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

Group 8: Trenton Potter, Helen Du, Ruilin Wang

# Catalogs

# Motivations

1. Understand what drives implied volatility surface changes
2. Find a latent factor structure
3. Understand how the factors evolve over time
4. How are these factors determined other market inputs
5. Can we use this factor structure to better predict option price

# Options Overview

## Definition

- Financial contract giving the **right, not obligation** to buy or sell an asset at a set price by a certain date
- 2 types of Contracts: **Call and Put**
  - Call: **Buy** the underlying asset at the **strike price**
    - Intrinsic Value=max(S-K, 0)
  - Put: **Sell** the underlying asset at the **strike price**
    - Intrinsic Value: max(K-S, 0)
- 2 types of Options: European and American Options

## Option Greeks

- **Delta**: Price sensitivity to underlying price changes
  - Call Delta: 0 to 1; Put Delta: -1 to 0
- **Gamma**: Delta sensitivity
  - High Gamma → Delta changes quickly → unstable hedge
- **Theta**: Time decay
- **Vega**: Implied Volatility impact
- **Rho**: Interest rate impact
  - Impact is more relevant for long-dated options

## Terms

- Key Terms:
  - **Premium**: Cost to purchase the option
  - **Strike Price**: Agreed price to buy/sell the asset
  - **Expiration Date**: Last valid date to exercise the option
  - **ITM (In the Money)**: Option has intrinsic value
  - **At-the-Money (ATM)**: strike ≈ spot price
  - **OTM (Out of the Money)**: No intrinsic value
  - **Implied Volatility:** Market's forecast of future volatility, derived from option prices

## Models

- Option Pricing Models:
  - Black Scholes Model
  - Binomial Tree
  - Monte Carlo Simulation
  - Finite Difference Method
  - Local & Stochastic Volatility Models
  - Analytical Approximations (Barone-Adesi & Whaley)

# Option Pricing: Black Scholes Model Overview

## Overview

- Developed by **Black, Scholes, and Merton** in 1973
- Calculates the Fair Value of European Call and Put Option
- Based on the idea of no-arbitrage and replicating portfolio
- Uses a risk-neutral world: Investors are indifferent to risk
- Foundation for modern option pricings
- Basis for calculating Implied Volatility

## Formula

$$C = S_0 * N(d_1) - K * e^{-r*T} * N(d_2)$$

$$d_1 = \frac{\left[\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) * T\right]}{\sigma * \sqrt{T}}, \qquad d_2 = d_1 - \sigma * \sqrt{T}$$

## Assumptions

- Lognormal distribution of asset prices - not negative
- Constant volatility of the underlying asset
- Constant risk-free interest rate
- No dividends during the life of option
- No Transaction cost or taxes

## Variables

- $S_0$ = Current Stock Price
- $K$ = Strike Price
- T = Time to expiration
- $r$ = Risk-free rate
- $\sigma$ = volatility of returns
- N(·) = Cumulative Standard Normal Distribution

# Option Price Depends on Implied Volatility

In pricing models **like Black-Scholes**, most inputs— Stock Price, Strike Price, Time, Interest rate—are **known** and fixed. The only unknown—and most sensitive input—is **IV**. A small change in IV can lead to a **large change in option price**, especially for ATM options (high Vega). Therefore, **IV is often the dominant factor in determining short-term price fluctuations** in options.

## When IV Matters MOST

- **Near events** (earnings, Fed decisions)
- **Short-term options**: IV shifts change premium dramatically
- **Volatility trading**: entire strategy is IV-centric (e.g., straddles)

## When IV Matters LESS

- **Long-Deep ITM/OTM options**: lower Vega → less IV impact
- **Short dated options**: Theta and Gamma dominate
- **Unhedged strategies**: price moves may dominate over IV shifts

**Challenges**

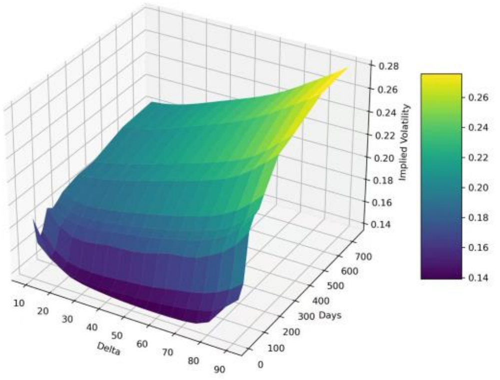
- Not Observable (Changes Depend on Supply/demand, events, and model inputs)
- No Single ”True” IV (smile/skew)
- Need Smooth and Predictable Surface

**Prediction**

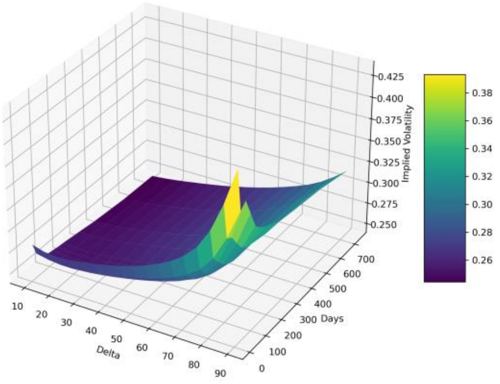
- Detect under/overpriced options
- Execute volatility arbitrage, Design IV-driven strategies (e.g., trading vega)
- Improve **hedging precision**

# Option Markets Are High-Dimensional

APPL: 2017 February 3rd Option Surface




APPL: 2020 February 3rd Option Surface




**Market Frictions & Noise**

- IV quotes **may be endogenous** or noisy due to market microstructure risk, demand-supply imbalance in certain strikes, etc.
- Option chain spreads not symmetric

**Dimensionality & Structure**

- Nearby IV points are naturally correlated
  - Creates **multicollinearity** -> Ill-conditioned matrices
- Covariance Patterns change during events
- IV Surface = Strike * Expiration days * IV

**Solutions & Tools**

- **Principal Component Analysis / IPCA**
  - Reduce dimensions
- Instrumented PCA
  - Extracts exogenous latent factors
  - Mapping to time-varying Latent factors

# Decomposition: PCA vs. Incremental PCA

| PCA | IPCA |
|---|---|
| 1 Large Table (e.g. 5x100) in one Batch | 1 Wanting to estimate r = C G F + e given r & c |
| 2 5D Plot transforms into 5x5 Covariance Matrix | 2 Guess G, then regress for F. |
| 3 Calculate Eigenvalues and Eigenvectors | 3 With new guess for F, regress better G |
| 4 Order Eigenstuff from Big to Small | 4 Repeat until F & G converge to stable answers |
| 5 Up to us picking how many eigenvectors and go back to plot those eigenvectors | 5 Use F & G to predict r given inputs C |

# PCA Implementation

## STEP 1: Data Extracting, Cleaning, Wrangling

- Options from OptionMetrics, it have Expiration Days, dates, delta, implied volatility, ticker, put, and call; 3-year of data 2015 - 2018
  - Extracted tickers (e.g. Apple) options, groupedby date, delta, and days.
  - Then we find the difference of the next day IV under same delta and expiration dates
- Forward price from Option Metrics

**PCA uses matrix format:**

| Date | (30, 10) | (30, 25) | ... | (730, 90) |
|---|---|---|---|---|
| 2023-01-02 | 0.01 | -0.02 | ... | 0.03 |
| 2023-01-03 | 0.02 | -0.01 | ... | 0.02 |

## STEP 2: PCA Algorithm

- X is lagged features – Lagged IV (t-1 and t-2 diff), Paste IV Level, and forward price of a stock
- Y is difference of the n and n+1 IV under same delta and expiration dates
- Train and Test split 80/20, without shuffling due to time-series characteristics. Time-series CV, spliting into 5 expanding folds
- Pricipal Components (from none to 10), Ridge Regression alpha (0, 10, 20)

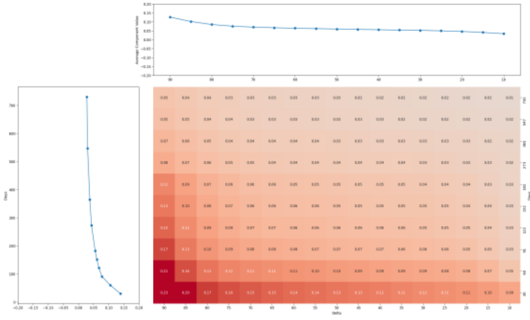
## STEP 3: PCA Results

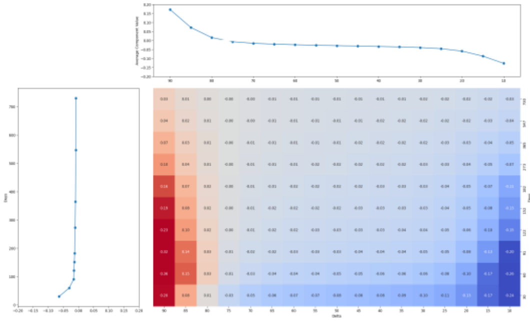
- Compute Vega-weighted error
- OOS R^2 and RMSE
- Pick the hyperparameter set with the lowest average error

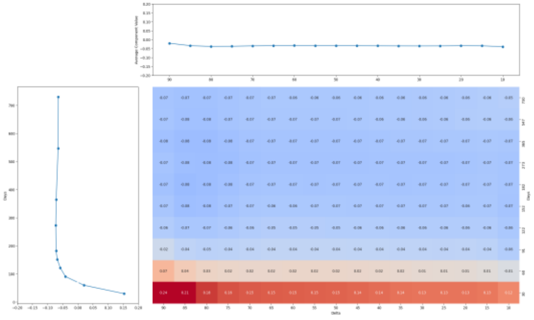
# Prediction Results: PCA GRAPH

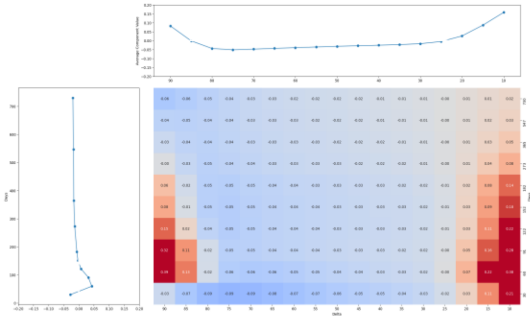
CHICAGO BOOTH
The University of Chicago Booth School of Business

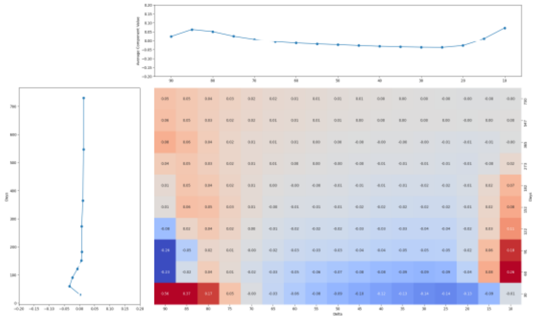
**Best results found with no dimension reduction when running a penalized multi-variate regression.**











# IPCA Implementation

## STEP 1: Data Extracting, Cleaning, Wrangling

- Data is from OptionMetrics too
- But presented in different way than PCA

📌 X: shape = (100 × 50, 7)

| Date | (Days, Δ) | IV_level | Vega | Log Fwd | Log Fwd Δ | Days | Delta |
|---|---|---|---|---|---|---|---|
| 2023-01-02 | (30, 10) | 0.21 | 11.5 | 4.51 | 0.01 | 30 | 10 |
| 2023-01-02 | (60, 25) | 0.19 | 13.0 | 4.51 | 0.01 | 60 | 25 |
| 2023-01-03 | (30, 10) | 0.22 | 12.0 | 4.52 | 0.01 | 30 | 10 |

📌 Y: shape = (100 × 50, 1)

| Date | (Days, Δ) | ΔIV |
|---|---|---|
| 2023-01-02 | (30, 10) | +0.01 |
| 2023-01-02 | (60, 25) | -0.02 |
| 2023-01-03 | (30, 10) | +0.02 |
| ... | ... | ... |

## STEP 2: IPCA Algorithm

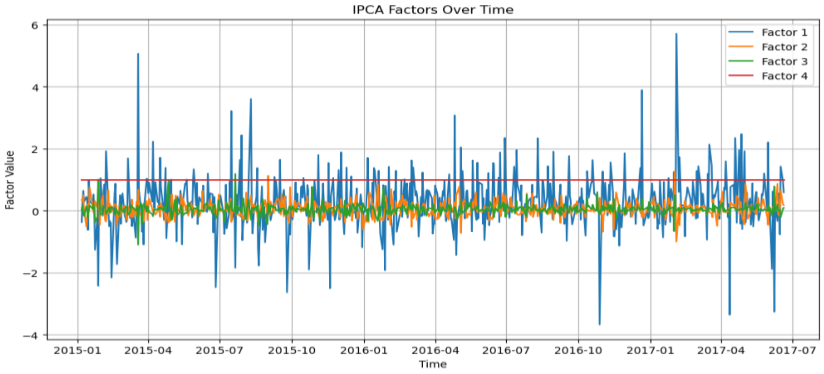
- X is lagged features – Lagged IV (t-1 and t-2 diff), Vega, and forward price of a stock
- Y is difference of the n and n+1 IV under same delta and expiration dates
- Use alternating least squares to solve for the factors and the loadings

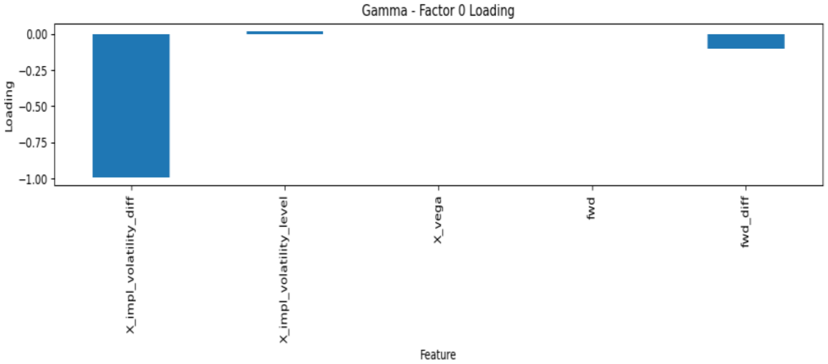
## STEP 3: IPCA Results

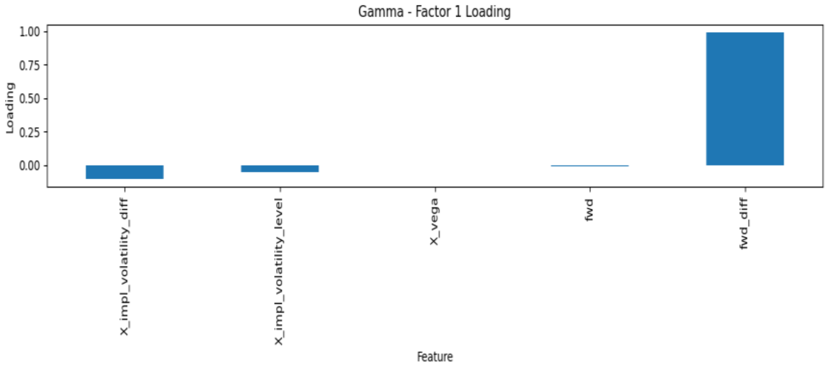
- Compute Vega-weighted error
- OOS $R^2$ and RMSE
- Pick the hyperparameter set with the lowest average error

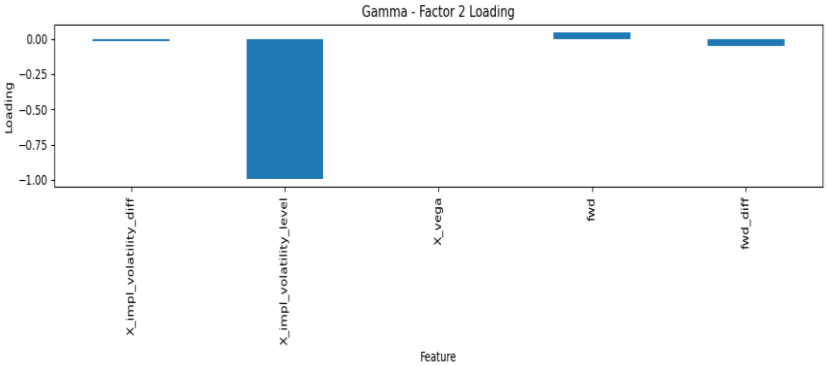
# Prediction Results: IPCA GRAPH

**First loading heavily on day-on-day implied volatility change, second on forward change, third on IV level. No clear ordering unlike PCA.**



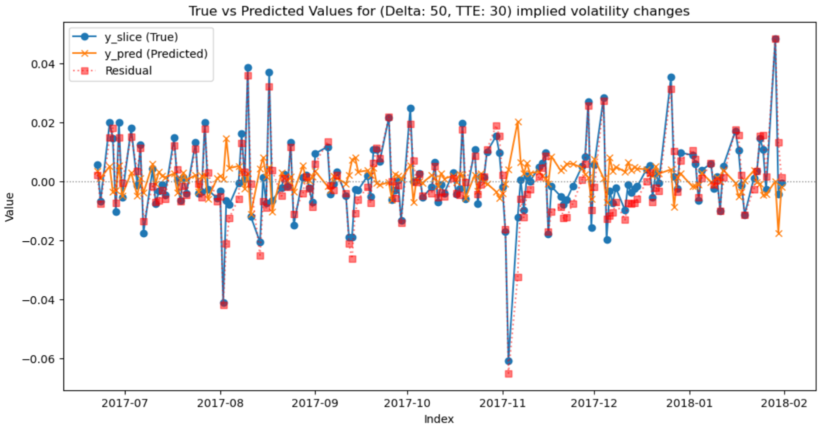



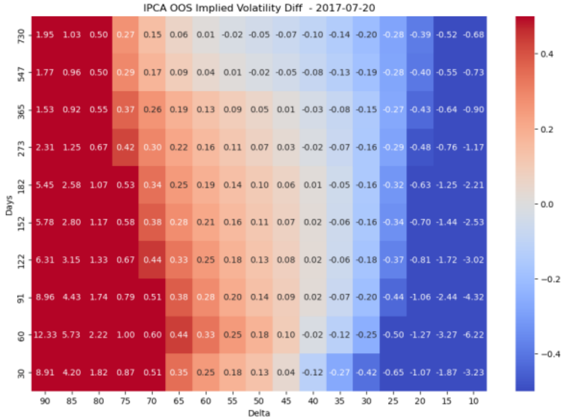



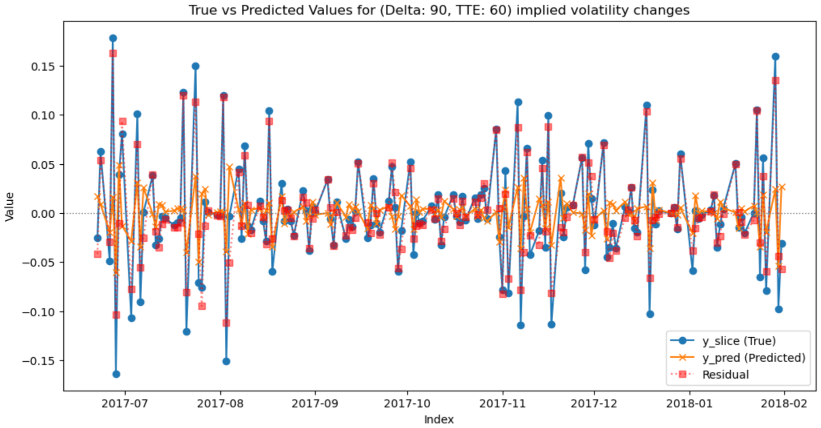



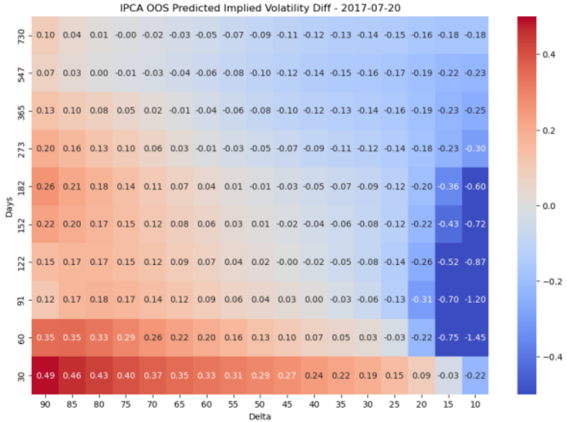
# Prediction Results: IPCA GRAPH

**Sample predictions from IPCA mirror shapes found in the components of PCA. Estimates are muted relative to the variance of IV.**



# Prediction Results: STATISTICS

We are capturing **a lot of variances** with IPCA. OOS R squared and RMSE are **positive.** The RMSE OOS is lower in IPCA than PCA but not overly differentiate**. IPCA OOS R squared has tremendous improvement.**

| | IS $R^2$ | OOS $R^2$ | IS RMSE | OOS RMSE |
|---|---|---|---|---|
| PCA | 0.016 | 0.001 | 0.0125 | 0.0112 |
| IPCA | -0.0234 | 0.200 | 0.0127 | 0.0101 |

# Discussion

**Interpretability**

- More interpretability: adding more covariates (index levels, macro indicators, other equities)

**Modeling Nonlinearity**

- IPCA assumes a linear relationship between factors and outcomes
  - Kernel IPCA
  - Autoencoders or transformer-based factor models
  - FPCA (Functional PCA)

**Hyperparameter Optimization**

- Number of factors (n_factors) & regularization strength (alpha) are tested via basic grid
  - Bayesian optimization (skopt, optuna) for more efficient and principled tuning

**Regime Awareness**

- IV evolves under different market regimes (calm vs crisis)
  - Use regime-switching IPCA
  - Cluster time periods and apply separate IPCA models per regime

# EXTRA: Simple Strategy

## Assumptions

- Assume continuous free hedging, no transaction costs
- One-day Holding period, buy today sell tomorrow.
- Quantities are equal to the price changes.

## Conclusion

- Returns largely in the deep-out-of-the-money IV ranges; this corresponds with relatively high variance
- Adjusting for return on a given strike X expiry, pseudo-Sharpe gives the best returns in ~6 month far-downside puts.