# Chapter 6

# Data-Driven Sequential Learning for Total Hedging Risk

As we have discussed in section 2.1, in reality, one often want to hedge until the expiry of the option. Total risk measures the hedging error from a dynamic hedging strategy in the entire hedging horizon. While mininimizing local risk has the effect of limiting the total hedging error, we investigate here whether minimizing the total hedging risk directly can lead to better total risk mimization strategy. From the model perspective, such enhancement is achievable. Indeed, recently a deep hedging model was proposed to minimize the total option hedging risk evaluated at the option expiry [28]. It is shown that using a RNN to represent the hedging position model can be a computationally efficient framework to determine the optimal hedging function when the market is incomplete, e.g., under the transaction cost [28].

Although minimizing the total hedging risk, which is the hedging portfolio value at the expiry $T$, is more desirable, there are several major obstacles in obtaining enough market data to build a data-driven total hedging model due to the fact that options listed in the exchanges often have fixed expiry dates (e.g., once a month for S&P500 index options). The deep hedging model [28] is only built on synthetic data. Due to the lack of market data needed to build a model, applying the deep hedging model [28] can be challenging in real world applications.

In this chapter, we provide a technique to deal with issue of lack of market data. In addition, an sequential total risk hedging model $\text{GRU}_{\text{TOTAL}}$ is introduced to minimize a discrete total risk hedging objective. We then build the total risk hedging model $\text{GRU}_{\text{TOTAL}}$ based on the data augmentation technique we propose and demonstrate its effectiveness and the performance of the total risk hedging model $\text{GRU}_{\text{TOTAL}}$ using real market data experiments. The goal in this chapter is to learn a hedging model for hedging option from a total hedging horizon of $N_H$ days to expiry.

## 6.1 Total Risk Hedging Model

In this section, we describe the total risk hedging model. Figure 6.1 depicts the proposed total risk GRU hedging model $\text{GRU}_{\text{TOTAL}}$, which is an extension from the local hedging model $\text{GRU}_{\delta}$ we proposed in Chapter 4. This model uses the sequential features, which encode information of two consecutive re-balancing time steps. In addition, the model uses the hedging position from the previous re-balancing time as the input. The output hedging position is used as the input for the next re-balancing time.

Following the discrete total risk definition in section 2.1.2, consider a hedging portfolio which is composed of:

- A short position on option $V_{t,T,K}$.
- $\delta^{M}_{t,T,K}$ shares of $S_t$.
- An amount in a risk-free bank account $B_t$.

As discussed in section 2.1.2, the final hedging portfolio value at $T$ is:

$$\text{Risk}^{total}_{t_0,T,K} = \sum_{j=0}^{N_{rb}-1} \left\{ \left[ \frac{S_{t_{j+1}}}{D(t_{j+1},T)} - \frac{S_{t_j}}{D(t_j,T)} \right] \delta^{M}_{t_j,T,K} \right\} + \frac{V_{t_0,T,K}}{D(t_0,T)} - V_{T,T,K} \tag{6.1.1}$$

where $D(t,T) = e^{-r(T-t)}$ is the discount factor and $\{t_0, t_1, \ldots, t_{N_{rb}-1}\}$ is the set of rebalacing time. Note that, for testing scenarios, following the scenario construction procedure ~~as in Algorithm 7~~, ~~we can guarantee that~~ $V_{t_0,T,K}$, $V_{T,T,K}$, and $S_t$ in equation (6.1.1) are all directly from market ~~instead of model~~. The details of testing scenario is presented in Algorithm 7 in Appendix E.

Now, consider at a rebalancing time $t \in \{t_0, t_1, \ldots, t_{N_{rb}-1}\}$, with a strike $K$, and an expiry $T$. Assume we have computed the hedging position at the previous rebalancing time $t - \Delta t$: $\delta_{t-\Delta t,T,K}$. Let $\Delta t_d$ denote the time interval for sequential information recording. In the subsequent empirical study, the interval $\Delta t_d$ equals one-day. ~~Please note~~ Note that this setting is the same for local risk model described in Chapter 4. We denote the sequential features recording the daily history for hedging the option with expiry $T$ and strike $K$ as

$$\mathbf{Y}^{T,K}_t = \left[ \mathbf{y}^{T,K}_{t-N\Delta t_d}, \ldots, \mathbf{y}^{T,K}_t \right]$$

For notational simplicity, we denote $\check{t}_i = t - (N+1-i)\Delta t_d$ with $i = 1, \ldots, N+1$, we thus have:

$$\mathbf{Y}^{T,K}_t = \left[ \mathbf{y}^{T,K}_{\check{t}_1}, \ldots, \mathbf{y}^{T,K}_{\check{t}_{N+1}} \right]$$

The vector $\mathbf{y}^{T,K}_{\check{t}_i} \in \mathbb{R}^{d_s}$ has $d_s$ features at time $\check{t}_i$ in the input sequential feature where $d_s$ is the dimension of the sequential feature $\mathbf{Y}^{T,K}_t$, and $N+1$ is the length of the sequential feature sequence.

We set $N = \frac{\Delta t}{\Delta t_d}$. Thus $\mathbf{Y}_t^{T,K}$ contains the sequential information in between two consecutive rebalancing time $t$ and $t - \Delta t$. The encoder transforms information from the sequential feature $\mathbf{Y}_t^{T,K}$ to a fixed-sized vector $\widehat{\mathbf{h}}_E$ and the decoder makes the final prediction based on both $\widehat{\mathbf{h}}_E$ and the previous hedging position $\delta^M_{t-\Delta t,T,K}$. The overall structure of the proposed model is illustrated in Figure 6.1. Note that for the initial hedging date $t = t_0$, the previous hedging position is set as 0.

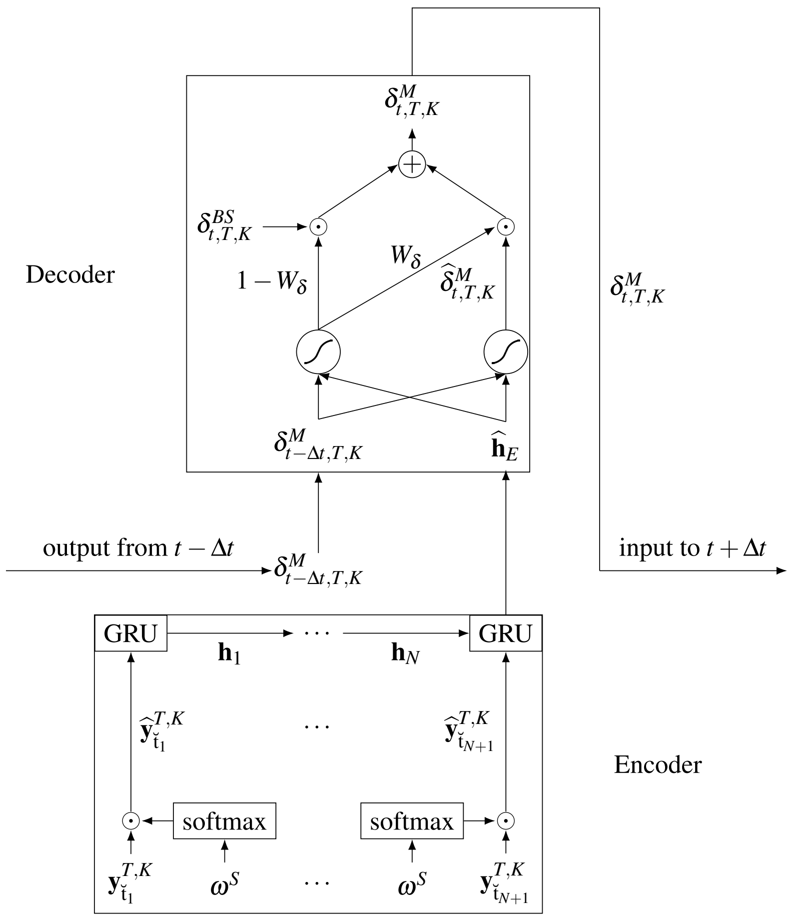



Figure 6.1: $\mathrm{GRU}_{\mathrm{TOTAL}}$: GRU encode-decoder total hedging model. The encoder summarizes the time series $\mathbf{Y}_t^{T,K} = \left[\mathbf{y}^{T,K}_{\check{t}_1}, \ldots, \mathbf{y}^{T,K}_{\check{t}_{N+1}}\right]$ as a succinct vector $\widehat{\mathbf{h}}_E$. The decoder outputs the hedging position based on the vector $\widehat{\mathbf{h}}_E$ and the previous hedging position $\delta^M_{t-\Delta t,T,K}$ observed at the hedging time $t$. More specifically, in the decoder, a candidate output $\widehat{\delta}^M_{t,T,K}$ is firstly produced. The final output $\delta^M_{t,T,K}$ is computed based on the linear combination of BS delta $\delta^{BS}_{t,T,K}$ and the candidate output $\widehat{\delta}^M_{t,T,K}$. The combination weight is determined by $W_\delta$. The feature weight $\omega^L$ is used to compute weighted sequential feature $\widehat{\mathbf{y}}^{T,K}_{\check{t}}$. The weighting acts as a feature selection process. Each edge in the graph has an arrow on it, pointing from a node whose output is used by the node pointed by the arrow as an input. The output $\delta^M_{t,T,K}$ at $t$ is used as the input for next step $t + \Delta t$.

### 6.1.1 The Difference Between GRU$_\delta$ And GRU$_{\text{TOTAL}}$

Comparing GRU$_\delta$ in Figure 4.1 and GRU$_{\text{TOTAL}}$ in Figure 6.1, it is clear that the model structures between GRU$_\delta$ and GRU$_{\text{TOTAL}}$ are similar. The more details of the model structure of GRU$_{\text{TOTAL}}$ is discussed in Appendix D.

Below we firstly discuss the differences between GRU$_{\text{TOTAL}}$ and GRU$_\delta$:

- The most significant difference between GRU$_{\text{TOTAL}}$ and GRU$_\delta$ is that GRU$_\delta$ is built on minimizing discrete local risk as in section 2.1.1 while GRU$_{\text{TOTAL}}$ is built based on minimizing the discrete total risk as in in section 2.1.2. The details of training objective and training procedure for GRU$_{\text{TOTAL}}$ are in later section 6.1.2 and 6.1.5.
- In GRU$_\delta$, we have a local features vector $\mathbf{x}_t^{T,K} \in \mathbb{R}^{d_l}$ which records local information at the hedging time $t$ for hedging the option with expiry $T$ and strike $K$.
- In GRU$_{\text{TOTAL}}$, we no longer retain the local features vector $\mathbf{x}_t^{T,K} \in \mathbb{R}^{d_l}$ as the input to the model GRU$_{\text{TOTAL}}$. We only retain the sequential feature $\mathbf{Y}_t^{T,K}$ which contains all the sequential information between two consecutive rebalancing time $t$ and $t-\Delta t$. We feel that the inclusion of $\mathbf{x}_t^{T,K} \in \mathbb{R}^{d_l}$ is a redundant design since the sequential feature $\mathbf{Y}_t^{T,K}$ contains the local information at the hedging time $t$ for hedging the option with expiry $T$ and strike $K$. This redundant design is retained in Chapter 4 because we want to demonstrate the importance of sequential learning by comparing the performance GRU$_\delta$, which contains the feature extraction part for both sequential features and local features, and NN$_\delta$, which is a simple model that only retains the local feature part.
- Since the analytical formula for the variance-optimal total risk hedging [164] and spline total risk minimization formulation [50] both demonstrate the dependence of the current hedging position on the past hedging position, we include previous hedging position $\delta^M_{t-\Delta t,T,K}$ as the input to compute the current hedging position $\delta^M_{t,T,K}$.

### 6.1.2 Training Objective For GRU$_{\text{TOTAL}}$

Assume that we are given a set of hedging scenarios identified by the expiry date $T_i$ and strike $K_i$:

$$\{Scenario(T_1,K_1),\ldots,Scenario(T_M,K_M)\}$$

A natural total risk hedging loss function is the mean squared error:

$$MSE_{total} = \frac{1}{M}\sum_{i=1}^{M}\left(\text{Risk}^{total}_{t_0^i,T_i,K_i}\right)^2$$

where $\text{Risk}^{total}_{t_0,T,K}$ is defined as in equation (6.1.1), $t_0^i = T_i - \frac{100}{250}$, and $M$ is the number of hedging scenarios. A more appropriate criteria however is the relative hedging error instead of the absolute total hedging error since we are mixing scenarios of different strikes together (i.e., the scenarios include near-money, in-the-money, and out-of-the money options. The absolute hedging error for in-the-money options tend to be much bigger than those of near-the-money options and out-of-money options). Additionally, MSE is sensitive to the existence of outliers. Furthermore, Coleman et al. [50] demonstrated that the linear loss function can produce better hedging performance. Therefore, in training the GRU$_{\text{TOTAL}}$, we use the following objective:

1- norm error

$$Obj_{total} = \sum_{i=1}^{M} \left| \text{Rel}^{total}_{t_0^i, T_i, K_i} \right| \tag{6.1.2}$$

where the relative total hedging error is defined as:

$$\text{Rel}^{total}_{t_0,T,K} = \frac{D(t_0, T_i)\text{Risk}^{total}_{t_0,T,K}}{V_{t_0,T,K}} \tag{6.1.3}$$

### 6.1.3 Market Data Augmentation

As discussed in section 2.1.2, the final hedging portfolio value at $T$ is:

$$\text{Risk}^{total}_{t_0,T,K} = \sum_{j=0}^{N_{rb}-1} \left\{ \left[ \frac{S_{t_{j+1}}}{D(t_{j+1},T)} - \frac{S_{t_j}}{D(t_j,T)} \right] \delta^M_{t_j,T,K} \right\} + \frac{V_{t_0,T,K}}{D(t_0,T)} - V_{T,T,K} \tag{6.1.4}$$

where $D(t,T) = e^{-r(T-t)}$ is the discount factor and $\{t_0, t_1, \ldots, t_{N_{rb}-1}\}$ is the set of rebalacing time. The $t_0$ in this thesis is set to be $N_H$ days to the expiry where $N_H$ is a constant.

Suppose we have $M$ hedging scenarios, a data driven model can be built by minimizing total hedging loss function such as MSE:

$$MSE_{total} = \frac{1}{M} \sum_{i=1}^{M} \left(\text{Risk}^{total}_{t_0^i,T_i,K_i}\right)^2$$

Clearly, in order to build an effective data-driven total risk hedging model, one needs to gather a sufficient number $M$ of sample paths.

For local risk hedging, each data instance is uniquely determined by the triplet $\{t, T, K\}$. We only need the market option prices of strike $K$ and expiry $T$ to be available on both $t$ and $t + \Delta t$ to gather a data instance for the data-driven local hedging model. Therefore, purely building a model based on market prices for local risk hedging is achievable. When we build the local risk hedging model, the time to maturity $\tau = T - t$ is not fixed.

For total risk hedging, if we assume the starting hedging date $t_0$ is $N_H$ days to expiry $T$, each data instance is then uniquely determined by the duplet $\{T, K\}$. If we only rely on market prices to build the total risk model, we will face the following challenges:

1. Options listed in an exchange only have fixed expiry dates and only a few strikes are listed every day. In other words, the number of unique duplet $\{T, K\}$ in actual market is small. For example, the expiration date for the standard S&P 500 index option is fixed at the third Friday of each month.[1] Therefore, if we only use market available expiration dates, the number of training scenarios will be severely limited.

2. In addition, the option with specific strike $K$ and expiry $T$ may not be traded on every trading date during its lifetime. Requiring the market option prices of strike $K$ and expiry $T$ to be available on entire hedging horizon $[t_0, T]$ is unrealistic especially for in-the-money and out-of-money options. Therefore, we will have to rely on certain parametric models calibrated to market prices to compute the necessary derived features (e.g., option sensitivity), which is used as the input of the data-driven model, when there are not market prices for the specific combination of $T$ and $K$ on some trading dates.

3. Lastly, under an ideal setting, one should use non-overlapping underlying asset paths to generate training and testing hedging scenarios as one often observe autocorrelation in financial time series. In reality, using non-overlapping underlying asset path to generate training and testing scenarios is not practical. For instance, assuming we are building a data-driven model for hedging 3 months until expiry, with non-overlapping underlying asset paths, 20 years of market data can provide only 80 different non-overlapping underlying asset paths, making building a data-driven model difficult.

In order to overcome above challenges, we ~~propose~~ use the following remedy. to augment training data

1. Overlapping underlying asset paths are allowed.

2. A no-arbitrage surface is calibrated on each business day to match the market prices. We will query the calibrated surface to obtain the option prices and option sensitivities when the corresponding market data is not available.

3. Instead of using only market available expiration dates, we assume every business day can be the expiration date of the options. The option prices and option sensitivities for these newly added expiry dates will come from querying the no-arbitrage surface.

Therefore, we greatly increase the number of training scenarios, since now the hedging scenarios can have more combinations of $T$ and $K$ even if the combinations of $T$ and $K$ are not directly observed in the market. We create a parametrization of the price surface that is **arbitrage-free**. This is done by using SABR model to match the market volatility smiles and we use an arbitrage-free interpolation based on Local Volatility Function (LVF) model to interpolate the SABR model value between expiries.

[1]Starting from 2016, CBOE also introduces weekly S&P 500 index options which expired on Monday, Wednesday and Friday of each week. The data we gathered is up to 2015-08-31, so our experiments in this thesis does not include those data.

calibrated

As an example, in Figure 6.2, we show the ~~resulting~~ price surface and implied volatility surface for SP500 index call options on 2012-01-04. In Figure 6.2, we use $\tau = T - t$ to denote the time to maturity.

We refer an interested reader to Appendix §E.3 for details of no arbitrage surface calibration.

### 6.1.4 Training and Testing Data Construction

~~In this section, we will discuss the training and testing data construction for building the total hedging model with the surfaces constructed from SABR and LVF model.~~ The training and testing data set are the collection of hedging scenarios. Each of the hedging scenarios is identified uniquely by a duplet of expiry date and strike $(T, K)$. [2] Let $t_0$ be the initial date to set up the hedging portfolio and $\mathbf{t}_B = \{t_0, \ldots, T\}$ be the set of all business days in between the initial date $t_0$ and expiry date $T$. Each of the hedging scenarios is the collection of the following time series identified uniquely by $(T, K)$:

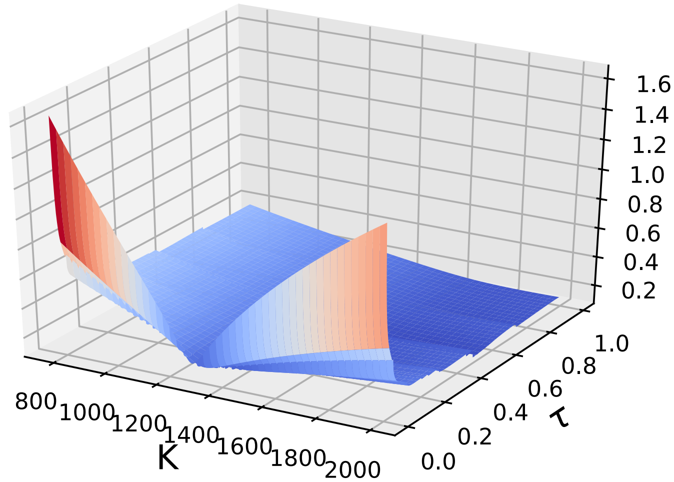
- $\{S_t | \forall t \in \mathbf{t}_B\}$: the the time-series of underlying prices with $t_0$ as the initial date and $T$ as the expiry date.
- $\{V_{t,T,K} | \forall t \in \mathbf{t}_B\}$: the time-series of option value for a hedging scenario identified by $(T, K)$.
- $\{\mathbf{y}_t^{T,K} | \forall t \in \mathbf{t}_B\}$: the time-series of feature vectors for a hedging scenario identified by $(T, K)$.

#### 6.1.4.1 Construction of Training Scenarios

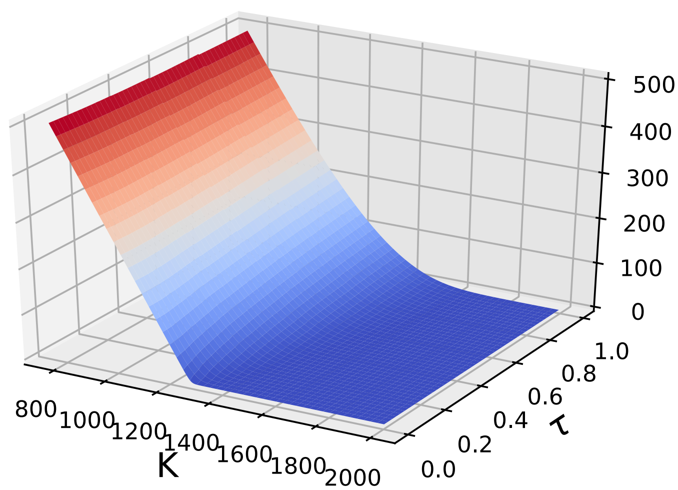
Based on the calibration process ~~we~~ discussed above, we summarize the procedure for generating training scenarios for hedging $N_H$ business days as the following.

1. Each of the hedging scenarios is identified uniquely by a duplet of expiry and strike $(T, K)$. We ~~need to~~ construct the time series corresponding to the duplet as the input to our model. For a training scenario, we do not require the option prices for the duplet $(T, K)$ to exist in market. We will query the surface constructed by Algorithm 5 for each time $t$: $V^t_{model}(T, K)$ for option values and compute the associated option related sensitivities, which are used to construct the feature vector for the model. In this way, we greatly increase the number of training scenarios.
2. The starting date $t_0$ to set up the initial hedging portfolio is $N_H$-business days away from the expiry date $T$. In this thesis, $N_H$ is set to be 100. We assume there are 250 business

[2] If we fixed the gap between two rebalancing time to be $\Delta t$ and we fix the number of times we rebalance the hedging portfolio to be $N_{rb}$, then given an expiry $T$, one can easily deduct the rebalance time $\{t_0, t_1, \ldots, t_{N_{rb}-1}\}$ . Therefore, we can uniquely define a hedging scenario just by $T$ and $K$.




(a) Implied Volatility Surface




(b) Price Surface

Figure 6.2: The price surface and implied volatility surface calibrated to SP500 index call options on 2012-01-04. The smile is more pronounced for the shorter maturities than the longer maturities, which is consistent with observations from various studies [34, 157] using market data. Interested readers can also refer to [157] for some mathematicaly explaination on why the smile becomes more and more flattened as $\tau$ increases.

days in a year. Denote $T^{mkt}_{t,max}$ as the maximum expiry listed ~~in exchange~~ on CBOE at time $t$. We further comment that we always have $T^{mkt}_{t,max} - t > 100/250$ in market for all the business dates we include in the experiments. Therefore, no volatility extrapolation is needed.

3. We check all the market observed strikes for all the business dates between $t_0$ and $T$. Let $K^{mkt}_{max}(t_0, T)$ be the maximum of strikes we observed in market between $t_0$ and $T$. The grid of strikes for the outputting option values with expiry date $T$ is defined as: $\mathbf{K}_{grid}(t_0, T) = \{0 = K_0 < K_1 < \cdots < 2 * K^{mkt}_{max}(t_0, \widehat{T})\}$ where $K_i - K_{i-1} = \Delta K, i \geq 1$. In this chapter, we set $\Delta K = 5$ for experiments on S&P 500 index options which is consistent with the S&P500 index option strike specification in real market [113].

A detailed description of training scenario generation is given in Algorithm 6 in Appendix §**??**.

#### 6.1.4.2 Construction of the Testing Scenarios

The key differences between the testing scenarios and the training scenarios are:

- For testing scenarios, we use real market prices to initialize the hedging portfolio at $t_0$ and the strikes and expiries are real market strikes and expiries.
- For testing scenarios, we use real option market data to construct the time-series whenever associated option market data is available. Otherwise, we will query the calibrated option value function.

  you don't query for testing data?
- For training scenarios, we use model prices to initialize the hedging portfolio. The strikes and expiries do not necessarily have to be are real market strikes and expiries.
- For training scenarios, we always query the parametrization of the option value we calibrated to construct the time series. Please note that since we calibrate the models to match the market prices, when market prices are available, the model prices will be very close to the market prices.

A summary of the construction of testing scenarios is as the following. A detailed algorithm is given in Algorithm 7.

1. A testing expiry date $T$ must be a real expiry date that exists in market. Before 2016, For the S&P 500 index options listed in Chicago Board Options Exchange (CBOE) the expiration dates are the third Fridays of each month. After 2016, more expiration dates are introduced in CBOE.

2. The date $t_0$ to set up the initial hedging portfolio is $N_H$-business days away from $T$.

3. On the starting date $t_0$, we can obtain all market option prices for the expiry $T$ and we have a grid of market strikes for expiry $T$ on date $t_0$: $\mathbf{K}^{mkt}_{grid}(t_0,T) = \{K^{mkt}_{t_0,T,1},\dots,K^{mkt}_{t_0,T,N_K}\}$. Note that we have market options prices for all $K \in \mathbf{K}^{mkt}_{grid}(t_0,T)$ at time $t_0$.

4. On each business day $t$ in between $t_0$ and $T$, an arbitrage free surface is constructed $\{V^t_{model}(T,K)\}_{T,K}$ using Algorithm 5. When there is no market price $V^{mkt}_{t,T,K}$ on time $t$ for $K \in \mathbf{K}^{mkt}_{grid}(t_0,T)$, we will query $\{V^t_{model}(T,K)\}_{T,K}$ to obtain option prices and option sensitivities. Note that, with this approach, for instance, a part of the time series of option prices for a testing scenario can be real market prices while the other part can be model prices from the parametrization obtained following the calibration process in section E.3.

#### ~~6.1.4.3 Construction of Training, Testing and Validation Data Sets~~

How often update?

each model update has time horizon = NH days ?

For the experiments in this chapter, we are hedging for a relatively long period (100 business days) until the expiry. Empirically, we have observed that if we do not update the data-driven model during the hedging period, the performance from the data-driven hedging would be much worse than the performance of the traditional parametric hedging models such as hedging with delta produced by Black-Scholes implied volatility. This is not surprising since market can drastically change during the hedging period which is relatively long and we cannot assume one set of parameter for a data-driven model to be effective for such long period. Lastly, by comparing the performance of local risk model $\mathrm{DKL_{SPL}}$, which is updated on a monthly basis, with those of $\mathrm{NN}_\delta$ and $\mathrm{GRU}_\delta$ ,which are updated on a daily basis in Chapter 5, we can already see the effectiveness of more frequent update. Therefore, the training and validation data set are updated as we move from one rebalancing date to another rebalancing date. We denote the total risk hedging model as $\mathrm{GRU_{TOTAL}}$. The detailed description of the model $\mathrm{GRU_{TOTAL}}$ will be discussed in the following section 6.1.

~~A summary of the construction procedure is given as the following~~. The detailed procedure of the construction of training, testing and validation dataset and the overall model building procedure is given in Algorithm 8 in Appendix **??**.

### 6.1.5 Training Procedure For $\mathrm{GRU_{TOTAL}}$

We initialize the GRU parameters using the same procedure as in section 4.2.1 and we pretrain the $\mathrm{GRU_{TOTAL}}$ similarly as in section 4.2.3. Early stopping is used as the regularization techniques. We reserve the validation set to determine when to stop ~~the~~ training. We train $\mathrm{GRU_{TOTAL}}$until the trust region algorithm~~s~~ (i.e., Algorithm 1) stops and slect the best performing model parameters on validation set based on the the total risk objective (6.1.2). The parameters for the trust region algorithm ~~is~~ are in Table 4.1. Early stopping is used as ~~the~~ regularization, which is the same as in chapter 4. The overall model building procedure is given in Algorithm 8. in Appendix Ref

## 6.2 Total Discrete Hedging Performance Comparison Using S&P 500 index Options

Using the S&P 500 (European) index option market data from September 1, 1996 to August 31, 2015[3], we compare the total hedging performance of different hedging strategies. We evaluate the total hedging performance using the following 5 criteria:

1. The mean absolute value of the relative hedging error:

$$Mean_{(t_0,T,K)}\left(\left|\mathrm{Rel}^{total}_{t_0,T,K}\right|\right)$$

   for all the testing scenarios.

2. The 95% Value-at-Risk (VaR) of the relative total hedging error $\mathrm{Rel}^{total}_{t_0,T,K}$

3. The 95% Conditional-Value-at-Risk (CVaR) of the relative total hedging error $\mathrm{Rel}^{total}_{t_0,T,K}$

4. The 99% Value-at-Risk (VaR) of the relative total hedging error $\mathrm{Rel}^{total}_{t_0,T,K}$

5. The 99% Conditional-Value-at-Risk (CVaR) of the relative total hedging error $\mathrm{Rel}^{total}_{t_0,T,K}$

### 6.2.1 Data and Experimental Setting

The sequential inputs to $\mathrm{GRU}_{\mathrm{TOTAL}}$, $\mathbf{Y}^{T,K}_t$, at a rebalancing time $t$, are the time series recorded daily from previous rebalancing time $t-\Delta t$ to current rebalancing time $t$ for the following features:

The number of hidden states, for the single-layer GRU encoder, the neural network outputting $\widehat{\delta}^M_{t,T,K}$, and the neural network outputting $W_\delta$ in Figure 6.1, are all set to be 5. Specifically, we compare with the following methods,

- $\mathrm{GRU}_{\mathrm{TOTAL}}$: the model shown in Figure 6.1 and is trained with the total risk objective (6.1.2).
- Bartlett: Barlett corrective delta based on (2.3.18),
- BS: BlackScholes delta based on the implied volatility

The hedging period is fixed to be 100 business days. We have two different hedging frequencies: weekly and monthly hedging. For weekly hedging, we rebalance every 5 business days so the number of rebalacing times is $N_{rb} = 20$. For monthly hedging, we rebalance every 20 business days so the number of rebalacing times is $N_{rb} = 5$.

[3] The option historical data from OptionMetric [147] started on September 1, 1996. Due to the limits of data license, we only have access to OptionMetric up to August 31, 2015.

rebalancing

| |
|---|
| Option price |
| BlackScholes implied volatility |
| BlackScholes delta |
| BlackScholes vega |
| Bartlett delta |
| Time to expiry |
| S&P 500 index price |
| VIX index price |
| Moneyness $S/K$ |
| Minimum variance delta $\delta_{MV}$ (2.3.5) |
| Strike $K$ |

Table 6.1: Sequential features for $\mathrm{GRU_{TOTAL}}$ and $\mathrm{GRU^{LOCAL}_{TOTAL}}$ at time $t$ are the time series of features listed in this table. The time series are constructed according to the procedures as in Algorithm 6 and Algorithm 7.

### 6.2.2 Call Option Total Hedging Comparison

In this subsection, we present the results for call options. We show the hedging performance for Near-The-Money(NTM), In-The-Money(ITM), Out-of-The-Money(OTM) separately. Note that we are not training models for NTM, ITM and OTM separately. We still train the model using all training set. The NTM, OTM, and ITM scenarios are classified based on the Black-Scholes delta at the initial date $t_0$ where we set up the hedging portfolio: $\delta^{BS}_{t_0,T,K}$. For call option, the criteria is:

- NTM: $0.3 \leq \delta^{BS}_{t_0,T,K} < 0.7$
- ITM: $0.7 \leq \delta^{BS}_{t_0,T,K} < 0.95$
- OTM: $0.05 \leq \delta^{BS}_{t_0,T,K} < 0.3$

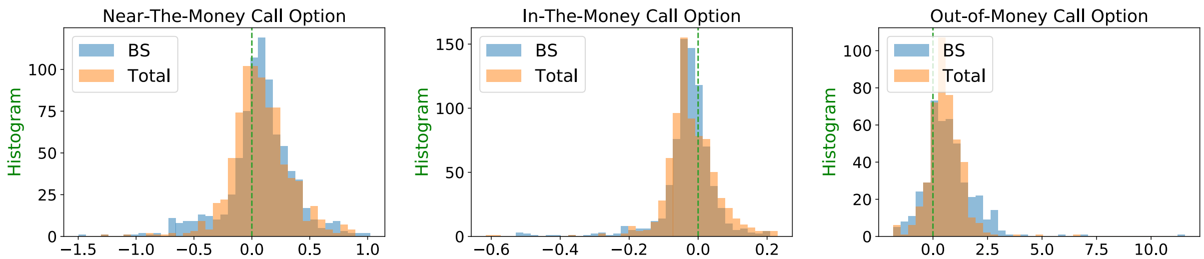
$0.95 \leq \delta^{BS}_{t_0,T,K} \leq 1$ $\quad$ $0. \leq \delta^{BS}_{t_0,T,K} \leq 0.05$

We omit the testing scenarios for deep in-the-money and deep out-of-the money options due to the fact that they are highly illiquid in market and their market quotes are highly unreliable. Also, the deep in-the-money and deep out-of-the money scenarios are deleted from training set and validation set.

- ~~Deep ITM: $0.95 \leq \delta^{BS}_{t_0,T,K} < 1.0$~~
- ~~Deep OTM: $0.0 \leq \delta^{BS}_{t_0,T,K} < 0.05$~~

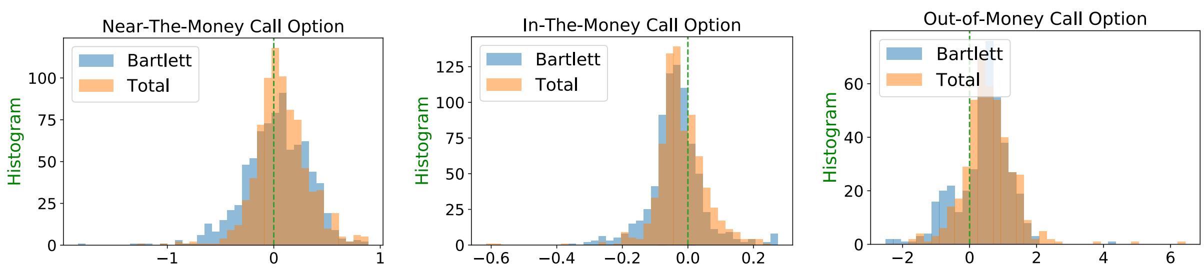
#### 6.2.2.1 Call Option Weekly Hedging Comparison

In Table 6.2, we demonstrate the results on weekly hedging call options. Furthermore, in Figure 6.3, and Figure 6.4, we compare the distribution of the relative hedging error ~~of~~ using $\mathrm{GRU_{TOTAL}}$ with ~~the distributions of~~ corresponding distributions from using the BS model and the Bartlett model respectively.




(a) (b) (c)

Figure 6.3: Comparing total risk hedging model $\mathrm{GRU}_{\mathrm{TOTAL}}$ and BS Model on weekly hedging S&P 500 call options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.




(a) (b) (c)

Figure 6.4: Comparing total risk hedging model $\mathrm{GRU}_{\mathrm{TOTAL}}$ and Bartlett model on weekly hedging S&P 500 call options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.

| | | Near-The-Money | In-The-Money | Out-of-The-Money |
|---|---|---|---|---|
| Mean Abs Relative Error | $\text{GRU}_{\text{TOTAL}}$ | **0.1927** | 0.0571 | **0.7344** |
| | Bartlett | 0.2347 | 0.0641 | 0.7383 |
| | BS | 0.2198 | **0.0531** | 0.9706 |
| VaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | **0.2827** | **0.1121** | **0.5298** |
| | Bartlett | 0.4836 | 0.1656 | 0.9841 |
| | BS | 0.4823 | 0.1523 | 0.8603 |
| CVaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | **0.4721** | **0.1865** | **1.0003** |
| | Bartlett | 0.7103 | 0.2192 | 1.4232 |
| | BS | 0.7009 | 0.2724 | 1.2299 |
| VaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | **0.5301** | **0.1976** | 1.5077 |
| | Bartlett | 0.778 | 0.2654 | 1.6152 |
| | BS | 0.7171 | 0.3653 | **1.3363** |
| CVaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | **0.8205** | 0.3261 | 1.6090 |
| | Bartlett | 1.0827 | **0.2883** | 2.1225 |
| | BS | 1.0040 | 0.4712 | **1.6074** |

Table 6.2: Summary of weekly hedging S&P 500 call options (testing set) for 100 business days with total hedging evaluation criteria described in section 6.2. Please note that the total hedging evaluation in this table assumes we are at the sell-side of the option trading.

From Table 6.2, we can see that, $\text{GRU}_{\text{TOTAL}}$ performs better than the other methods in terms of mean absolute relative hedging error for NTM and OTM scenarios, except that BS model performs slightly better than $\text{GRU}_{\text{TOTAL}}$ for ITM scenarios. The difference between $\text{GRU}_{\text{TOTAL}}$ and BS model is small for ITM scenarios. Additionally, $\text{GRU}_{\text{TOTAL}}$, in most of cases, perform the best in terms of VaR and CVaR, indicating that $\text{GRU}_{\text{TOTAL}}$ performs better in reducing the tail loss, ~~The~~ with exceptions ~~is given as the following~~ given below:

- BS model performs best in terms of VaR(99%) and CVaR(99%) for OTM scenarios. This is an interesting observation. From 6.3 (c), We also note that the extreme tail on the profit side from BS model on OTM scenarios is actually longer than the other three models, indicating a larger probability of getting profit. However, we notice that the difference in CVaR(99%) for OTM scenarios between $\text{GRU}_{\text{TOTAL}}$ and BS model is small.

- Bartlett model performs the best in terms of CVaR(99%) for ITM scenarios. However CVaR(99%) of $\text{GRU}_{\text{TOTAL}}$ is only slightly worse than that of the Bartlett model meanwhile $\text{GRU}_{\text{TOTAL}}$ perform still the best in terms of the VaR(99%).

Another interesting observation is that Bartlett delta actually performs worse than BS delta in most of the cases as shown in Table 6.2. We suspect that this is due to the fact that SABR model was originally designed for modeling interest rate derivatives, the time to maturity for which is usually ~~bigger~~ longer than one-year, and it is less suitable to model option surface with extreme short time to maturity[35]. For weekly hedging, we have used SABR model to produce Bartlett delta with extremely small time to maturity, e.g., 5/250 for the last rebalancing time. Notice that in chapter 5 when we compare models on local risk criteria, options with time-to-expiry less than 14 days are removed from the data set. Therefore, we did not notice this phenomenon.

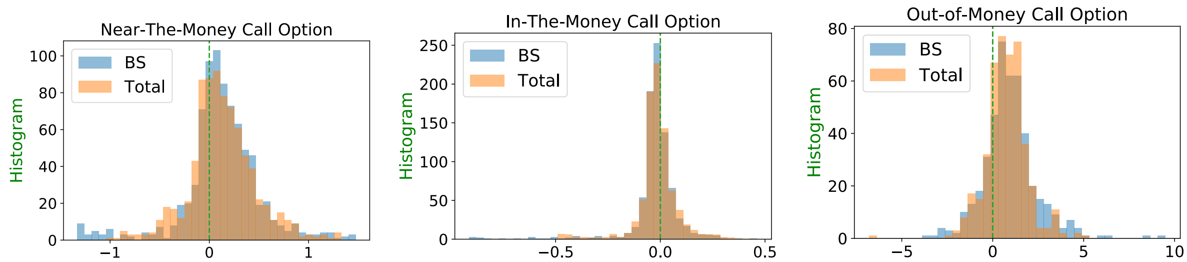
#### 6.2.2.2 Call Option Monthly Hedging Comparison

In Table 6.3, we demonstrate the results on monthly hedging call options. Furthermore, in Figure 6.5 and Figure 6.6, we compare the distribution of the relative monthly hedging error with the distributions ~~of~~ from the BS model and Bartlett model respectively.

From Table 6.3, we can see that, $\text{GRU}_{\text{TOTAL}}$ performs better than the other methods in terms of mean absolute relative hedging error for NTM and ITM scenarios. Bartlett method performs best in terms of mean absolute relative hedging error for OTM scenarios. In terms of VaR and CVaR, by comparing Table 6.3 and Table 6.2, $\text{GRU}_{\text{TOTAL}}$ is less dominant in monthly hedging call options than in weekly hedging call options. Bartlett delta produces best VaR(99%) and CVaR(99%) for NTM scenarios. However, from Table 6.3 we can also see, the performance of $\text{GRU}_{\text{TOTAL}}$ is very close to best performance even if $\text{GRU}_{\text{TOTAL}}$ is not the dominant model in terms of certain criteria.

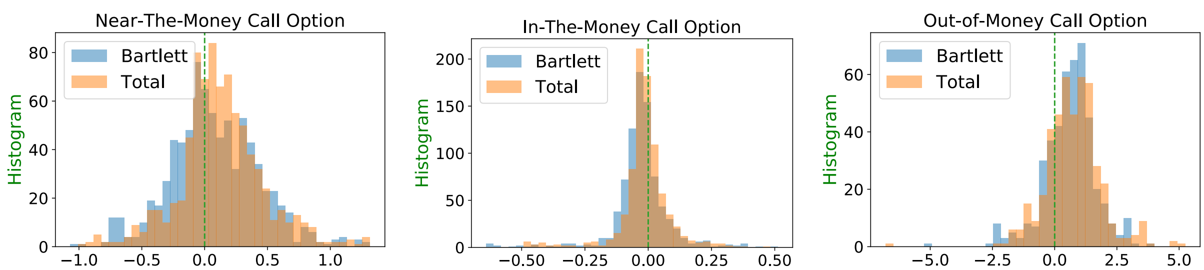
| | | Near-The-Money | In-The-Money | Out-of-The-Money |
|---|---|---|---|---|
| Mean Abs Relative Error | $\text{GRU}_{\text{TOTAL}}$ | **0.2643** | **0.0633** | 1.0479 |
| | Bartlett | 0.282 | 0.073 | **0.9674** |
| | BS | 0.2865 | 0.0655 | 1.3248 |
| VaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | **0.4102** | **0.1472** | **1.0842** |
| | Bartlett | 0.4775 | 0.1611 | 1.1936 |
| | BS | 0.5115 | 0.1554 | 1.3442 |
| CVaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | **0.6073** | **0.3125** | **1.6658** |
| | Bartlett | 0.6680 | 0.3372 | 2.0221 |
| | BS | 0.9735 | 0.4380 | 2.0016 |
| VaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | 0.7752 | **0.4300** | **1.7567** |
| | Bartlett | **0.7201** | 0.4815 | 2.3799 |
| | BS | 1.2384 | 0.6058 | 2.8419 |
| CVaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | 0.8692 | **0.4627** | **2.7536** |
| | Bartlett | **0.8090** | 0.5725 | 2.8797 |
| | BS | 1.2864 | 0.7859 | 3.0839 |

Table 6.3: Summary of monthly hedging S&P 500 call options (testing set) for 100 business days with total risk hedging evaluation criteria described in section 6.2. The total hedging evaluation in this table assumes we are on the sell-side of the option trading.




(a) (b) (c)

Figure 6.5: Comparing total risk hedging model $GRU_{TOTAL}$ and BS model on monthly hedging S&P 500 call options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.




(a) (b) (c)

Figure 6.6: Comparing total risk model $GRU_{TOTAL}$ and bartlett model on monthly hedging S&P 500 call options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are at the sell-side of the option trading.

### 6.2.3 Put Option Total Risk Hedging Comparison

In this subsection, we present the results for put options. We again show the hedging performance for Near-The-Money(NTM), In-The-Money(ITM), Out-of-The-Money(OTM) separately. The NTM, OTM, and ITM scanerios are classified based on the Black-Scholes delta at the initial date $t_0$ where we set up the hedging portfolio: $\delta^{BS}_{t_0,T,K}$. For put option, the criteria is:

- NTM: $-0.3 \geq \delta^{BS}_{t_0,T,K} > -0.7$
- ITM: $-0.7 \geq \delta^{BS}_{t_0,T,K} > -0.95$
- OTM: $-0.05 \geq \delta^{BS}_{t_0,T,K} > -0.3$ $\quad 0 \geq \delta^{BS}_{t_0,T,K} > -0.05$

$-0.95 > \delta^{BS}_{t_0,T,K} > -1$

We omit the testing scenarios for deep in-the-money and deep out-of-the-money options due to the fact that they are highly illiquid in market and their market quotes are highly unreliable. Also, the deep in-the-money and deep out-of-the money scenarios are deleted from training set and validation set.

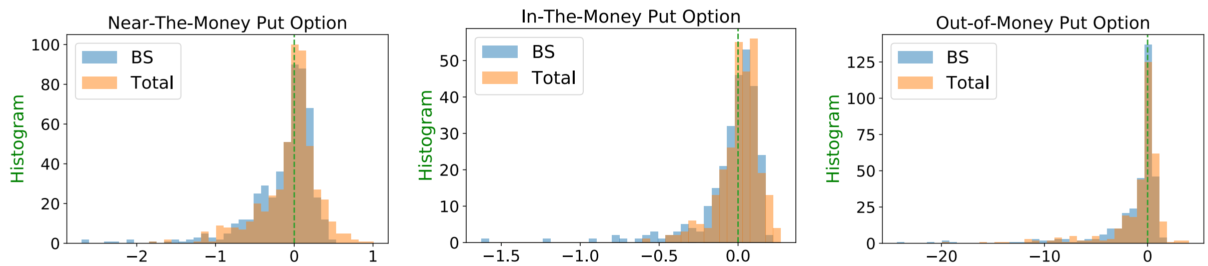
- Deep OTM: $0.0 \geq \delta^{BS}_{t_0,T,K} > -0.05$
- Deep ITM: $-0.95 \geq \delta^{BS}_{t_0,T,K} > -1.0$

#### 6.2.3.1 Put Option Weekly Hedging Comparison

In Table 6.4, we demonstrate the results on ~~monthly~~ weekly hedging put options. Furthermore, in Figure 6.8 and Figure 6.7, we compare the distribution of the relative hedging error of $\text{GRU}_{\text{TOTAL}}$ with the distributions of the relative hedging error of the BS model and the Bartlett model respectively.

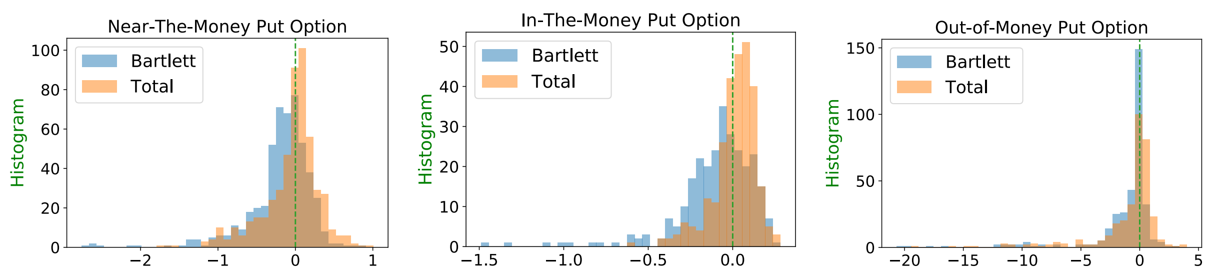
From Table 6.4, we can see that, $\text{GRU}_{\text{TOTAL}}$ performs better for weekly hedging put options in terms of most of the criteria for NTM, ITM, and OTM options. There is one exceptions: Bartlett delta performs slightly better than $\text{GRU}_{\text{TOTAL}}$ for OTM scenarios in terms of mean absolute relative hedging error.

Another interesting observation is that, for put options, the loss tail of the relative hedging distribution is significantly longer than call option. We suspect this is due to the fact that selling put options during market crisis period can lead to significant losses as the original OTM options can become ITM in a short period of time, especially when we are getting closer to the expiry. For weekly hedging put options, it is worth to note that the tail loss, which is measured by VaR and CVaR, from $\text{GRU}_{\text{TOTAL}}$ is significantly smaller than those from the BS model and the Bartlett model.




(a) (b) (c)

Figure 6.7: Comparing total risk hedging model $\mathrm{GRU}_{\mathrm{TOTAL}}$ and BS model on weekly hedging put options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.




(a) (b) (c)

Figure 6.8: Comparing total risk hedging model $\mathrm{GRU}_{\mathrm{TOTAL}}$ and Bartlett model on weekly hedging put options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.

| | | Near-The-Money | In-The-Money | Out-of-The-Money |
|---|---|---|---|---|
| Mean Abs Relative Error | $GRU_{TOTAL}$ | **0.2535** | **0.0965** | 1.5356 |
| | Bartlett | 0.2993 | 0.167 | **1.4815** |
| | BS | 0.2773 | 0.1227 | 1.7109 |
| VaR (95%) | $GRU_{TOTAL}$ | **0.8124** | **0.2364** | **7.2478** |
| | Bartlett | 0.9374 | 0.5133 | 8.5614 |
| | BS | 0.8854 | 0.4274 | 8.7374 |
| CVaR (95%) | $GRU_{TOTAL}$ | **1.0475** | **0.3452** | **10.9438** |
| | Bartlett | 1.4781 | 0.8078 | 12.2226 |
| | BS | 1.4812 | 0.7236 | 13.3299 |
| VaR (99%) | $GRU_{TOTAL}$ | **1.1138** | **0.3763** | **11.7573** |
| | Bartlett | 1.6118 | 0.99 | 12.2933 |
| | BS | 1.7625 | 0.7979 | 19.0822 |
| CVaR(99%) | $GRU_{TOTAL}$ | **1.3597** | **0.4616** | **15.1555** |
| | Bartlett | 2.3355 | 1.2264 | 17.4385 |
| | BS | 2.2831 | 1.1347 | 20.6413 |

Table 6.4: Summary of weekly hedging S&P 500 put options (testing set) for 100 Business days with total risk hedging evaluation criteria described in section 6.2. Please note that the total hedging evaluation in this table assumes we are on the sell-side of the option trading.

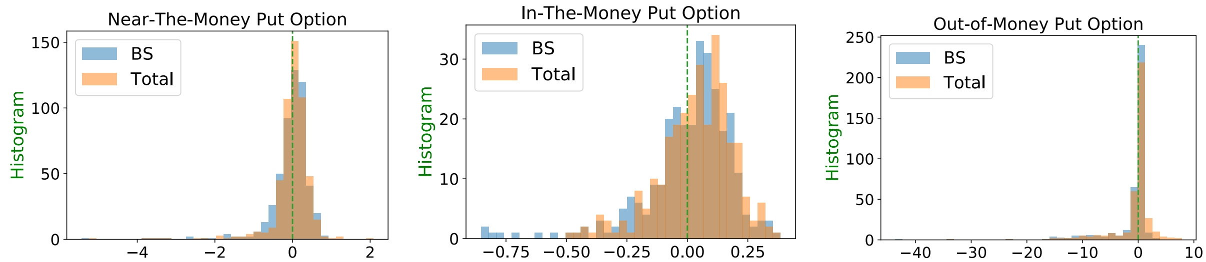
#### 6.2.3.2 Put Option Monthly Hedging Comparison

In Table 6.5, we demonstrate the results on monthly hedging put options. Furthermore, in Figure 6.9 and Figure 6.10, we compare the distribution of the relative hedging error ~~of~~ using $GRU_{TOTAL}$ with the distributions ~~of~~ using the BS model and the Bartlett model respectively.

From Table 6.5, we can see that, $GRU_{TOTAL}$ is still the dominant method for monthly hedging put options in terms of most of the criteria for NTM, ITM, and OTM scenarios. However, by comparing Table 6.4 and Table 6.5, we can see $GRU_{TOTAL}$ is less dominant in monthly hedging than in weekly hedging. In certain case, Bartlett methods can perform much better. For instance, for NTM scenarios, the CVaR (95%) and CVaR (99%) from Bartlett method is significantly better than the other two comparing methods. Another interesting observation is that the $GRU_{TOTAL}$ produces longer tail for NTM and OTM scenarios on the profit side while for ITM scenarios, $GRU_{TOTAL}$ has a much shorter tail on the loss side.

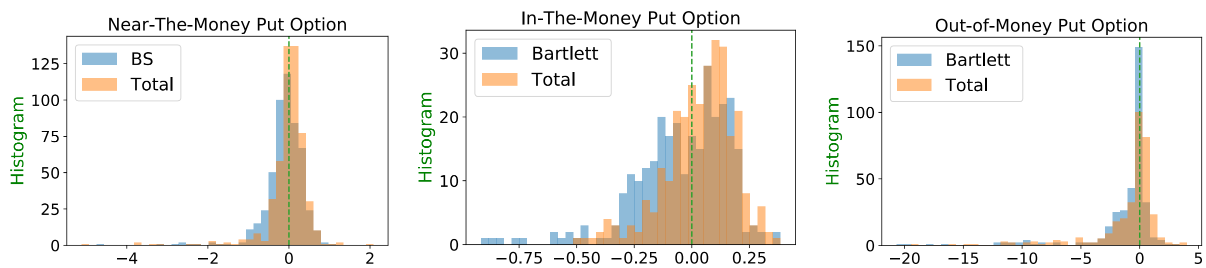
### 6.2.4 Comparison to Local Risk Hedging Model

Since $GRU_{\delta}$ is built on top of pure market data while $GRU_{TOTAL}$ is built on top of augmented market data, we will not compare the performance of $GRU_{\delta}$ and $GRU_{TOTAL}$ directly. Since, here we are more interested in the effect of the choice of objective functions. Therefore, we ~~define~~ compute a new comparing model $GRU_{TOTAL}^{LOCAL}$. The model structure of $GRU_{TOTAL}^{LOCAL}$ is exactly the same as




(a) (b) (c)

Figure 6.9: Comparing total risk hedging model $\text{GRU}_{\text{TOTAL}}$ and BS model on monthly hedging put options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.




(a) (b) (c)

Figure 6.10: Comparing total risk hedging model $\text{GRU}_{\text{TOTAL}}$ and Bartlett model on monthly hedging put options (testing set) in terms of the distribution of the relative hedging portfolio value at the expiries as in equation (6.1.3). The distribution in this figure assumes we are on the sell-side of the option trading.

| | | Near-The-Money | In-The-Money | Out-of-The-Money |
|---|---|---|---|---|
| Mean Abs Relative Error | $\text{GRU}_{\text{TOTAL}}$ | **0.2986** | **0.1240** | 1.7639 |
| | Bartlett | 0.3205 | 0.1583 | **1.7383** |
| | BS | 0.3224 | 0.1342 | 1.8482 |
| VaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | **0.7395** | **0.2562** | **8.5602** |
| | Bartlett | 0.8370 | 0.3583 | 9.0303 |
| | BS | 0.7768 | 0.3088 | 9.7018 |
| CVaR (95%) | $\text{GRU}_{\text{TOTAL}}$ | 1.7761 | **0.3577** | **13.3160** |
| | Bartlett | **1.5710** | 0.6016 | 14.4425 |
| | BS | 1.8682 | 0.5401 | 16.1024 |
| VaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | **2.1792** | **0.4121** | 15.2323 |
| | Bartlett | 2.1925 | 0.7728 | **15.0144** |
| | BS | 2.5569 | 0.7583 | 15.8393 |
| CVaR (99%) | $\text{GRU}_{\text{TOTAL}}$ | 3.4001 | **0.4509** | **20.6503** |
| | Bartlett | **2.9164** | 0.8463 | 24.1757 |
| | BS | 3.5486 | 0.8109 | 26.4941 |

Table 6.5: Summary of monthly hedging S&P 500 put options for 100 business days with total hedging evaluation criteria described in section 6.2. The total hedging evaluation in this table assumes we are on the sell-side of the option trading.

$\text{GRU}_{\text{TOTAL}}$ which is as in Figure 6.1. The only difference is $\text{GRU}_{\text{TOTAL}}$ is built on minimizing objective (6.1.2), while the $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is built on minimizing a local risk objective.

More specifically, for an expiry $T$ and a strike $K$, at a rebalancing time $t_j$ ,we have

$$\begin{aligned} \Delta V_{t_j,K,T} &= D(t_0,t_{j+1})V_{t_{j+1},K,T} - D(t_0,t_j)V_{t_j,K,T} \\ \Delta S_{t_j} &= D(t_0,t_{j+1})S_{t_{j+1}} - D(t_0,t_j)S_{t_j} \\ D(t,T) &= e^{-r(T-t)} \end{aligned}$$

$$t_j = t_0 + j\Delta t;\ j = 0,\ldots,N_{rb}-1;\ \ t_0 = T - N_{rb}\Delta t\ .$$

The objective for the $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is therefore:

$$Obj_{Local} = \sum_{i=1}^{M} \sum_{t\in \mathbf{t}^i_{RB}} |\Delta V^{mkt}_{t,T^i,K^i} - \Delta S_t \delta^M_{t,T^i,K^i}| \tag{6.2.1}$$

where $\mathbf{t}^i_{RB} = \{t^i_0,\ldots,t^i_{N_{rb}-1}\}$ is the set of rebalancing dates for the $i$-th hedging scenarios with expiry $T^i$ and initial date $t^i_0$. The model structure for $\text{GRU}_{\text{TOTAL}}$ and $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is the same which is further discussed in Appendix D. The same set of hedging scenarios are used as the training, testing and validation data sets. The training procedure is also the same as indicated in Algorithm 8. The only difference is the objective function used in training.

The detailed comparison between $\text{GRU}_{\text{TOTAL}}$ and $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is given in Appendix E. Here we summarize the major results:

- For weekly hedging call options, $\text{GRU}_{\text{TOTAL}}$ performs better than $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ in terms of most of total risk measures. However, in terms of tail loss reduction, the improvement from $\text{GRU}_{\text{TOTAL}}$ over $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is less significant.
- For monthly hedging call options, $\text{GRU}_{\text{TOTAL}}$ improve over $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ in terms of the mean absolute relative error slightly. The $\text{GRU}_{\text{TOTAL}}$ and $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ perform roughly the same in terms of tail loss measured by VaR and CVaR.
- For weekly hedging put options, $\text{GRU}_{\text{TOTAL}}$ still performs better than $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ in terms of reducing the mean absolute relative error for ITM and OTM scenarios and the performance for NTM scenarios is similar. On the other hand, in terms of tail loss measured by VaR and CVaR, the reduction from $\text{GRU}_{\text{TOTAL}}$ over $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is significant.
- For monthly hedging put options with NTM and ITM scenarios, we achieve better mean absolute relative error from $\text{GRU}_{\text{TOTAL}}$. For OTM scenarios, $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ performs better in terms of mean absolute relative error. The tail loss measured by VaR and CVaR for NTM scenarios is roughly the same for $\text{GRU}_{\text{TOTAL}}$ and $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$. The tail loss from $\text{GRU}_{\text{TOTAL}}$ for OTM scenarios is slightly better than $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$. The tail loss from $\text{GRU}_{\text{TOTAL}}$ for ITM scenarios is significantly better than $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$.

As we have already discussed in section 2.1.2, with the assumption of zero interest rate, the *discrete total hedging risk* which is defined as:

$$\text{Risk}_{t_0,T,K}^{total} = \sum_{j=0}^{N_{rb}-1} \left\{ \Delta S_{t_j} \delta_{t_j,T,K} - \Delta V_{t_j,T,K}^{mkt} \right\} = \sum_{j=0}^{N_{rb}-1} \text{Risk}_{t_j,T,K}^{local} \tag{6.2.2}$$

In other words, *discrete total hedging risk* is the summation of the *discrete local hedging risk* evaluated at discrete rebalancing time $\{t_0, t_1, \ldots, t_{N_{rb}-1}\}$. As a consequence, building a model reducing the discrete local hedging risk will reduce the discrete total hedging risk as well. Therefore, it is not surprising to see that $\text{GRU}_{\text{TOTAL}}^{\text{LOCAL}}$ is still competitive in terms of total risk measurements.

# Chapter 7

# Conclusion and Future Work

directly from market prices

In this thesis, we have proposed a direct kernel hedging model $DKL_{SPL}$ and a novel encoder-decoder model, $GRU_{\delta}$, for discrete local risk option hedging. The $DKL_{SPL}$ is our first exploration on computing a data-driven local hedging model without estimating a option pricing model. $GRU_{\delta}$ is proposed to further improve direct data-driven local risk hedging using machine learning techniques. $GRU_{\delta}$ consists of an encoder, which generates a concise representation of the past market information. The decoder uses the Black-Scholes delta as a pre-trained model and utilizes a gate to generate a predicative hedging model, combining the pre-trained delta model and the outputs from the encoder. Feature selections are implemented through normalized weights embedded in the model training. In addition, a data instance adaptive Huber loss function is incorporated for robustness, with the error from the pre-trained Black-Scholes delta model for that instance as the thresholding parameter.

price

Using the S&P 500 index and the index option data, from January 1, 2004, to August 31st, 2015, we assess and compare hedging performance of the $GRU_{\delta}$ and $DKL_{SPL}$ with other hedging strategies in terms of local risk criteria. For weekly and monthly hedging, computational results demonstrate that performance of the proposed $GRU_{\delta}$ significantly surpasses that of the MV model, SABR-Bartlett, regularized spline kernel model $DKL_{SPL}$, all of which perform significantly better than the BlackScholes model with implied volatility in terms of local risk hedging criteria (6.2.1). In addition, the $DKL_{SPL}$ also outperforms MV model in terms of weekly and monthly hedging results. We further demonstrate that the encoder for the sequential features plays a significant role in $GRU_{\delta}$, since removing the encoder deteriorates hedging performance. Lastly, by comparing the weekly and monthly hedging performance from the $GRU_{c}$, for which we remove the output gate and training with MSE, and $GRU_{\delta}$, we demonstrate that the output gate, robust loss function and also play significant roles in $GRU_{\delta}$.

We further demonstrate that the daily hedging performance of the proposed $GRU_{\delta}$ also surpasses that of the MV hedging method, LVF and SABR corrective methods (implemented in [112]), data-driven regularized spline kernel network model $DKL_{SPL}$, and SABR-Bartlett. In addition, $DKL_{SPL}$ also outperforms MV hedging method , LVF and SABR corrective methods (implemented in [112]).

In addition, from the testing hedging performance, we assess feature importance in $\text{GRU}_\delta$ for the S&P 500 index option hedging. The monthly average feature weights identify the underlying as the most important local feature and the past implied volatility sequence as the most important sequential feature during normal market periods.

To assess whether minimizing the total risk directly can lead to better performing total risk minimization strategy, in the context of data-driven hedging, we extend $\text{GRU}_\delta$ to $\text{GRU}_{\text{TOTAL}}$ for multi-step total risk hedging. To deal with the challenges of acquiring enough market option information for building data-driven hedging models, we augment the market data using SABR model and local volatility model. Using the S&P 500 index option market data from January 1st, 2000 to August 31st, 2015, we compare the weekly and monthly hedging performance of the proposed total risk hedging model $\text{GRU}_{\text{TOTAL}}$ with the sequential data-driven local risk hedging model $\text{GRU}^{\text{LOCAL}}_{\text{TOTAL}}$, which adopts the same model structure but is trained with a local risk training objective, the Black-Scholes delta with implied volatilities, and the SABR-Bartlett delta. We measure the total risk hedging performance, which is evaluated on the expiries of the options. We demonstrate the effectiveness of the total risk hedging model $\text{GRU}_{\text{TOTAL}}$ in reducing the sell-side loss tail risk for both put and call options. We also confirm that the total hedging model $\text{GRU}_{\text{TOTAL}}$ often leads to better hedging performance in terms of total hedging criteria when compared with $\text{GRU}^{\text{LOCAL}}_{\text{TOTAL}}$, SABR-Bartlett method, and Black-Scholes model. However, alternative local risk model $\text{GRU}^{\text{LOCAL}}_{\text{TOTAL}}$ remain competitive in controlling the total hedging risk.

The main objective of this research is to assess hedging performance of strategies learned directly from the historical time series of the market option price and underlying price, using machine learning methods. Hedging performance comparisons between data-driven models $\text{DKL}_{\text{SPL}}$, $\text{GRU}_\delta$, $\text{GRU}_{\text{TOTAL}}$,and $\text{GRU}^{\text{LOCAL}}_{\text{TOTAL}}$ and the classical option hedging based on parametric model calibration suggest that the data driven learning hedging can be a viable alternative to the classical methods, potentially leading to better hedging performance.

In terms of limitation of this research, we compare the data-driven models mostly with parametric models available in academic literature. We understand that the actual industry practice may not apply the parametric models in the same way as we did in this thesis for hedging derivatives. While it would also be interesting to compare hedging methods actually adopted in the financial industry, limitation in accessing industry practice makes it difficult to conduct such a study.

Additionally, comparing with the calibration of the traditional parametric pricing models on vanilla index options, the learning process for the data-driven hedging model is less computationally efficient. However, one should notice that once the model learning of the data-driven model is done, the outputing process of the hedging position is computaionally efficient. In practice, one can tra the model after the business trading hours and use the trained model to produce the hedging position efficiently during the business trading hours.

Lastly, the learning process requires certain amount of historical data. For calibrating a parametric model, one usually needs much less data and can build the model based on market data observed on spot and compute the sensitivity as hedging position accordingly. Therefore, our proposed data-driven models are not directly applicable to the illiquid derivative markets.

train

For extending our work, we note that there are several directions:

- In this thesis, we rely on market data to generate the hedging scenarios. The volatility surface is calibrated from market data and underlying price paths are extracted from market with overlapping period. Our models based on a neural network approach have less parameters compared with other applications of the deep learning techniques, given the relative scarcity of available historical data. For future work, one can explore how we can use machine learning techniques to generate hedging scenarios so that we can combine artificial scenarios with real scenarios. This will enable us to build more complex model for hedging. Indeed, for this direction, there are already several attempts. For example, Bergeron et al. [18] apply the variational autoencoders [121] on generating synthetic volatility surface that are indistinguishable from those observed historically. Pardo and López [149] apply the Generative Adversarial Networks (GANs) [91] on learning the underlying structure inherent to the dynamics of financial series and acquiring the capacity to generate scenarios that share many similarities to those seen in the historic time series.
- In this thesis, we use S&P 500 index options for experimental comparison. It will be interesting to explore effectiveness of the data-driven models on more complex derivatives such as basket options where the dimensionality of the underlying is higher.
- In this thesis, we have not included transaction cost into our models. A more realistic model should include the effect of the transaction cost as in [28].
- In this thesis, for total hedging model, we assume we rebalance every 5 business days or 20 business days. In reality, we do not have to fix the interval length between two rebalancing times. We can extend the model so that the data-driven model determines when is the best time to rebalance the hedging portfolio.