PAPER • OPEN ACCESS

Empirical performance of interpolation techniques in risk-neutral density (RND) estimation

To cite this article: H Bahaludin and M H Abdullah 2017 *J. Phys.: Conf. Ser.* **819** 012026

View the article online for updates and enhancements.

Empirical performance of interpolation techniques in risk-neutral density (RND) estimation

H Bahaludin, M H Abdullah

Department of Computational and Theoretical Sciences, Kulliyyah of Science, International Islamic University Malaysia, Bandar Indera Mahkota Campus, Jalan Sultan Ahmad Shah, 25200 Kuantan, Pahang, Malaysia

hafizahbahaludin@iium.edu.my, mimihafizah@iium.edu.my

Abstract. The objective of this study is to evaluate the empirical performance of interpolation techniques in risk-neutral density (RND) estimation. Firstly, the empirical performance is evaluated by using statistical analysis based on the implied mean and the implied variance of RND. Secondly, the interpolation performance is measured based on pricing error. We propose using the leave-one-out cross-validation (LOOCV) pricing error for interpolation selection purposes. The statistical analyses indicate that there are statistical differences between the interpolation techniques:second-order polynomial, fourth-order polynomial and smoothing spline. The results of LOOCV pricing error shows that interpolation by using fourth-order polynomial provides the best fitting to option prices in which it has the lowest value error.

1. Introduction

The volatility function technique is considered as a nonparametric approach to extract the RNDs [1] . This method is based on the volatility smile function. In this approach, the estimated RNDs are calculated using second derivative of option prices with a condition of continuum option prices [2]. The condition of continuum option prices is impossible to exist in the real market. Thus, previous studies proposed various interpolation techniques to overcome this condition of continuum option prices. Interpolation techniques play a major part in deriving the true density from option prices.

In beginning, the second-order polynomial interpolation was introduced to smooth a volatility smile [3]. It has been proposed to fitting the implied volatility instead of using the option prices directly due to the non-linearity problem [4]. Similar to [3], a second-order polynomial is used by [5] but in an implied volatility-delta space to interpolate the implied volatility function. The delta can be defined as the rate of change of the option prices with respect to the price of the underlying asset [6]. In a different study, a cubic spline interpolation is used to interpolate the implied volatility in an implied volatility-strike space [7]. The cubic spline is preferred because of the flexibility in controlling the shape of a volatility smile. A smoothing spline intepolation was developed by [1] that combining [5,7] approached with certain modifications. This approach introduced a vega of options as the weight of each distribution together with a smoothing parameter [1]. Recently, drawback of the cubic spline in RNDs estimation has been pointed out by [8] and proposed a fourth-order polynomial interpolation on an implied volatility – strike price space to generate the continuum of option prices.

To date, there are several types of interpolation techniques provided by previous literature of RNDs. However, previous studies do not have an agreement to which interpolation techniques should

be used in RNDs estimation. Thus, our study is largely motivated to know if there are any differences among the interpolation techniques with regards to RNDs estimation. To the best of our knowledge, this study is the first to analyze the statistical differences between the interpolation techniques namely second-order polynomial, fourth order polynomial and smoothing spline which are used to extract the RNDs. This paper contributes to the empirical finance literature by applying the cross-validation technique to determine the performance of each interpolation technique. The leave-one-out cross validation (LOOCV) of option prices error is chosen to select the appropriate interpolation which gives the lowest value error.

This paper is organized in the following ways. The second section presents the formula of Black-Scholes-Merton model, and the third section presents the theory of risk-neutral density. Fourth section lists the procedures to estimate the risk-neutral density using interpolation techniques. Section five describes the model evaluation and section six presents the data. Seventh section discusses the results and the final section presents the concluding remarks.

2. Black-Scholes-Merton model and implied volatility
This study uses the Black-Scholes-Merton (BSM) model to price the option. The formula of BSM model is as follows:

$$c = S_0 e^{-qT} N(d_1) - K e^{-rT} N(d_2)$$
$$p = K e^{-rT} N(-d_2) - S_0 e^{-qT} N(-d_1)$$

(1)

where

$$d_1 = \frac{\ln(S_0/K) + \left((r-q) + \sigma^2/2\right)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}$$

The function $N(\cdot)$ is the cumulative probability distribution function for a standardised normal distribution. The variables c and p are the European call and European put prices, respectively; S_0 is the price of the underlying asset; K is the strike price; r is the continuously compounded risk-free rate; σ is the underlying asset price volatility; and T is the time to maturity of the option.

The volatility of underlying asset price in the BSM formula cannot be derived directly. Thus, this study implements the bisection method algorithm to get the implied volatility by using the inverse formula of the BSM model.

3. Risk-neutral density from option prices
Price of a call option is given by the discounted value of expected payoff on the maturity date, T with respect to the risk-neutral probability:

$$C(T,K) = e^{-rT} \int_0^\infty (S_T - K, 0) f(S_T) dS_T$$

(2)

where C is the European call price, S_T is the price of the underlying asset; K is the strike price; r is the continuously compounded risk-free rate and $f(S_T)$ is the RND function.

The RND can be extracted from the second derivative of an option price with continuous strike prices [9]. From equation (2) the partial derivative with respect to the strike price, K yields:

$$\frac{\partial C(T,K)}{\partial K} = -e^{-rt} \int_K^\infty f(S_T) dS_T = e^{-rt}[1 - F(K)]$$

(3)

Solving the risk neutral distribution, $F(K)$, gives

$$F(K) = e^{rT} \frac{\partial C}{\partial K} + 1$$

(4)

Then, the second derivative with respect to K in equation (4) gives a risk-neutral density function at K:

$$f(K) = e^{rT} \frac{\partial^2 C(T,K)}{\partial K^2} \tag{5}$$

Numerically, equations (3) and (5) can be approximated using the following equations:

$$F(K_N) \approx e^{rT} \left[\frac{C_{n+1} - C_{n-1}}{K_{n+1} - K_{n-1}} \right] + 1 \tag{6}$$

$$f(K_n) \approx e^{rt} \frac{C_{n+1} - 2C_n + C_{n-1}}{(\Delta K)^2} \tag{7}$$

4. Estimation of risk-neutral density procedures

This subsection explains the procedure for the RNDs estimation by using smoothing spline and polynomial interpolation.

4.1. Extract RNDs using a smoothing spline interpolation

Firstly, the call prices are translated into implied volatilities by using the inverse of the BSM model. Then, the strike prices are converted into option delta. The delta estimation is based on the at-the-money implied volatility in order to maintain the same ordering with the strike prices [1,9,10]. According to [6], the delta of call options can be calculated as follows :

$$\Delta_C = \frac{\partial C}{\partial S} = e^{-qT} N(d_1) = e^{-qT} \Phi \left(\frac{\ln S_0 - \ln K + \left(r - q + \frac{\sigma_{ATM}^2}{2} \right) T}{\sigma_{ATM} \sqrt{T}} \right) \tag{8}$$

where Φ is the standard normal cumulative distribution function.

At this stage, the call prices-strike prices spaces are fully converted into implied volatility–delta spaces. Then, the smoothing spline is used to interpolate the implied volatility by using the following equation:

$$\sigma_i(\Delta) = a_i + b_i(\Delta - \Delta_i) + c_i(\Delta - \Delta_i)^2 + d_i(\Delta - \Delta_i)^3 \tag{9}$$

The parameter of equation (9) can be estimated by minimizing the objective function as stated in equation (10)

$$\min_{\Theta} \left\{ \sum_{i=1}^{N} w_i \left(\sigma_i - \hat{\sigma}_i(\Delta_i, \Theta)^2 + \lambda \int_0^{+\infty} \sigma''(\Delta; \Theta)^2 d\Delta \right) \right\} \tag{10}$$

In line with [1], vega of options is calculated and is to represent the weight of parameter (w_i) and the value of λ is set to 0.99. Once it is fitted, extract the 1000 points of implied volatility from the implied volatility function. Last but not least, convert the 1000 point of implied volatility into the call prices using the BSM model and the option deltas are converted into strike prices. Finally, the RNDs are obtained by using the numerical calculation based on equation (7).

4.2. Extract RNDs using a polynomial interpolation

The procedures of the second and fourth order polynomials in RNDs estimation are basically referred to [3,8] . The polynomial interpolation takes place in the implied volatility-strike prices spaces. Thus, the call prices are converted into implied volatilities. The implied volatility- strike space is fitted using

the second-order or fourth order polynomials. The implied volatility functions can be expressed as follows:

i.   Second-order polynomial : $\sigma(K\,|\,\theta)=a+bK+cK^2$ where $\theta=(a,b,c)$

ii.   Fourth-order polynomial : $\sigma(K\,|\,\theta)=a+bK+cK^2+dK^3+eK^4$ where $\theta=(a,b,c,d,e)$

The 1000 points of implied volatility are extracted from the fitted curve and are transformed back to call prices by using the BSM model. Finally, the RNDs are calculated based on equation (7).

5. Model Evaluation

This subsection describes the interpolation performance measurement. Firstly, statistical analysis is applied to examine if the interpolation techniques provide any significant differences in RND estimation. Secondly, the leave-one-out cross-validation (LOOCV) is used to examine the best fitting among the interpolation approaches.

5.1. Statistical analyses

The performance of the interpolation approaches in RNDs estimation is tested based on the following hypotheses:

Hypothesis 1 (h1) : *The interpolation by using second and fourth order polynomials do not provide any statistical difference in RND estimation.*

Hypothesis 2 (h2) : *The interpolation by using fourth order polynomial and smoothing spline do not provide any statistical difference in RND estimation*

Hypothesis 3 (h3) : *The interpolation by using second order polynomial and smoothing spline do not provide any statistical difference in RND estimation*

The implied mean and the implied variance of RNDs are used as an input for statistical analyses. The implied mean of RNDs give an expectation the price of the underlying asset price in future. The implied variance of RNDs indicates the deviation or uncertainty of the evolvement of the underlying asset price. The hypothesis for the implied mean is tested using standard *t*-test and the implied variance is tested using *F*-test.

The implied mean and the implied variance can be calculated as follows:

i.   Implied mean   : $\mu=\int_{\infty}^{\infty}Kf(K)dK$

ii.   Implied variance   : $\sigma=\sqrt{\int_{\infty}^{\infty}(K-\mu)^2 f(K)dK}$

where K is the strike price and $f(K)$ is the RND corresponding to the strike price.

5.2. Cross-validation

Basically the whole data set is fitted using the interpolation approaches. Then, the interpolation approach that provides the lowest error is chosen when compared between the data set and the fitted values. However, this approach may lead to over fitting the dataset and gives a false conclusion of a particular interpolation that provides the best fitting with the lowest error. In addition, it does not exhibit the true performance of the interpolation approach when involves a prediction of a new value.

Thus, cross-validation is used to detect and prevent the overfitting case.[11] is the first to propose a model evaluation by using a cross-valuation method. Briefly, the cross-validation deleted some of the data before an interpolation starts. The deleted data is then used to get the predicted value

using an interpolation equation. The cross-validation can be divided into three categories namely the simple cross-validation with holdout method, the K-fold cross-validation and the leave-one-out cross-validation (LOOCV). LOOCV is highly recommended as an appropriate model to use as it gives an unbiased estimator to represent the performance of a model [12].

However, the LOOCV method is not widely used in finance literature. This study is inspired by the seminal work of [13] that uses the LOOCV method to examine the performance of a spline in a volatility surface. Therefore, this study adopts the concept of LOOCV in order to determine the ability of each interpolation in RNDs estimation.

There are several steps in LOOCV estimation. Let N be the number of observation data. Firstly, remove the i-th observation data and then data is fitted corresponding to the particular interpolation. Re-evaluate the i-th data by using the model estimated on $N-i$ data. This method requires intensive computation since it is fitted N times. The LOOCV estimation can be calculated as follows:

$$CV_n = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{11}$$

where $e_i = \left(y_i - \hat{y}_i\right)^2$.

Note that the main objective of this study is to generate a continuum of option prices by using the fitted implied volatility. The difference between the market prices for options and the option prices based on the fitted model is considered as an approximation error. Following [13], the error is calculated as

$$e_i = \left(C_i - \hat{C}_i\right)^2 \tag{12}$$

where C_i is the market price of call options and $\hat{C}_i$ is the price of call options based on the fitted model.

6. Data
Our study uses dataset on Dow Jones Industrial Average (DJIA) index options and is obtained from OptionData.net. The dataset covers the period from January 2009 until December 2015. The interest rate used is based on the London Interbank Offer Rate (LIBOR) for 1-month. The price of the underlying asset, interest rate and the dividend yield are retrieved from the DataStream. Only out-of-the-money (OTM) of call and put options is considered in this study due to liquidity reasons [8,10,14–22]. In line with [8,15,23], the average of bid-ask quotes is used to represent the closing price of options. This study only considers options with a one-month maturity according to the Chicago Board Options Exchange (CBOE) calendar.

There are several restrictions applied on the dataset before the final dataset is obtained. Firstly options with bid or ask quotes equal to zero are eliminated and to ensure that the ask quotes are greater than the bid quotes. Secondly, option prices that violate the arbitrage condition are excluded. Lastly, only options with the lowest delta value less than or equals to 0.25 and the highest delta value equals or greater than 0.75, are used. This condition is applied to ensure the RNDs extracted exhibits the true density. The final sample consists of 83 set of options with a one-month maturity.

7. Results and Discussions
This subsection explains the result of the analyses. and is divided into two parts. The first part is the explanation of statistical analyses based on the implied mean and the implied variance. The second part explains the LOOCV option price error.

Options prices that is available on 22 February 2010 is used to illustrate the computational work explained in Section 4. Figure 1 illustrates the steps in RNDs estimation for the three types of interpolation. The interpolation of volatility smile is presented in the first row, followed by the graph of call prices based on the fitted implied volatility and in the last row is the RNDs estimation. These computations are applied to the whole 83 set of options. The polynomial interpolation is taken place in

an implied volatility-strike prices space. On other hand, the interpolation using smoothing spline is taken place in an implied volatility-delta space. The graph for each step corresponding to each interpolation does not display an outstanding differences and it leads us to further analyze these interpolation techniques using statistical analyses.
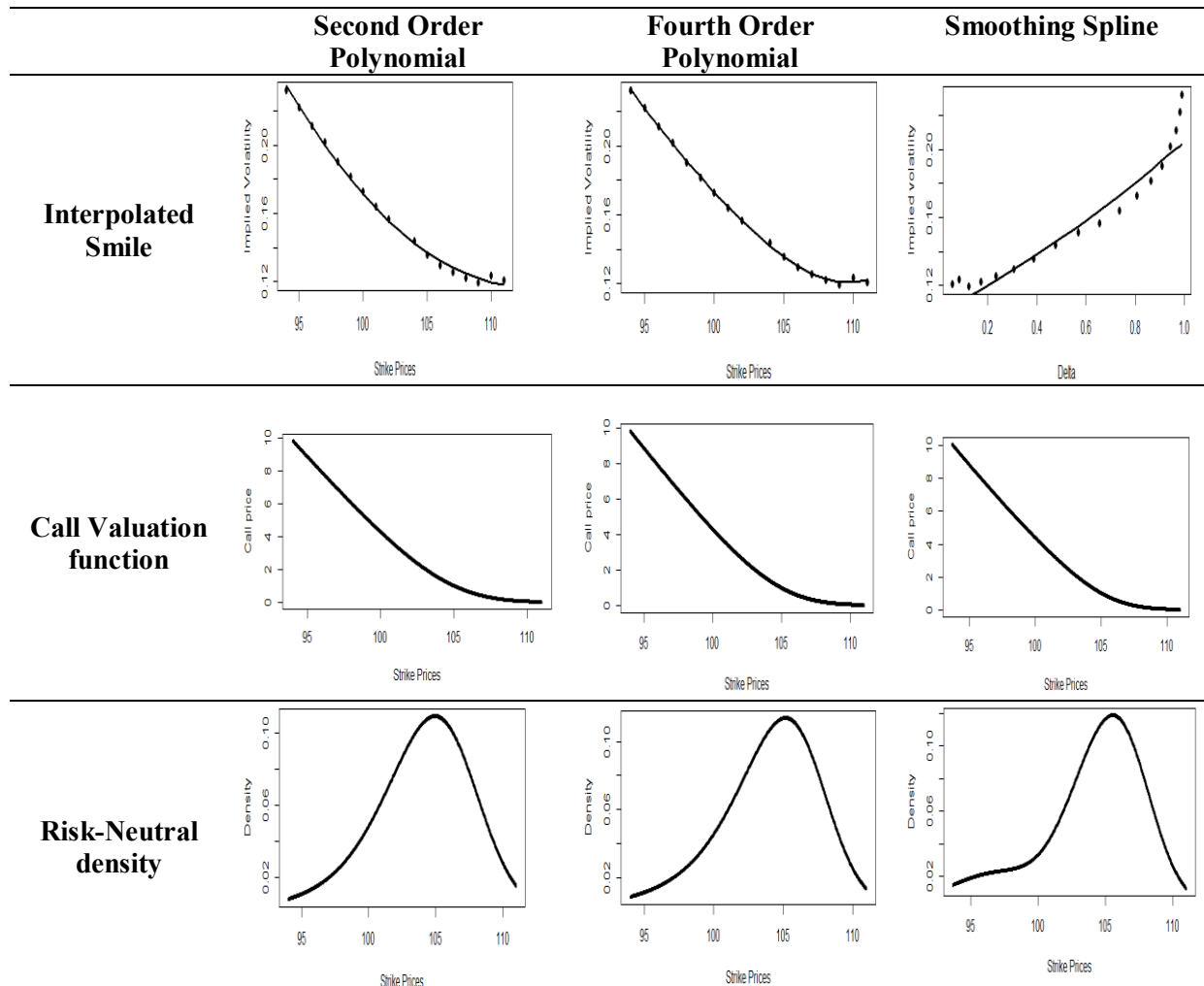

Figure 1. Illustration of computation using DJIA index options prices, February 22, 2010

7.1. Statistical Moments

Implied mean and implied variance for the DJIA index options with a one-month maturity are presented in Figure 2 and Figure 3, respectively. All interpolation method do not clearly exhibit any differences. Then, the performances of the interpolations are tested using statistical analysis. Table 1 reports the statistical differences results of the implied mean and the implied variance according to the hypotheses mentioned in Section 4.1. Majority of the cases clearly reject the null hypotheses of equality of the implied mean and the implied variance[1]. Therefore, it can be concluded that all the interpolation approaches have statistical significances. It implies that the interpolation approach affect the RNDs estimations.

---

[1] Due to the limited space, the complete results for the statistical test are available to the reader upon request
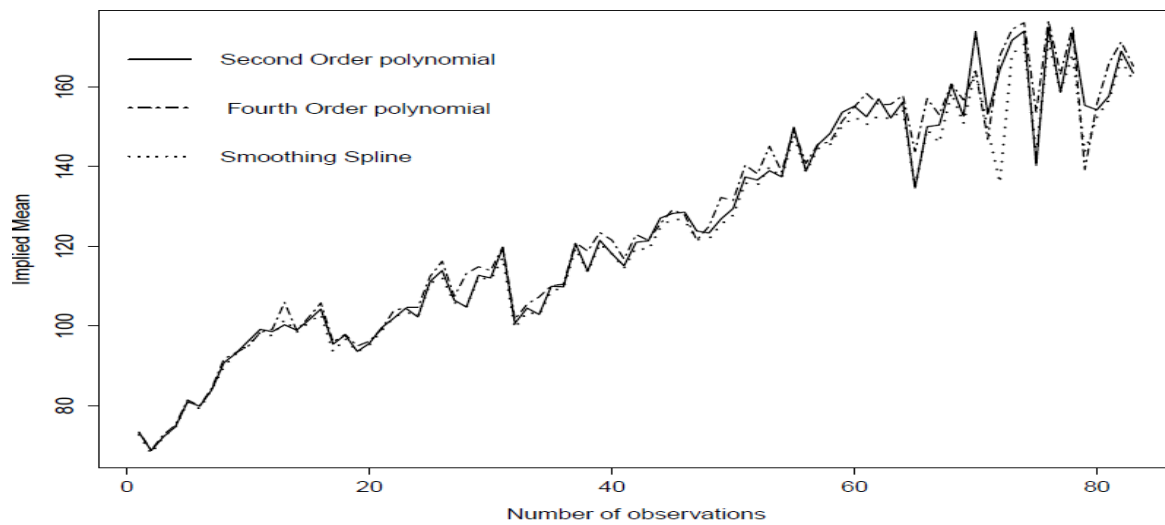
Figure 2. Implied mean
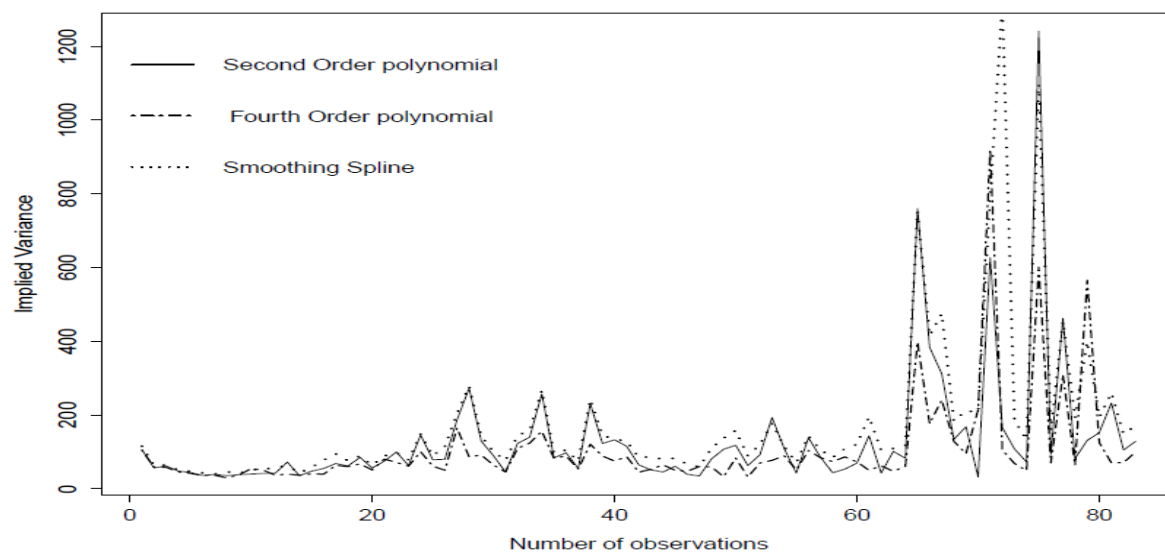
Figure 3. Implied variance

Table 1. Statistical analyses for implied mean and implied variance for DJIA index options from January 2009 until December 2015

Hypothesis	Comparison of interpolation		*t*-test for implied mean	*F*-test for implied variance
h1	second-order polynomial	fourth-order polynomial	62 cases reject h1	63 cases reject h1
h2	fourth-order polynomial	Smoothing spline	73 cases reject h2	78 cases reject h2
h3	second-order polynomial	Smoothing spline	57 cases reject h3	65 cases reject h3

7.2. LOOCV pricing error

The previous results only convey that each interpolation has different effect towards the RNDs estimation without pointing out which of the interpolation approach is the most appropriate. This study further analyzes which interpolation provides the best fitting so that the call prices generated has the lowest approximation error. Thus, the LOOCV approach is used to examine the performance of the interpolations. The ability of interpolation approach in general is measured based on various error metrics namely root mean square error (RMSE), mean absolute percentage option pricing error (MAPE), and mean absolute option pricing error (MAE). The results are presented in Table 2.

Overall, the interpolation of fourth-order polynomial has the lowest value compared to the two interpolation approaches. This finding is in agreement with [8] in which a higher order polynomial is better than a lower order polynomial. Even though the smoothing spline is widely used in the previous literature [1,10,24], however it shows the highest error of option pricing based on LOOCV techniques.

Table 2. LOOCV pricing error of the option prices corresponding to the interpolation approach

	RMSE	MAPE	MAE
Second-order polynomial	1.364	11.633	1.008
Fourth-order polynomial	1.221	10.863	0.929
Smoothing spline	1.895	11.243	1.222

8. Conclusions

The interpolation approaches play a crucial role in the volatility function techniques for the RNDs estimation. This study statistically and empirically analyze the performance of the interpolation approaches namely second-order polynomial, fourth-order polynomial and smoothing spline. Results show that each interpolation has significance difference among each other. Each interpolation provides different value of RNDs estimation. Further, the leave-out-one cross-validation (LOOCV) was conducted to determine which interpolation is appropriate to be used. The finding shows that the interpolation of fourth order polynomial outperformed the second-order polynomial and smoothing spline in which it gives the lowest value of LOOCV pricing error.

References
[1]   Bliss R R and Panigirtzoglou N 2002 *J. Bank. Financ.* **26** 381–422
[2]   Breeden D T and Litzenberger R H 1978 *J. Bus.* **51** 621–51
[3]   Shimko D 1993 *Risk* **6** 33–7
[4]   Brunner B and Hafner R 2003 *J. Comput. Financ.* **7** 75–106
[5]   Malz A M 1997 *J. Deriv.* **5** 18–36
[6]   Hull J C 2009 *Options, futures and other derivatives* ed Pearson International Edition (New Jersey: Pearson Prentice Hall)
[7]   Campa J M, Chang P H K and Reider R L 1998 *J. Int. Money Financ.* **17** 117–60
[8]   Figlewski S 2010 Estimating the Implied Risk-Neutral Density for the US Market Portfolio *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle* pp 323–56
[9]   Liu X, Shackleton M B, Taylor S J and Xu X 2007 *J. Bank. Financ.* **31** 1501–20

[10]   Bliss R R and Panigirtzoglou N 2004 *J. Finance* **59** 407–46
[11]   Stone M 1974 *J. R. Stat. Soc.* **36** 111–47
[12]   Yang Y 2007 *Ann. Stat.* **35** 2450–73
[13]   Orosi G 2012 *J. Deriv. Hedge Funds* **18** 361–76
[14]   Aït-sahalia Y and Lo A W 1998 *J. Finance* **53** 499--547
[15]   Birru J and Figlewski S 2011 *J. Financ. Mark.* **15** 151–80
[16]   Christo P and Pan X N 2014 *Equity Portfolio Management Using Option Price Information*
[17]   Grith M, Hardle W K and Schienle M 2012 Nonparametric Estimation of Risk-Neutral
         Densities *Handbook of Computational Finance* (Springer Berlin Heidelberg) pp 277–305
[18]   Ivanova V and Puigvert Gutiérrez J M 2014 *J. Bank. Financ.* **48** 210–23
[19]   Kempf A, Korn O and Sassning S 2015 *Rev. Financ.* **19** 467–90
[20]   Kostakis A, Panigirtzoglou N and Skiadopoulos G 2011 *Manage. Sci.* **57** 1231–49
[21]   Puigvert-Gutiérrez J M and De Vincent- Humpreys R 2012 *Eurasian Econ. Rev.* **2** 1–31
[22]   Santos A and Guerra J 2014 *J. Futur. Mark.* **0** 1–24
[23]   Nikkinen J 2003 *Int. Rev. Financ. Anal.* **12** 99–116
[24]   Bu R and Hadri K 2007 *Econom. J.* **10** 216–44