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

January 22, 2019

Agenda

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}: market option price
V_{BS}^{-1} : 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 BS assumption
- Dependence of volatility on underlying asset price

Variants of 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:

- Heston Model

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

- Many stochastic volatility models do not have analytical formula for pricing and hedging function.

Minimum Variance Approach

Considering the the dependence of imply volatility on asset price:

- The Minimum Variance (MV) delta:

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

- The authors [1]propose:

$$\frac{\partial \sigma_{imp}}{\partial S} = \frac{a + b\delta_{BS} + c\delta_{BS}^2}{S\sqrt{T}} \quad (1)$$

a, b and c are the parameter 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]: 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

Problem 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 compute the hedging position.

Direct Data-driven Approach

The direct data-driven approach is

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

$$\Delta V_i : \text{ the change of option value in data instance } i$$
$$\Delta S_i : \text{ the change of asset price in data instance } i$$

Real Data Hedging Experiments

- Data: S&P 500 index option from Jan 2007 and Aug 2015
- Model Calibration:
  - SABR: daily calibration
  - LVF: $\frac{\partial \sigma_{imp}}{\partial S}$ from implied volatility surface
  - MV: Use a 36 months time window to train
  - $\mathrm{DKL}_{\mathsf{SPL}}$: Use a 36 months time window to train. Models are separately calibrated for different Black-Sholes delta range.

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

S&P 500 Call Options

Delta	SABR (%)	LVF (%)	MV (%)	$\mathrm{DKL}_{\mathsf{SPL}}$ (%) Leave-One-Out[1] Traded	All
0.1	42.1	39.4	42.6	**44.1**	**44.4**
0.2	35.8	33.4	36.2	**37.8**	**38.1**
0.3	31.1	29.4	30.3	**33.1**	**33.6**
0.4	28.5	26.3	26.7	**30.9**	**31.3**
0.5	27.1	24.9	25.5	**30.0**	**30.4**
0.6	25.7	25.2	25.2	**29.3**	**29.8**
0.7	25.4	24.7	25.8	**28.4**	**30.2**
0.8	24.1	23.5	25.4	22.5	**28.0**
0.9	16.6	**17.0**	16.9	8.1	12.7
Overall	25.7	24.6	25.5	**31.3**	**26.8**

Table: S&P 500 Call Option Daily Hedging: bold entry indicating best Gain

[1] For each month, the penalties for models are determined by leave-one-out cross validation.

Data-Driven Kernel Learning Framework

Data-Driven Kernel Learning Framework [4] suffers from several drawbacks:

- Computationally expensive
- Limited number of variables
- No feature selection
- Not suitable for time series

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

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 (1)

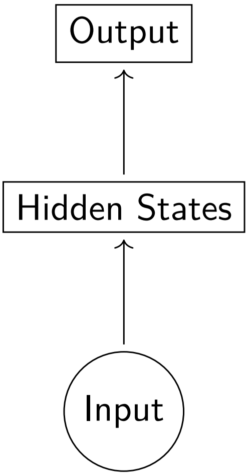


Figure: Neural Network

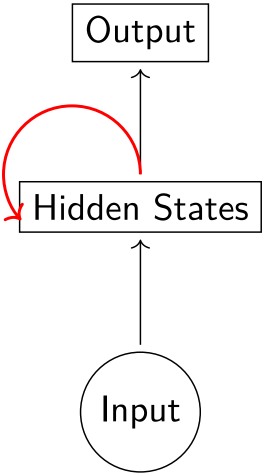
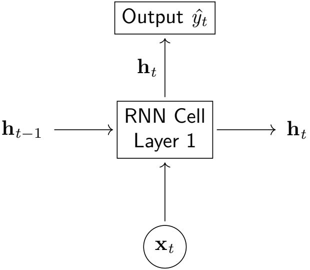


Figure: Recurrent Neural Network

Recurrent Neural Network (2)



In each RNN cell:

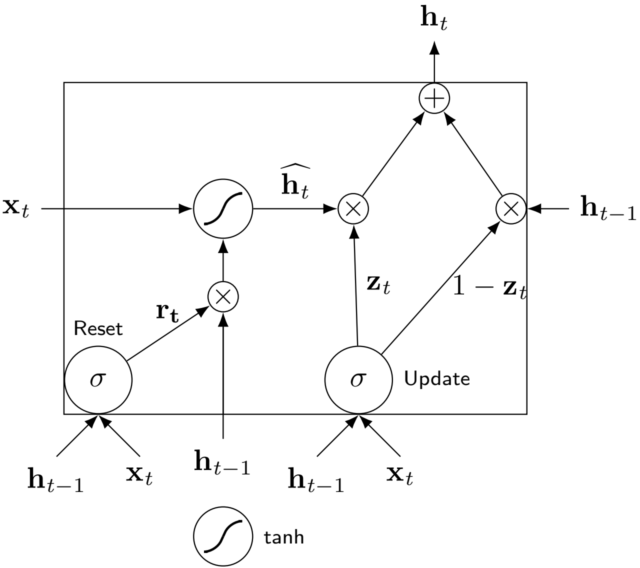
$$\mathbf{h}_t = f_{act}(\boldsymbol{W}_{hx}\mathbf{x}_t + \boldsymbol{W}_{hh}\mathbf{h}_{t-1} + b_h)$$

$$\hat{y}_t = f_{out}(\boldsymbol{W}_{yh}\mathbf{h}_t + b_y)$$

- The original RNN model suffers from the problem of vanishing gradients.
- Gated Recurrent Unit (GRU) [6]model is introduced to combat vanishing gradients through a gating mechanism.

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

Gated Recurrent Unit (GRU) Model



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

Features

When outputting the hedging position for a single data instance on a specific date, we have gathered some local features $\mathbf{x}_L \in \mathbb{R}^d$ and some 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 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}

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 σ_{imp}.
3. BS delta δ_{BS}.
4. BS gamma γ.
5. BS vega $vega_{BS}$.
6. Moneyness S/K.

Feature Weighting (1)

- Unnormalized feature weighting vector $LW \in \mathbb{R}^d$ for the raw local features $\mathbf{x}_L \in \mathbb{R}^d$.
- Normalized feature weight:

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

Thus:

$$\sum_{i=1}^{d} \omega_i = 1$$

- Weighted local feature vector:

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

Feature Weighting (2)

- Unnormalized the feature weighting vector $SW \in \mathbb{R}^D$ for the sequential features $\mathbf{X} \in \mathbb{R}^{D \times N}$.

$$SW = [SW_1, \ldots, SW_D]$$

- Normalized feature weight:

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

Thus:

$$\sum_{i=1}^{D} \alpha_i = 1$$

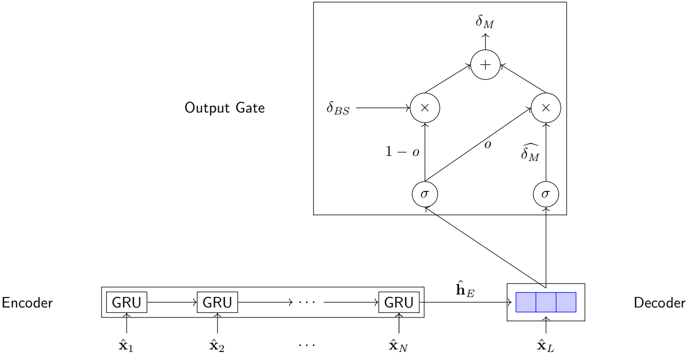
David R. Cheriton School of Computer Science, University of Waterloo

Feature Weighting (2)

Recall that $\mathbf{x_t} \in \mathbb{R}^D$ is a vector recording the D features at a specific time step t in the input sequential feature $\mathbf{X}$ and $\mathbf{x^k} \in \mathbb{R}^N$ is a vector recording the k-th sequential feature. The weighted input sequences $\hat{\mathbf{X}}$ is:

$$\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_N] = \begin{bmatrix} (\hat{\mathbf{x}}^1)^\top = \alpha_1(\mathbf{x}^1)^\top \\ (\hat{\mathbf{x}}^2)^\top = \alpha_2(\mathbf{x}^2)^\top \\ \vdots \\ (\hat{\mathbf{x}}^D)^\top = \alpha_D(\mathbf{x}^D)^\top \end{bmatrix}$$

Encoder-Decoder Model



Robust Losses and Output Gate (1)

- Hedging Loss for BS delta of a data instance i

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

- Hedging Loss for delta from the proposed model of a data instance i

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

- Mean Squared Loss: $L_S = \frac{1}{m}\sum_{i=1}^{m} loss_i^2$
- Modified Huber loss:

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

$$L_H = \frac{1}{m}\sum_{i=1}^{m} L(loss_i)$$

Robust Losses and Output Gate (2)

The candidate output is computed by a single layer feedforward network:

$$\widehat{\delta_M} = \sigma(\mathbf{v}_{out}^T \ tanh(\boldsymbol{U}^{out}\widehat{\mathbf{h_E}} + \boldsymbol{W}^{out}\hat{\mathbf{x}}_L + \mathbf{b}^{out})) \tag{2}$$

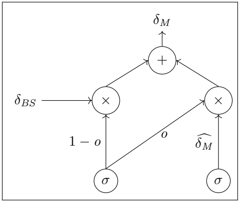
The output gate value is computed by another single layer feedforward network:

$$o = \sigma(\mathbf{v}_{Gate}^T \ tanh(\boldsymbol{U}^{Gate}\widehat{\mathbf{h_E}} + \boldsymbol{W}^{Gate}\hat{\mathbf{x}}_L + \mathbf{b}^{Gate}))$$

The final output from the model is :

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

Output Gate



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

Robust Losses and Output Gate (3)

The intuitions for the robust losses and output gate are:

- We focus more on the data instances where $|loss_i| \leq |BSloss_i|$ can be easily achieved.
- When $|loss_i| > |BSloss_i|$, we penalize the difference between $|loss_i|$ and $|BSloss_i|$ and force the outputs of these data instances to be closer to the associated δ_{BS} .
- The output gate enables the model to directly output δ_{BS}

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

Training and Regularization

- Trust region method is used as the optimization technique.
- Early stopping is used as the regularization technique.
- A small portion of the training set is used as the validation set to determine when to stop the training procedure.
- The model is updated on daily basis.

Call Option Daily Hedging

Delta	MV (%)	SABR(%)	LVF(%)	Data-Driven Model: DKL_{SPL} (%) Traded	Data-Driven Model: DKL_{SPL} (%) All	Data-Driven Model: DRNN (%) Traded	Data-Driven Model: 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**

Put Option Daily Hedging

Delta	MV (%)	SABR(%)	LVF(%)	Data-Driven Model			
				DKL_{SPL} (%)		DRNN (%)	
				Traded	All	Traded	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**

Call Option Weekly Hedging and Monthly Hedging

Delta	Data-Driven Model DKL_{SPL}(%) Traded	All	DRNN(%) 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 DKL_{SPL} (%) Traded	All	DRNN (%) 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)

Put Option Weekly Hedging and Monthly Hedging

Delta	Data-Driven Model			
	$\mathrm{DKL}_{\mathrm{SPL}}$(%)		DRNN(%)	
	Traded	All	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			
	$\mathrm{DKL}_{\mathrm{SPL}}$ (%)		DRNN (%)	
	Traded	All	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)

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

Where does the improvement come from? (1)

- Remove the sequential learning part

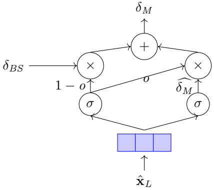


Figure: DNN

- Remove the output gate

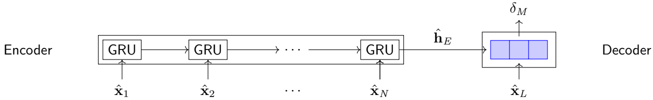


Figure: $DRNN_C$

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

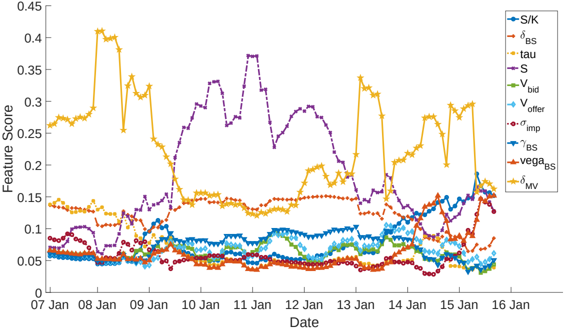
Where does the improvement come from? (2)

Delta	Data-Driven Model(%)							
	Weekly				Monthly			
	MV	$\mathrm{DKL}_{\mathrm{SPL}}$	DNN	DRNN	MV	$\mathrm{DKL}_{\mathrm{SPL}}$	DNN	DRNN
0.1	26.3	38.9	35.6	**47.8**	13.5	22.7	29.7	**53.9**
0.2	21.6	29.0	36.4	**48.5**	16.4	23.5	38.4	**51.7**
0.3	20.1	23.5	38.6	**48.5**	17.9	24.0	40.2	**50.2**
0.4	18.1	20.8	38.7	**45.9**	16.9	21.0	38.6	**47.8**
0.5	16.0	19.9	42.3	**46.6**	15.2	13.5	36.3	**44.5**
0.6	12.1	17.3	43.4	**44.8**	12.7	14.3	36.0	**44.6**
0.7	8.1	16.8	**45.6**	43.9	5.9	6.1	30.2	**35.3**
0.8	3.7	12.5	39.6	**37.7**	-1.2	5.3	22.3	**24.8**
0.9	2.4	6.2	**26.3**	16.4	-1.8	4.1	**21.1**	10.5
Overall	15.1	20.2	39.9	**43.7**	13.4	16.3	35.4	**44.5**

Where does improvement come from? (3)

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**

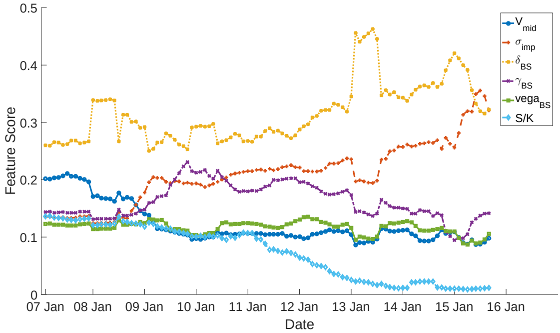
Feature Score (1)



(a) Local Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

Feature Score (2)

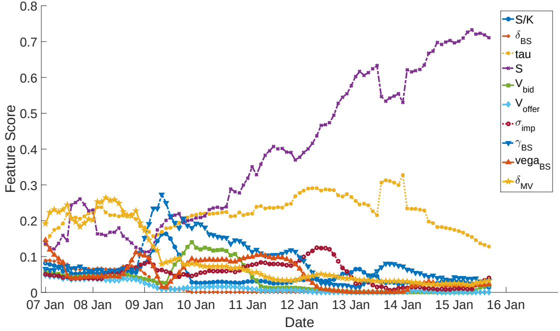


(a) Sequential Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

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

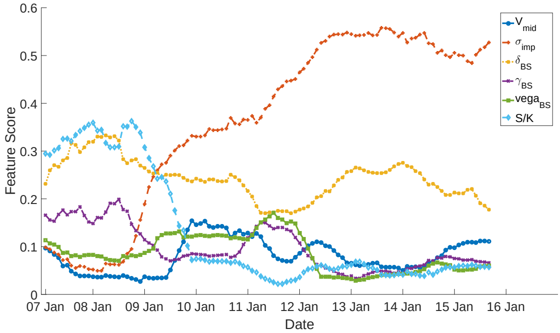
Feature Score (3)



(a) Local Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

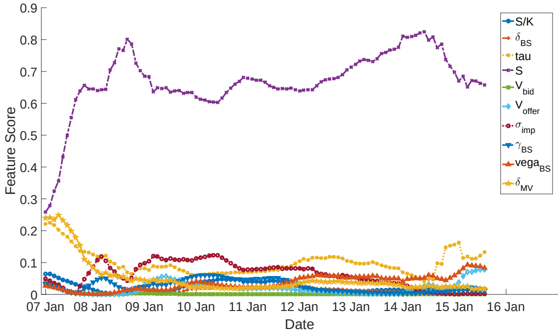
Feature Score (4)



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

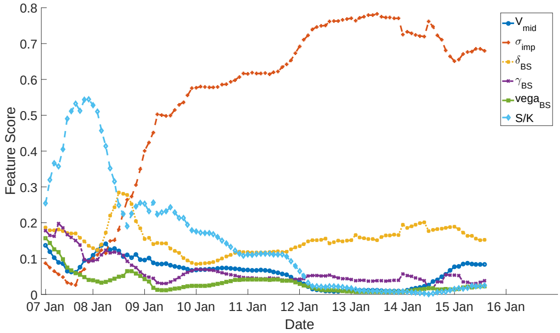
Feature Score (5)



(a) Local Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

Feature Score (6)



(a) Sequential Features

Figure: Feature Score of S&P500 Call Option (Daily Hedging)

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 losses and robust model design are also beneficiary.

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

Q & A

Thank you very much!

Any Questions?