.g., 90-95%) of total variance.
- **Horn's Method:** Select  $K$  by comparing the eigenvalues from real data against eigenvalues derived from random data matrices.
- **Information Criteria (AIC/BIC):** These statistical criteria penalize model complexity, offering a balance between complexity and fit.
- **Cross-Validation:** Using out-of-sample predictive performance to determine an optimal  $K$  that generalizes well to future periods.

In addition, economic interpretation of principal components strongly influences the selection decision. Components with clear market interpretations are typically preferred. Below are portfolio optimization results for the same expected return vector, but varying levels of  $K$ .







Complementary regularization methods, such as ridge regression or shrinkage estimators, may further stabilize numerical outcomes alongside PCA.

### Sharpe Ratio and Validation in PCA-based Covariance Selection

The Sharpe ratio is a key metric used to evaluate portfolio performance, defined as:

$$\text{Sharpe Ratio} = \frac{\mu^\top w}{\sqrt{w^\top \Sigma w}}$$

Cross-sectionally, increasing the number of principal components  $K$  mechanically leads to higher Sharpe ratios, as richer covariance structures allow portfolios to exploit subtle differences in expected returns. However, this improvement does not necessarily indicate better out-of-sample performance, as real-world expected returns are inherently uncertain and estimated with error.

Thus, practical PCA application emphasizes time-series validation techniques, such as rolling-window validation or out-of-sample backtests, to assess the robustness of Sharpe ratio improvements. These methods evaluate how effectively portfolios optimized with different  $K$  values perform against future, unforeseen market conditions, allowing traders to select a  $K$  value that offers stability and predictive robustness.

Economic interpretability also plays an essential role. Components with a strong economic rationale are typically favored, even if statistically marginal. Ultimately, balancing these statistical measures with economic intuition and

rigorous time-series validation ensures the chosen  $K$  enhances actual trading performance.