Abstract

A variety of methods have been devised to extract risk-neutral density (RND) from European option prices. Some methods focus on the pricing ability of RND and formulated the extraction as an optimization problem. Whereas the others have stronger theoretical groundings and approximate the RND using actual option prices. In a review by *Bahra (1997),* several extraction methods have been investigated in the FX market. In this paper, we present some results of implementation in the HSI option market and identify the difficulty of implementing the methods using the option data released by HKEX. In addition, we proposed a modified parametric method to improve the pricing ability of existing parametric methods. We then compare the performance of the methods in terms of interpretability of moments and pricing accuracy.

Introduction

Option prices can provide rich information on the forward-looking distribution of terminal underlying asset price. Such distribution, called Risk-Neutral Density (RND), can be used to track changes in expected moments of terminal price and gauge changes in market sentiment. Several approaches to estimate the RND are mentioned in a comprehensive review by *Bahra (1997)*, and the purpose of this study is to implement these methods in the Heng Seng Index market, propose new methods for RND estimation and compare their performance in terms of interpretability of moments and pricing accuracy.

I. Source of Data

Since May 2021, Hong Kong Stock Exchange has been releasing daily statistics of Heng Seng Index option. The daily information includes closing price of European options (quotation / last price), Implied volatility, open interest, high and low price. However, some of these prices are not traded prices (but official quotation price) and all of them are rounded to the nearest integer (rounded upward to 1 if the price is below 1). This creates several problems in our implementation, which will be discussed later.

II. Breeden and Litzenberger Approach

In a risk-neutral world, options are priced by their discounted expected payoffs at maturity using a risk-neutral density.

$$c(K,T) = e^{-rT} \int_{-\infty}^{\infty} q(S_T) Max(S_T - K, 0) dS_T \tag{1}$$

Where $c(K,T)$ denotes price of European call option with strike K and maturity T, $q(S_T)$ is the risk-neutral density of terminal price S_T.

A consequence of the pricing formula is that we can obtain the risk-neutral density from option prices. To illustrate, take derivatives of the call price function $c(K,T)$ with respect to strike K twice:

$$\frac{\partial c}{\partial K} = -e^{-rT}\int_K^{\infty} q(S_T)dS_T \tag{2}$$

$$\frac{\partial^2 c}{\partial K^2} = e^{-rT}q(K) \tag{3}$$

Hence, the second order partial derivatives with respect to strike K of the option price function yield us the risk-neutral density at $S_T = K$.

The Breeden and Litzenberger approach capture this idea and approximates this second order derivatives with prices of butterfly spread:

$$\frac{\partial^2 c}{\partial K^2} \approx \frac{\frac{\partial c}{\partial K}|_{K=K} - \frac{\partial c}{\partial K}|_{K=K-\Delta K}}{\Delta K}$$

$$\approx \frac{\frac{\left(c(K+\Delta K,\mathrm{T}) - c(K,\mathrm{T})\right)}{\Delta K} - \frac{\left(c(K,\mathrm{T}) - c(K-\Delta K,\mathrm{T})\right)}{\Delta K}}{\Delta K}$$

$$= \frac{B(K,\Delta K,T)}{\Delta K^2} \tag{4}$$

Where $B(K,\Delta K,T)$ denotes price of butterfly spread with centered strike K, strike differential ΔK and maturity T.

Similarly, by put-call parity, put butterfly spread can also be used to approximate risk-neutral density.

$$B(K,\Delta K,T) = \left(c(K+\Delta K,\mathrm{T}) - c(K,\mathrm{T})\right) - \left(c(K,\mathrm{T}) - c(K-\Delta K,\mathrm{T})\right)$$

$$= c(K+\Delta K,\mathrm{T}) - 2c(K,\mathrm{T}) + c(K-\Delta K,\mathrm{T})$$

$$= p(K+\Delta K,T) - 2p(K,\mathrm{T}) + p(K-\Delta K,\mathrm{T}) + S_0 - (K+\Delta K)e^{-rT} - 2S_0$$
$$+ 2Ke^{-rT} + S_0 - (K-\Delta K)e^{-rT}$$

$$= p(K + \Delta K, T) - 2p(K, \mathrm{T}) + p(K - \Delta K, \mathrm{T}) \tag{5}$$

One problem of this approach is that the obtained risk-neutral densities are separated at discrete points, which is not ideal. To avoid this, we can obtain risk-neutral probabilities of terminal price S_T falling around the centered strike K:

$$P_{\Delta K}(S_T = K) = P(K - 0.5\Delta K \leq S_T \leq K + 0.5\Delta K)$$

$$= \int_{K-0.5\Delta K}^{K+0.5\Delta K} q(S_T)\, dS_T \tag{6}$$

Which can be approximated as a rectangular area using the risk-neutral densities:

$$P_{\Delta K}(K) = \int_{K-0.5\Delta K}^{K+0.5\Delta K} q(S_T)\, dS_T$$

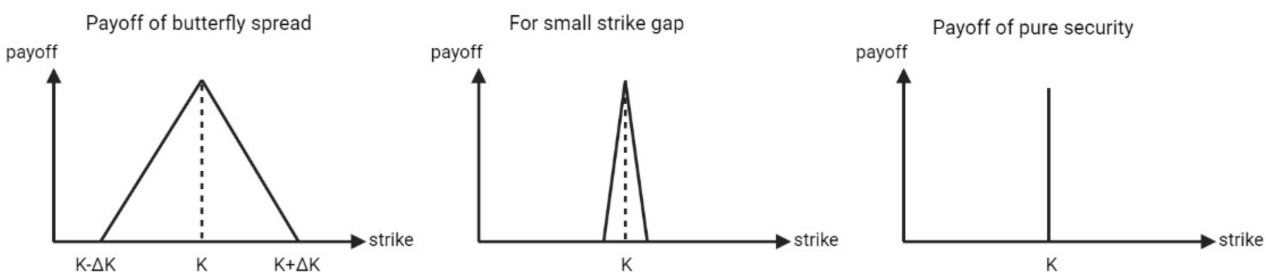
$$\approx q(K) * \Delta K$$

$$\approx \frac{c(K + \Delta K, \mathrm{T}) - 2c(K, \mathrm{T}) + c(K - \Delta K, \mathrm{T})}{\Delta K} * e^{rT}$$

$$= \frac{B(K, \Delta K, T)}{\Delta K} * e^{rT} \tag{7}$$

This risk-neutral probability $P_{\Delta K}$ is easier to work with in pricing European option prices.

Figure 1: Illustration of the approximation in payoff diagrams:



As shown above, the approach can also be interpreted as approximating the price of pure securities (which are proportional to their corresponding risk-neutral density) with the price of butterfly spreads. $As\, \Delta K$ tends to 0, payoff of butterfly spread (middle diagram) converges to payoff of pure security that pays ΔK in state $S_T = K$ (right diagram). The smaller the strike differential ΔK, the better the approximation. Hence, the approach works best when there is a continuum of strikes (or very close strikes) (*Bahra, 1997)*.

In practice, strikes are offered for every 100-1000 HSI points. As a result, we only get a rough estimation of the risk-neutral density. In addition, strike differentials ΔK are not constant across moneyness. Some strikes are not offered for some maturities. Below situations may occur in practice:

1. Strike differential ΔK changes across moneyness. Most common ΔKs for HSI options are 200 and 400 HSI points.
2. For close maturities, odd value strikes (i.e., $\Delta K = 100$) maybe offered at extreme moneyness.
3. For distant maturities, ΔK is usually 1000 HSI points at extreme moneyness.

Example is shown below for 2021/12/06 MAY-22 maturity HSI options. Notice that the strike differential ΔK changes from 100 HSI points to 200 HSI points at 20000.

Table 1: Option Strikes for MAY-22 maturity on 2021/12/06:

Strike	Call Price	Put Price
…	…	…
19700	3834	177
19800	3742	184
19900	3651	194
20000	3561	203
20200	3382	222
20400	3204	244
20600	3029	268
…	…	…

To handle this, we can apply data interpolation on option prices. However, since we want to retain data authenticity in this approach, we form groups of options with the same ΔK, and estimate the risk-neutral density using a different ΔK for each group. Some options may belong to two groups, and their price is used twice in the estimation, such as $K = 20000$ in the example.

After calculating the spread prices, we notice that the price of call butterfly spread and put butterfly spread are different, some are even negative or zero. If there are no arbitrage opportunities, they should have identical price for the same centered strike K and same ΔK as they have identical payoff at maturity. The abnormal prices could be due to the following reasons:

1) Some of the prices are just quotations which are not traded price
2) Some of the traded prices are last price
3) HKEX performed data cleaning on the prices such as rounding to integers, which distort the true price structure

Example of 2021/12/06, FEB-22 maturity, butterfly spread prices (ΔK is 200 points for this group of spread) are shown below:

Table 2: Butterfly spread prices on 2021/12/06 for FEB-22 options:

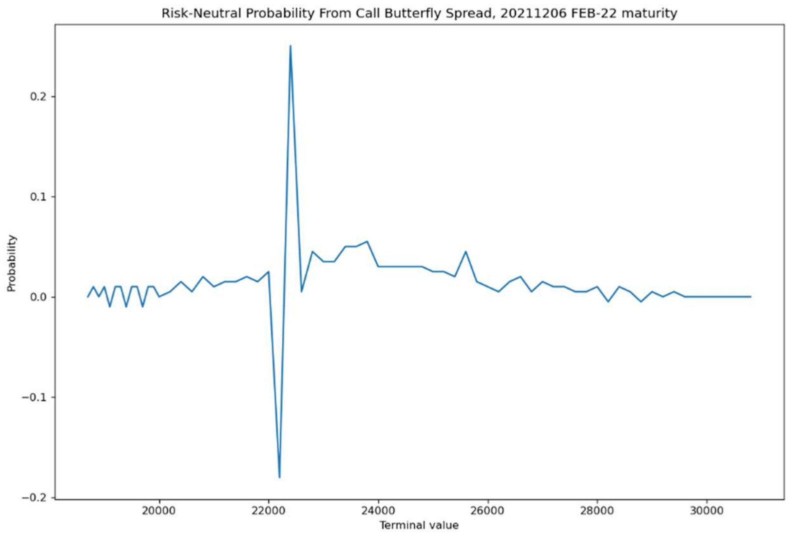
Center strike K	Call BF spread	Put BF spread
...	...	...
22000	5	1
22200	**-36**	7
22400	**50**	5
22600	1	6
22800	9	5
23000	7	9
23200	7	7
23400	10	3
23600	10	**25**
23800	11	**-9**
24000	6	**22**
...	...	...

To handle the anomalies in price, we will use only call butterfly spread. Although an average of 2 is a possible solution, it may complicate the problem.

We observe that for negative spread prices, prices nearby are usually abnormally high, which form a spike in the price-strike diagram. These negative prices imply negative probabilities $P_{\Delta K}$: [1]

$$P_{\Delta K}(S_T = K) \approx \frac{B(K, \Delta K, T)}{\Delta K} * e^{r_f T} < 0 \tag{8}$$

Figure 2: Negative risk-neutral probability due to negative butterfly spread prices:



Although it could be that the RND is indeed "spiky", the spike can also be a result of mispricing (i.e., quotation / last price deviates from true price). For example, if a call option with strike K is overpriced in our data, the following will occur:

$$B(K, \Delta K, T) = c(K - \Delta K) + c(K + \Delta K) - 2\, c(K) \quad \text{is underpriced}$$

$$B(K - \Delta K, \Delta K, T) = c(K - 2\Delta K) + c(K) - 2\, c(K - \Delta K) \quad \text{is overpriced}$$

$$B(K + \Delta K, \Delta K, T) = c(K) + c(K + 2\Delta K) - 2\, c(K + \Delta K) \quad \text{is overpriced}$$

This can form a spike at $S_T = K$. Combining with the fact that there is negative price in

---

[1] r_f is assumed to be 1-year HIBOR rate, T is calculated as $\frac{\text{trading days}}{251}$

the region, they are likely formed by inaccurate prices provided by HKEX.

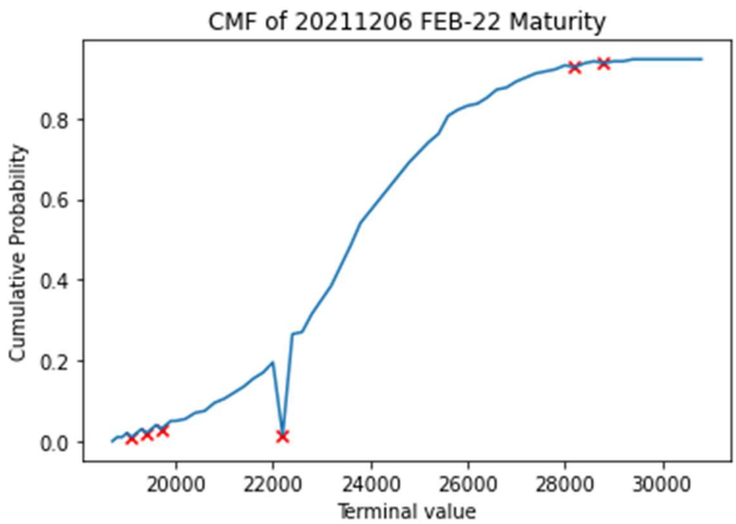
II-B. Enforcing monotonicity in CMF using PCHIP interpolation

To deal with the negative butterfly spread prices, we enforce monotonicity in the cumulative mass function (CMF) by removing the "incorrect" prices that lead to arbitrage and replace them with interpolated values. We first estimate the CMF through cumulative sum of the $P_{\Delta K}(S_T)$:[2]

$$P(S_T \leq X) = \int_0^X q(S_T)\, dS_T \approx \sum_{\{s:s<X,\ s\in A\}} P_{\Delta K}(s)$$
$$= \sum_{\{s:\, s<X,\ s\in A\}} \frac{B(s, \Delta K, T)}{\Delta K} * e^{rT} \tag{9}$$

Then, we remove the decreasing values in the CMF (i.e., negative spread prices):

Figure 3: Decreasing values in CMF due to negative butterfly spread prices:

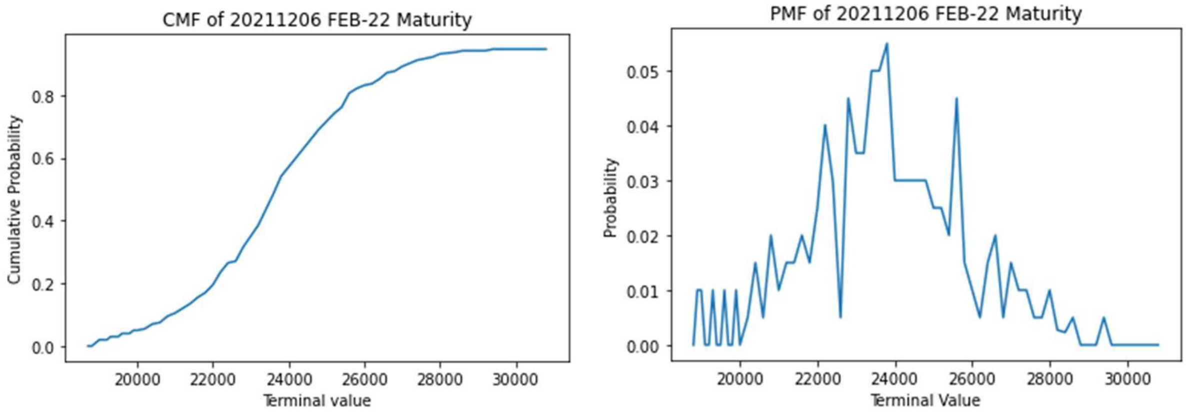


Next, we replace them with interpolated values using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method. The PCHIP method is desirable for this problem as it ensures monotonicity and nonlinearity between knots. Finally, we convert it back to probabilities $P_{\Delta K}(S_T)$, and obtain the following result:

---

[2] A is the set of centered strike of call butterfly spreads

Figure 4: Corrected CMF and the corresponding PMF



One drawback of this data correction is that we only deal with the negative spread prices while retaining the other possibly overvalued/undervalued ones.

In addition, the risk-neutral probabilities $P_{\Delta K}(s)$ obtained do not sum to one:

$$\sum_{\{s:s<X,\ \ s\in A\}} P_{\Delta K}(s) = \sum_{\{\mathrm{s}:\ \mathrm{s}<\mathrm{X},\ \ \mathrm{s}\in\mathrm{A}\}} \frac{\mathrm{B}(\mathrm{s}, \Delta\mathrm{K}, \mathrm{T})}{\Delta\mathrm{K}} * \mathrm{e}^{\mathrm{rT}} \neq 1 \tag{10}$$

We need to determine how to attribute the "missing mass" to the edges where the densities are not estimated (*Neuhaus, 1995)*. In our implementation, we simply assign them to the existing mass:

$$P_{\Delta K}(S_T = k) = \frac{P_{\Delta K}(S_T = k)}{\sum_{x\in A} P_{\Delta K}(x)} \tag{11}$$

III. Density Estimation Using Kernel Regression on Histogram

The Breeden and Litzenberger approach yield us an estimation of risk-neutral density at discrete points (centered strike K of our butterfly spreads), which is undesirable. If we assume that the true RND is "smooth", we can use Nadaraya–Watson kernel regression to estimate the density in a continuum of strikes. The Nadaraya–Watson kernel estimator of risk-neutral density at $S_T = x$ is defined as[3]:

$$\widehat{m}(x) = \frac{\sum_{\{s \in A\}} K\left(\frac{x-s}{w}\right) \ \hat{q}(s)}{\sum_{\{s \in A\}} K\left(\frac{x-s}{w}\right)} \tag{12}$$

Where K is the kernel function, w is a bandwidth. We use a gaussian kernel in the implementation:

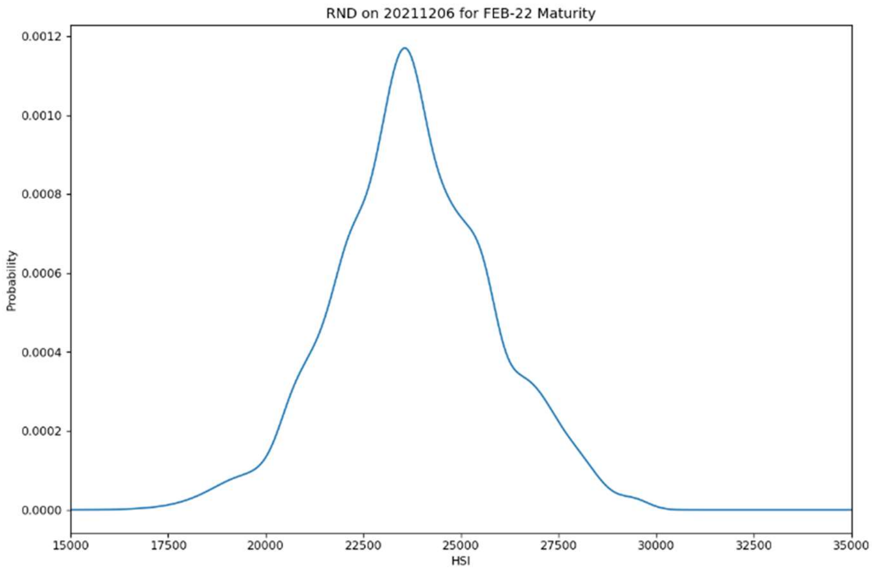
$$K\left(\frac{x-s}{w}\right) = \frac{1}{\sqrt{2\pi}} \ e^{-\frac{\left(\frac{x-s}{w}\right)^2}{2}} \tag{13}$$

$\hat{q}(s)$ is the estimated density obtained from the Breeden and Litzenberger approach:

$$\hat{q}(s) \ \approx e^{rT} \frac{B(s, \Delta K, T)}{\Delta K^2} \qquad (3) \ \& \ (4)$$

The kernel estimator yields us an estimate of risk-neutral density at any terminal value. An example of the RND extracted from FEB-22 call option prices on 2021/12/06 using a bandwidth w of 400 HSI points is shown below:

Figure 5: RND using kernel regression method for FEB-22 maturity on 2021/12/06:



One drawback of this approach is that it requires the selection of bandwidth w, but in practice the difference is small when w ranges from 200 to 600 HSI points. Another

[3] A is the set of centered strike of call butterfly spreads

drawback is that the estimation error is high, which means the RND has poor pricing performance, which will be discussed later

IV. Price Interpolation Approach

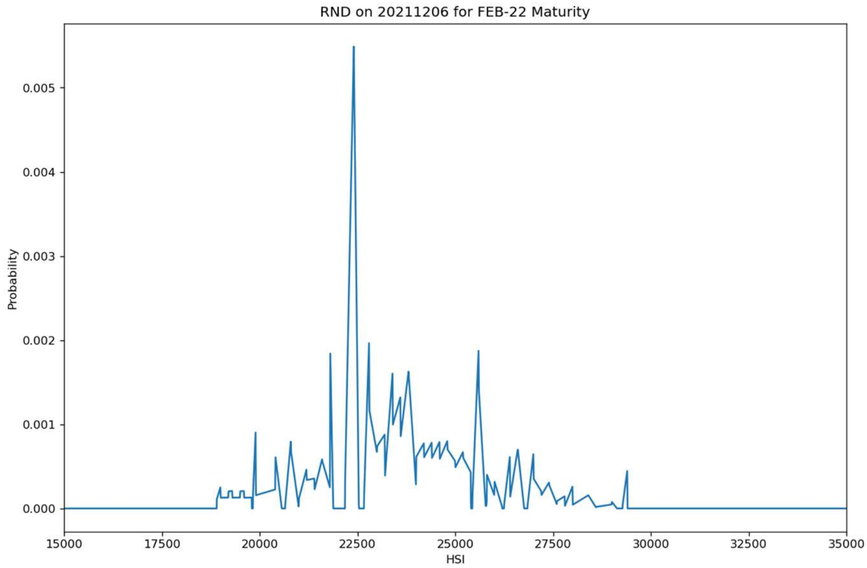
One difficulty of implementing the Breeden and Litzenberger Approach is the lack of continuum of strikes. As *Bahra (1997)* mentioned, we can interpolate option price with respect to strike and obtain a continuous RND.

Since the price-strike curve of European option is convex and monotonic, we use the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method to interpolate option prices between strikes.

Before interpolation, we remove the “wrong” option prices that lead to arbitrage opportunities in order to avoid negative density. Because an overvalued call option with strike K can lead to negative butterfly spread price with centered strike K, we identify the centered strike of butterfly spreads that have negative prices and exclude the call option with that strike.

Then, we interpolate the option price and calculate the risk-neutral density using the same procedure as before. An example of the RND extracted from FEB-22 call option prices on 2021/12/06 is shown below:

Figure 6: RND using price interpolation method for FEB-22 maturity on 2021/12/06:



One major drawback of this approach is that the interpolated prices are “illusory” - they are not real prices. In addition, the risk-neutral density obtained varies with the interpolation method used , since the risk-neutral densities between knots are the second order derivatives of the interpolation function. In addition, one major drawback of the PCHIP interpolation method is that the second order derivatives are not continuous at knots, meaning that the risk-neutral density may “jump” at knots. This leads to “spikes” in the extracted RND.

V. Implied Volatility Interpolation Approach

An alternative method for interpolating option price is to interpolate the Black Scholes implied volatility (IV) curve and use the Black-Scholes formula to transform the IV curve into a price-strike curve (Shimko, 1993).

However, as previously mentioned, some option prices in the HKEX dataset lead to arbitrage opportunities, and hence their IV are also “wrong”. Interpolating the IV curve will retain these errors and lead to negative risk-neutral probabilities. Therefore, we experimented with using the Nadaraya–Watson kernel estimator to interpolate and smooth the IV curve. The estimator is defined as :

$$\widehat{m}(x) = \frac{\sum_{\{s \in A\}} K\left(\frac{x-s}{w}\right)\sigma_s}{\sum_{\{s \in A\}} K\left(\frac{x-s}{w}\right)} \tag{14}$$

σ_s is the implied volatility of option with strike s. w is a bandwidth, and $K(s)$ is a kernel function. We use the gaussian kernel defined previously (equation 13) in this approach and set bandwidth w to be 400 HSI points.

By transforming the IV curve to option prices through the Black-Scholes formula and take second order derivatives with respect to strike, we obtain an RND. Although the kernel regression "smooth out" the "wrong" prices that lead to arbitrage, the density can still be negative. We apply the same techniques in section II-B to remove the negative values. The result is shown below:

Figure 7: Interpolated IV curve using kernel regression on 2021/12/06:

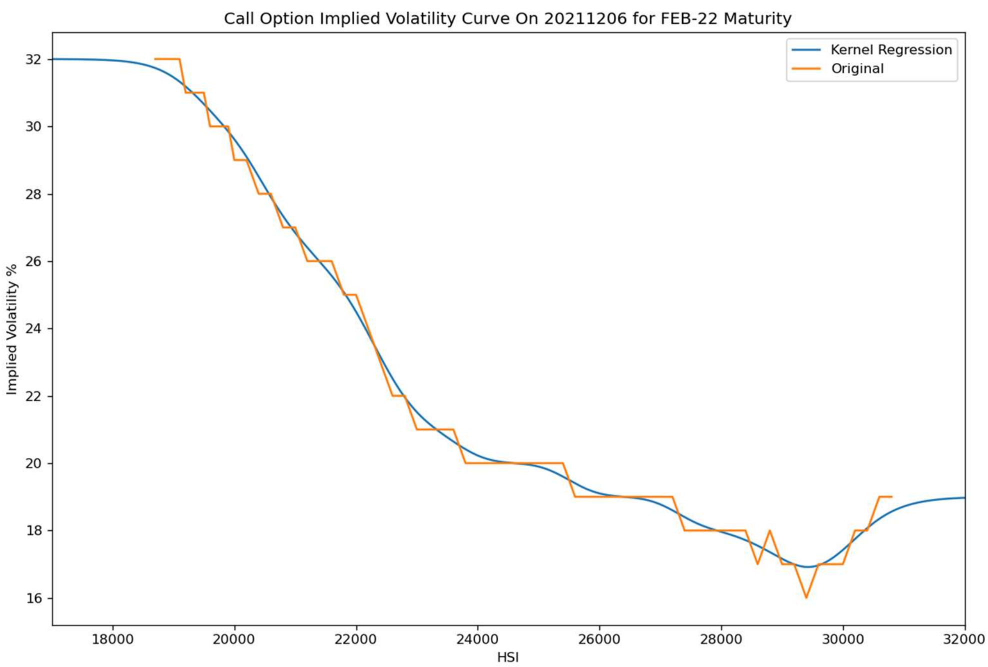
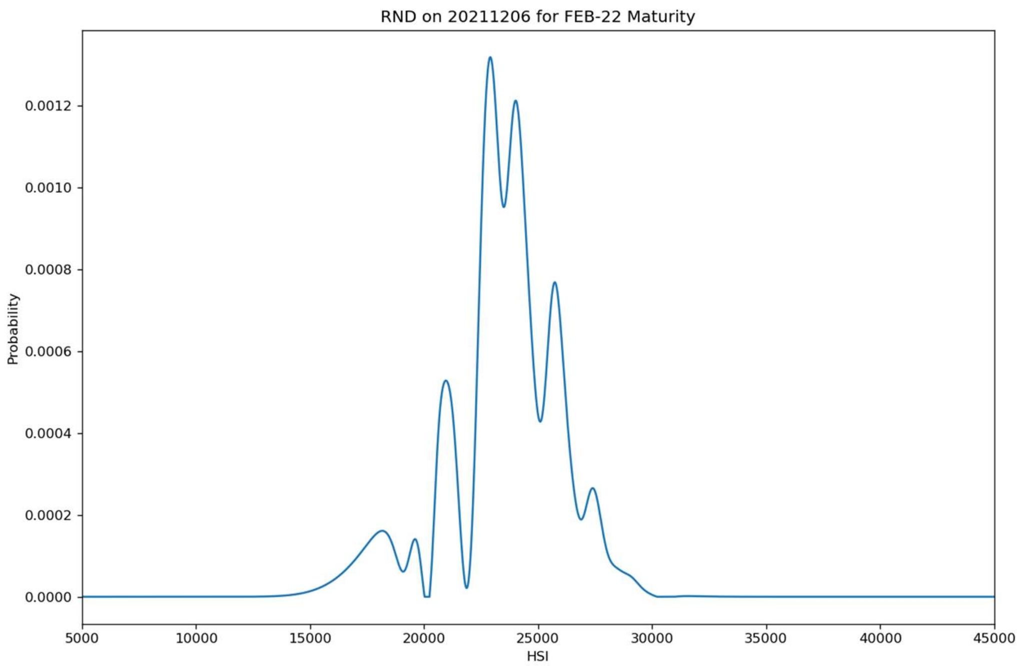


Figure 8: RND using IV interpolation method for FEB-22 maturity on 2021/12/06:



VI. Parametric Lognormal Approach (Squared Error)

In a Black-Scholes world, asset price at maturity follows a lognormal distribution with the following form:

$$f(x, S_0, r, \sigma, T) = \frac{1}{x\sigma\sqrt{T}\sqrt{2\pi}}\, e^{-\frac{\left(\ln\left(\frac{x}{S_0}\right)-\left(r-\frac{1}{2}\sigma^2\right)T\right)^2}{2\sigma^2 T}} \tag{15}$$

Since option prices are the discounted expected payoff at maturity using RND, to "fit" the lognormal distribution to option prices, we estimate the parameters by solving the below minimization problem:[4]

$$Min_{r,\sigma} \sum_{\{s:s\in C\}} (c_*(s,T) - \hat{c}(s,T))^2 + \sum_{\{s:s\in D\}} (p_*(s,T) - \hat{p}(s,T))^2 \tag{16}$$

$$subject\ to \quad \begin{cases} |r| < 20\% \\ \sigma > 10\% \end{cases}$$

[4] C is the set of call option strikes in dataset. D is the set of pall option strikes in dataset.

Where $\widehat{c_i}$, $\widehat{p_i}$, are defined as:

$$\hat{c}(s,T) = e^{-r_f T} \sum_{i=0}^{11000} f(5000 + i * 5\,, S_0, r, \sigma) * Max(5000 + i * 5 - s, 0) * 5 \tag{17}$$

$$(\approx\ e^{-r_f T} \int_0^{\infty} Max(x - s, 0) *\ f(x, S_0, r, \sigma)\, dx)$$

$$\hat{p}(s,T) = e^{-r_f T} \sum_{i=0}^{11000} f(5000 + i * 5\,, S_0, r, \sigma) * Max(s - 5000 + i * 5, 0) * 5 \tag{18}$$

$$(\approx\ e^{-r_f T} \int_0^{\infty} Max(s - x, 0) *\ f(x, S_0, r, \sigma)\, dx)$$

$$c_*(s,T) \text{ is the call option price for strike } s, \text{maturity } T \text{ in dataset}$$

$$p_*(s,T) \text{ is the put option price for strike } s, \text{maturity } T \text{ in dataset}$$

In other words, we are minimizing the estimated total squared error. The constraints are imposed to prevent overfitting and to shorten the optimization time. We use the Constrained Optimization BY Linear Approximation (COBYLA) algorithm to solve the constrained optimization problem.

In theory, to avoid arbitrage opportunities in the forward contract, mean of RND should equal to the forward price $e^{r_f T} S_0$ (Bahra, 1997), which means that $r = r_f$[5]. We could include this in our optimization by adding a penalty term $S_0 e^{rT} - S_0 e^{r_f T}$ to the cost function. But we notice that the resulting RND has unsatisfactory pricing performance when we enforce its mean to be close to the forward price.

Example of extracted RNDs from 2021/05/03 to 2021/05/31 using the lognormal parametric approach is shown below:

[5] We use the 1-year HIBOR rate as the risk-free rate

Figure 9: RNDs across maturities using parametric lognormal on 2021/05/03:

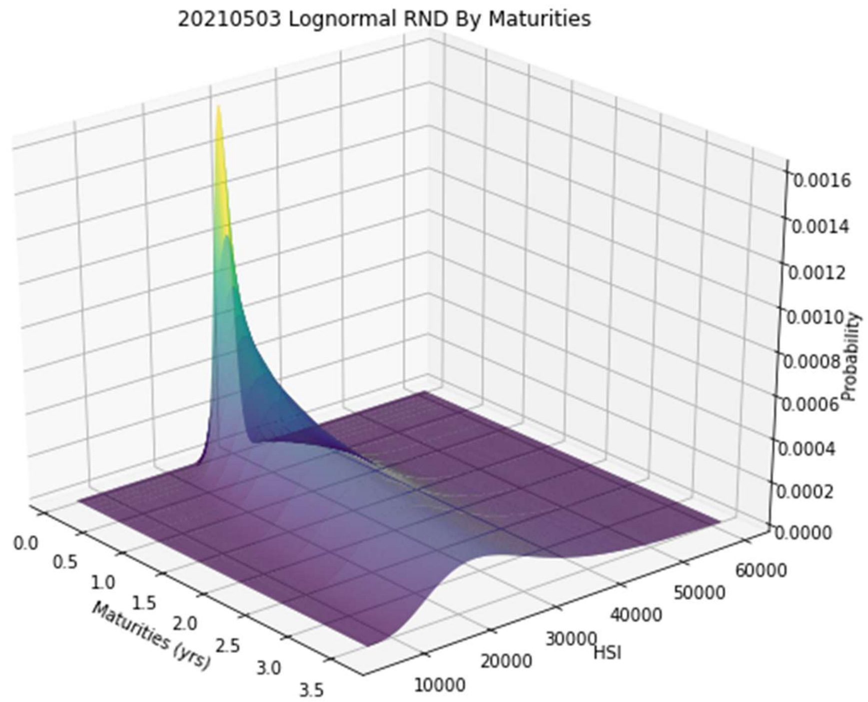
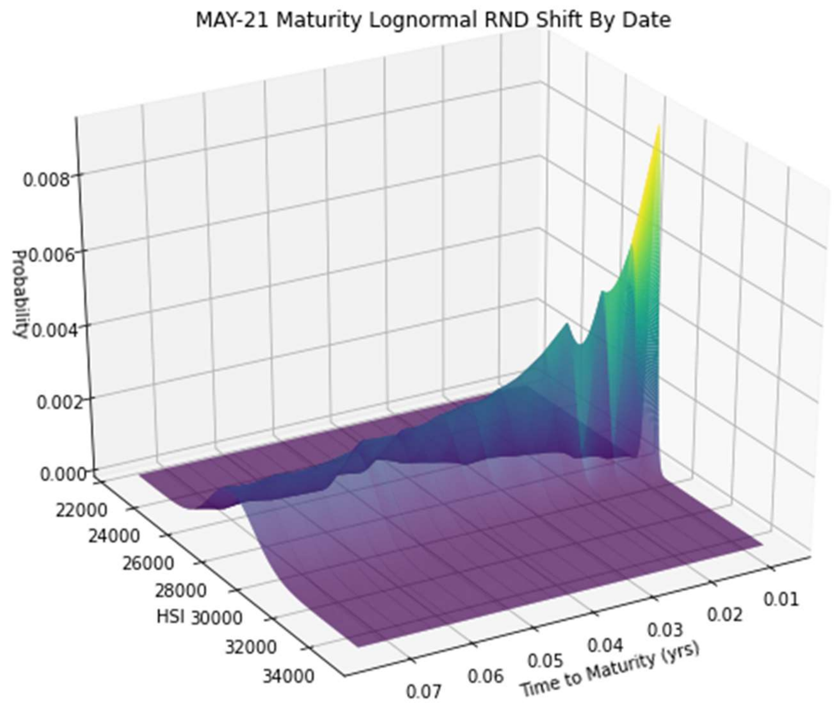


Figure 10: Shift in MAY-21 RND across time using parametric lognormal:



As shown above, variance of RND increases with maturity T. RNDs of more distant maturities are flatter. This shows that variance of terminal value increases with maturity T. In addition, for the same maturity, variance of extracted RND also decreases with the time-to-maturity.

VII. Parametric Lognormal Approach (Absolute Error)

One problem of minimizing squared error in the HSI option dataset is that it overfits the "wrong prices". For example, prices of OTM options are rounded upward to 1 if they are trading in cents. Moreover, some prices can lead to arbitrage opportunities as shown in the negative spread prices, because they could be quotation or last price rather than the latest tradable price. The squared error cost function is sensitive to these outliers and tends to overfit them.

To mitigate this problem, we propose to replace the squared error with absolute error, which is more robust to outliers (Pontius, Thontteh & Chen, 2008). We also propose to exclude the options with 1HKD price except the first one in the optimization problem, since their price is more distorted[6]. For example, options highlighted in red are excluded on 2021/09/07:

Table 3: Option prices that are excluded in the optimization on 2021/09/07

Strike	Call Price
30800	1
31000	1
31200	1
31400	1
31600	1
31800	1
32000	1

[6] If the true price of an ITM option is 1500.5 HKD, HKEX would round its price to 1501, which means 0.033% overvaluation in the price data. If the true price of an OTM option is 0.6 HKD, HKEX would round its price to 1 HKD, which means 66% overvaluation in the price data.

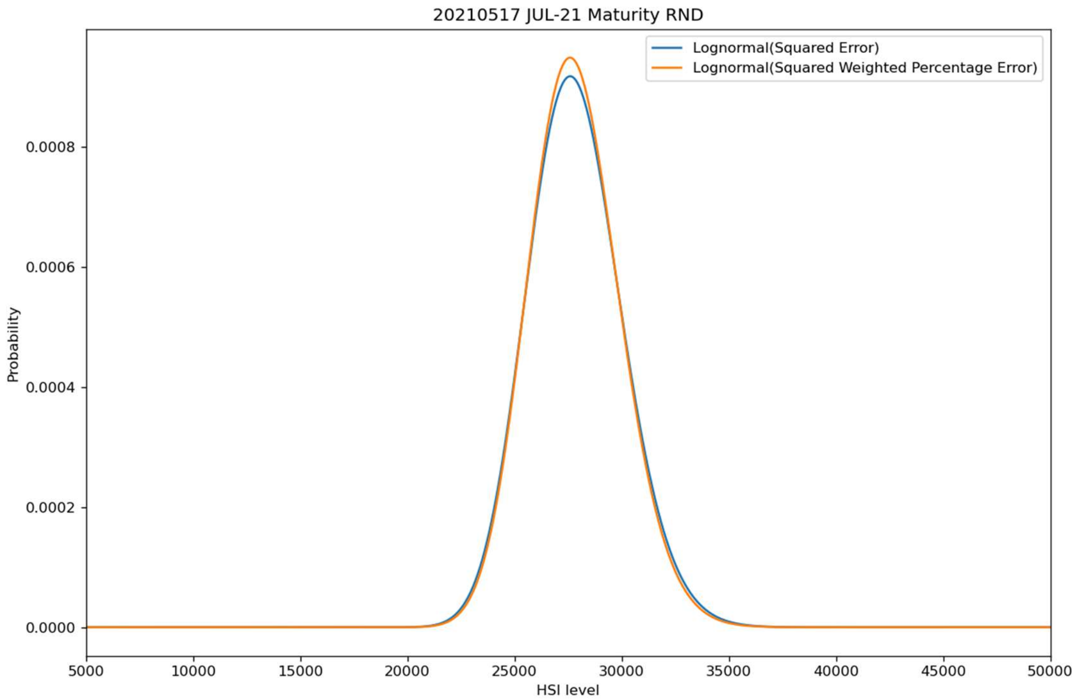
Finally, the optimization problem is defined as[7]:

$$Min_{r,\sigma} \sum_{\{s:s\in C^*\}} |c_*(s,T) - \hat{c}(s,T)| + \sum_{\{s:s\in D^*\}} |p_*(s,T) - \hat{p}(s,T)| \tag{19}$$

$$subject\ to \quad \begin{cases} |r| < 20\% \\ \sigma > 10\% \end{cases}$$

Below is a comparison of the RND extracted using squared error and absolute error on 2021/05/17 JLY-21 options. The two RND are very similar, but the absolute error approach yields lower average absolute percentage error[8].

Figure 11: Comparison of RNDs on JUL-21 maturity, 2021/09/17



[7] $\hat{c}(s,T)$, $\hat{p}(s,T)$, $c_*(s,T)$, $p_*(s,T)$ are defined as the same in section VI, C^* is the set of call option strikes, excluded strikes with 1HKD price except the first one, D^* is the set of put option strikes, excluded strikes with 1HKD price except the first one

[8] $average$ absolute percentage error $= \left(\sum\left|\frac{\text{predicted price}}{\text{actual price}} - 1\right|\right) / \ N$

Figure 11: Average absolute percentage error of two approaches on JUL-21 call options, 2021/05/17

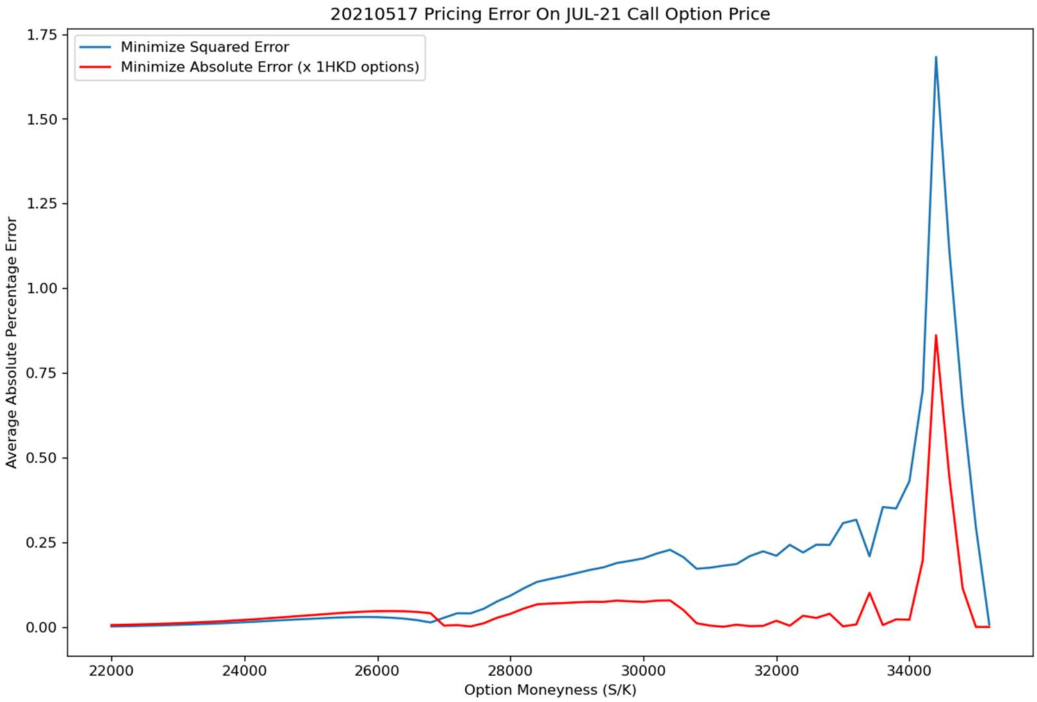
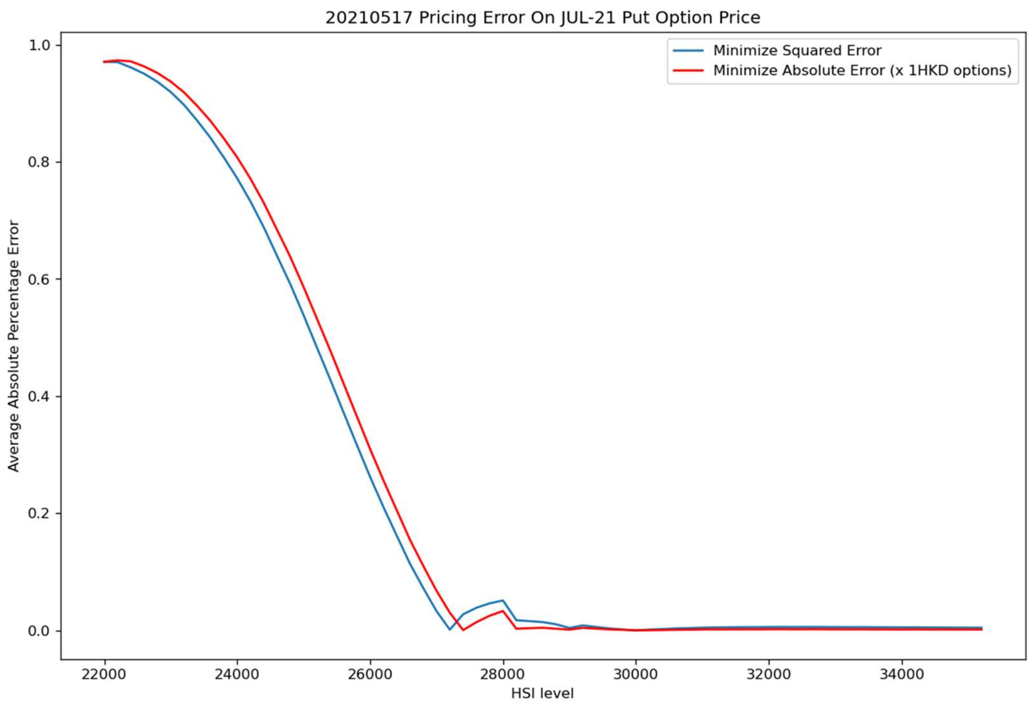


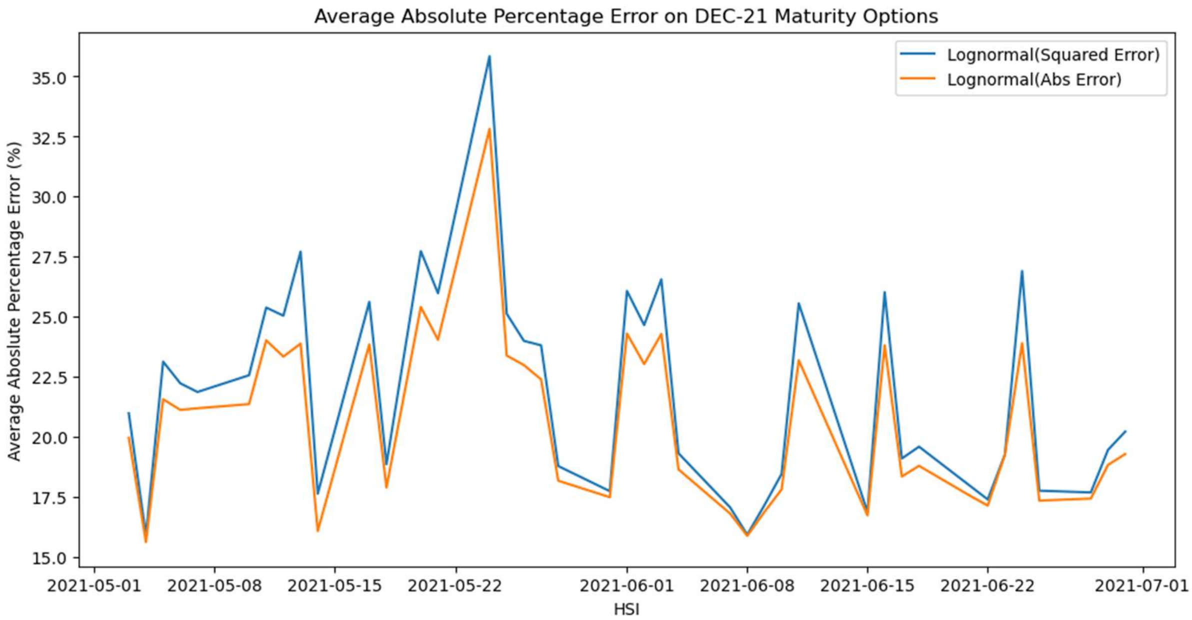
Figure 12: Average absolute percentage error of two approaches on JUL-21 put options, 2021/05/17



Generally speaking, the absolute error approach gives better pricing performance on the call options[9], but perform worse on the put options. However, the gain in performance on call options is much larger than the lose in performance on put options.

Moreover, we compare the average absolute percentage pricing error of the two approaches from 2021/05/01 to 2021/07/01 and found that this approach yields better pricing performance on most days[10]. Below is an example on DEC-21 options (both call & put):

Figure 13: Average absolute percentage error on DEC-21 options



VIII. Mixture Lognormal Approach (Squared Error)

One major drawback of the parametric lognormal approach is that it assumes zero skewness premium. In practice, skewness premium could exist and skew of the RND can

[9] Pricing performance is defined as $\text{absolute percentage error} = |\frac{\text{predicted price}}{\text{actual price}} - 1|$

[10] Average absolute percentage error is defined as the average absolute percentage error across all options

deviate from that of the lognormal distribution[11] (Bahra, 1997).

The mixture lognormal approach mitigates this problem by providing freedom to capture the skewness and the kurtosis of the true underlying RND. This means that a mixture lognormal parametric model can capture the non-constant Implied volatility smile curve (Bahra, 1995).

In our implementation, we use a mixture of two lognormal distributions. The optimization problem to estimate the parameters is defined as follows[12]

$$Min_{r_1,r_2,\sigma_1,\sigma_2} \sum_{\{s:s\in C\}} (c_*(s,T) - \hat{c}(s,T))^2 + \sum_{\{s:s\in D\}} (p_*(K_i,T) - \hat{p}(s,T))^2 \tag{20}$$

$$subject\ to \quad \begin{cases} |r_1|, |r_2| < 20\% \\ \sigma_1, \sigma_2 > 10\% \\ 0 \le w \le 1 \end{cases}$$

$$\hat{c}(s,T) = e^{-r_f T} \sum_{i=0}^{11000} g(5000 + i * 5, S_0, r_1, r_2, \sigma_1, \sigma_2, w) * Max(5000 + i * 5 - s, 0) * 5 \tag{21}$$

$$\hat{p}(s,T) = e^{-r_f T} \sum_{i=0}^{11000} g(5000 + i * 5, S_0, r_1, r_2, \sigma_1, \sigma_2, w) * Max(s - 5000 + i * 5, 0) * 5 \tag{22}$$

$$\begin{aligned} g(x, S_0, r_1, r_2, \sigma_1, \sigma_2, w) \\ = w\left(\frac{1}{x\sigma_1\sqrt{T}\sqrt{2\pi}} e^{-\frac{\left(\ln\left(\frac{x}{S_0}\right)-\left(r_1-\frac{1}{2}\sigma_1^2\right)T\right)^2}{2\sigma_1^2 T}}\right) \\ + (1-w)\left(\frac{1}{x\sigma_2\sqrt{T}\sqrt{2\pi}} e^{-\frac{\left(\ln\left(\frac{x}{S_0}\right)-\left(r_2-\frac{1}{2}\sigma_2^2\right)T\right)^2}{2\sigma_2^2 T}}\right) \end{aligned} \tag{23}$$

Some examples of the extracted RND using this approach are shown below.

---

[11] Zero skewness premium means that the probability of an OTM call being ATM at maturity is the same as that of an equally OTM put being ATM at maturity (Bahra, 1997).

[12] $c_*(s,T)$, $p_*(s,T)$, C, D are defined as the same in section VI.

Notice that it is left-skewed on some dates:

Figure 14: 2-Component Mixture Lognormal RND of DEC-21 Options

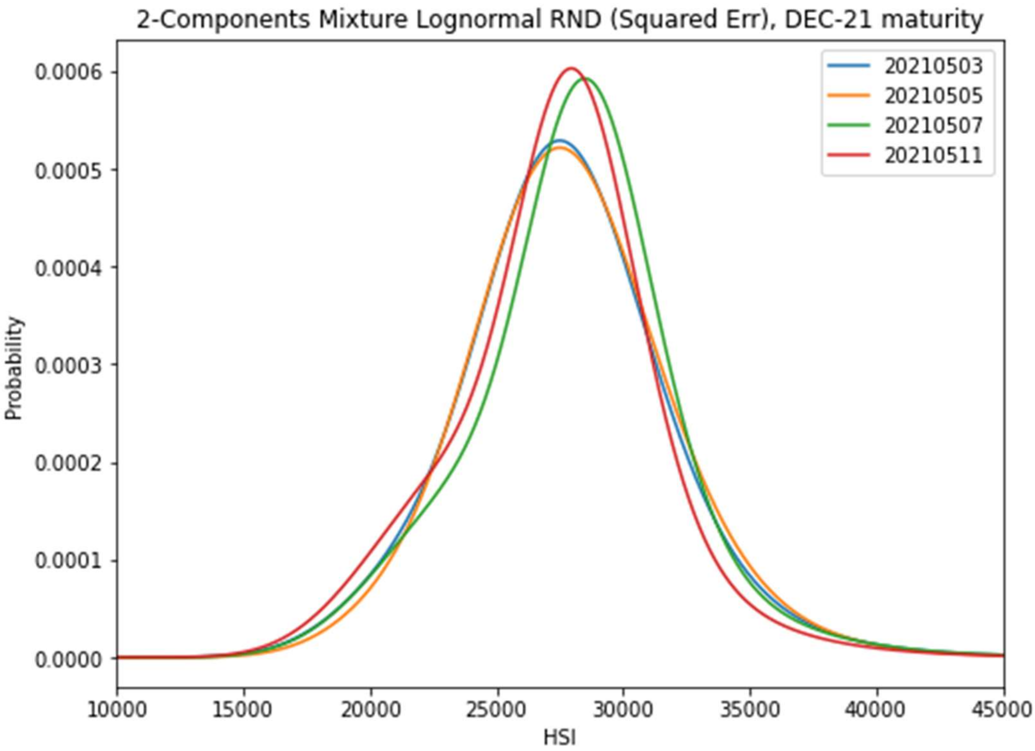


Figure 15: 2-Component Mixture Lognormal RND (Squared Error) on 2021/05/03

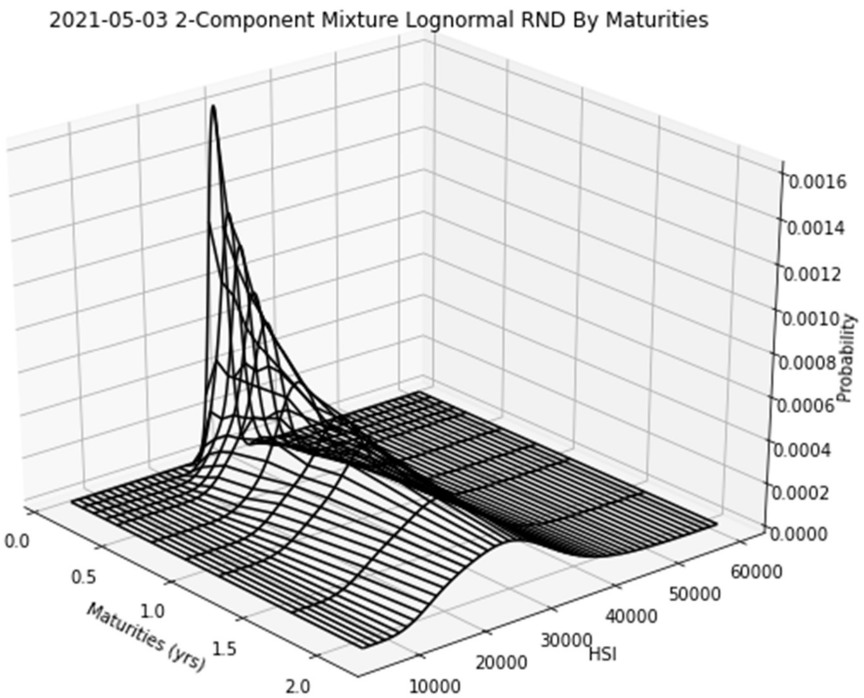
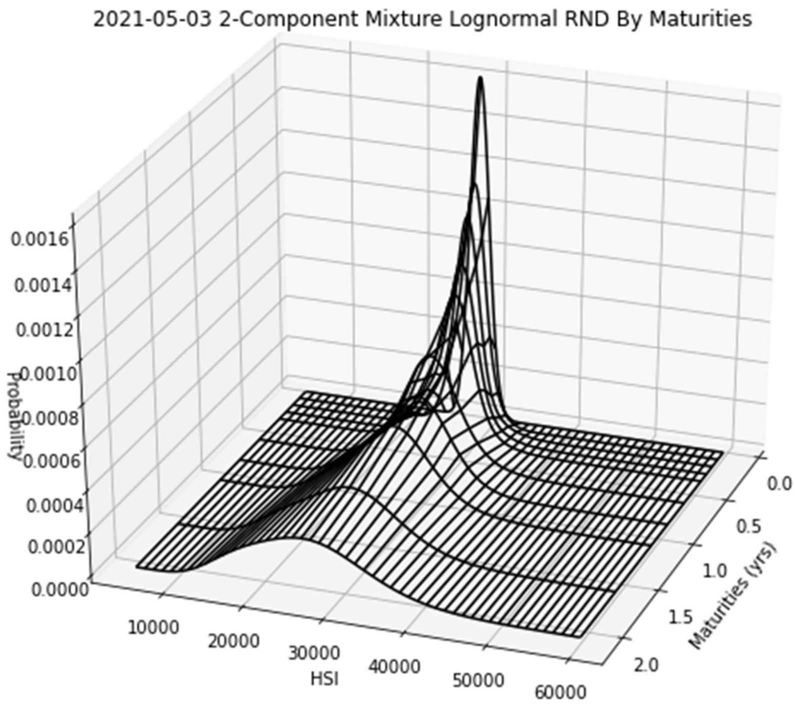


Figure 16: 2-Component Mixture Lognormal RND (Squared Error) on 2021/05/03



IX. Mixture Lognormal Approach (Absolute Error)

To mitigate the problem described in section VII, we replace the cost function in section VIII with absolute error, and excluded most options with 1 HKD price except the first one[13]:

$$Min_{r_1,r_2,\sigma_1,\sigma_2} \sum_{\{s:s\in C^*\}} |c_*(s,T) - \hat{c}(s,T)| + \sum_{\{s:s\in D^*\}} |p_*(s,T) - \hat{p}(s,T)| \tag{24}$$

$$subject\ to \quad \begin{cases} |r_1|, |r_2| < 20\% \\ \sigma_1, \sigma_2 > 10\% \\ 0 \le w \le 1 \end{cases}$$

Some examples of the RNDs extracted using this approach, and a comparison of RNDs using two different cost functions are shown below. Notice that the two approaches generate quite different RND on the same day.

---

[13] $\hat{c}(s,T)$, $\hat{p}(s,T)$ $c_*(s,T)$, $p_*(s,T)$,C^*, D^* are defined as the same in section VII.

Figure 17: 2-Component Mixture Lognormal (Abs Error) RND of DEC-21 Options

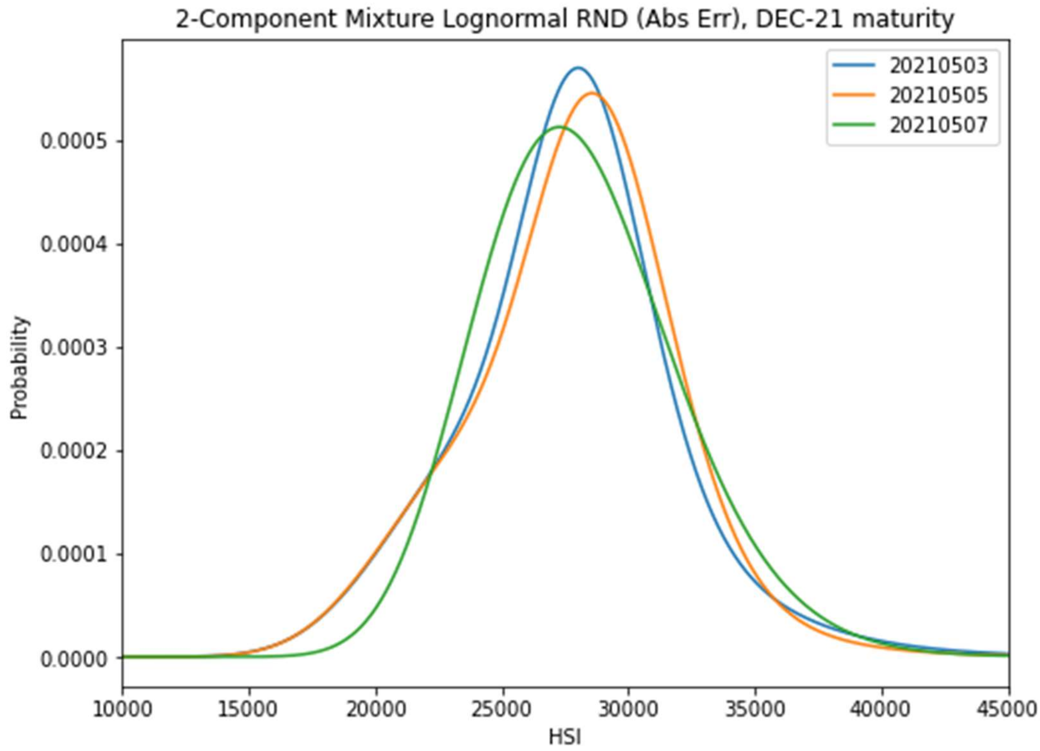


Figure 18: 2-Component Mixture Lognormal RND (Absolute Error) on 2021/05/03

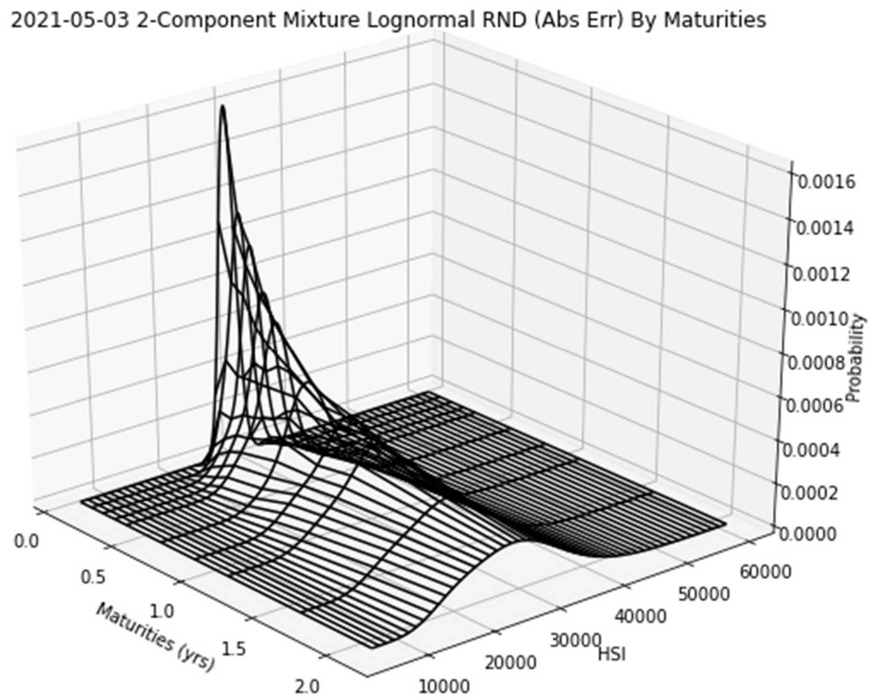
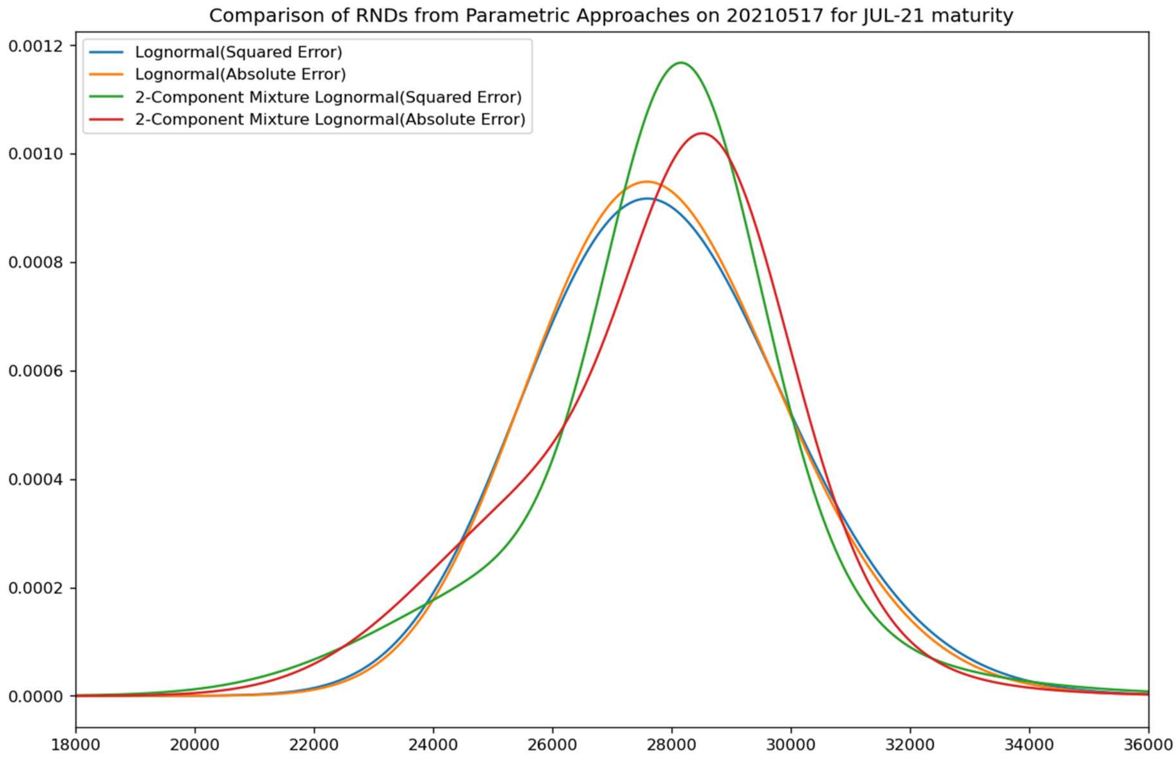


Figure 19: Comparison of Parametric Approaches for JUL-21 Maturity on 2021/05/03



X. Comparison and Summary

Section A. Moments Comparison

In practice, RNDs extracted from different approaches can be very different, since they have very different purpose. The parametric approaches derive RNDs that can accurately price options by solving an optimization problem. Whereas the Breeden and Litzenberger approach has a stronger theoretical grounding and uses option prices to approximate second order derivatives with respect to strike of option price function in order to estimate RND. In this section, we compare the first 4 moments (mean, variance, skewness, kurtosis) extracted from the methods, and evaluate which method generates the most interpretable moments for gauging market sentiments. In this section, we do not include the risk-neutral density obtained from the Breeden and Litzenberger approach since they are discrete. However, kernel regression on density from the Breeden and Litzenberger approach is evaluated as a proxy. We use the below formula to approximate the centered moments of return:

$$S(i) = 5000 + i * 5 \tag{25}$$

$$\hat{\mu} = \int_0^{\infty} \left(\frac{S_T}{S_0} - 1\right) q(S_T) dS_T \approx \sum_{i=0}^{11000} \hat{q}\big(S(i)\big) * \left(\frac{S(i)}{S_0} - 1\right) * 5 \tag{26}$$

$$\widehat{\sigma^2} = \int_0^{\infty} \left(\left(\frac{S_T}{S_0} - 1\right) - \hat{\mu}\right)^2 q(S_T) dS_T \approx \sum_{i=0}^{11000} \hat{q}\big(S(i)\big) * \left(\left(\frac{S(i)}{S_0} - 1\right) - \hat{\mu}\right)^2 * 5 \tag{27}$$

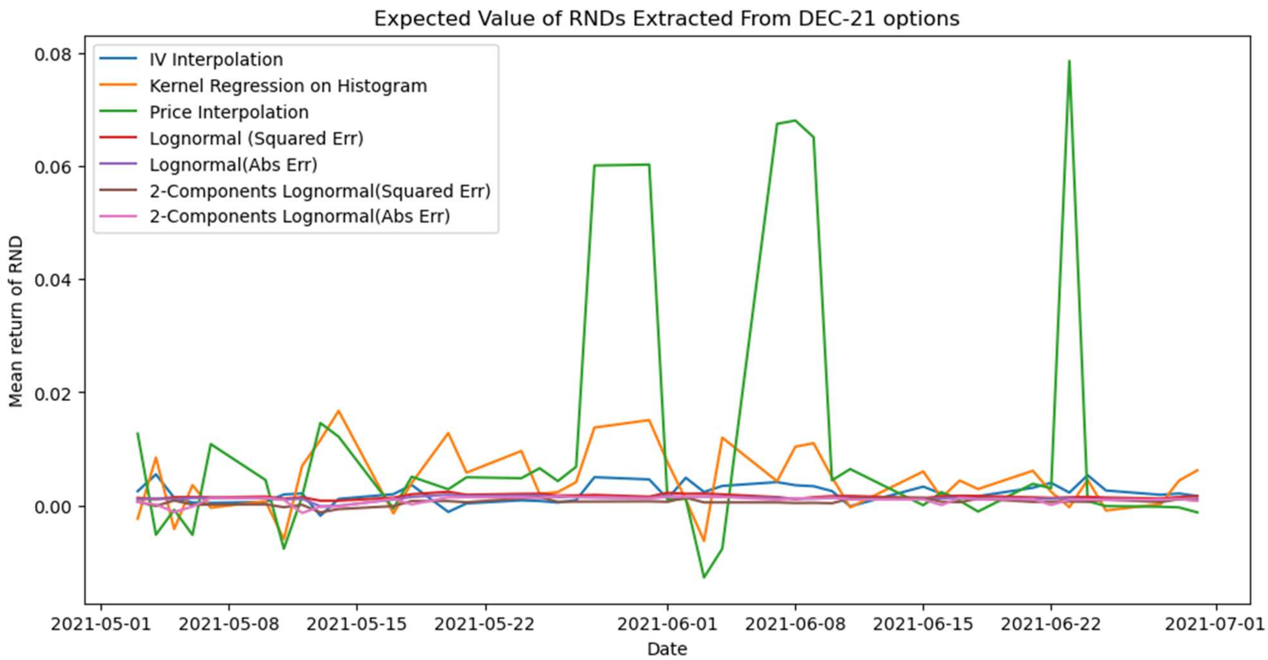
$$\widehat{\mu_3} = \int_0^{\infty} \left(\frac{\left(\frac{S_T}{S_0} - 1\right) - \hat{\mu}}{\hat{\sigma}}\right)^3 q(S_T) dS_T \approx \sum_{i=0}^{11000} \hat{q}\big(S(i)\big) * \left(\frac{\left(\frac{S(i)}{S_0} - 1\right) - \hat{\mu}}{\hat{\sigma}}\right)^3 * 5 \tag{28}$$

$$\widehat{\mu_4} = \int_0^{\infty} \left(\frac{\left(\frac{S_T}{S_0} - 1\right) - \hat{\mu}}{\hat{\sigma}}\right)^4 q(S_T) dS_T \approx \sum_{i=0}^{11000} \hat{q}\big(S(i)\big) * \left(\frac{\left(\frac{S(i)}{S_0} - 1\right) - \hat{\mu}}{\hat{\sigma}}\right)^4 * 5 \tag{29}$$

A-1: Mean of Extracted RNDs

Overall, mean of extracted RNDs are quite different. The price interpolation method is not robust as the mean shoots up occasionally. In addition, the parametric methods give unrealistic means (implied r_f) that are close to 0[14]. Whereas methods that are based on butterfly spread prices (Kernel regression and price interpolation method) give more realistic means, which are around $1\% - 2\%$[15]. In this regard, the parametric methods are less preferable, as they generate RNDs that imply $r_f = 0$. The IV interpolation, kernel regression and price interpolation method are more preferable to gauge the market implied r_f, even though their mean fluctuates a lot.

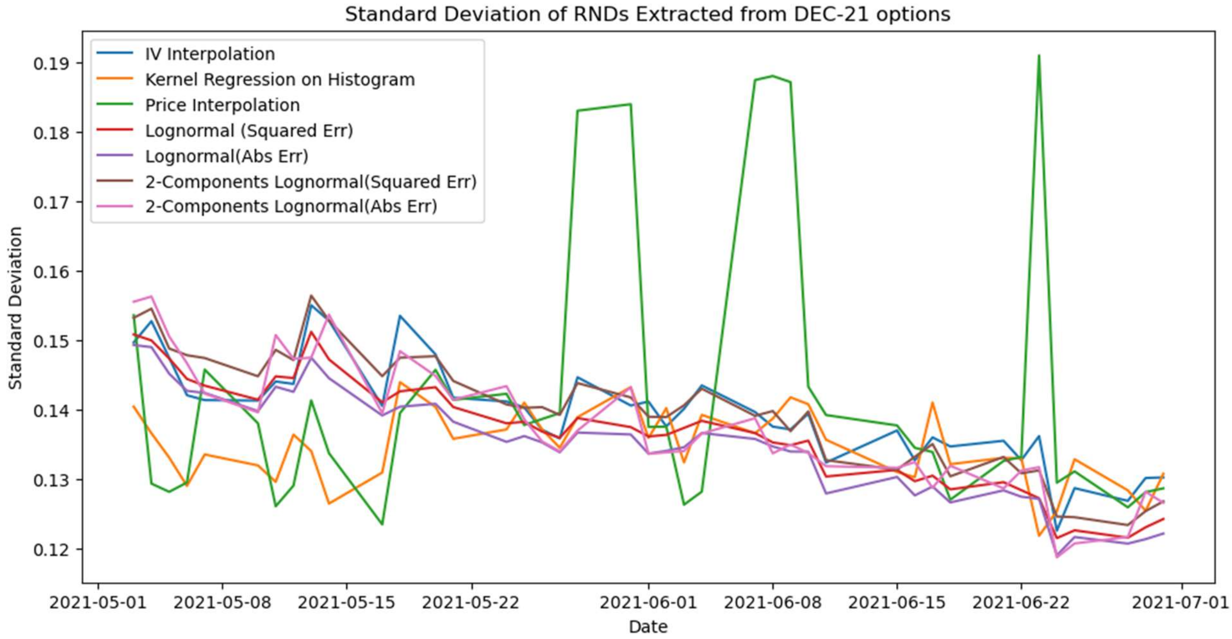
Figure 20: Extracted mean from different approaches



[14] Note that this is not annualized mean return

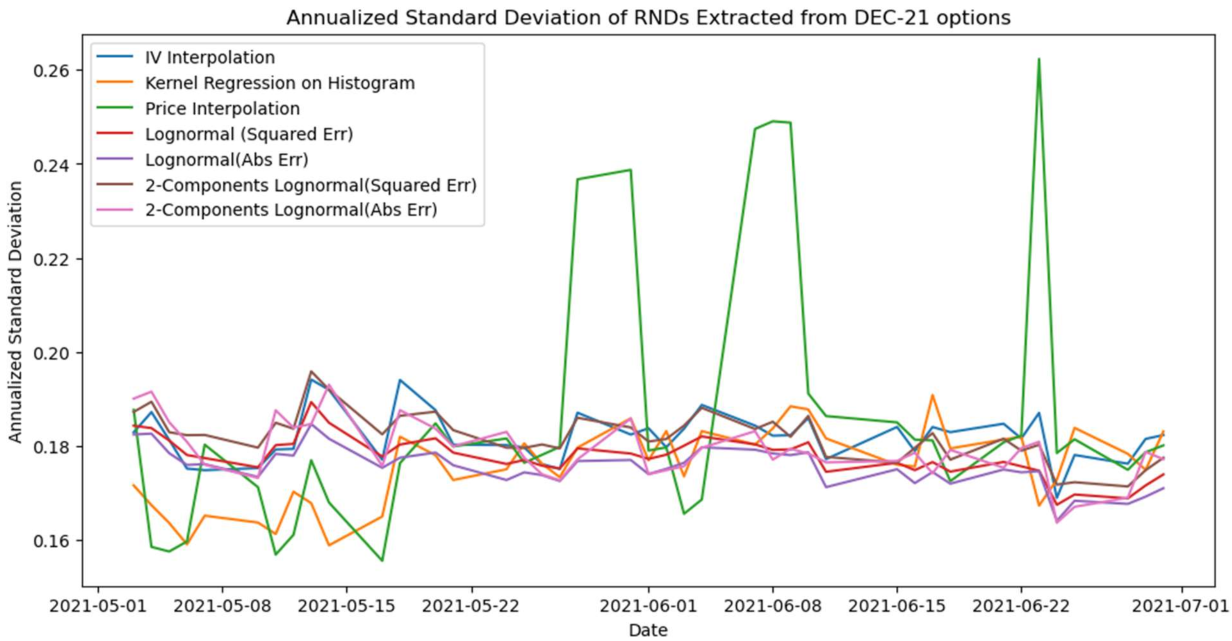
[15] The mean was around 1% for the two methods in May, when the time-to-maturity T for the DEC-21 options was around 0.5.

Figure 21: Extracted standard deviation for DEC-21 maturity



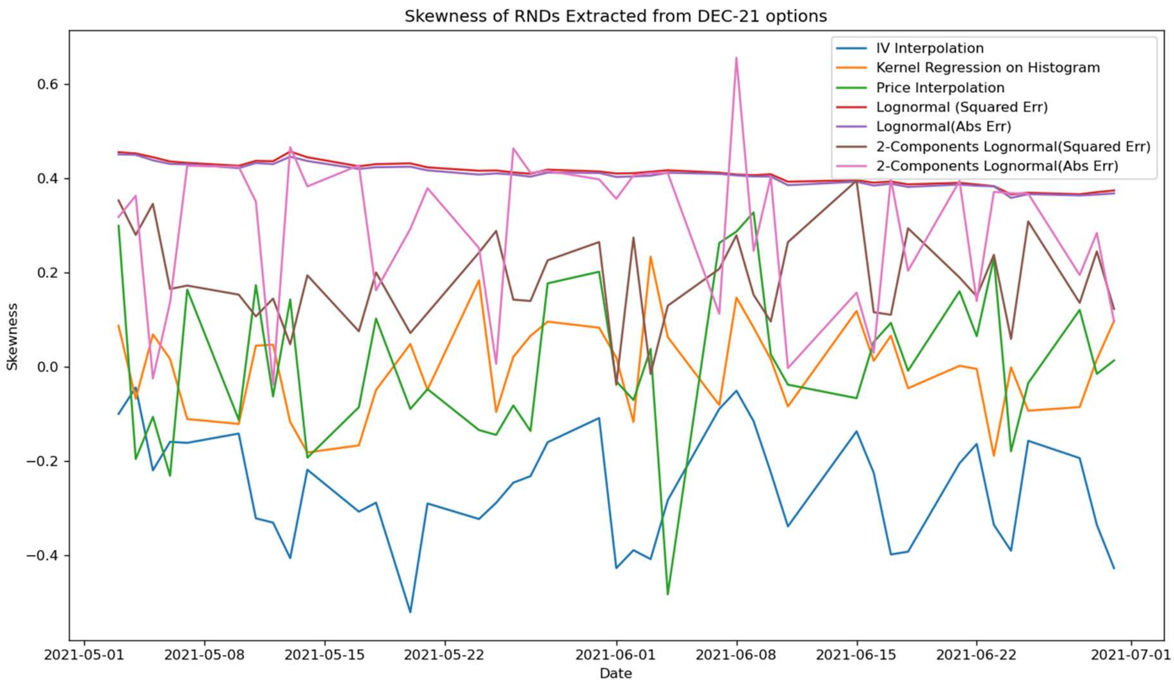
All methods give similar variance which decreases with the time-to-maturity T. Similar to the extracted mean, the price interpolation method is not robust as the variance shoots up occasionally. We annualized the extracted standard deviation and obtain the following:

Figure 22: Extracted annualized standard deviation for DEC-21 maturity



As shown above, annualized standard deviation of HSI index implied by the RNDs is around 18% during the two-month period. The parametric methods and IV Interpolation method seem to be quite robust in extracting the second moments as the daily variation of the extracted annualized standard deviation is small. This extracted annualized standard deviation does not have to equal the historical standard deviation, and it could be used to gauge the market expectation of future variance.

Figure 23: Skewness of RNDs Extracted for DEC-21 maturity

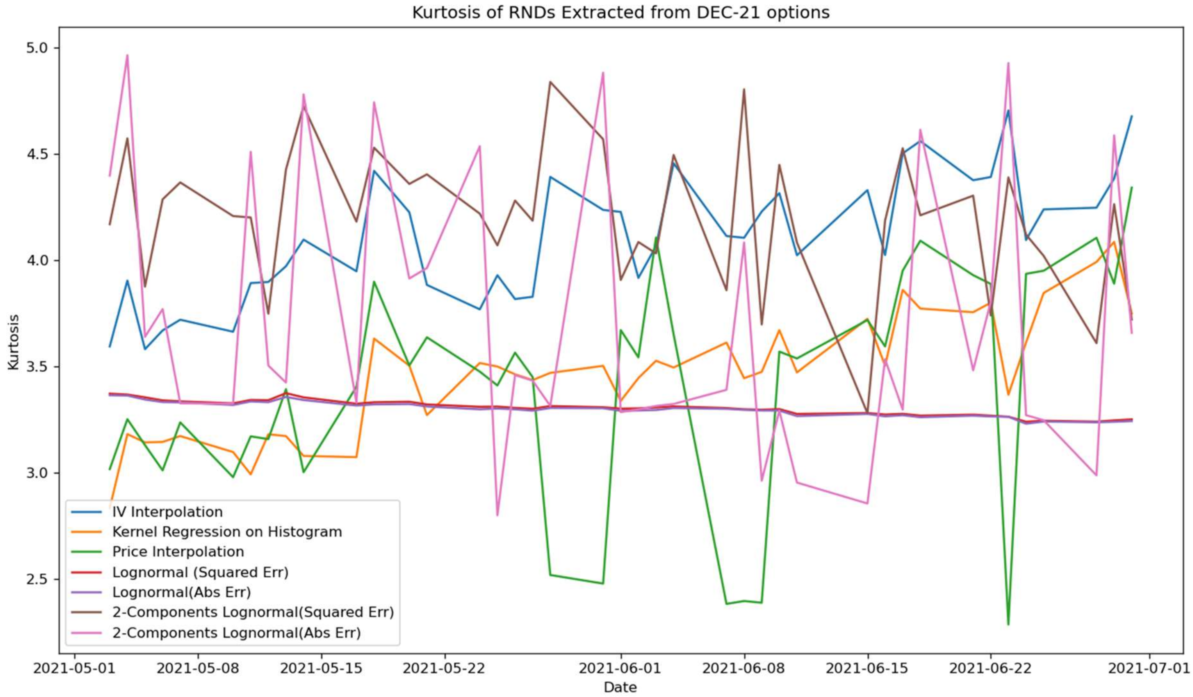


Lognormal distribution is positively skewed and hence the skewness of RNDs extracted from the parametric lognormal approaches are always positive. RNDs extracted from the IV interpolation method are always negatively skewed, as the method retains the Implied volatility patterns of the options[16]. In this regard, the Implied volatility Interpolation method maybe more preferable, and it should have better pricing accuracy.

[16] IV interpolation method interpolates the Implied volatility smile curve directly

Most of the other methods give fluctuating results around 0 which make it hard to evaluate true sign of skewness of the true RND.

Figure 24: Kurtosis of RNDs Extracted for DEC-21 maturity



All methods except the price interpolation method have positive excess kurtosis (kurtosis above 3) over the 2-month period, meaning that their RNDs are fat-tailed.

The IV interpolation method and 2-components lognormal (Squared Error) method are the most leptokurtic over the period. Since the main contributions to the volatility smile curve of options are skewness and kurtosis (Bahra, 1997), the higher kurtosis of the two methods could mean that they can best retain the original implied volatility structure of option prices.

Generally speaking, IV interpolation method and kernel regression method give the "nicest" extracted kurtosis, as the daily variation is low, and we can see the trend more clearly (blue & orange line is trending upward). For other methods, the extracted kurtosis is too volatile that it is hard to infer any meaningful trend (and use it to gauge market sentiments).

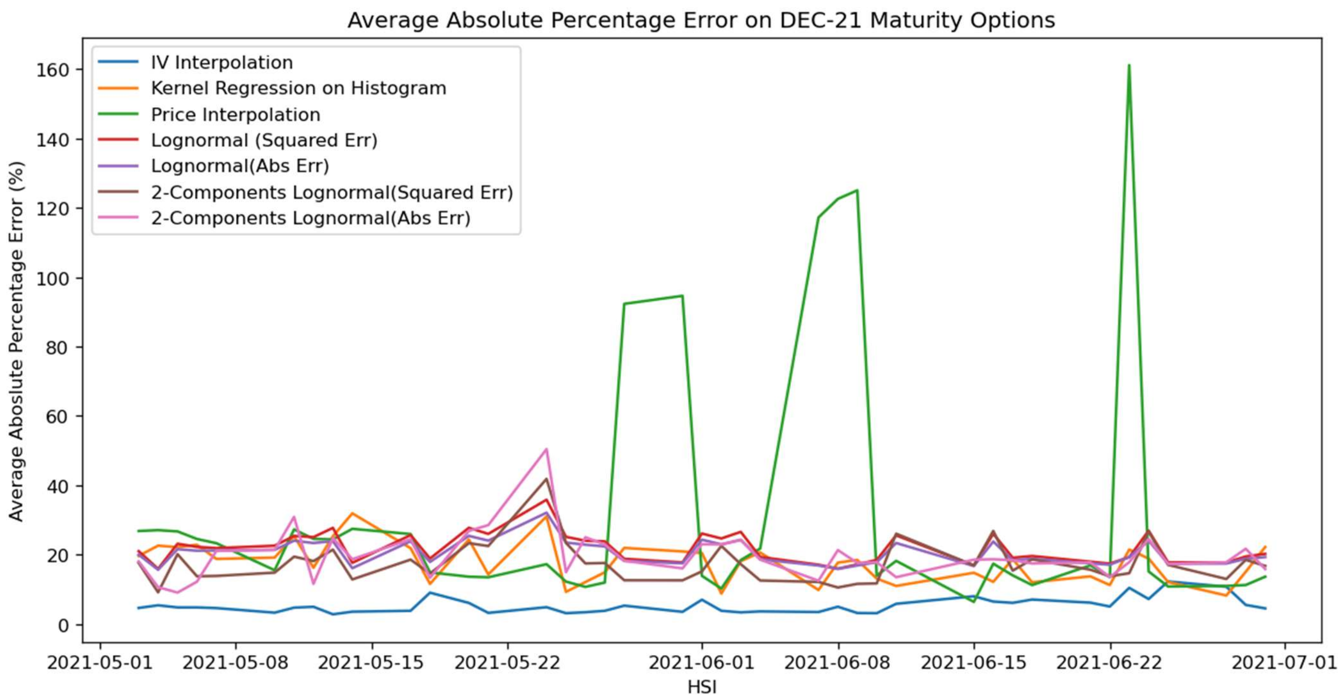
Section B. Pricing Accuracy Comparison

We compare the pricing performance of the extracted RNDs in terms of average absolute percentage error E[17]:

$$E = \frac{\sum_{\{s:s\in C\}} \left|\frac{c_*(s,T)}{\hat{c}(s,T)} - 1\right| + \sum_{\{s:s\in D\}} \left|\frac{p_*(s,T)}{\hat{p}(s,T)} - 1\right|}{N} \tag{30}$$

We estimated RNDs using different approaches from 2021/05/01 to 2021/07/01 and use the daily RNDs to evaluate the pricing error on each day. We tested the pricing error for three different maturities: DEC-21, MAR-22 and JUN-22. Below are the results:

Figure 25: Daily average absolute percentage error on DEC-21 maturity options



[17] $c_*(s,T)$, $p_*(s,T)$, C, D are defined as the same in section VIII, $\hat{c}(s,T)$, $\hat{p}(s,T)$ are the estimated option prices of strike s, maturity T using RND.

Figure 26: Daily average absolute percentage error on MAR-22 maturity options

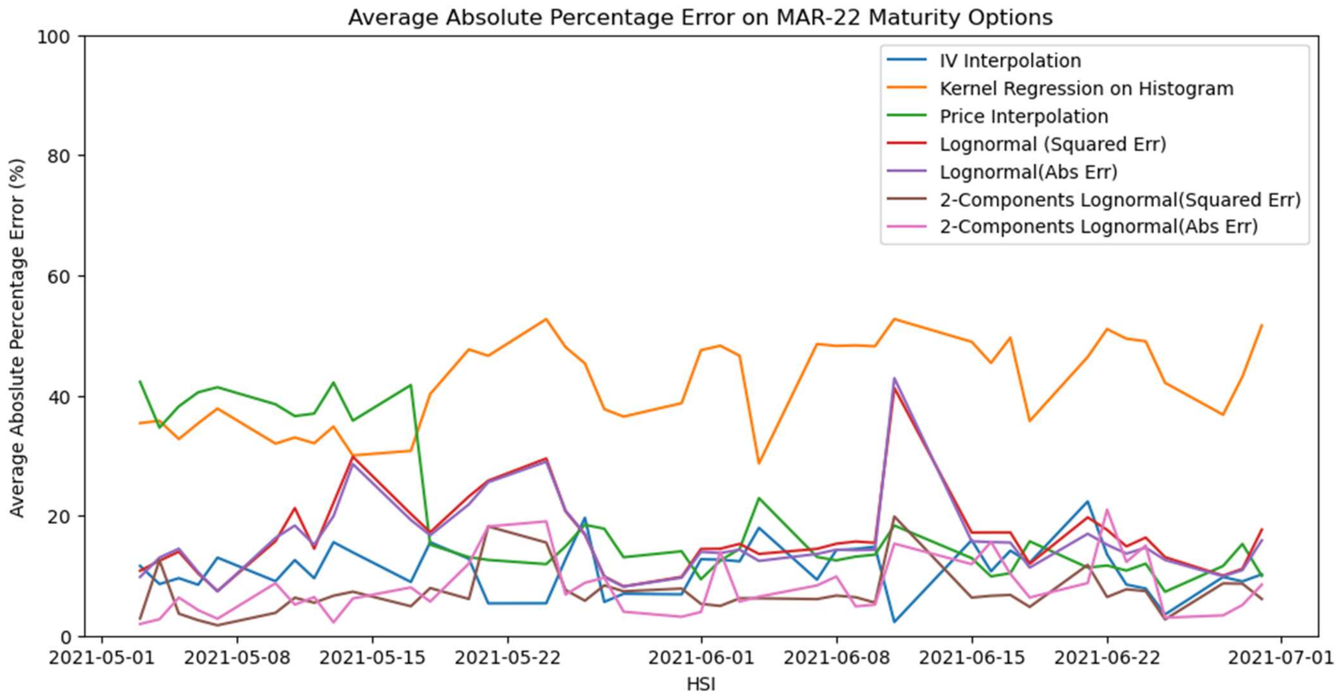
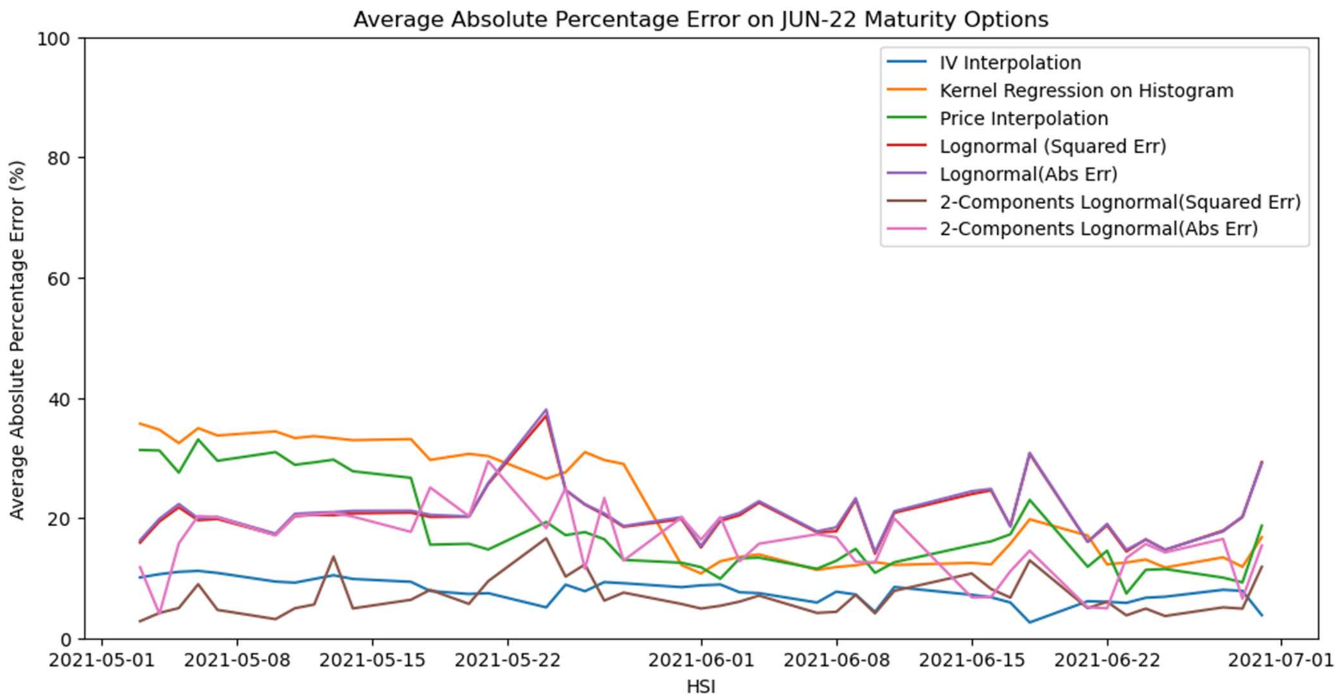


Figure 27: Daily average absolute percentage error on JUN-22 maturity options



Overall, the IV interpolation method and 2-components lognormal parametric methods (both absolute error and squared error) have the best pricing accuracy. The daily average absolute percentage error was only 3-5% for DEC-21 options and 5-10% for MAR-22 and JUN-22 options during the 2-month period for the three methods. This could be due to the fact that they are able to capture the skewness and kurtosis of the true underlying RND. This means that they are able to capture the implied volatility pattern of the options, and hence price options more accurately.

Although the price interpolation method did well for some periods, it is not robust as the error shoots up occasionally. The lognormal parametric approaches are mediocre in terms of pricing accuracy compared with the mixture lognormal approaches. Lastly, the kernel regression method has very different pricing performance on different maturities.

Section C. Summary

Overall, the Implied volatility interpolation method is the best method to extract RND from European option prices in the HSI option market. It generates moments that are interpretable, and its pricing accuracy is high. And more importantly, it retains the original implied volatility structure of option prices, as the method directly interpolate the IV curve.

The Breeden and Litzenberger Approach is intuitive but there are many difficulties in implementation. As the option prices released by the HKEX may not be the latest tradeable price, and they are rounded to integers, negative butterfly spread prices may occur in practice, which requires special treatments. In addition, a continuum of options does not exist in practice, which reduces the approximation accuracy of the method.

The price interpolation is easy to implement in practice, but it requires selection of an interpolation method which could greatly change the resulting RND. Furthermore, the method is not robust using PCHIP interpolation method as it may lead to "poor" RNDs occasionally which have low pricing accuracy and volatile moments.

The parametric methods using a lognormal density is also intuitive as it aligns with the consequence of Black-Scholes model. However, it is always positively skewed, and it assumes a single implied volatility for all strikes. On the other hand, the parametric methods using a mixture of two lognormal distributions are better because they can capture the skewness and kurtosis of the true underlying RND, which are the main contributions to the implied volatility curve (Bahra, 1997). Because of this, the mixture approaches also yield better pricing accuracy.

References

Bahra, Bhupinder, Implied Risk-Neutral Probability Density Functions from Option Prices: Theory and Application (July 1997). *Bank of England* Working Paper No 66.

Neuhaus, Holger, The Information Content of Derivatives for Monetary Policy: Implied Volatilities and Probabilities (1995). *Bundesbank* Series 1 Discussion Paper No. 1995,03E.

Pontius, R.G., Thontteh, O. & Chen, H. Components of information for multiple resolution comparison between maps that share a real variable. *Environ Ecol Stat* 15, 111–142 (2008).

Shimko, D, Bounds of Probability (1993). *RISK*, Vol.6, No.4.