

Modeling Market Geometry

An Engineering Approach to Implied Volatility in Options

F.J. Penagos

¹ Duke University, Pratt School of Engineering, Durham, NC, USA
e-mail: felipe.jaramillo@duke.edu

*

**

² **Integrating Stochastic Search, Dimensional Analysis, and Options.** Research conducted under the academic supervision of Prof. Henri P. Gavin (Chair, Civil and Environmental Engineering) and Prof. David Ye (Professor in the Master of Interdisciplinary Data Science and Mathematics at Duke) as part of final and independent projects.

Received May 5, 2025

ABSTRACT

This paper unites optimization theory, applied mathematics, and quantitative finance within an interdisciplinary engineering framework for modeling and trading based on the implied volatility surface of options. Beginning with a stochastic optimization engine designed to test robust decision-making under uncertainty, we extend the methodology to the geometry of financial markets—treating the volatility surface as a dynamic manifold whose deformations encode market expectations. Through Gaussian smoothing, cubic spline fitting, and principal component analysis (PCA), the research extracts low-dimensional representations of volatility dynamics and integrates them into a quantitative strategy validated within the QuantConnect research environment. The result is a unified synthesis of engineering optimization and financial signal modeling that demonstrates how stochastic methods, penalty-based objective functions, and surface analysis can jointly reveal structure and predictability within the apparent randomness of markets.

Key words. optimization under uncertainty – applied mathematics – implied volatility surface – stochastic search – principal component analysis – quantitative finance – QuantConnect – financial engineering – machine learning – gradient boosting – statistics

1. Introduction

Stock exchanges serve purposes beyond institutional wealth generation. We consider financial markets and their instruments to offer a uniquely rich environment for applying analytical and computational methods across disciplines. They are dynamic, adaptive systems—continuously influenced by stochastic processes, feedback effects, and collective behavior—making them ideal laboratories for testing ideas from optimization, chaos and game theory, and data science. Studying such systems requires tools from engineering and applied mathematics, computer science, psychology, and other domains. In this research, we combine data modeling, probabilistic thinking, and robust optimization under uncertainty.

Interdisciplinary Motivation.

The motivation for this research project stems from the desire to synthesize principles from **engineering**, **mathematics**, **computer science**, and **finance** to analyze and optimize complex decision systems. The behavior of financial markets resembles that of nonlinear dynamical systems—where control, noise, and structure interact in high-dimensional spaces. Engineering contributes methodology for system analysis and optimization; mathematics provides the theoretical foundation for modeling randomness and constraints; computer science

supplies the computational power for iterative and stochastic search; and finance frames these tools within an uncertain, information-driven environment for practical application.

Technical Overview.

At its core, the objective is to design and analyze a system that identifies, quantifies, and optimizes **signals extracted from the implied volatility surface**—a multidimensional function describing option contracts' pricing for a stock's (the "underlying" asset) expected behavior. By integrating methods from probability, numerical analysis, and machine learning, the study aims to construct a theoretically grounded approach for volatility-based trading decisions. While the paper assumes familiarity with the topics discussed, we will briefly review key concepts.

Optimization forms the theoretical backbone of this work. It is defined as **the process of finding the best possible solution or outcome from all available options, often through computational methods**. In finance, optimization problems typically maximize expected return subject to risk or capital constraints, as pioneered by Markowitz (1952). Beyond static mean-variance theory, modern quantitative strategies increasingly rely on *dynamic optimization*—minimizing cost functions that evolve over time, often under uncertainty and with penalization to enforce smoothness or robustness. The same principles apply to calibrating volatility surfaces and determining optimal "quality weights" in multidimensional signal spaces (expanded upon later).

Options contracts are agreements granting the buyer the right, but not the obligation, to buy or sell an asset (usually a stock, the underlying) at a predetermined price within a specified timeframe. The

* Department of Civil and Env. Engineering, Duke University

** Department of Mathematics, Duke University

*** For modules 6-7, I worked alongside *Shiya Ye*, *Tianai Wang*, and *Ben Greene* in Math585 course at Duke. As such, they are co-authors of the research developed in QuantConnect

Black–Scholes framework (Black & Scholes 1973) laid the foundation for pricing derivatives under continuous-time stochastic processes. In its classical form, the price of a European call option $C(S, t)$ satisfies the partial differential equation:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0, \quad (1)$$

where S is the underlying asset price, σ its volatility, and r the risk-free rate of return. Under assumptions of constant volatility and lognormal asset dynamics, this equation admits a closed-form solution:

$$C = SN(d_1) - Ke^{-rT}N(d_2), \quad (2)$$

60 where $N(\cdot)$ denotes the cumulative distribution function (CDF) of a standard normal distribution, K is the strike price, and

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}. \quad (3)$$

While the Black–Scholes model provides mathematical elegance, its assumptions fail to capture intrinsic characteristics of real markets, such as heavy tails, volatility clustering, and the presence of smiles and skews in implied volatility. Subsequent research introduced **stochastic volatility models** and surface-based formulations treating **implied volatility** (IV) as a nonlinear function of strike and time to maturity, providing a more flexible representation of market expectations (Hull 2018).

70 In tandem, this project leverages **machine learning** methods to uncover latent structure in high-dimensional financial data. Specifically, **principal component analysis (PCA)** reduces dimensionality and identifies dominant modes of variation (used here to explain volatility surface changes). Mathematically, PCA decomposes the covariance matrix $\Sigma = \frac{1}{n}X^T X$ of a centered data matrix X into orthogonal eigenvectors and eigenvalues, capturing maximal variance along successive axes. This allows complex, correlated surface deformations to be represented through a few principal components, simplifying both analysis and optimization.

80 The **implied volatility surface** itself represents the market's expectation of future volatility based on current option prices. It is a 3D surface with option strike prices on one axis, maturities on another—and the implied volatility "level" traditionally on the z -axis; forming a 3D landscape whose structure (or changes in structure over time) encodes what we believe are opportunities for arbitrage.

Methodological Overview

The methodological foundation originates from a supervised research project conducted by the author under the academic mentorship of **Prof. Henri P. Gavin, Ph.D.**, Chair of Civil and Environmental Engineering at Duke University. Prof. Gavin's extensive **MATLAB library for optimization and probabilistic modeling** provided the computational environment for developing and testing optimization routines. The project's objective was to explore **stock exchange algorithm performance using optimization** through *stochastic search methods and penalty-based objective functions*. These experiments established a framework for evaluating trade-offs between robustness and efficiency, serving as the conceptual and methodological basis for the present study. The next section expands on this foundational research.

100 In summary, this project treats the volatility surface not merely as a pricing tool, but as a **signal manifold**. Our hypothesis posits that its gradients, curvatures, and temporal evolutions encode information about market sentiment and structural shifts. Dimensionality reduction techniques—particularly principal component analysis (PCA) (Jolliffe & Cadima 2016)—enable extraction of dominant factors governing surface dynamics. These dominant factors can be interpreted and forecasted using machine-learning techniques. Furthermore, once normalized and combined through a weighted "quality equation," these factors form the basis for a trading signal whose parameters can be optimized using penalized objective functions. The resulting model connects theory and application by blending optimization under uncertainty with financial signal extraction.

Therefore, after conducting this process—assuming the IV surface can predict movements in the underlying, and assuming our model correctly captures the dominant factors—we should be able to generate a profitable theoretical trading algorithm.

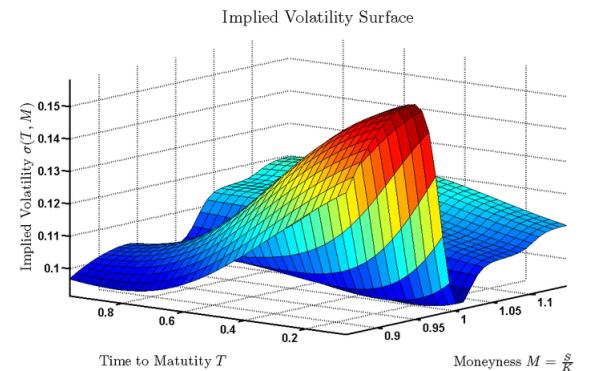


Fig. 1: **Illustrative Implied Volatility Surface.** The surface represents the implied volatility $IV(M, T)$ as a function of *strike price* (moneyness) and *time to maturity*. Typical market surfaces exhibit a pronounced downward skew in equity options, reflecting higher implied volatilities for low strikes (out-of-the-money puts) and flattening across longer maturities. The curvature and temporal evolution of this surface constitute the core informational manifold analyzed in this study.

2. Methodological Origins

To understand the development of our volatility-surface framework, we must first revisit the optimization research foundation upon which it was built—it played a major role in justifying our approach.

2.1. Context and Supervision

This optimization framework originated in a course project conducted by the author under the academic supervision of **Prof. Henri P. Gavin** (Duke CEE). The implementation leverages his MATLAB optimization library, including `ORSOpt.m` (Optimized Step-Size Random Search), to search rugged, non-smooth design spaces with penalty handling for constraints.

2.2. Problem Setting: Quantifying "Quality"

The decision to buy or sell stock is considered both an art and a science. We aimed for a quantitative approach, claiming such decisions should be based on a stock statistic we call "**quality**" Q . Intuitively, stocks with higher quality are better purchase candidates, while those with lower quality are better candidates for sale. For this research, we assumed the quality measure of stock s on day d ($Q_{d,s}$) could be a weighted sum of (1) the fractional change in stock price, (2) the rate of fractional change in stock price, and (3) the volatility of stock price over a recent period.

Letting $p_{d,s}$ denote the closing price of stock s on day d , a smoothed series $\bar{p}_{d,s}$ is computed over a window N (weighted average), from which we form *kinematic* features:

$$\dot{\bar{p}}_{d,s} \equiv \bar{p}_{d,s} - \bar{p}_{d-1,s}, \quad \ddot{\bar{p}}_{d,s} \equiv \dot{\bar{p}}_{d,s} - \dot{\bar{p}}_{d-1,s},$$

and a recent-window volatility estimate $v_{d,s}$ (e.g., standard deviation of fractional changes over N days). The *quality* of stock s on day d is then a weighted function:

$$Q_{d,s} = q_1 \dot{\bar{p}}_{d,s} + q_2 \ddot{\bar{p}}_{d,s} + q_3 v_{d,s}, \quad (4)$$

with design variables $\mathbf{q} = (q_1, q_2, q_3)$ and smoothing window N . To build cohesive trading logic, we must also consider buy/sell thresholds and the fraction of cash to invest. These functionalities are controlled by parameters (f_c, B, S) ; the full design vector for our project is:

$$\theta = (N, q_1, q_2, q_3, f_c, B, S, W),$$

where W (an optional parameter) governs a **weighted moving-average** (WMA) confirmation rule. While we will explain the mathematics behind a WMA shortly, this parameter computes the mean of data points (prices) by assigning greater weights to more recent values, emphasizing current trends.

Mathematics of the Parameters

Here we present a more detailed description of the features in our equation.

Each component of the quality equation *Eq. (4)* quantifies a distinct aspect of short-term price behavior, smoothed over a recent time window of N trading days. The coefficients q_1 , q_2 , and q_3 are weighting parameters to be optimized through `ORSOpt.m`, representing the relative importance of momentum, acceleration, and volatility, respectively. The smoothed price $\bar{p}_{d,s}$ of stock s on day d is computed as an exponentially weighted moving average:

$$\bar{p}_{d,s} = \sum_{i=1}^N w_i p_{(d-i),s}, \quad (5)$$

where the weights w_i assign greater emphasis to recent observations:

$$w_i = \frac{\exp(-5i/N)}{\sum_{i=1}^N \exp(-5i/N)}. \quad (6)$$

These exponentially decaying weights ensure responsiveness to current trends while maintaining numerical smoothness. Figure 2 illustrates

typical weight decay profiles for $N = 5$, $N = 10$, and $N = 20$, while Figure 3 shows how N influences smoothed prices and their derivatives.

The first derivative $\dot{\bar{p}}_{d,s}$ represents the *fractional change* of the smoothed price:

$$\dot{\bar{p}}_{d,s} = \frac{\bar{p}_{d,s} - \bar{p}_{(d-2),s}}{\bar{p}_{d,s} + \bar{p}_{(d-2),s}}, \quad (7)$$

while the second derivative $\ddot{\bar{p}}_{d,s}$ captures the *rate of change of that fractional change*:

$$\ddot{\bar{p}}_{d,s} = \frac{\dot{\bar{p}}_{d,s} - \dot{\bar{p}}_{(d-2),s}}{2}. \quad (8)$$

Finally, the daily volatility $v_{d,s}$ is defined as the running coefficient of variation over the last N days:

$$v_{d,s} = \frac{1}{\bar{p}_{d,s}} \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left[p_{(d-i),s} - \left(\frac{1}{N} \sum_{j=1}^N p_{(d-j),s} \right) \right]^2}. \quad (9)$$

This term measures local dispersion relative to the smoothed mean price, thereby encoding short-term uncertainty. Together, $(\dot{\bar{p}}_{d,s}, \ddot{\bar{p}}_{d,s}, v_{d,s})$ define the input features for Eq. (4), with q_1 , q_2 , and q_3 optimized to yield the most robust and profitable portfolio responses.

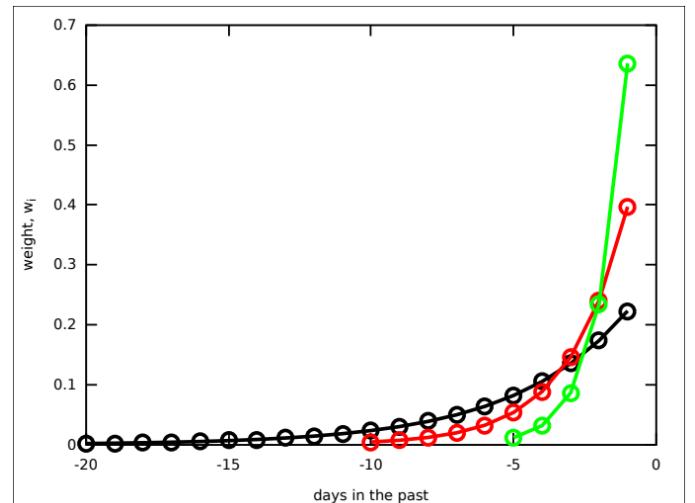


Fig. 2: **Exponential weighting functions.** Illustration of exponential decay profiles for smoothing windows of $N = 5, 10, 20$. More recent observations receive exponentially greater weights.

In short, the objective is to maximize profitability.

2.3. Objective, Constraints, and Optimization Algorithm

The optimization goal was to determine the parameter vector $\theta = (N, q_1, q_2, q_3, f_c, B, S, W)$ that maximizes terminal portfolio value on test (unseen) data while maintaining stability under uncertainty.

Each element governs a distinct behavioral dimension: N controls the smoothing horizon; (q_1, q_2, q_3) weight momentum, acceleration, and volatility; f_c determines the cash-allocation fraction; (B, S) set the buy/sell thresholds; and W defines the window size of the weighted moving average filter used for confirmation.

The backtesting engine `exchange_analysis.m` returns the terminal value $V(\theta)$ after T simulated trading days along with diagnostics such as drawdown and turnover. A robustness-adjusted objective was defined as:

$$\max_{\theta \in \Theta} \widehat{V}(\theta) = V(\theta) - \lambda_1 \text{DD}(\theta) - \lambda_2 \text{Turnover}(\theta), \quad (10)$$

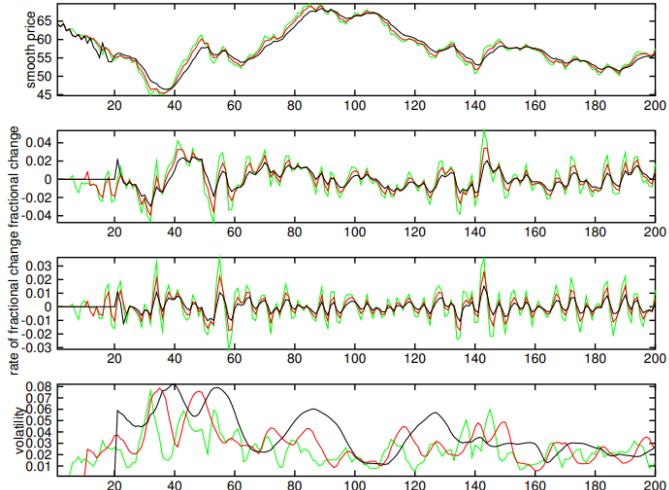


Fig. 3: **Effect of the smoothing window N .** Demonstration of how the number of trading days used in the exponential moving average affects (a) the smoothed price $\bar{p}_{d,s}$, (b) its fractional change $\bar{p}_{d,s}$, and (c) the rate of fractional change $\ddot{p}_{d,s}$.

Constraints. In equation Eq. (10), DD is the maximum drawdown penalty and Turnover discourages excessive trading activity (due to transaction costs).

Drawdown is the percentage loss in portfolio value relative to its highest peak since the algorithm began.

Additional constraints are system-rooted and imposed as simple bounds, e.g., $q_i \in [-1, 1]$ and $f_c \in [0, 1]$, noting that quality equation weights must remain within this range, along with operational rules prohibiting shorting (selling borrowed stocks) and enforcing conservative re-entry after liquidation.

2.4. Optimization Method: From SQP to Random Search

Several algorithms were evaluated for solving Eq. (10).

Sequential Quadratic Programming (SQP) proved unsuitable because the objective surface was highly non-convex and non-differentiable.

The Nelder–Mead Algorithm (NMA), while gradient-free, relies on simplex geometry—an algorithm using geometric constructs to search for optima iteratively. Therefore, it struggled to scale efficiently with the problem’s dimensionality. NMA is powerful, but in higher-dimensional spaces, its performance becomes unreliable. Consider that with only our initial design parameters, we already had 8 dimensions; moreover, stock-price data combined with our parameters yielded highly non-smooth, noisy, and irregular objective functions. We’ll expand on this ahead.

Therefore, we adopted ORSopt.m—the **Optimized Random Search** method from Prof. Gavin’s MATLAB library—which adaptively samples the design space with variable step sizes and stochastic perturbations. ORSopt excels on rugged landscapes: it explores broadly initially, refines locally near promising regions, and integrates penalty functions directly into its acceptance criteria, providing a practical global-search compromise between stochastic exploration and local refinement.

2.5. Trading Logic and Risk Adaptation

A trade was initiated when the quality signal $Q_{d,s}$ exceeded the buy threshold B and remained above its weighted moving average (WMA) over window W ; positions were reduced or closed when $Q_{d,s}$ fell below the sell threshold S or underperformed its WMA. Capital was allocated proportionally to the available cash fraction f_c across active positions.

If a drawdown limit was breached over consecutive days, the port-

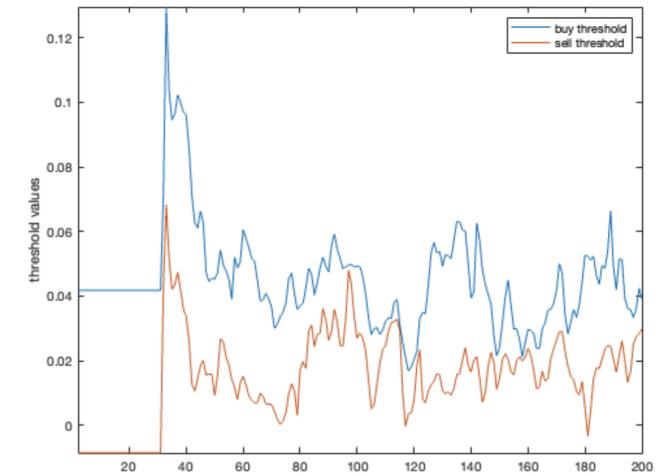
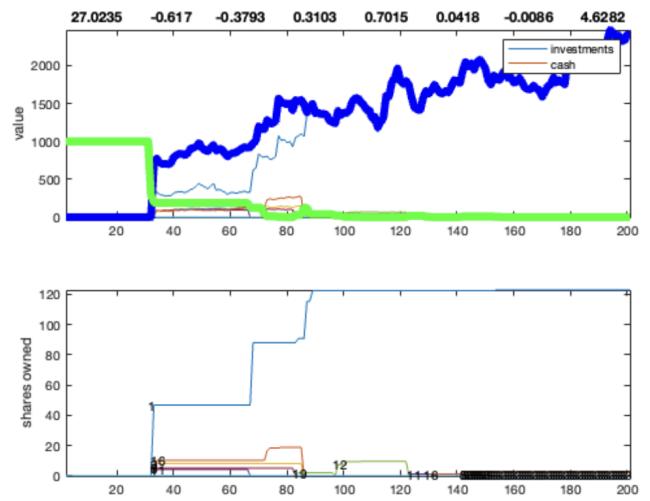


Fig. 4: **Optimized Random Search with penalties.** Conceptual overview of our performance using ORSopt.m: adaptive step-size proposals, constraint limitations, and the Q-equation. Here you can evaluate portfolio value, stocks owned over time, and threshold evolution.

folio liquidated decaying positions and dynamically tightened (B, S), thereby re-entering the market more cautiously. This adaptive feedback mechanism stabilized exposure and prevented over-commitment during volatility spikes.

2.6. Optimization Results and Surface Smoothing

The resulting objective landscape was markedly *rugged*, exhibiting many sharp, isolated peaks—typical of noisy financial systems. Initial optimizations achieved formidable returns but displayed sensitivity to small parameter perturbations, suggesting local overfitting. This would not only indicate high sensitivity to the model’s underlying assumptions but also imply poor performance on unseen data. We had to address this to improve our algorithm.

To remedy this, we computed a parameter-grid map of $V(q_i, q_j)$ while holding other parameters fixed and applied a two-dimensional Gaussian filter:

$$\tilde{V}(q_i, q_j) = (V * G_\sigma)(q_i, q_j), \quad G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

where G_σ smooths high-frequency noise without distorting large-scale trends. Intuitively, this replaces each point on the surface with a weighted average of its neighbors, producing a more continuous "terrain" and revealing globally consistent optima.

Regions remaining stable under small random perturbations of θ were classified as **robust optima**, while narrow peaks were penalized as likely overfit. Figures 5 and 6 visualize the effect: after smoothing, multiple local maxima collapse, leaving broad, generalized optima with reduced variance under perturbation.

In short, although our data was not 3-dimensional, we used partial derivatives to model changes in Portfolio Value while shifting only two parameters at a time. We repeated this process for all parameter combinations and visualized the surfaces created by plotting the landscape of our final portfolio value. Ultimately, we used a mathematical technique called a Gaussian Filter to smooth these surfaces and identify points where optima remain consistent even with slight stochastic perturbations (i.e., optimal weights for which parameters still yield high values even on unseen stocks).

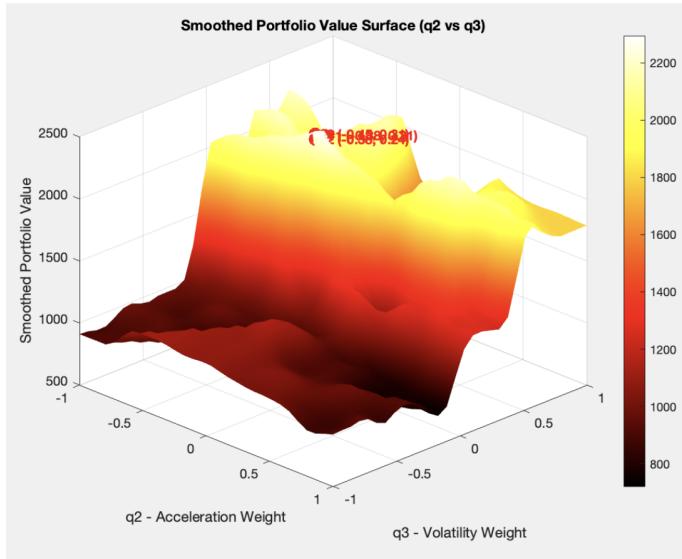


Fig. 5: **Parameter landscape.** Terminal value surface over (q_2, q_3) at fixed (N, q_1, f_c, B, S, W) after Gaussian smoothing.

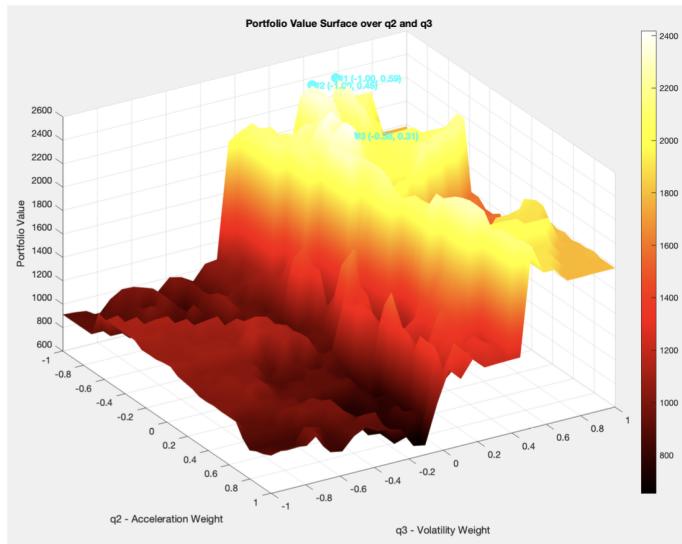


Fig. 6: **Non-smoothed Surface Example**

2.7. Bridge to the Volatility-Surface Strategy

The lessons from this experiment—smoothing rugged objective landscapes, penalizing sensitivity, and selecting globally stable optima—along with deriving an equation to quantify stock "quality" and perfecting it through optimization, directly informed the next phase of this research.

In the volatility-surface framework, the implied-volatility manifold replaces price dynamics, and the same principles of dimensional reduction (performed here with Principal Component Analysis), penalized optimization, and stochastic perturbation analysis are applied to construct interpretable trading signals.

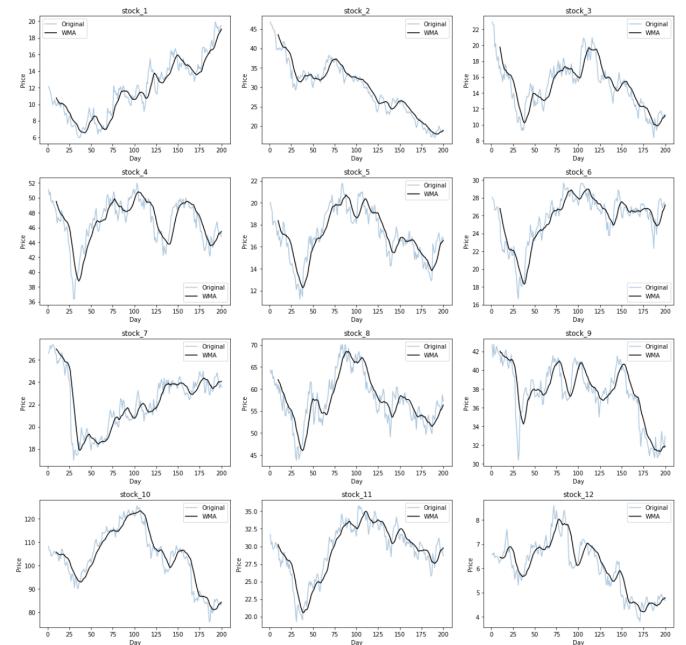


Fig. 7: **Effect of WMA on Stock Price Time-Series**

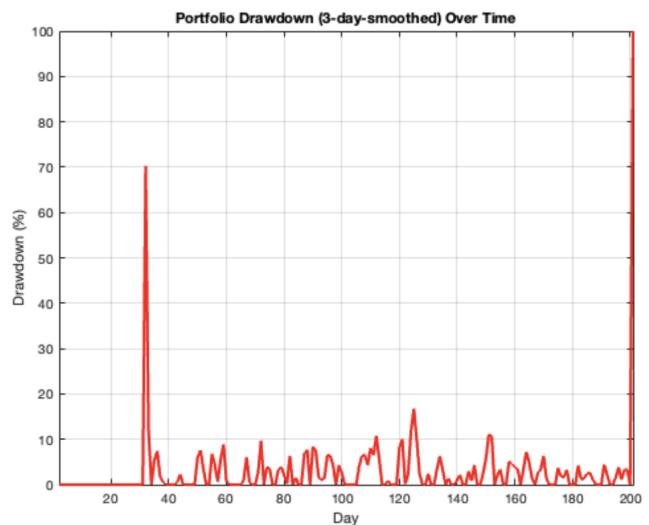


Fig. 8: **Drawdown.** Days 27 and 200 represent entry of the trading algorithm and manual liquidation to finalize the observation period

3. Volatility Surface Trading – The Core Module

3.1. Conceptual Introduction

The transition from equity optimization to volatility modeling represents an expansion in dimensionality and abstraction. Whereas the first framework optimized discrete stock-level parameters, the volatility-surface approach treats the entire implied volatility manifold as a continuous function encoding collective market expectations.

Formally, the **implied volatility surface** $IV(K, T)$ defines the market-implied volatility of options as a function of strike price K and time to maturity T . Each point on this surface reflects the volatility input that, when substituted into the Black–Scholes pricing equation, reproduces the market price of an option.

Unlike realized volatility—which measures historical price variation—implied volatility is inherently forward-looking. Thus, its geometry across strikes and maturities captures information about perceived tail risk, leverage effects, and term structure. Empirically, equity option markets exhibit a characteristic **downward skew**: options with lower strikes (out-of-the-money puts) trade at higher implied volatilities, reflecting the market's asymmetric fear of downside shocks.

Moneyness The term *moneyness* describes the relationship between an option's strike price K and the current price of its underlying asset S . An option is **in the money (ITM)** if exercising it would yield immediate profit (e.g., $S > K$ for a call, or $S < K$ for a put). It is **at the money (ATM)** when $S \approx K$, and **out of the money (OTM)** when exercising would be unprofitable (e.g., $S < K$ for a call, $S > K$ for a put). Moneyness provides a natural horizontal coordinate for the implied-volatility surface, standardizing volatility comparisons across strikes.

We reinterpret this surface as a **signal manifold**. Rather than serving solely as an input to pricing models, the volatility surface becomes a dynamic, information-rich object whose gradients, curvatures, and temporal evolutions may anticipate shifts in realized volatility. This viewpoint generalizes our earlier "quality" optimization framework:

Where before we quantified the *quality* of a stock through smoothed price-based statistics, we now aim to quantify the *quality of volatility dynamics as a trading signal through structural deformations of the IV surface itself*.

Our objective, therefore, is not to price options, but to **extract predictive structure** from the evolving surface—reducing its infinite-dimensional behavior to a tractable, low-dimensional representation. Subsequent sections describe how we construct the volatility surface via cubic spline interpolation and decompose its daily variations using **Principal Component Analysis (PCA)**, thereby building a signal-driven trading framework that leverages these factors to time volatility exposure under strict delta-neutral risk control.

3.2. Data Preparation and Research Environment

We implemented and validated the volatility-surface methodology directly within **QuantConnect's Research Environment**, using its proprietary, institution-grade options database. QuantConnect is an algorithmic trading platform enabling users to design, test, and deploy trading strategies across multiple asset classes using historical and live market data. This platform provided daily snapshots of complete option chains—including mid implied volatilities, bid–ask quotes, strikes, expirations, and underlying spot levels—for all selected underlyings. Once validated in research mode, the workflow was tested on QuantConnect's **deployment environment**, ensuring the end-to-end pipeline translated consistently from theoretical development to live execution.

Market Liquidity

Each daily option chain is represented as a collection of tuples. Because raw quotes can contain noise or microstructure irregularities, we restricted the dataset to contracts with positive bid–ask spreads, nonzero

open interest, and expirations between 10 and 45 days to maturity—this last filter focused on higher liquidity contracts.

To compare across underlying prices and obtain a more explanatory implied volatility surface, we focused on moneyness rather than raw strike price:

$$M = \frac{K}{S_t} - 1, \quad 350$$

so that $M = 0$ denotes the at-the-money (ATM) point, $M < 0$ corresponds to in-the-money puts (out-of-the-money calls), and $M > 0$ to out-of-the-money puts (in-the-money calls). This reparameterization linearizes the smile near the origin and facilitates interpolation across underlyings with different spot prices.

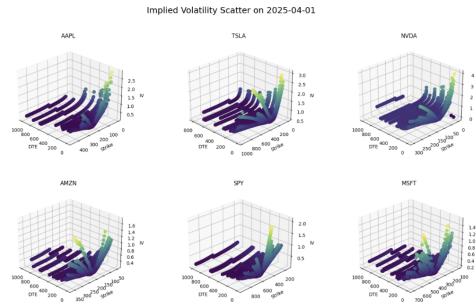


FIGURE 1. Raw Option Chains Data

Fig. 9: **Raw option-chain data.** Scatter of implied volatility across moneyness and days-to-expiry for one underlying.

Formally, for each trading day t , the fitted surface satisfies:

$$\begin{aligned} IV_t^{\text{spline}}(m, \tau) = \arg \min_{f \in C^2} & \left\{ \sum_{i=1}^{N_t} (IV_i - f(m_i, \tau_i))^2 \right. \\ & \left. + \lambda \iint \left[\left(\frac{\partial^2 f}{\partial m^2} \right)^2 + \left(\frac{\partial^2 f}{\partial \tau^2} \right)^2 \right] dm d\tau \right\}. \end{aligned} \quad (11)$$

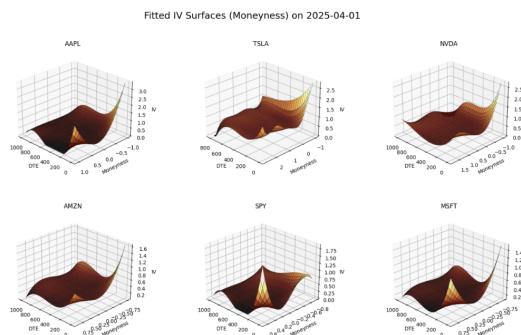


Fig. 10: **Spline-smoothed surface.**

While the spline produces a continuous representation over the full (m, τ) plane, our subsequent analysis concentrates on the most liquid segment of the surface—maturities between 10 and 45 DTE and moneyness $m \in [-0.25, 0.25]$. This region coincides with high trading volume and reliable quotes.

Note that focusing on this liquid band also mitigates extrapolation risk: the spline is not forced to infer volatility behavior in far-OTM or ultra-long-dated contracts, where pricing errors and market frictions dominate.

Having established a smooth, arbitrage-consistent surface representation, the next methodological step is to quantify how these surfaces

370 evolve over time. Section 3.3 explains how we decomposed intra-day surface dynamics via Principal Component Analysis (PCA), translating complex deformations into interpretable signals.

4. Volatility Surface Momentum: From Theory to Strategy

Having established the methodological foundation in optimization and signal construction, we now extend those principles to the implied volatility (IV) surface itself. The central idea is that the shape of the IV surface—its level, skew, and curvature—encapsulates market expectations of future uncertainty. If these geometric properties evolve smoothly over time, then their directional shifts may reveal exploitable momentum in volatility. In this section, we formalize how such structural changes are quantified, modeled, and ultimately translated into trading signals.

4.1. Stocks of Interest

For each underlying asset [AAPL, TSLA, NVDA, AMZN, SPY], we obtained end-of-day option chains including strikes K , maturities τ , bid/ask quotes, and implied volatilities. After filtering out contracts with missing data, negligible open interest, or unrealistic pricing, the remaining quotes were projected onto a continuous surface using a **polynomial smoothing spline**:

$$390 \quad IV_t^{\text{spline}}(M, \tau) = \text{Spline}(M, \tau),$$

where $M = K/S_t - 1$ denotes moneyness and τ the days-to-expiration. The spline minimizes the weighted sum of squared residuals and second-derivative penalties, enforcing continuity in both slope and curvature. This non-parametric interpolation, akin to the framework of Fengler (2005), preserves realism in local option prices while eliminating microstructural noise, particularly across far-OTM or ultra-short maturities.

To standardize analysis across maturities, we extract from each daily surface the slice corresponding to predetermined i calendar days to expiration:

$$S_t(M) = IV_t^{\text{spline}}(M, i),$$

spanning moneyness in $[-0.25, 0.25]$. These fixed-DTE cross-sections approximate horizons of one week, one trading month, a fiscal period, and one trading year—aligning with the volatility measure implicit in the VIX index.

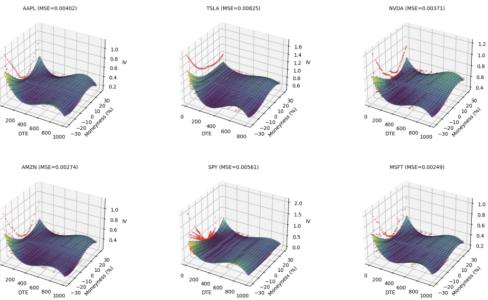


Fig. 11: Principal component modes of the IV surface. The first component corresponds to a uniform level shift, the second to a tilt (skew), and the third to curvature (smile) adjustments.

4.2. Dimensional Decomposition and Feature Extraction

Each daily slice $S_t(m)$ defines a one-dimensional curve representing the market's instantaneous volatility smile. To describe how this shape changes through time, we apply **Principal Component Analysis**

(PCA) to the stacked sequence of slices $\{S_t(m)\}_{t=1}^T$. Let the daily deviations be $\Delta S_t = S_t - S_{t-1}$, from which the covariance matrix:

$$\Sigma = \frac{1}{T-1} \sum_t \Delta S_t^\top \Delta S_t$$

is decomposed into eigenvalues and eigenvectors:

$$\Sigma e_i = \lambda_i e_i, \quad i = 1, 2, 3.$$

The first three eigenmodes e_1, e_2, e_3 capture over ninety percent of total variance and correspond to intuitive surface deformations:

Level, Skew, and Smile.

Their temporal amplitudes are obtained as:

$$PC_i^t = e_i^\top \Delta S_t, \quad i = 1, 2, 3,$$

and their daily changes,

$$\Delta PC_i^t = PC_i^t - PC_i^{t-1},$$

represent the momentum of those deformations.

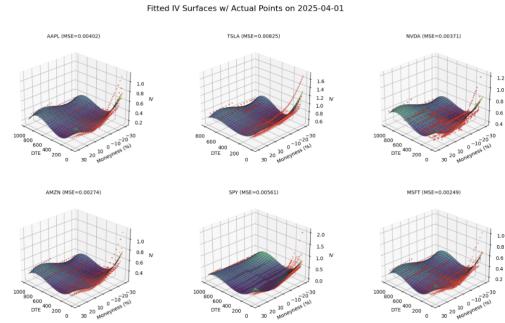


Fig. 12: Fit accuracy. Option-chain points (red) overlaid on the spline-smoothed IV surface for NVDA. Residual errors remain minimal except at extreme strikes.

To reinforce signal persistence, a short-term average of level change is also computed:

$$\text{RollingLevel}_t = \frac{1}{3}(\Delta PC_1^t + \Delta PC_1^{t-1} + \Delta PC_1^{t-2}),$$

along with the curvature adjustment:

$$\kappa_t = \left. \frac{\partial^2 IV_t^{\text{spline}}(m, 30)}{\partial m^2} \right|_{m=0} - \left. \frac{\partial^2 IV_{t-1}^{\text{spline}}(m, 30)}{\partial m^2} \right|_{m=0}.$$

Finally, we compute the **volatility-risk premium**:

$$430 \quad \text{VRP}_t = IV_{\text{ATM}}^t - RV_t, \quad RV_t = \sqrt{\frac{252}{k} \sum_{i=0}^{k-1} r_{t-i}^2},$$

as the difference between current at-the-money implied volatility and backward-looking realized volatility, interpreted as the market's compensation for volatility risk.

In summary, our feature construction is initially given by:

PCA Momentum (deltas):

$$\Delta PC_i^t = PC_i^t - PC_i^{t-1}, \quad i = 1, 2, 3\dots$$

Curvature Momentum (second derivative change in moneyness):

$$\kappa_t = \left. \frac{\partial^2 IV_t(m, \bar{d})}{\partial m^2} \right|_{m=0} - \left. \frac{\partial^2 IV_{t-1}(m, \bar{d})}{\partial m^2} \right|_{m=0}$$

or optionally averaged over moneyness values.

Volatility Risk Premium: Let realized volatility over the past k days be:

$$\text{RV}_t = \sqrt{\frac{252}{k} \sum_{i=0}^{k-1} r_{t-i}^2}, \quad r_t = \log(S_t) - \log(S_{t-1})$$

Then the **IV-RV spread** is:

$$\Delta\text{IV}-\text{RV}_t = \text{IV}_{\text{ATM}}^t - \text{RV}_t$$

This spread is often interpreted as the market's pricing of volatility risk (premium): a positive IV-RV spread implies options are expensive relative to realized uncertainty, creating potential opportunities for short-volatility trades.

4.3. Our Constructed Model - Prediction Targets

At this point, the experiment strongly resembles its foundational parent—the Q-equation—whereby parameters are weighted analytically to achieve optimality. Namely, we must test whether the feature vector:

$$[\Delta\text{PC}_1^t, \Delta\text{PC}_2^t, \Delta\text{PC}_3^t, \kappa_t, \Delta\text{IV}-\text{RV}_t]$$

is predictive of **future near-or-ATM IV movement** over a short horizon (e.g., 5 days):

4.3.1. Intuition

This hypothesis reflects the belief that **forward-looking market expectations**, encoded in the shape and dynamics of the IV surface, can serve as predictive signals for future volatility or option profitability.

4.4. Volatility Positioning and Payoff Mechanics

Having defined the features driving volatility momentum, we translate them into two complementary trade structures capturing opposite sides of the implied volatility premium: the *long volatility straddle* and the *short volatility strangle*. It is vital to examine these payoff mechanics as these positions form the mechanical vehicle of our trading system and embody the two distinct return regimes observed in options markets through our lens.

Long Volatility: The Straddle

A long volatility position is initiated by purchasing an at-the-money (ATM) call and put of the same expiration, forming a **straddle**. This position is long both gamma and vega¹, meaning it benefits from large realized price movements or from expansion in implied volatility after entry. In effect, we pay the premium upfront and gain convex exposure to the underlying's magnitude of motion:

$$\Pi_{\text{long vol}} = +\Gamma, +\nu, -\Theta.$$

Profit arises through two pathways: (i) a substantial directional move pushing one leg deep in-the-money, generating gamma gains, or (ii) a broad upward shift in implied volatilities lifting the entire straddle's value (vega gains). Time decay (Θ) acts as a continuous cost, hence positions are exited promptly once the volatility expansion or price move is realized.

Short Volatility: The Strangle

The short-volatility counterpart involves selling a pair of out-of-the-money options—one call and one put—roughly 5%–10% away from the current price, forming a **strangle**. This trade structure is short

¹ Option "Greeks" come from academia—they are the sensitivities of an option's price to changes in key market factors. Delta measures sensitivity to the underlying asset's price, Gamma to Delta's rate of change, Vega to volatility, Theta to time decay, and Rho to interest rates

gamma and vega but earns positive theta, profiting when volatility remains subdued and both options decay in value:

$$\Pi_{\text{short vol}} = -\Gamma, -\nu, +\Theta.$$

As long as realized volatility stays below the implied level sold, the collected premium can be retained. The position benefits further from gradual compression in implied volatilities, allowing the strangle to be bought back at a lower price.

For our algorithm, because such exposure can be vulnerable to large underlying moves, delta hedging is performed intraday, and positions are force-closed if losses exceed a pre-set drawdown threshold.

500

4.5. Algorithmic Execution and Signal Translation

In live implementation, these payoff mechanics are enacted systematically through QuantConnect's trading framework. The algorithm continuously monitors the PCA-derived signals and executes trades only when the volatility surface deformation exhibits a statistically significant directional bias.

When the model forecasts an upward jump in IV ($\Delta\text{Level}_t \geq 0$ or threshold), it enters a long-volatility position by purchasing the ATM straddle corresponding to the optimally configured DTE contract. Delta neutrality is enforced at entry, and positions are liquidated under any of the following: a +20% profit, a -14% loss, or 15 trading days elapsed—whichever occurs first. If instead the forecast indicates volatility compression ($\Delta\text{Level}_t < 0$), the strategy sells a 5% OTM strangle and applies the same profit, loss, and duration constraints. The resulting exposure is short vega and short gamma but positive theta, with ongoing delta hedges to maintain neutrality whenever $|\Delta| > 0.25$.

This symmetry allows the algorithm to dynamically toggle between volatility-buying and volatility-selling modes based on surface momentum, ensuring consistent participation across volatility regimes.

4.6. Empirical Assessment of our Models for Realized Volatility

To validate whether the extracted features genuinely anticipate realized volatility, we model next-day volatility as a nonlinear function of the PCA delta vectors and their rolling averages:

$$\text{RV}_{t+1} = f(\Delta\text{PC}_t, \Delta\text{PC}_{t-1}, \Delta\bar{\text{PC}}_t) + \varepsilon_t,$$

where $f(\cdot)$ is estimated using gradient-boosting, a popular machine learning technique for time-series forecasting. This specification captures both contemporaneous and lagged interactions among level, skew, and smile dynamics. Compared with simple linear regressions ($R^2 \approx 0.07$), nonlinear models achieved markedly higher explanatory power ($R^2 \approx 0.8$), confirming that short-term IV-surface deformations encode predictive structure that linear models fail to detect.

Linear Regression $R^2: 0.0713$
Gradient Boosting $R^2: 0.817$
Samples used: 280

Fig. 13: **Nonlinear Relationships Captured.** Gradient boosting models reveal curved interactions between PCA-based features and subsequent realized volatility.

The regression results identify two consistent drivers of short-term volatility evolution: (1) the *change in level* and its *three-day rolling average*, both highly significant (F-statistics of 189 and 116, respectively); and (2) the *smile* and *skew* components, which exhibit secondary influence. Previous values of these components showed little explanatory power, suggesting that volatility momentum is inherently short-lived.

510

520

530

540

540 4.7. Threshold Calibration and Optimization Landscape

Rather than relying on arbitrary cutoffs, we estimate empirical thresholds via randomized quantile search over the standardized feature distributions. This optimization balances signal strength against sample size to maximize the joint Sharpe ratio of all trades. The optimal quantiles typically fell between the 60th and 70th percentiles:

$$\theta_{\text{Level}} = 0.64, \quad \theta_{\text{RollLev}} = 0.67, \quad \theta_{\text{Skew}} = 0.62, \quad \theta_{\text{Smile}} = 0.54.$$

The resulting decision rule activates long-volatility trades when all four features exceed their respective positive thresholds and short-volatility trades when they fall below the negatives.

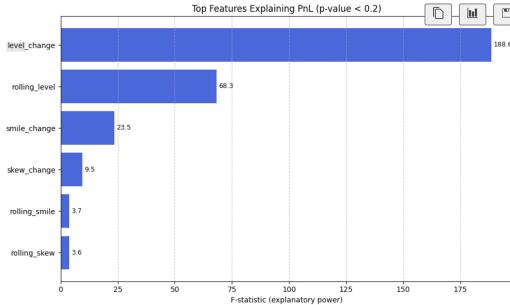


Fig. 14: **Feature Ranking.** Ranked importance of features based on F-scores in the volatility prediction model.

550 A projection of the parameter space onto the two most influential features—level change and skew change—yields a smooth *optimization landscape*, highlighting regions of consistent profitability versus those prone to overfitting. This visualization not only aids in interpreting model robustness but also establishes confidence intervals for live-trading threshold stability.

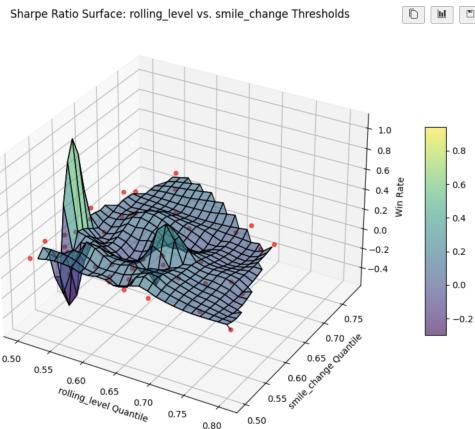


Fig. 15: Parameter Landscape w.r.t. Performance.

4.8. Risk Architecture and Execution Discipline

The resulting strategy operates under a three-tiered risk framework: (1) trade-level controls (stop-loss, take-profit, time limit); (2) portfolio-level exposure caps (maximum three concurrent positions, 0.5% capital per trade); and (3) systemic controls (delta-neutral enforcement, curvature filters, and liquidity screens). This design ensures that exposure remains tightly constrained even during volatility shocks, while the curvature filter ($|\kappa_t - \kappa_{t-1}| > 0.01$) suppresses signals in stagnant regimes.

In aggregate, these components constitute a coherent bridge between theory and implementation: a volatility-surface model that not only diagnoses the geometry of implied expectations but operationalizes it into a live, hedged trading system.

4.9. Live Trading

During one week of live trading, the algorithm executed six trades (three option structures). Cumulative P&L was near zero, consistent with expectations for such a short horizon: the model's profitability emerges only over extended cycles where volatility regimes can evolve. Comparative analysis confirms that longer evaluation periods (e.g., In-Sample and OOS A) yield smoother compounding and higher reliability than brief deployment tests.

4.10. Stress Testing

Stress scenarios illuminate the model's structural behavior under extreme volatility shifts. The COVID-19 period produced mixed outcomes: early gains were offset by subsequent reversals as volatility normalized faster than the model adapted. Event-driven shocks such as the 2025 tariff disruption resulted in minor losses and low PSR, underscoring the difficulty of capturing ultra-short-lived bursts of uncertainty. These findings suggest that while the model excels in sustained volatility regimes, additional mechanisms for high-frequency shock adaptation could further enhance resilience.

5. Discussion

The Volatility Surface Momentum framework provides an interpretable, data-driven approach to volatility trading. By decomposing daily IV-surface movements into orthogonal components—level, skew, and smile—it captures structured information about option-market sentiment and uses it to modulate exposure between long- and short-volatility regimes. Because positions are delta-neutral, the strategy's PnL originates purely from changes in volatility, not directional market drift.

Empirical quantile thresholds and curvature filters ensure that trades occur only when the entire surface signals a coherent move, reducing noise from isolated strikes or illiquid segments. This "broad consensus" requirement among components functions as a natural confirmation filter and improves entry precision.

5.1. Interpretation of Surface Dynamics

The PCA decomposition clarifies how surface deformations map to market psychology: a strong positive Δ Level typically denotes a generalized repricing of risk, a steepening Δ Skew reflects demand for downside protection, and a pronounced Δ Smile indicates renewed interest in tail hedges. Requiring agreement across these components effectively filters local distortions, ensuring trades respond only to genuine regime transitions rather than transient quoting anomalies.

5.2. Relation to Prior Work

Our empirical findings echo the volatility-momentum literature. Egloff et al. (2010) documented that dynamically switching between long and short variance exposure enhances capture of the variance-risk premium. Likewise, Martin & Gao (2021) emphasize the importance of timing the volatility carry trade—selling implied volatility only when its trend is declining. Our PCA-based method constitutes an option-level realization of those ideas, identifying when the entire surface—not merely ATM volatility—suggests compression or expansion.

6. Limitations

Despite encouraging results, several limitations deserve attention. First, the spline-fitted surface and PCA decomposition rely on smooth historical data; both can be sensitive to sparse or noisy option chains. Second, the backtest assumes frictionless hedging and execution at mid-quotes. In practice, transaction costs, bid-ask spreads, and hedge-adjustment frequency could materially reduce performance—especially for short-volatility trades requiring frequent rebalancing.

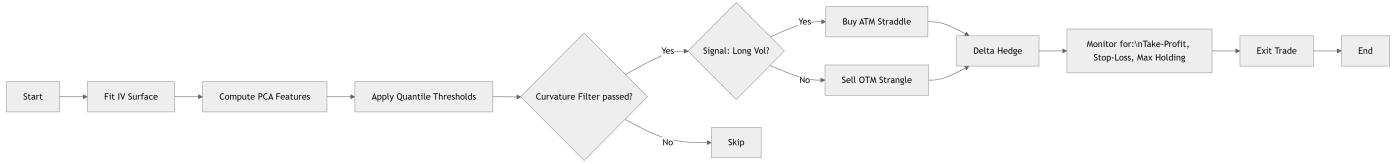


Fig. 16: **Signal-to-Execution Pipeline.** Overview of how volatility surface dynamics propagate through feature extraction, signal generation, and risk-managed trade execution.

Period	Time	Ann. Ret.	Sharpe	DD	WR	PSR
In-Sample	4y	22.74%	1.23	39.4%	56%	3.83%
OOS A	6m	(3.35)%	(0.71)	17.5%	46%	17.9%
OOS B	1y	130.27%	1.87	9.8%	43%	64.39%
OOS C	1y	24.82%	1.17	16.6%	46%	50.44%
Stress: COVID	3m	(0.41)%	(0.17)	17.3%	48%	26.37%
Stress: 2008	8m	1.23%	0.74	4.21%	50%	44.47%
Live Trading	1w	1.0%	—	0.1%	—	49.2%

Table 1: Performance metrics across test periods. Ann. Ret. = Annualized Return, DD = Maximum Drawdown, WR = Win Rate, PSR = Probabilistic Sharpe Ratio.



Fig. 17: **In-Sample Backtest.** Equity curve and cumulative return over four-year calibration period.

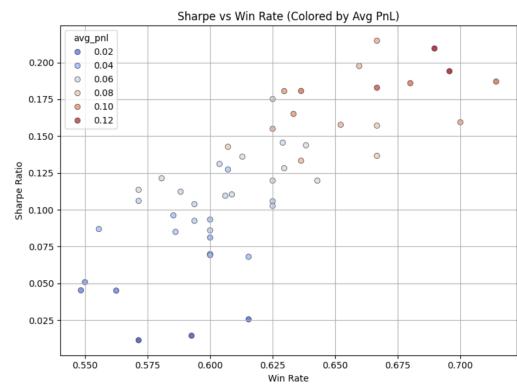


Fig. 19: **Trade Efficiency Metrics.** Empirical relationship between win rate and Sharpe ratio under varying quantile thresholds.



Fig. 18: **Out-of-Sample Backtests.** Representative equity curves under distinct market regimes.

Third, using fixed quantile thresholds across the full horizon introduces potential regime bias. Adaptive calibration tied to macro-volatility indicators may improve responsiveness to structural shifts.

Practical implementation also requires robust data management and margin oversight. Option liquidity varies by strike and maturity; while



(a) COVID



(b) 2008 Crash

Fig. 20: **Stress-Period Backtests.** Strategy response under abrupt volatility spikes and regime transitions.

ATM straddles are typically deep, OTM strangles may suffer from wider spreads. Moreover, short-option positions demand margin allocation and tight tail-risk control. Future work could address these concerns through margin-adjusted position sizing, adaptive hedging frequency, and cross-asset volatility hedges.

7. Conclusion and Further Work

This study presents a systematic framework for trading volatility through the geometric evolution of the implied-volatility surface. By combining cubic-spline fitting with PCA-based feature extraction, the model detects structural momentum in volatility level, skew, and cur-

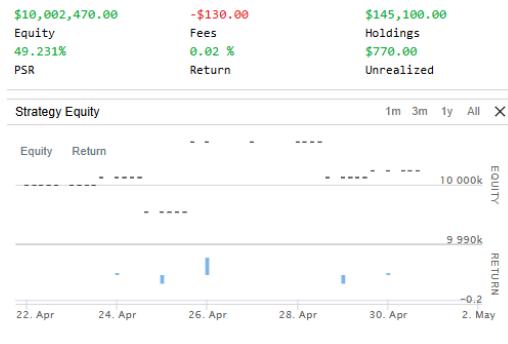


Fig. 21: Live-Trading Snapshot. Short-term performance aligns with expectation of near-flat results over limited data.

640 vature, translating it into delta-neutral option positions. The approach delivers interpretable, regime-adaptive exposure and demonstrates consistent profitability across diverse backtesting horizons.

Future research should explore dynamic thresholding and alternative dimensionality-reduction techniques—such as autoencoders or dynamic factor models—to capture non-linear or time-varying surface effects. Incorporating realized-volatility forecasts or order-flow information may further refine entry timing. Most importantly, rigorous live testing under realistic transaction-cost and margin assumptions will be essential to assess scalability and robustness.

650 In essence, the Volatility Surface Momentum framework bridges statistical signal extraction and practical options trading, illustrating how quantitative structure—when disciplined by risk and interpretability—can convert latent information embedded in volatility surfaces into actionable trading intelligence.

References

- Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1), 101–143. Available at: <http://www.jstor.org/stable/1262727>
- 660 Black, F., & Litterman, R. (1991). Global asset allocation with equities, bonds, and currencies. Fixed Income Research, Goldman Sachs. Available at: https://people.duke.edu/~charvey/Teaching/BA453_2006/Black_Litterman_GAA_1991.pdf
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Bliss, R. R., & Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The Journal of Finance*, 59(1), 407–446. Available at: <https://pages.stern.nyu.edu/~dbackus/Disasters/BlissPanigirtzoglou%20JF%2004.pdf>
- 670 Clewlow, L., & Skiadopoulos, G. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1), 45–60. Available at: https://www.academia.edu/46962528/Dynamics_of_implied_volatility_surfaces
- Cont, R., & Da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1), 45–60.
- Coval, J. D., & Shumway, T. (2001). Expected option returns. *The Journal of Finance*, 56(3), 983–1009.
- Derman, E., & Kani, I. (1994). Riding on a smile. *Risk*, 7(1), 32–39.
- Dupire, B. (1993). Pricing and hedging with smiles. Working paper. Available at: https://globalriskguard.com/resources/deriv/pric_hedg_with_smile.pdf
- Egloff, D., Leippold, M., & Wu, L. (2010). The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, 45(5), 1279–1310. Available at: <http://www.jstor.org/stable/27919564>
- Fengler, M. R. (2005). *Semiparametric Modeling of Implied Volatility*. Springer Finance. Berlin: Springer-Verlag.
- Hull, J. C. (2018). *Options, Futures, and Other Derivatives* (10th ed.). Upper Saddle River, NJ: Pearson Education.
- 690 Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91.
- Martin, I. W. R., & Gao, C. (2021). Volatility, valuation ratios, and bubbles: An empirical measure of market sentiment. *SSRN Electronic Journal*.
- Maylath, A. (2020). Understanding volatility performance with PCA. *Quantitative Insights*, Medium. Available at: <https://medium.com/quantitative-insights/understanding-volatility-performance-with-pca-b89f797f3237>
- Rustamov, O., Aliyev, F., Ajayi, R., & Suleymanov, E. (2024). A new approach to build a successful straddle strategy: The analytical option navigator. *Risks*, 12(7), 113.
- Sachs, G., Derman, E., Kani, I., & Zou, J. (1995). The local volatility surface: Unlocking the information in index option prices. Quantitative Strategies Research Notes, Goldman Sachs. Available at: https://emanuelderman.com/wp-content/uploads/1996/06/gs-local_volatility_surface.pdf
- Skiadopoulos, G., Hodges, S., & Clewlow, L. (2000). The dynamics of the S&P 500 implied volatility surface. *Review of Derivatives Research*, 3(3), 263–282.
- Sylla, A., & Villa, C. (2000). Measuring implied volatility surface risk using principal components analysis. *Lecture Notes in Statistics*, 157, 131–148.