

IPCA On Implied Volatility Surface in Predicting Short-Term Option Prices

Authors: Trent Potter, Frank Wang, Helen Du

Institution: Booth School of Business

Class: Machine Learning In Finance

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.

Predicting the movement of implied volatility for a given option contract is of particular importance for constructing profitable delta-hedged strategies. Since a delta-hedged option position is largely insulated from movements in the underlying asset, its profit and loss is primarily driven by changes in implied volatility—commonly referred to as “vega risk.” Accurate forecasts of IV dynamics thus enable traders to systematically capture volatility risk premia and exploit mispricings in the options market, forming the basis for sophisticated volatility arbitrage and relative value strategies.

There are several challenges to overcome when predicting implied volatility for a given asset class:

- **High dimensionality:** Individual equities can have thousands of traded options across numerous expirations and strikes, creating co-linear 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.
- **Endogeneous, non-closed-form estimation:** Computational methods for backing out even simple Black-Scholes IV don’t have closed form solutions. More advanced stochastic and local volatility models are even more challenging to run computationally.

We compare two simple linear approaches to predicting next-day implied volatility changes for AAPL options:

- **PCA with Penalized Linear Regression:** Intuitive decomposition of the surface provides interpretable components, but little OOS R^2 is achieved.
- **Instrumented Principal Components Analysis:** More robust R^2 in next day prediction, but more complex and nuanced interpretation.

IPCA Predictive Results

We evaluated IPCA using different numbers of factors (plus a constant) and report both in-sample (IS) and out-of-sample (OOS) performance:

Factors (+ Constant)	IS R^2	IS RMSE	OOS R^2	OOS RMSE
1	0.6492	0.0089	0.2939	0.0119
2	0.7910	0.0069	-0.1199	0.0150
3	0.8830	0.0052	-0.1474	0.0152

The best out-of-sample performance is achieved with a single factor plus a constant, yielding an OOS R^2 of 0.2939 and OOS RMSE of 0.0119. Adding more factors improves in-sample fit but leads to overfitting and worse OOS results.

The Dimensionality Challenge and Data Representation

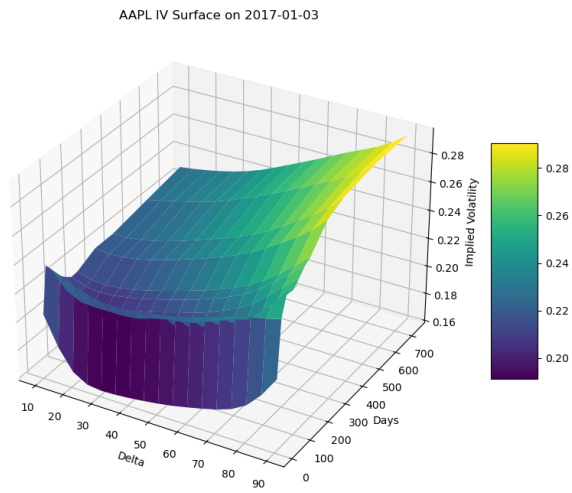
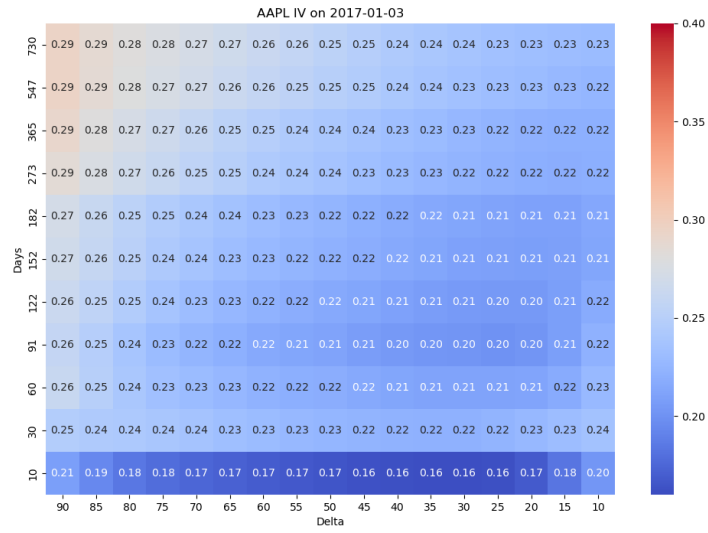
A critical step in modeling IV surfaces is converting observed market data (options listed by discrete strikes and expirations) into a structured, less-sparse representation. First translating for strikes and expiration dates to Δ and TTM makes observations comparable across different time periods. We 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:

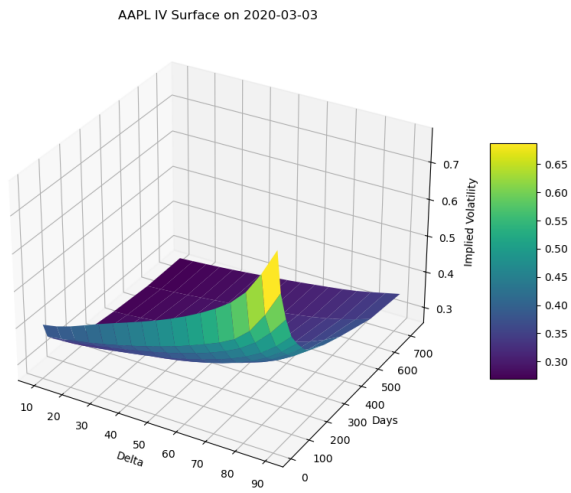
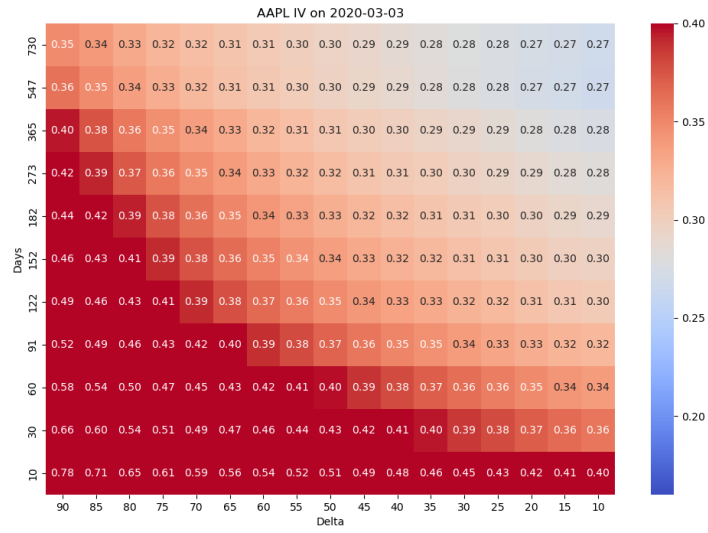
- **Functional Representation:** Model IV as a continuous function over $(\text{delta}, \text{TTM})$. Estimation of this \mathbb{R}^2 valued function.
- **Grid-Based Sampling:** Generate a discrete sampling grid and populate points through interpolation or kernel-weighted averaging. Yields indexing over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

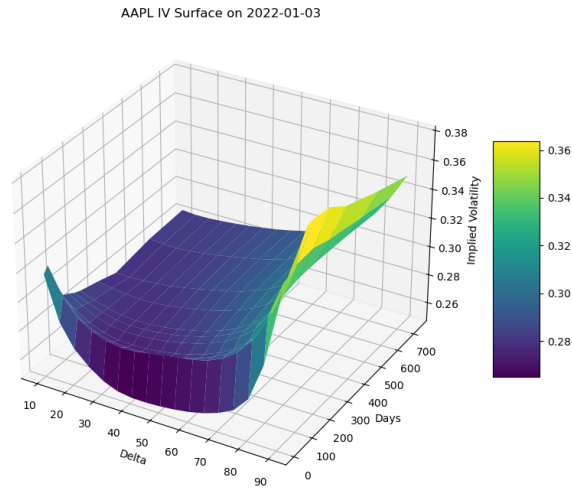
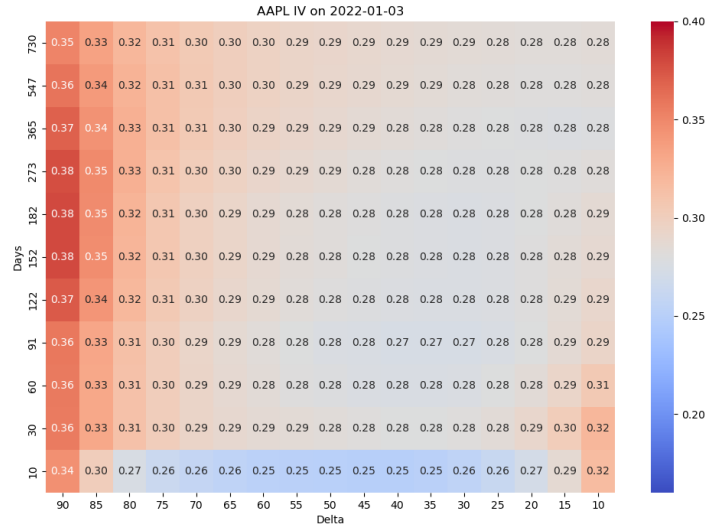
Given practical constraints, this report adopts the latter method, constructing a standardized, interpolated grid of IV observations across delta and TTM. While this discretized representation simplifies covariance calculation, we recognize it involves potential interpolation errors and sacrifices structural 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 at various times are visualized below:







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
- σ : 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 σ , 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.

The OptionMetrics dataset conveniently provides precomputed implied volatility (IV) surface values for each option contract, eliminating the need for manual inversion of the pricing model. Specifically, OptionMetrics employs an inverse Cox, Ross, Rubinstein (CRR) binomial tree search algorithm, which incorporates expected dividend payments to accurately reflect the early exercise premium present in American-style options. This approach ensures that the reported IVs are consistent with observed market prices and the actual contract specifications, allowing us to work directly with a robust, market-consistent IV surface as the foundation for our analysis.

Vega Weighting the Surface

To appropriately weight residuals in price-space, we scale the predicted changes in implied volatility by the Black-Scholes Vega corresponding to each option. This approach effectively translates IV forecasts into their monetary impact on option prices, mirroring the sensitivity of a delta- and rho-hedged options portfolio to volatility fluctuations. While this methodology abstracts away from the practical frictions and transaction costs inherent in continuous hedging, it provides a theoretically sound framework for evaluating predictive performance in a manner that is economically meaningful for risk management and trading applications.

While the Black-Scholes framework assumes European-style options, most exchange-traded equity options in the U.S. are American-style, allowing for early exercise. However, for deep in-the-money calls and puts, American options can exhibit price deviations from their European counterparts due to the early exercise feature—particularly for puts (cash premium) and calls (dividend exceeds extrinsic value), where early exercise may be optimal. In contrast, for out-of-the-money options (the primary focus of this analysis), the early exercise premium is negligible, and American and European prices converge. Therefore, throughout this report, we use European vega calculations as a practical and sufficiently accurate approximation for out-of-the-money options, acknowledging that any pricing discrepancies due to early exercise are minimal in this regime.

Prediction with PCA

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.

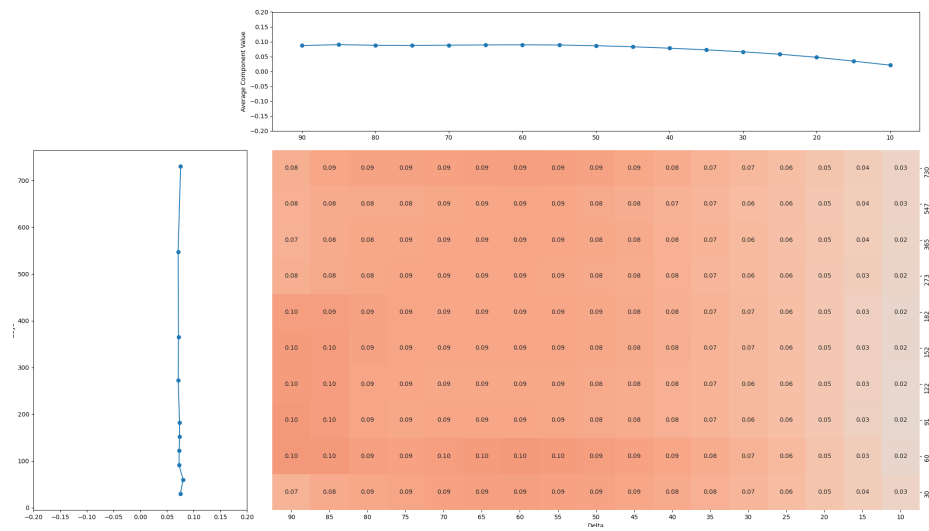


Figure 1: PC 1 - Level Factor

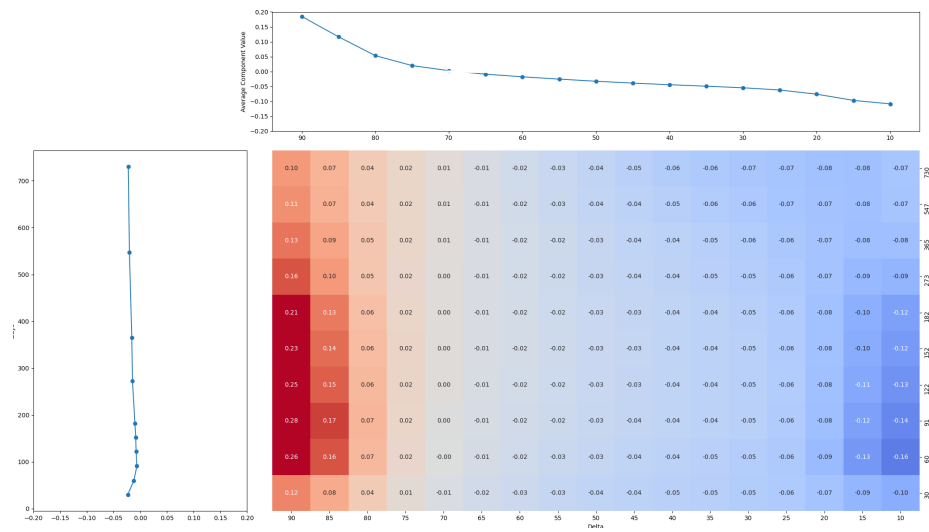


Figure 2: PC 2 - Skew Factor

PC 3 (Term Structure Slope Factor)

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

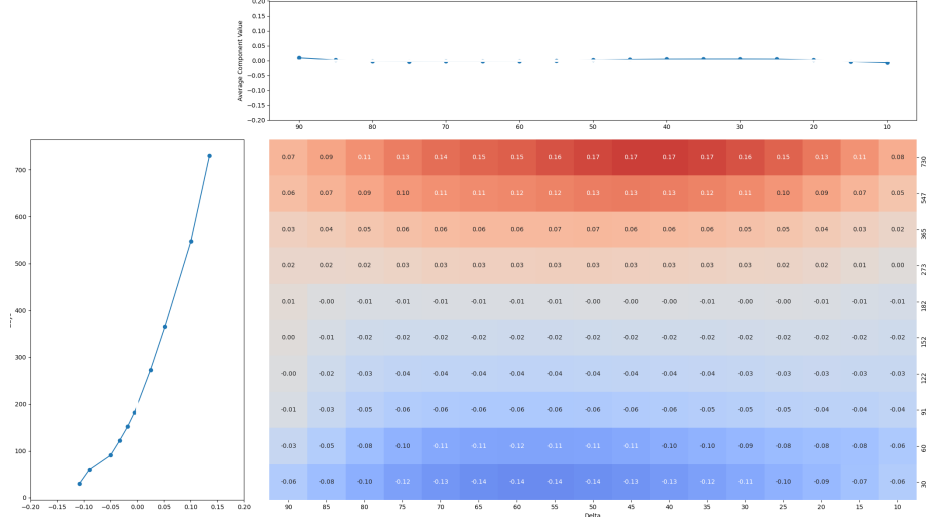


Figure 3: PC 3 - Term Structure Slope Factor

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.

PCA Results

Applying PCA followed by ridge regression to predict next-day changes in the AAPL implied volatility surface yielded modest but positive results. The optimal configuration used five principal components and a ridge penalty of 1.0, achieving an out-of-sample (OOS) R^2 of 0.0249 and an OOS RMSE of 0.0138. While the in-sample (IS) R^2 was slightly higher at 0.0416, the overall R^2 remained low, indicating that the linear model captures only a small fraction of the predictable variation in IV changes. Nevertheless, the positive OOS R^2 suggests that the model does extract some signal from the principal components, particularly those corresponding to broad level shifts and skew in the IV surface. The results highlight both the challenge of forecasting short-term IV dynamics and the value

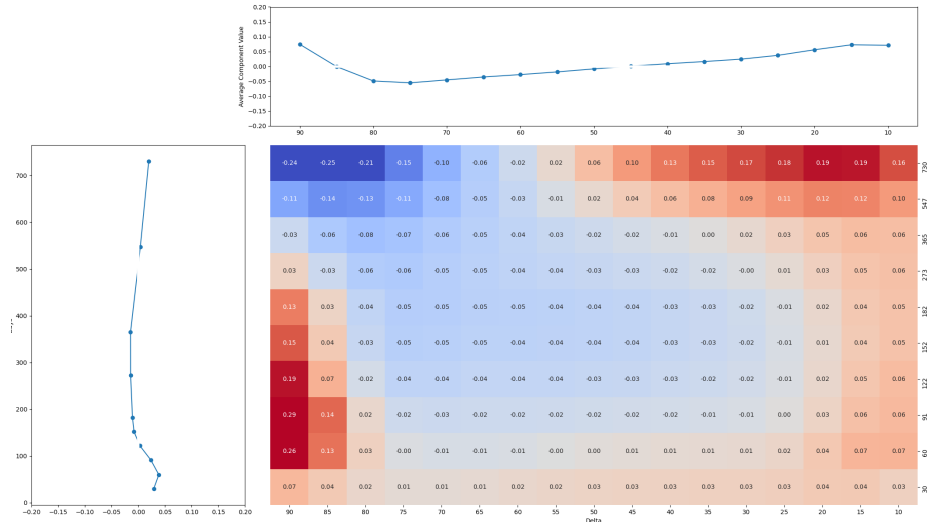


Figure 4: PC 4 - Decoupled Across Term Structure Factor

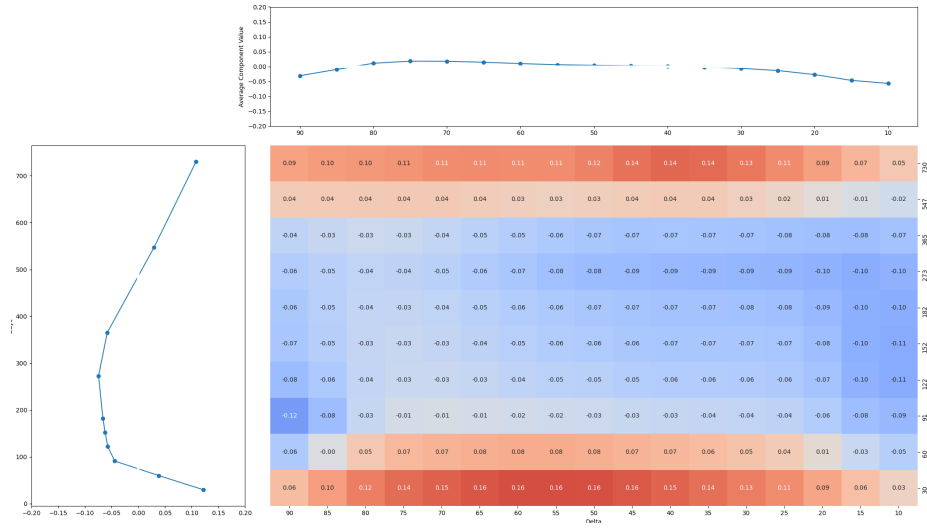


Figure 5: PC 5 - Term Structure Curvature Factor

of dimensionality reduction for stabilizing predictions, even if the incremental predictive power is limited in this setting.

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

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

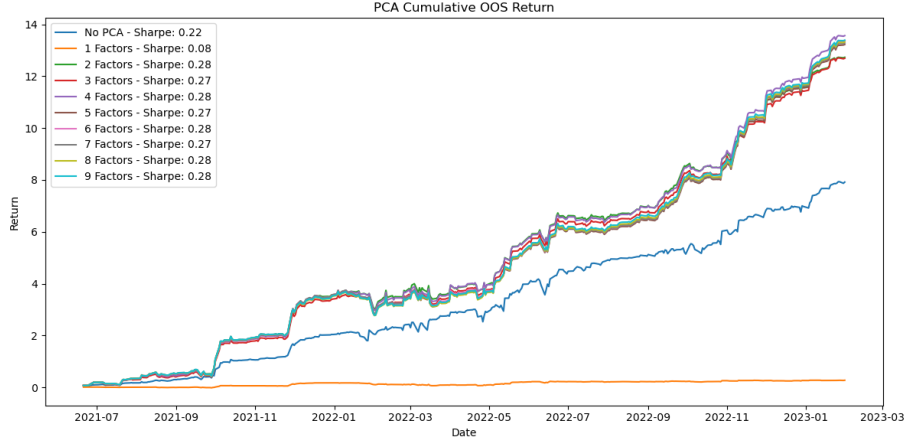


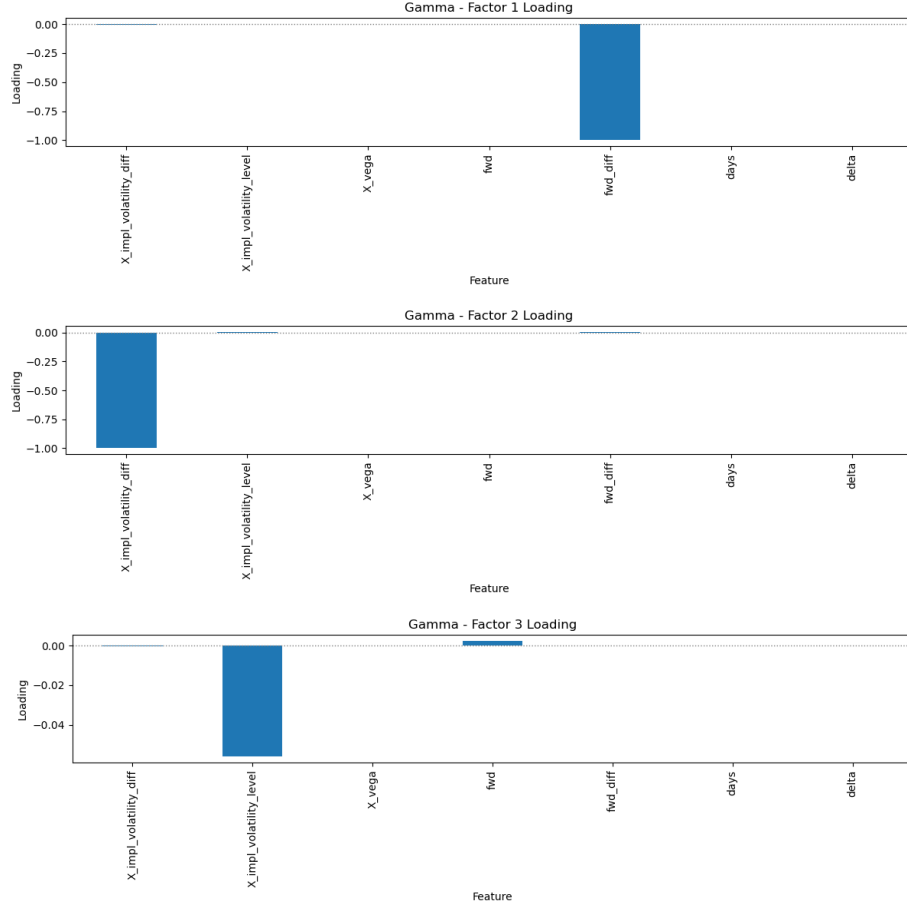
Figure 6: PCA Out-of-Sample Return

IPCA Results

Interpretation of the Gamma Loading Matrix

The Gamma loading matrix in IPCA represents how each point on the implied volatility surface loads onto the latent factors. Unlike standard PCA, where loadings are static, IPCA allows these loadings to be instrumented by observable characteristics, such as delta, vega, time-to-maturity, IV changes, and more. This results in a more flexible and interpretable mapping between the IV surface and the underlying factors. In our analysis, the most information rich component is the 1-day implied volatility diff with a negative weight suggesting a mean-reversion factor. In the 2-factor + constant model, our 2nd factor consists of the log-forward difference, implying negative correlation between IV changes and spot changes.

The following figures illustrate the Gamma loading matrix for the first three IPCA factors:



Time-Varying Factors

A key advantage of IPCA is its ability to estimate factors that evolve dynamically over time, conditioned on observable characteristics. The extracted factors display clear temporal patterns, often aligning with major market events and volatility regimes.

Prediction Methodology and Results

To evaluate predictive performance, we regress next-day changes in the IV surface on the time-varying IPCA factors, using an expanding window to estimate parameters with a mean factor. The model is trained on a subset of the data and tested out-of-sample, with performance measured by R^2 and RMSE. The IPCA approach achieves its best out-of-sample performance with a single factor plus a constant, yielding an OOS R^2 of 0.2939 and OOS RMSE of 0.0119. Adding more factors improves in-sample fit (up to IS R^2 of 0.8830 with three factors),

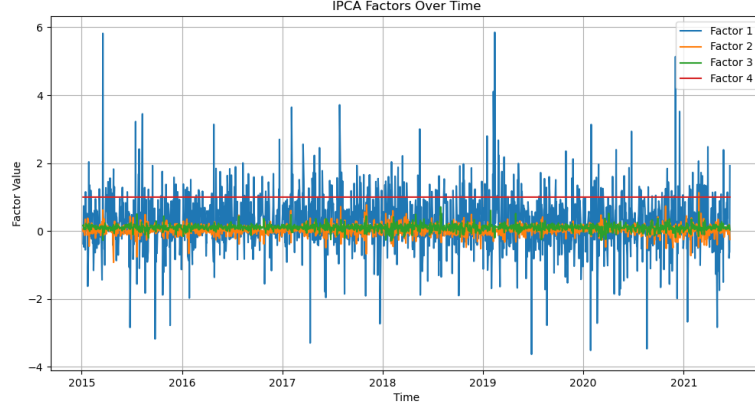


Figure 7: IPCA Time-Varying Factors

but leads to overfitting and negative OOS R^2 . These results demonstrate that IPCA's flexible factor structure and dynamic loadings provide a meaningful edge in forecasting short-term IV changes, supporting its use in volatility trading and risk management applications.

A representative prediction for a given IV change might have the following time-series characteristics:

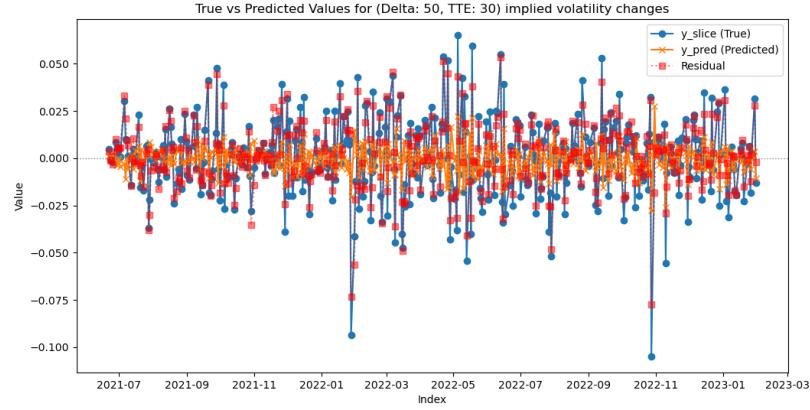


Figure 8: IPCA Out-of-Sample Prediction (50 train, 30 test)

And the following delta by maturity structure:

The average IV “return”, defined as $IV_{pred} * IV_{actual}$ had the following

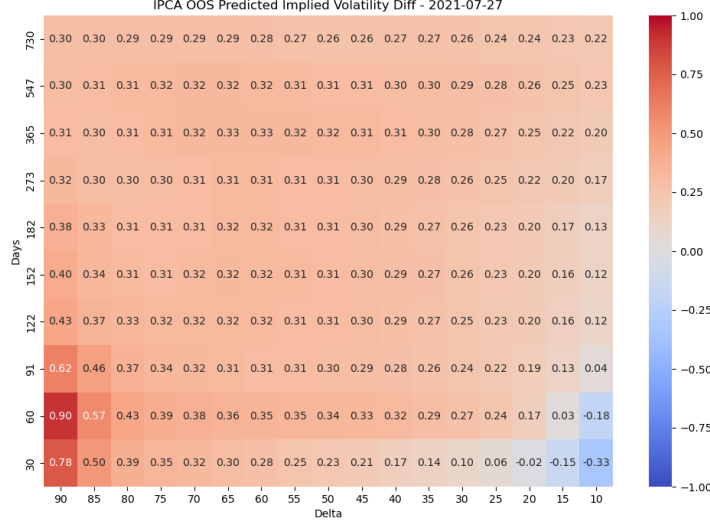


Figure 9: IPCA Out-of-Sample Prediction (2021-07-27)

mean, std, and Sharpe Ratio:

Next steps

Looking forward, several promising directions can further enhance the predictive modeling and practical utility of implied volatility surface analysis:

1. Inclusion of Multiple Asset Classes:

Expanding the analysis to incorporate a broader universe of asset classes—including index options (e.g., SPY), volatility derivatives (e.g., VIX), and a diverse set of single-name equities—would allow for a more comprehensive understanding of volatility dynamics across markets. Cross-asset modeling can reveal commonalities and idiosyncrasies in volatility risk premia, as well as enable the construction of relative value strategies that exploit mispricings between related instruments.

2. Functional Surface Modeling:

Current grid-based approaches treat each point on the IV surface independently, without imposing any functional structure or smoothness across delta or expiry dimensions. Future work could leverage methodologies that explicitly model the IV surface as a continuous function—such as spline-based methods, Gaussian processes, or other functional data analysis techniques. These approaches can better capture the inherent relationships and smoothness in the surface, improving both interpretability and predictive accuracy.

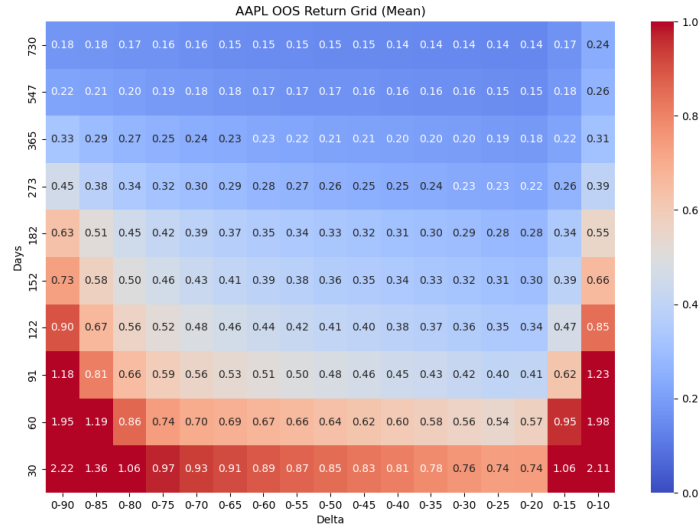


Figure 10: IPCA Out-of-Sample Return Grid

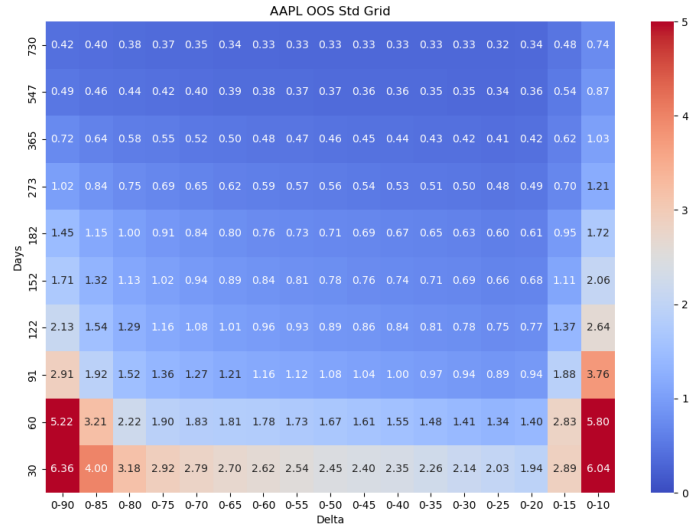


Figure 11: IPCA Out-of-Sample Standard Deviation Grid

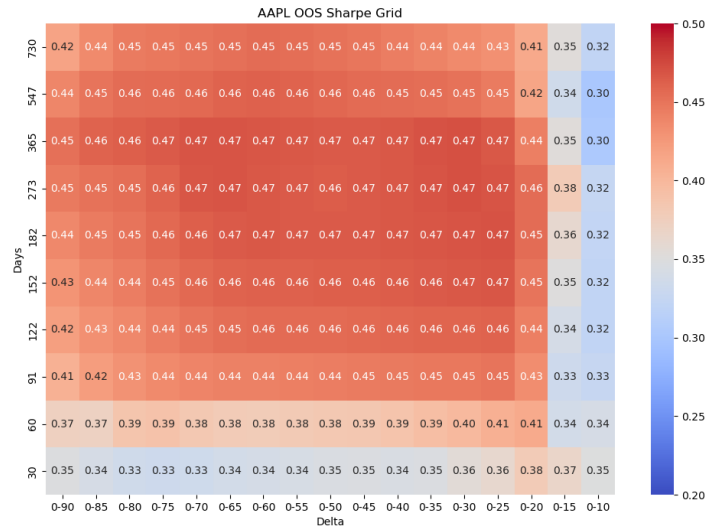


Figure 12: IPCA Out-of-Sample Sharpe Ratio Grid

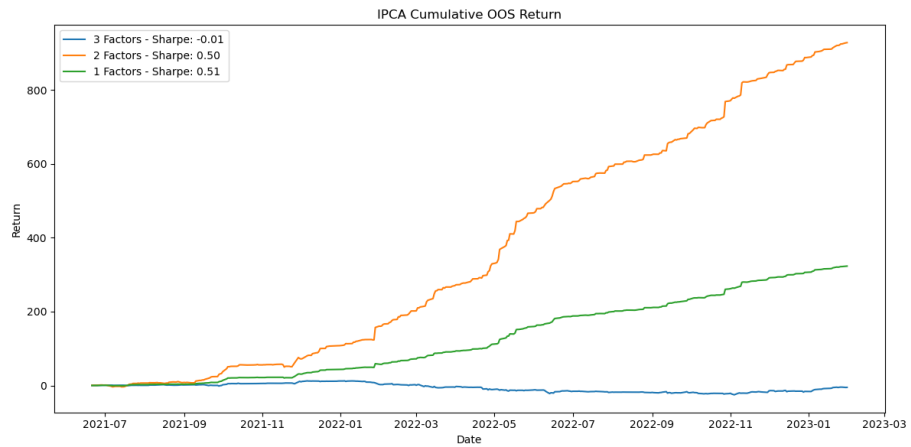


Figure 13: IPCA Cumulative Out-of-Sample Return

3. Nonlinear Factor Models and Auto-Encoders:

While linear factor models like PCA and IPCA provide interpretability and computational tractability, they may fail to capture complex, nonlinear dependencies present in the data. Incorporating nonlinear dimensionality reduction techniques—such as auto-encoders or variational auto-encoders—can uncover richer latent structures in the IV surface. These methods can be further integrated with the IPCA framework, allowing for instrumented, nonlinear factor extraction that adapts to observable characteristics and market regimes.

4. Direct Price-Space Analysis:

A key limitation of IV-based analysis is that it assumes a static mapping between delta-moneyness and option contracts, whereas in practice, the set of traded contracts migrates across the surface as market conditions evolve. This can lead to slippage between predicted and realized P&L when trading on IV signals. Future research should focus on modeling and predicting option price changes directly, accounting for the dynamic rebalancing required to maintain exposure at fixed deltas or maturities. Such an approach would provide a more realistic assessment of strategy performance and risk in live trading environments.

By pursuing these directions, future work can address both methodological and practical challenges, ultimately leading to more robust, generalizable, and actionable models for volatility prediction and trading.