On the Pricing of Capped Volatility Swaps using Machine Learning Techniques

Eva Verschueren

January 31, 2023
PARTY 2023, Valencia, Spain.

KU LEUVEN
LRISK
assenagon

Joint work with Dr. Stephan Höcht (Assenagon) & Prof. Wim Schoutens (KU Leuven).

Outline of the presentation

1. Introduction
2. Data
3. Modeling
4. Results
5. Conclusion

Introduction

Definition

What is a capped volatility swap?

A **capped volatility swap** is a forward contract on an asset's capped, annualized, realized volatility, over a fixed period of length T, with payoff structure

$$\text{Notional} \times [\min(\text{Cap Level},\ \sigma_R) - K].$$

- σ_R is the annualized, realized volatility of an asset S, over a period of length T, calculated as

$$\sigma_R = 100 \times \sqrt{\frac{252}{T} \sum_{t=1}^{T} \left(\ln \left[\frac{S_t}{S_{t-1}} \right] \right)^2}.$$

Definition

What is a capped volatility swap?

A **capped volatility swap** is a forward contract on an asset's capped, annualized, realized volatility, over a fixed period of length T,
with payoff structure

$$\text{Notional} \times [\min(\text{Cap Level}, \sigma_R) - K].$$

- σ_R is the annualized, realized volatility of an asset S, over a period of length T, calculated as

$$\sigma_R = 100 \times \sqrt{\frac{252}{T} \sum_{t=1}^{T} \left(\ln \left[\frac{S_t}{S_{t-1}} \right] \right)^2}.$$

Definition

What is a capped volatility swap?

A **capped volatility swap** is a forward contract on an asset's capped, annualized, realized volatility, over a fixed period of length T,
with payoff structure

$$\text{Notional} \times [\min(\text{Cap Level},\ \sigma_R) - K].$$

- The strike K is the fair price of the contract, determined at contract initiation.
- To limit the risk exposure of the issuer, a cap level equal to $2.5 \times K$ is set on σ_R.

Definition

What is a capped volatility swap?

A **capped volatility swap** is a forward contract on an asset's capped, annualized, realized volatility, over a fixed period of length T,
with payoff structure

$$\text{Notional} \times [\min(\text{Cap Level},\ \sigma_R) - K].$$

- The strike K is the fair price of the contract, determined at contract initiation.
- To limit the risk exposure of the issuer, a cap level equal to $2.5 \times K$ is set on σ_R.

Problem Statement

Pricing a capped volatility swap

- Volatility swaps provide a pure exposure to the volatility of the underlying asset and are therefore frequently traded.

- The contracts are traded over-the-counter, meaning that no price is readily available on exchange.

⚠ Prices from different pricing sources tend to deviate from time to time.

- We build our own pricing tool, to validate external prices.
  - In general,

$$\text{Price}_t = \text{DF}_t \times \mathbb{E}_t^{\mathcal{Q}}(\text{payoff});$$

  - current literature is highly focused on a model-based pricing approach;
  - we focus on a data-driven approach to tackle the nonlinear pricing problem.

Data

General Overview

2 types of data

- The data sets of type I consist of time series of prices, from initiation of the contract to settlement, on the same underlier.

	S&P 500	AAPL	JPM
Time span	Nov. 18 - Sept. 22	Nov. 18 - Sept. 22	Nov. 18 - Sept. 22
Number of contracts	432	228	194
Number of observations	72 889	35 440	26 924

- The data sets of type II consist of spot prices of volatility swap contracts at initiation, on different underliers.

	Index	Equity
Time span	Nov. 18 - No. 22	Jan. 17 - Nov. 22
Number of underliers	8	235
Number of observations	818	8 647

Response variable

From Price to IVOL/Strike

The price, at time t, of a running contract is a rather unintuitive number.

- Type 1

  We rewrite the price of a contract using the fact that it reflects both the already realized $(0 - t)$ and unrealized $(t - T)$ part of the contract:

$$\text{Price}_t = \text{DF}_t \times \left(\sqrt{\text{IVOL}_t^2 \times (1 - \text{Weight}_t) + \text{Accrued Vol}_t^2 \times \text{Weight}_t} - \text{Strike} \right).$$

  - Weight = proportion of the current lifetime of the swap
  - Accrued Vol = already realized historical volatility

- Type 2

  We model strike K, since this is determined such that price=0 at contract initiation.

Predictor variables

Market-implied moments

- We model the variables IVOL/Strike that are completely determined by the movements of the underlying asset S.

- We look at distributional features of S for possible predictor variables.

- Market-implied (MI) N-th moment of the risk-neutral distribution of S is estimated from quoted European vanilla option prices, via[1]

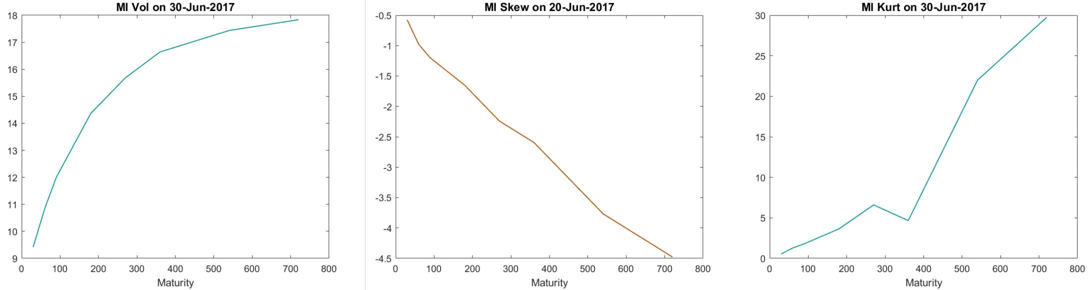
$$\mathbb{E}\left[\log\left(\frac{S_T}{S_0}\right)^N\right] = \cdots + \mathrm{e}^{rT}\int_0^{\kappa} \ldots EP(K,T)dK + \mathrm{e}^{rT}\int_{\kappa}^{\infty} \ldots EC(K,T)dK.$$

- We focus on summary statistics volatility, skewness and kurtosis.

[1] See Madan and Schoutens (2016)

Predictor variables

Market-implied moments on S&P500

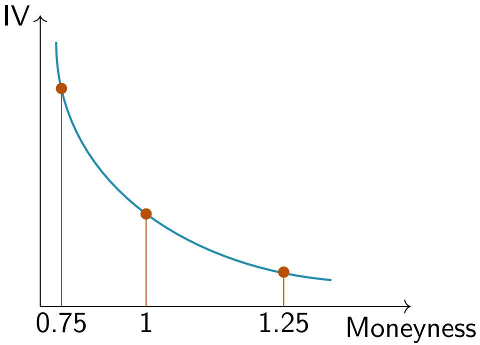


- We see large day-to-day instabilities, especially for the higher moments used in the calculation of skewness and kurtosis.
- We decide to focus solely on market-implied volatility and skewness.

Predictor variables

Implied volatility-based features

- We translate option prices into implied volatilities (IV) using Black-Scholes formula.
- We use[2]
  - MI Vol $\longleftrightarrow$ 100% IV;
  - MI Skew $\longleftrightarrow$ $\left(\frac{75-125}{100}\right)$% IV;
  - MI Kurt $\longleftrightarrow$ $\left(\frac{90-95}{100}\right)$% IV & $\left(\frac{110-105}{100}\right)$ % IV.



[2] See, e.g., Mixon (2011)

	S1	S2	S3	S4	S5**
Contract Specific Features*	✓	✓	✓	✓	✓
`MI Vol`	✓	✓			
`MI Skew`		✓			
`100% IV`			✓	✓	✓
$\left(\frac{75-125}{100}\right)$`% IV`				✓	✓
$\left(\frac{90-95}{100}\right)$`% IV` & $\left(\frac{110-105}{100}\right)$`% IV`					✓

*such as time to maturity.
**for S&P 500 and Type II data sets.

Modeling

Candidate Models

Gaussian Process Regression (GPR)

- Consider $(\boldsymbol{X}, y) = \{(\boldsymbol{x}_i, y_i) \mid i = 1, \ldots, n\}$, a set of observations.
- We assume

$$y = f(\boldsymbol{x}) + \varepsilon,$$

with Gaussian process $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$ and i.i.d. noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$.

A Gaussian process is a collection of random variables of which any finite set follows a multivariate Gaussian distribution. For a sample $(\boldsymbol{X}, \boldsymbol{f}) = \{(\boldsymbol{x}_i, f_i) | i = 1, \ldots, n\}$ generated from $f(\boldsymbol{x})$, it holds

$$\boldsymbol{f} \sim \mathcal{N}\left(m(\boldsymbol{X}) = \boldsymbol{0}, K(\boldsymbol{X}, \boldsymbol{X}) = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix} \right)$$

- GPR is a Bayesian method, combining the Gaussian process prior with observed data points.
- By definition[3], for each test set $\boldsymbol{X}_*$ of inputs and unknown $\boldsymbol{f}_*$, it holds

$$\begin{bmatrix} y \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 \mathbb{I} & K(\boldsymbol{X},\boldsymbol{X}_*) \\ K(\boldsymbol{X}_*,\boldsymbol{X}) & K(\boldsymbol{X}_*,\boldsymbol{X}_*) \end{bmatrix}\right),$$

such that

$$\boldsymbol{f}_* | \boldsymbol{X}_*, \boldsymbol{X}, y \sim \mathcal{N}\Big(K(\boldsymbol{X}_*,\boldsymbol{X}) \left[K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 \mathbb{I}\right]^{-1} y, \\ K(\boldsymbol{X}_*,\boldsymbol{X}_*) - K(\boldsymbol{X}_*,\boldsymbol{X}) \left[K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 \mathbb{I}\right]^{-1} K(\boldsymbol{X},\boldsymbol{X}_*) \Big)$$

- Point predictions for $\boldsymbol{f}_*$ correspond to the mean of this distribution.

[3] See e.g., Rasmussen and Williams (2006).

Candidate Models

Tree-based ML

In general, Regression Trees partition the predictor space, predicting the same value for each member of the constructed subsets.

Random Forest (RF)[4]

- Build a collection of deep regression trees on bootstrapped samples.
- At each split, only a selection of predictor variables is used as split candidates, to decorrelate the trees.

Gradient Boosting Machine (GBM)[5]

- Build a collection of simple regression trees, all fit sequentially, learning from the errors of predecessors.

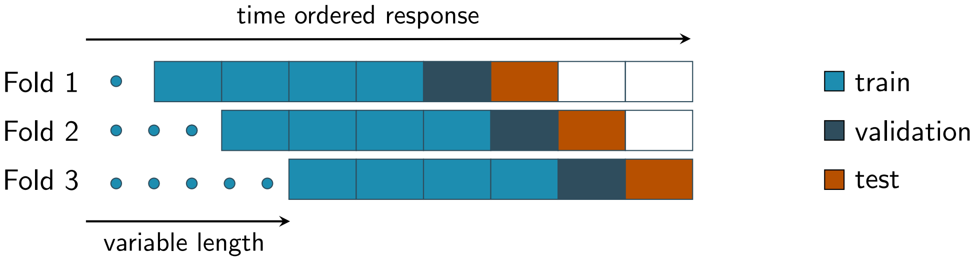
See [4]Breiman (2001) and [5]Friedman (2001).

Purged, Walk-forward Cross-Validation

We use purged walk-forward validation[6] for hyper parameter tuning and testing the model's out-of-sample performance.

- Purging = all observations overlapping in time are included in the same set.
- Walk-forward = train-validation-test window rolls forward in time.

⟶ Information leakage between training and validation/test set is limited.



[6]See de Prado (2018).

Test sets on S&P 500

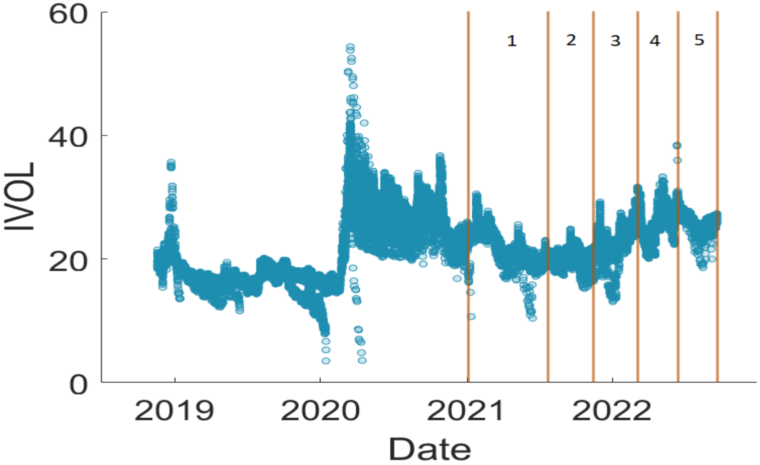


Figure: Visualization of the different test sets on data set type I of S&P 500.

Results

How to interpret the results?

- The models are evaluated using the Mean Absolute Error (MAE) of prediction:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{f}(\boldsymbol{x}_i)\right|,$$

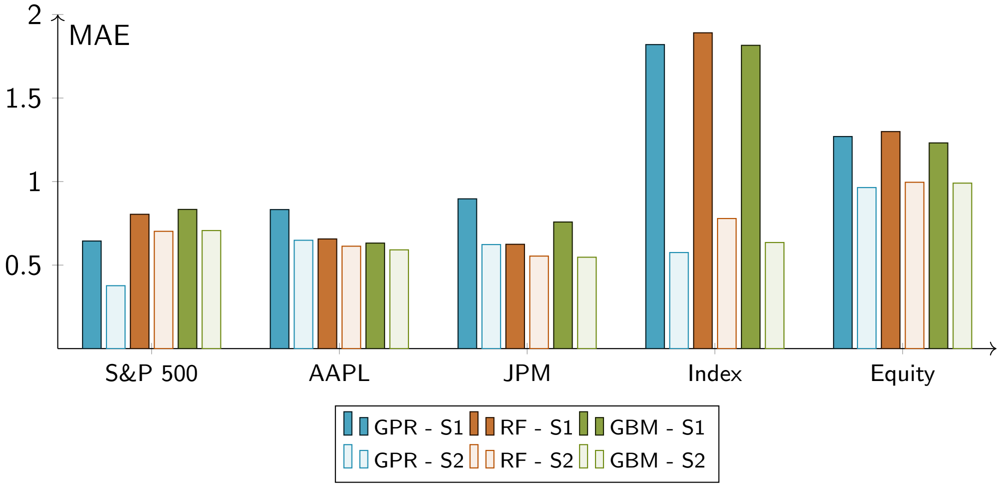
  with y_i the observed value for IVOL/Strike and $\widehat{f}(\boldsymbol{x}_i)$ the prediction by the model.

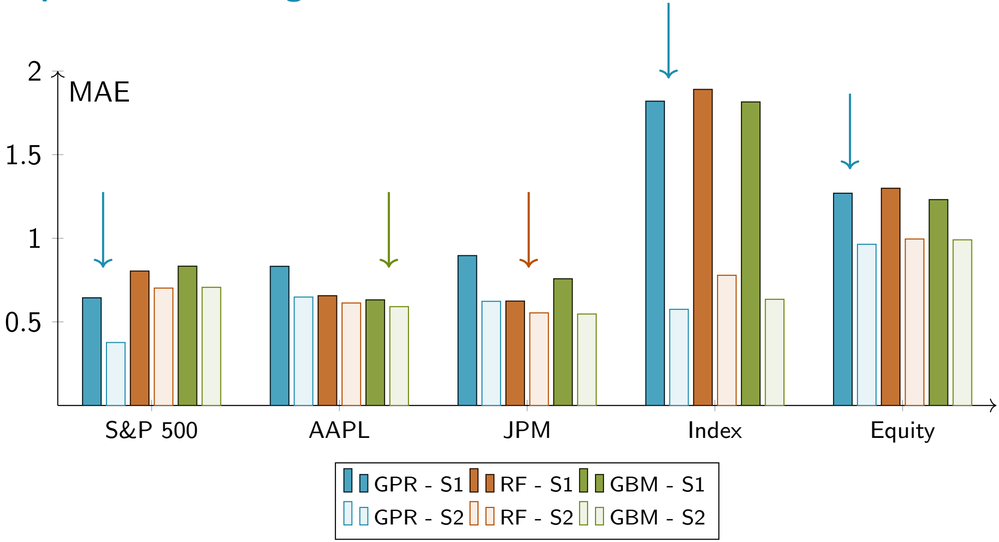
- When comparing different levels of IVOL/Strike, we use the Mean Relative Percentage Error (MRPE) of prediction:

$$\text{MRPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \widehat{f}(\boldsymbol{x}_i)\right|}{y_i}.$$

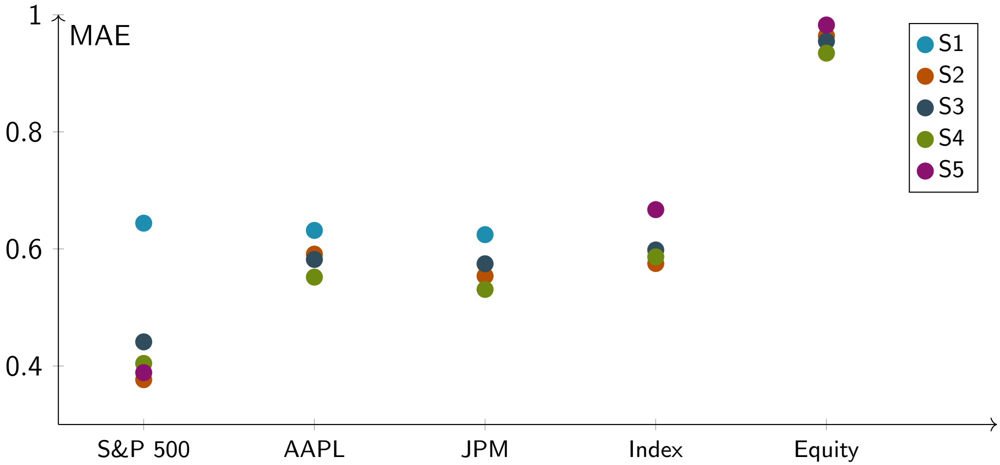
- The average is reported over the different test folds (out-of-sample performance).

Comparison of the Algorithms




2
1.5
1
0.5
MAE
S&P 500
AAPL
JPM
Index
Equity
GPR - S1
RF - S1
GBM - S1
GPR - S2
RF - S2
GBM - S2

Comparison of the Input Settings



- The best performing algorithm changes over the different data sets.
- In general, adding higher order information on skewness (and kurtosis) improves the predictive accuracy.
  - We prefer IV-based input settings, because of the complexity to calculate MI-moments.
  - Setting 4 is a safe choice, across the different data sets.

Type I vs Type II

	S&P 500	AAPL	JPM	Index	Equity
MRPE	0.0175	0.0183	0.0168	0.0224	0.0284

- Size of the data sets is much smaller.
- Type II data sets mixes various underliers.

General Findings

- The best performing algorithm changes over the different data sets.
- In general, adding higher order information on skewness (and kurtosis) improves the predictive accuracy.
  - We prefer IV-based input settings, because of the complexity to calculate MI-moments.
  - Setting 4 is a safe choice, across the different data sets.

Type I vs Type II

	S&P 500	AAPL	JPM	Index	Equity
MRPE	0.0175	0.0183	0.0168	0.0224	0.0284

- Size of the data sets is much smaller.
- Type II data sets mixes various underliers.

Conclusion

The pricing of capped volatility swaps, at any point in time, during the lifetime of the product, is an unsolved problem.

(1) While current literature is focused on model-based approximations of the price,

(2) we develop a data-driven pricing approach.

- This presentation deploys the use of various machine learning techniques within a tailored cross-validation setting.
- The results show that
  - the best performing algorithm changes over the different data sets;
  - the preference is to use IV-based features, both on volatility and skewness.

Thank you!

eva.verschueren@kuleuven.be

assenagon

Breiman, L. (1996). Bagging predictors. Machine Learning, 24:123-140.

Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

Davis, J., Devos, L., Reyners, S. and Schoutens, W. (2020). Gradient Boosting for Quantitative Finance. Journal of Computational Finance, 24(4).

Demeterfi, K., Derman, E., Kamal, M. and Zou, J. (1999). More Than You Ever Wanted To Know About Volatility Swaps. Quantitative Strategies Research Notes, Goldman Sachs, March, 1–52.

De Spiegeleer, J., Madan, D., Reyners, S. and Schoutens, W. (2018). Machine learning for quantitative finance: fast derivative pricing, hedging and fitting. Quantitative Finance, 18(10):1635-1643.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 29(5):1189–1232.

Gan, L., Wang, H. and Yang, Z. (2020). Machine learning solutions to challenges in finance: An application to the pricing of financial products. Technological Forecasting & Social change, 153.

Madan, D. and Schoutens, W. (2016). Applied Conic Finance. Cambridge University Press: Cambridge, UK.

Rujivan, S. and Rakwongwan, U. (2021). Analytically pricing volatility swaps and volatility options with discrete sampling: Nonlinear payoffvolatility derivatives. Communications in Nonlinear Science and Numerical Simulation, September, 100.

Kim, S-W. and Kim, J-H (2020). Volatility and variance swaps and options in the fractional SABR model. The European Journal of Finance, 26(17):1725-1745.

Issaka, A. (2020). Variance swaps, volatility swaps, hedging and bounds under multi-factor Heston stochastic volatility model. Stochastic Analysis and Applications, 38(5):856-874.