Data-Driven Models for Discrete Hedging Problem

Ke Nian
Supervisors: Prof.Yuying Li and Prof.Thomas.F.Coleman

David R. Cheriton School of Computer Science,
University of Waterloo,

Waterloo, Canada

May 1, 2019

Agenda

Introduction
Delta Hedging Variants
Stochastic Volatility Model
Minimum Variance Approach
Local Volatility Model
Data Driven Approach
Indirect Data-driven Approach
Direct Data-Driven Approach
Sequential Learning Framework
Motivation
Recurrent Neural Network
Encoder-Decoder Model
Results and Discussion
Summary
Appendix

Practitioner Black-Scholes (BS) Delta Hedging

- BS model:

$$\frac{dS}{S} = \mu dt + \sigma dZ$$

$$\sigma : \text{ Constant}$$

- Implied Volatility

$$\sigma_{imp} = V_{BS}^{-1}(V_{mkt}, .)$$

$$V_{mkt}\text{: market option price}$$
$$V_{BS}^{-1} : \text{ inverse of BS pricing function}$$

- BS Delta:

$$\delta_{BS} = \frac{\partial V_{BS}}{\partial S}$$

Problem with Black-Scholes Delta

Problem with the traditional Black-Scholes delta:

- Market violates Black-Scholes assumption
- Dependence of implied volatility on underlying asset price

Variants of delta hedging strategy:

- Stochastic Volatility Model
- Local Volatility Model
- Minimum Variance Approach
- Indirect Data-Driven Approach
- **Direct Data-Driven Approach**

Stochastic Volatility Model

Stochastic volatility models:

- Example: Heston Model

$$
\begin{aligned}
dS_t &= rS_t dt + \sqrt{\upsilon_t} S_t dW_t \\
d\upsilon_t &= \kappa(\overline{\upsilon} - \upsilon_t) dt + \eta \sqrt{\upsilon_t} dZ_t \\
dZ_t dW_t &= \rho dt
\end{aligned}
$$

Minimum Variance Approach

The correction for the dependence of implied volatility on asset price:

- The Minimum Variance (MV) delta:

$$\delta_{MV} = \frac{\partial V_{BS}}{\partial S} + \frac{\partial V_{BS}}{\partial \sigma_{imp}} \frac{\partial \sigma_{imp}}{\partial S}$$

- A parametric model [1] learned from market data:

$$\frac{\partial \sigma_{imp}}{\partial S} = \frac{a + b\delta_{BS} + c\delta_{BS}^2}{S\sqrt{T}}$$

a, b and c are the parameters to be fitted using market data.

[1] Hull, J. and White, A., "Optimal delta hedging for options." Journal of Banking and Finance 82 (2017): 180-190.

Local Volatility Model

The local volatility function (LVF) [2]: the volatility is a deterministic function of S and t.

$$\delta_{MV} = \frac{\partial V_{BS}}{\partial S} + \frac{\partial V_{BS}}{\partial \sigma_{imp}} \frac{\partial \sigma_{imp}}{\partial S}$$

Local volatility model can also be used to calculate the $\frac{\partial \sigma_{imp}}{\partial S}$.

[2] Coleman, T.F., Kim, Y., Li, Y. and Verma, A., 'Dynamic hedging with a deterministic local volatility function model,' Journal of risk, 4 ,1 (2001):63-89

Problem with Parametric Approach

Parametric approaches:

- Model mis-specification.
- Sub-optimal for discrete hedging problems.

Data-driven approaches:

- Minimum assumptions on S.
- Model is determined by market data.

Indirect Data-driven Approach

The indirect data-driven approach [3]can be summarized as following:

- Let X be the features from market.
  - Asset price S.
  - Strike Price K.
  - Time to expiration $T - t$.
- Determine the data-driven pricing function $V(X)$ using regression model.
- Compute

$$\delta_{ID} = \frac{\partial V(X)}{\partial S}$$

[3]Hutchinson, J.M., Lo, A.W. and Poggio, T., "A nonparametric approach to pricing and hedging derivative securities via learning networks." The Journal of Finance 49.3 (1994): 851-889.

Problem with Indirect Data-driven Approach

Problems with the indirect data-driven approach:

- Unnecessary intermediate procedure.
- Sub-optimal for discrete hedging.
- Model parameters depend on the asset price.

Direct data-driven approach can be more useful in practice.

- Customized hedging position function.
- Directly learn the hedging position.

David R. Cheriton School of Computer Science, University of Waterloo

Direct Data-driven Approach

The direct data-driven approach is

$$\min_f \left[\frac{1}{N} \sum_{i=1}^{N} (\Delta V_i - \Delta S_i f(X_i))^2 \right]$$

- ΔV_i : the change of option value in data instance i.
- ΔS_i : the change of asset price in data instance i.
- $f(X_i)$: option hedging position function.
- Data-Driven models outperform other delta hedging strategies [4].

[4] Nian, Ke, Thomas F. Coleman, and Yuying Li. "Learning minimum variance discrete hedging directly from the market." Quantitative Finance (2018): 1-14.

Data-driven Kernel Learning Approach

Kernel Learning Framework:

$$\min_{f \in RKHS} \left[\frac{1}{N} \sum_{i=1}^{N} (\Delta V_i - \Delta S_i f(x_i))^2 + \lambda \|f\|_K^2 \right]$$

Matrix Form:

$$\min_{\alpha} (DK\alpha - \Delta V)^T (DK\alpha - \Delta V) + \lambda \alpha^T K \alpha$$

Where D is the diagonal matrix with ΔS on its diagonal and K is the kernel matrix

Data-Driven Kernel Learning Framework

The data-driven kernel learning framework suffers from several drawbacks:

- Computationally expensive.
- Limited number of variables.
- No feature selection.
- Ignoring the auto-correlation in market data.

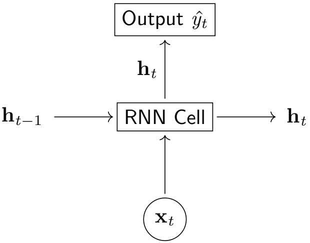
Volatility Clustering and Financial Time Series

Sequential learning framework may further improve the performance:

- Volatility clustering observed in the financial market.
- Autocorrelation between data instances near in time.
- Dependence of option pricing function on the past history of the underlying has been shown in GARCH models [5].

[5] Heston, Steven L., and Saikat Nandi "A closed-form GARCH option valuation model." The review of financial studies 13.3 (2000): 585-625.

Recurrent Neural Network



In each RNN cell:

$$\mathbf{h}_t = f_{act}(\boldsymbol{W}_{hx}\mathbf{x}_t + \boldsymbol{W}_{hh}\mathbf{h}_{t-1} + b_h)$$
$$\hat{y}_t = f_{out}(\boldsymbol{W}_{yh}\mathbf{h}_t + b_y)$$

- Vanishing gradients problem.
- Gated Recurrent Unit (GRU) [6]model is introduced to combat the problem of vanishing gradients.

[6]Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).

David R. Cheriton School of Computer Science, University of Waterloo

Features

When generating the hedging position, we use local features $\mathbf{x}_L \in \mathbb{R}^d$ and sequential features $\mathbf{X} \in \mathbb{R}^{D \times N}$:

$$\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}] = \begin{bmatrix} (\mathbf{x^1})^\top \\ (\mathbf{x^2})^\top \\ \vdots \\ (\mathbf{x^D})^\top \end{bmatrix} = \begin{bmatrix} x_1^1 & \ldots & x_N^1 \\ \vdots & \ldots & \vdots \\ x_1^D & \ldots & x_N^D \end{bmatrix}$$

Local Features

Local features $\mathbf{x}_L$ for the current day contains:

1. Moneyness S/K.
2. BS delta δ_{BS}.
3. Time to expiry τ.
4. Index close price S .
5. Option bid price V_{bid}.
6. Option offer price V_{offer}.
7. Implied volatility σ_{imp}.
8. BS gamma γ_{BS}.
9. BS vega $vega_{BS}$.
10. Minimum variance delta δ_{MV}

David R. Cheriton School of Computer Science, University of Waterloo

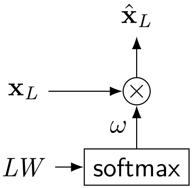
Sequential Features

Sequential features $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ recording the past history contains:

1. Option middle price V_{mid}.
2. Implied volatility σ_{imp}.
3. BS delta δ_{BS}.
4. BS gamma γ.
5. BS vega $vega_{BS}$.
6. Moneyness S/K.

David R. Cheriton School of Computer Science, University of Waterloo

Weighting Local Features

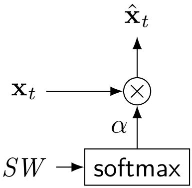


- $LW \in \mathbb{R}^d$: The unnormalized feature weighting vector for the local features $\mathbf{x}_L \in \mathbb{R}^d$.

$$\omega_i = \frac{exp(LW_i)}{\sum_{j=1}^{d} exp(LW_j)}, \; \sum_{i=1}^{d} \omega_i = 1$$

$$\hat{\mathbf{x}}_L = [\omega_1 \mathbf{x}_L^1, \ldots, \omega_d \mathbf{x}_L^d]$$

Weighting Sequential Features

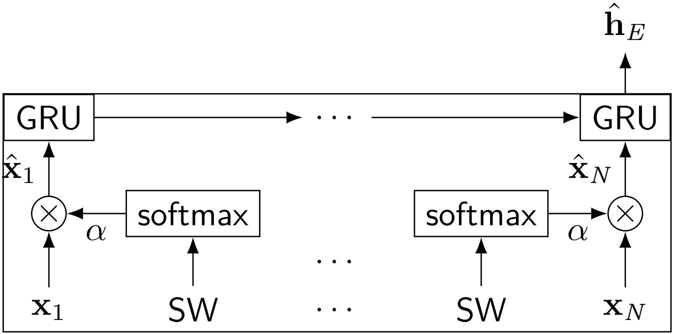


- $SW \in \mathbb{R}^D$:The unnormalized feature weighting vector for the sequential features $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}]$.
- $\mathbf{x}_t \in \mathbb{R}^D$: The feature vector in $\mathbf{X}$ at time step t

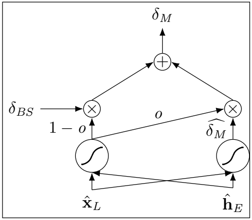
$$\alpha_i = \frac{exp(SW_i)}{\sum_{j=1}^{D} exp(SW_j)}, \ \sum_{i=1}^{D} \alpha_i = 1$$

$$\hat{\mathbf{x}}_t = [\alpha_1 \mathbf{x}_t^1, \ldots, \alpha_D \mathbf{x}_t^D]$$

Encoder Model



Decoder Model



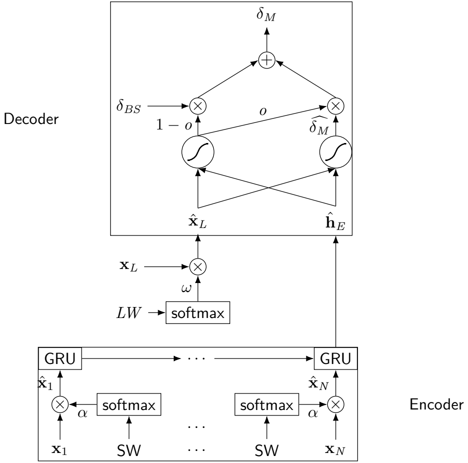
$$\widehat{\delta_M} = sigmoid(\mathbf{v}_{out}^T \ tanh(\boldsymbol{U}^{out}\hat{\mathbf{h}}_E + \boldsymbol{W}^{out}\hat{\mathbf{x}}_L + \mathbf{b}^{out}))$$

$$o = sigmoid(\mathbf{v}_{Gate}^T \ tanh(\boldsymbol{U}^{Gate}\hat{\mathbf{h}}_E + \boldsymbol{W}^{Gate}\hat{\mathbf{x}}_L + \mathbf{b}^{Gate}))$$

The final output is

- Call: $\delta_M = \widehat{\delta_M} \times o + \delta_{BS} \times (1-o)$.
- Put: $\delta_M = -\widehat{\delta_M} \times o + \delta_{BS} \times (1-o)$.

Encoder-Decoder Model



Objective Functions

- Hedging error for BS delta:

$$\eta_i = \Delta V_i - \Delta S_i \delta^i_{BS}$$

- Hedging error for the proposed model:

$$l_i = \Delta V_i - \Delta S_i \delta^i_M$$

- Mean squared loss: $L_S = \frac{1}{2m} \sum_{i=1}^m l_i^2$
- Huber loss:

$$L(l_i, \eta_i) = \begin{cases} \frac{1}{2} l_i^2, & \text{if } |l_i| \leq |\eta_i| \\ |\eta_i|(|l_i| - \frac{1}{2}|\eta_i|), & \text{otherwise} \end{cases}$$

$$L_H = \frac{1}{m} \sum_{i=1}^{m} L(l_i, \eta_i)$$

Training and Regularization

- Optimization technique: trust region method
- Regularization technique: early stopping
- Usage of validation set
- Daily update

Evaluation Criteria: Local Risk

The percentage increase in the effectiveness over the BS hedging:

$$Gain = 1 - \frac{SSE[\Delta V_i - \Delta S_i \delta^i]}{SSE[\Delta V_i - \Delta S_i \delta^i_{BS}]}$$

- SSE: sum of squared errors
- δ: hedging position computed from different models
- δ_{BS}: BS delta

David R. Cheriton School of Computer Science, University of Waterloo

Experimental Setting

- Data: S&P 500 index option from Jan 2007 to Aug 2015
- The models to be compared:
  - $\mathrm{DKL}_{\mathsf{SPL}}$: Direct data-driven kernel learning model.
  - MV: Minimum variance hedging formula.
  - LVF: Local volatility function model.
  - SABR: SABR stochastic volatility model.
  - DRNN: The proposed encoder-decoder model

Call Option Daily Hedging

Delta	MV (%)	SABR(%)	LVF(%)	Data-Driven Model DKL$_{\text{SPL}}$ (%) Traded	DKL$_{\text{SPL}}$ (%) All	DRNN (%) Traded	DRNN (%) All
0.1	42.1	39.4	42.6	47.1	48.6	32.3	33.8
0.2	35.8	33.4	36.2	37.8	40.0	33.7	36.4
0.3	31.1	29.4	30.3	34.1	35.1	34.1	**35.5**
0.4	28.5	26.3	26.7	32.3	32.0	**33.7**	**34.2**
0.5	27.1	24.9	25.5	29.3	29.4	**35.1**	**33.0**
0.6	25.7	25.2	25.2	29.9	28.4	**35.6**	**32.1**
0.7	25.4	24.7	25.8	29.0	26.8	**31.8**	**29.7**
0.8	24.1	23.5	25.4	25.9	24.7	**28.6**	**26.5**
0.9	16.6	17.0	16.9	17.7	13.9	**19.3**	**18.9**
Overall	25.7	24.6	25.5	31.3	26.0	**32.9**	**28.7**

- Performance will be slighted better than $\mathrm{DKL}_{\mathrm{SPL}}$.

Put Option Daily Hedging

Delta	MV (%)	SABR(%)	LVF(%)	Data-Driven Model $\mathrm{DKL}_{\mathrm{SPL}}$ (%) Traded	$\mathrm{DKL}_{\mathrm{SPL}}$ (%) All	DRNN (%) Traded	DRNN (%) All
-0.9	15.1	11.2	-7.4	8.6	13.6	15.1	**17.2**
-0.8	18.7	19.6	6.8	6.5	16.7	**23.2**	**28.5**
-0.7	20.3	17.7	9.1	10.6	19.8	**28.5**	**32.8**
-0.6	20.4	16.7	9.2	14.9	21.0	**28.3**	**33.9**
-0.5	22.1	16.7	10.8	22.5	23.1	**29.2**	**34.5**
-0.4	23.8	17.7	12.0	24.2	25.2	**29.9**	**34.7**
-0.4	27.1	21.7	16.8	27.7	28.3	**30.6**	**33.6**
-0.2	29.6	25.8	20.6	30.1	30.8	25.4	29.9
-0.1	27.5	26.9	17.7	29.1	31.2	18.7	21.4
Overall	22.5	19.0	10.2	23.4	23.2	**26.2**	**29.7**

- Performance will be slighted better than $\mathrm{DKL}_{\mathrm{SPL}}$.

Call Option Weekly Hedging and Monthly Hedging

Delta	Data-Driven Model			
	$\mathrm{DKL}_{\mathsf{SPL}}$(%)		DRNN(%)	
	Traded	All	Traded	All
0.1	38.9	38.3	**47.8**	**45.6**
0.2	29.0	26.9	**48.5**	**46.0**
0.3	23.5	25.3	**48.5**	**46.6**
0.4	20.8	24.3	**45.9**	**45.4**
0.5	19.9	22.8	**46.6**	**45.0**
0.6	17.3	19.5	**44.8**	**43.1**
0.7	16.8	17.7	**43.9**	**42.4**
0.8	12.5	12.3	**37.7**	**39.0**
0.9	6.2	5.1	**16.4**	**29.1**
Overall	20.2	17.1	**43.7**	**40.5**

Delta	Data-Driven Model			
	$\mathrm{DKL}_{\mathsf{SPL}}$ (%)		DRNN (%)	
	Traded	All	Traded	All
0.1	22.7	24.8	**53.9**	**39.4**
0.2	23.5	25.5	**51.7**	**48.3**
0.3	24.0	24.6	**50.2**	**49.1**
0.4	21.0	20.7	**47.8**	**48.3**
0.5	13.5	12.7	**44.5**	**47.6**
0.6	14.3	13.5	**44.6**	**47.4**
0.7	6.1	7.0	**35.3**	**42.9**
0.8	5.3	4.1	**24.8**	**34.1**
0.9	4.1	2.3	**10.5**	**19.9**
Overall	16.3	12.5	**44.5**	**42.3**

Table: Weekly(Left) and Monthly(Right)

- Performance will be significantly better than $\mathrm{DKL}_{\mathsf{SPL}}$.

Put Option Weekly Hedging and Monthly Hedging

Delta	Data-Driven Model $\mathrm{DKL}_{\mathrm{SPL}}$(%) Traded	All	DRNN(%) Traded	All
-0.9	10.1	7.3	**34.7**	**35.7**
-0.8	18.3	11.5	**44.2**	**45.1**
-0.7	20.2	16.3	**49.6**	**47.3**
-0.6	20.8	18.4	**51.3**	**49.6**
-0.5	22.4	21.2	**53.5**	**51.0**
-0.4	21.0	23.9	**53.2**	**51.2**
-0.3	22.2	26.1	**51.1**	**51.7**
-0.2	20.8	29.7	**46.3**	**51.8**
-0.1	19.2	29.1	**37.2**	**47.6**
Overall	20.4	20.3	**49.1**	**49.4**

Delta	Data-Driven Model $\mathrm{DKL}_{\mathrm{SPL}}$ (%) Traded	All	DRNN (%) Traded	All
-0.9	6.5	5.8	**32.6**	**33.1**
-0.8	6.1	7.8	**49.5**	**45.3**
-0.7	7.3	11.9	**52.4**	**46.3**
-0.6	10.3	9.5	**51.6**	**47.0**
-0.5	13.9	12.8	**51.4**	**46.7**
-0.4	15.6	16.7	**53.4**	**45.1**
-0.3	19.5	13.4	**48.4**	**40.7**
-0.2	20.6	18.4	**44.7**	**35.1**
-0.1	13.0	19.9	**26.8**	**25.3**
Overall	13.5	12.7	**49.5**	**41.2**

Table: Weekly(Left) and Monthly(Right)

- Performance will be significantly better than $\mathrm{DKL}_{\mathrm{SPL}}$.

Importance of Sequential Learning

- The RNN part is removed.

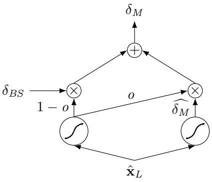


Figure: DNN

Delta	Data-Driven Model(%) Weekly DNN	Weekly DRNN	Monthly DNN	Monthly DRNN
0.1	35.6	**47.8**	29.7	**53.9**
0.2	36.4	**48.5**	38.4	**51.7**
0.3	38.6	**48.5**	40.2	**50.2**
0.4	38.7	**45.9**	38.6	**47.8**
0.5	42.3	**46.6**	36.3	**44.5**
0.6	43.4	**44.8**	36.0	**44.6**
0.7	**45.6**	43.9	30.2	**35.3**
0.8	39.6	**37.7**	22.3	**24.8**
0.9	**26.3**	16.4	**21.1**	10.5
Overall	39.9	**43.7**	35.4	**44.5**

- Performance will be decreased if we remove the RNN part.

A Simplified Model for Comparison

- Directly outputting δ_M without combining with δ_{BS}:

$$\delta_M = sigmoid(\mathbf{v}_{out}^T\ tanh(\boldsymbol{U}^{out}\hat{\mathbf{h}}_E + \boldsymbol{W}^{out}\hat{\mathbf{x}}_L + \mathbf{b}^{out}))$$

- The model is trained without early stopping.
- The objective function is fixed to be the mean squared loss.

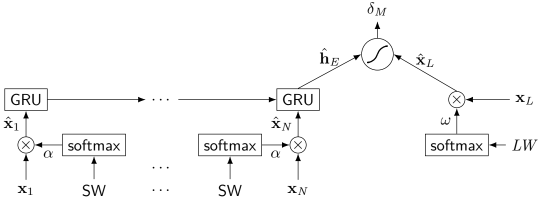


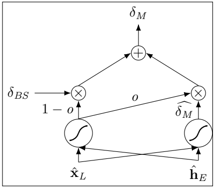
Figure: $DRNN_C$

Importance of Robust Model Design and Regularization

Delta	Data-Driven Model (%)			
	$DRNN_C$		DRNN	
	Weekly	Monthly	Weekly	Monthly
0.1	36.6	34.8	**47.8**	**53.9**
0.2	39.6	38.9	**48.5**	**51.7**
0.3	39.7	41.7	**48.5**	**50.2**
0.4	38.9	42.6	**45.9**	**47.8**
0.5	37.5	42.3	**46.6**	**44.5**
0.6	33.5	40.7	**44.8**	**44.6**
0.7	31.1	33.0	**43.9**	**35.3**
0.8	31.7	**26.3**	**37.7**	24.8
0.9	**28.7**	**17.3**	16.4	10.5
Overall	33.5	38.0	**43.7**	**44.5**

- Performance will be decreased if we use the simplified model $DRNN_C$.

Importance of Output Gate



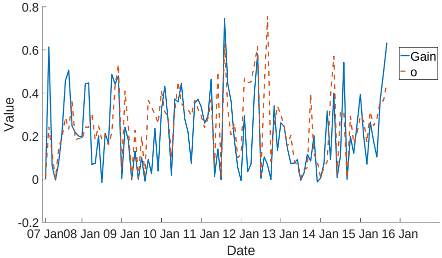
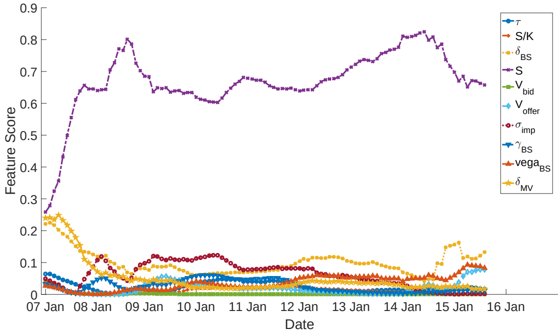


Figure: Call Option Monthly Hedging

- When o is close to 0, what the model output is close to δ_{BS}.

David R. Cheriton School of Computer Science, University of Waterloo

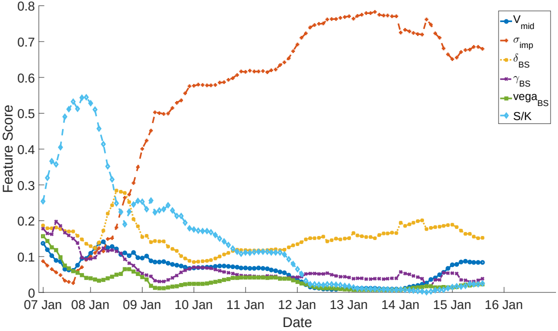
Local Feature Score (ω) for Monthly Hedging



(a) Local Features

Figure: Feature Score of S&P500 Call Option

Sequential Feature Score (α) for Monthly Hedging



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option

Summary

- Loosing assumption on the market dynamic is a good practise
- Data-driven approach can lead to better performance.
- Incorporating the information about the past history can further improve the hedging performance.
- Robust model design is also beneficiary.

What to do next?

- Extend the learning framework to multi-step total hedging problems.
- Use convolution neural network to extract features from volatility surface.

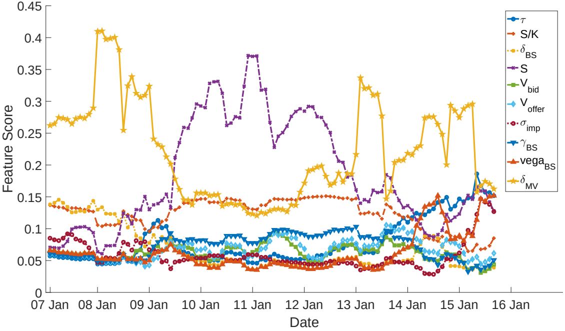
Q & A

Thank you very much!

Any Questions?

David R. Cheriton School of Computer Science, University of Waterloo

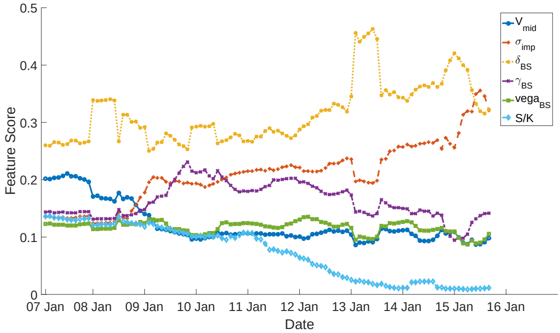
Local Feature Score (ω) for Daily Hedging



(a) Local Features

Figure: Feature Score of S&P500 Call Option

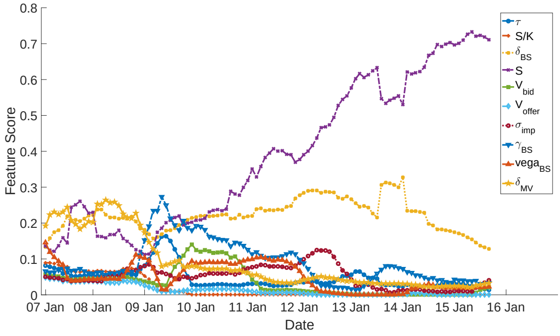
Sequential Feature Score (α) for Daily Hedging



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option

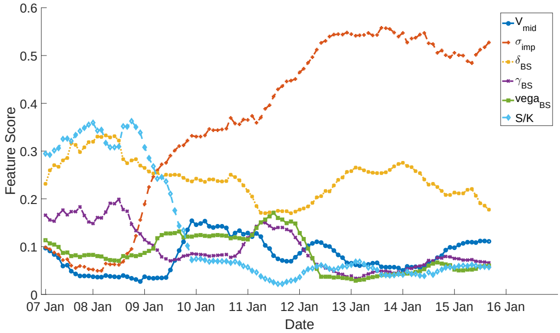
Local Feature Score (ω) for Weekly Hedging



(a) Local Features

Figure: Feature Score of S&P500 Call Option

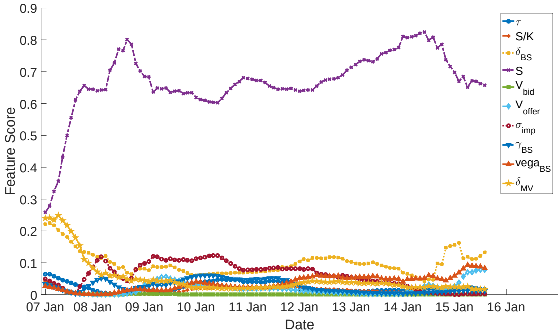
Sequential Feature Score (α) for Weekly Hedging



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option

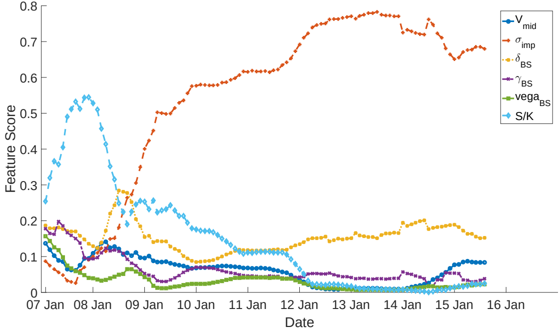
Local Feature Score (ω) for Monthly Hedging



(a) Local Features

Figure: Feature Score of S&P500 Call Option

Sequential Feature Score (α) for Monthly Hedging



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option