Original Article

Empirical performance of a spline-based implied volatility surface

Received (in revised form): 25th January 2012

Greg Orosi

is currently an Assistant Professor in the Department of Mathematics and Statistics at the American University, Sharjah. His research areas include computational finance and applications, numerical methods applied to derivative pricing, empirical performance of option pricing models, and nonparametric modeling.

Correspondence: Greg Orosi, Department of Mathematics and Statistics, American University of Sharjah,

Office: Nab 254, P.O. Box 26666, Sharjah, UAE

E-mail: gorosi@ucalgary.edu

ABSTRACT Since the crash of 1987, it has been observed by option market participants that implied volatilities for out-of-the-money options are higher than predicted by the constant volatility Black-Scholes (1973) model. Option prices also exhibit dependence on time to expiry. The collection of these implied volatilities across strike and maturity is known as the implied volatility surface (IVS). We propose a nonparametric spline-based representation of the IVS and evaluate its empirical performance. Our findings indicate that the proposed model significantly outperforms the best performing implied volatility model reported in the current literature for the purpose of pricing European-style S&P500 index options. We further contribute to the empirical finance literature by choosing a proper model evaluation criterion. By measuring the leave-one-out cross-validation model pricing error of the thin-plate spline-based model, we demonstrate that this superior performance is not the result of overfitting. Although we have previously shown that spline-based models have superior empirical performance, the models considered in this study have advantages over the previously considered models.

Journal of Derivatives & Hedge Funds (2012) 18, 361–376. doi:10.1057/jdhf.2012.1; published online 23 February 2012

Keywords: nonparametric modeling; implied volatility; volatility smile; index option pricing; empirical performance

INTRODUCTION

In their classic paper, Black and Scholes (1973) presented their famous option pricing formula. The model is based on the assumption that the stock price follows a geometric Brownian motion

with a constant drift and volatility. On the basis of no-arbitrage arguments, a riskless portfolio can be formed and a partial differential equation can be derived for option prices. Because of its simplicity, the Black-Scholes model and its modified versions are the most often used option pricing models in financial practice. However, modifications to the Black–Scholes model are necessary because empirical evidence indicates that the constant volatility Black–Scholes model exhibits systematic biases across strike and maturity. This can be observed by inverting the Black–Scholes formula to calculate implied volatilities. Dependence on strike is commonly referred to as volatility skew or smile, whereas dependence on time to expiration is commonly referred to as volatility term structure.

The shortcomings of the Black-Scholes model have led to a considerable amount of research to develop models that attempt to describe the dynamics of the underlying asset in terms of alternative distributions. However, these models have a series of limitations, such as difficulty of calibration, and are not currently used by many practitioners to price European-style options. Option traders are fully aware of the limitations of the Black-Scholes model, but rather than replacing the model, they have been able to adequately modify it in order to account for certain imperfections. The most prevalent practice of option market participants is to use an implied volatility surface (IVS) to price a set of European calls and puts for a given strike and maturity. The volatility surface is indispensable for option market makers who are required to provide a price for an option at given strike and expiry. If the particular option is liquid, the market maker can use the option's quoted price. However, option markets do not have liquid quotes for many options, and hence an interpolation tool is necessary to extract a price for a particular option from other liquid options.

Figlewski (2009) points out that this is done by the following procedure. First, option prices are converted to implied volatilities to obtain the IVS. Then, an interpolant is fitted to the IVS, typically a cubic spline or a low-order polynomial. Finally, the implied volatilities are converted back to option prices. Therefore, the effectiveness of the IVS as an interpolation tool can be assessed by measuring the option pricing error of the IVS.

The determination of the suitable parameterization of the IVS has implications that reach beyond option trading. A series of recent academic studies found that the parameters of the IVS contain important information on the underlying stock prices. For example, Duan and Wei (2009) found a strong relationship between the systematic risk proportion of stock returns in the spirit of the Capital Asset Pricing Model and the parameters (level and slope) of the IVS. Figlewski (2009) extracts the risk-neutral probability distribution from option prices to obtain insights about how information and risk preferences are incorporated into prices. Moreover, Jiang and Tian (2007) point out that the CBOE's implementation to estimate VIX index has important flaws, which lead to systematic biases. These biases can be eliminated by an interpolation/extrapolation method of the implied volatility function.

Practitioners commonly model the surface by regressing a quadratic polynomial as a function of time and strike to determine the coefficients in the model:

$$\sigma(K,T) = \beta_1 + \beta_2 K + \beta_3 K^2 + \beta_4 K T + \beta_5 T \tag{1}$$

where K is the strike price and T is the time to expiry. This model , known as the Practitoner's Black Scholes (PBS), is popular because it is relatively easy to calibrate and has strong in-sample and out-of-sample performance. For example, Christoffersen *et al* (2009) show that by



fitting a nonlinear least squares loss function (the objective function to be minimized) instead of an implied volatility loss function, the PBS can outperform the popular Heston model (Heston (1993)) and even an improved two-factor model. In fact, Christoffersen *et al* claim that this model is the best performing model in the current literature.

Although the quadratic surface performs well, most parametric models are unable to capture all the characteristics of the plotted implied volatilities. However, because of the typically small number of data points, using higher order polynomials (even cubics) can lead to overfitting and poor out-of-sample performance. To avoid overfitting while capturing most of the characteristics of the IVS, it is advantageous to consider a nonparametric spline-based representation.

A typical quadratic fit and the proposed thin-plate spline fit are shown in Figure 1 and Figure 3. Although the spline-based model seems to provide a better fit to the data than the quadratic model, it is impossible to determine by

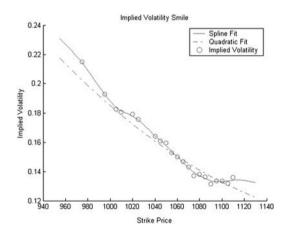


Figure 1: Implied volatilities with fitted spline and quadratic models for S&P500 index call options on 10 August 2004 with nearest expiry.

visual inspection whether the deviations from a quadratic fit are caused by noise in the data or by the actual structure of the IVS. Therefore, out-of-sample performance of these representations has to be examined to determine whether the actual structure of the implied volatility exhibits such a complex dependence as shown in Figure 1 by the spline-based fit.

Besides examining whether the performance of the quadratic practitioner's surface can be improved by a spline-based representation, studying the IVS has important implications for developing structural models. If the spline-based model depicted in Figure 1 is representative of the actual make-up of the IVS, then structural models should be able to reproduce it. However, even with jumps, one-factor stochastic volatility models are unable to reproduce the complex characteristics of the spline-based fit. For example, implied volatilities based on the Heston model parameters fitted to the data corresponding to Figure 1 are shown in Figure 2; even the two-factor models considered by Christoffersen et al (2009) are unable to reproduce all these features.

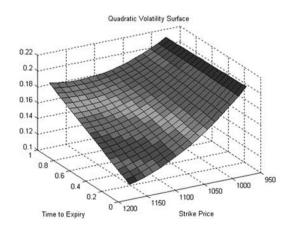


Figure 2: Quadratic surface fitted to S&P500 index call options on 10 August 2004.

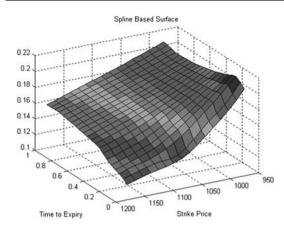


Figure 3: Proposed spline-based surface fitted to S&P500 index call options on 10 August 2004.

RELATED WORK

Although implied volatility-based representations, such as the Practitioner's Black-Scholes (PBS), work well for vanilla options, they can not be applied to the pricing of path dependent options because of the inconsistency of the constant volatility Black-Scholes model. To price exotic options consistently, a local volatility surface (introduced by Dupire (1994) and Derman and Kani (1994)) can be considered. In the field of option pricing, splines have been mostly employed to represent and regularize the local volatility surface.

Therefore, to understand the available choices to properly represent the IVS by splines, it is instructive to review the research in this area. In the local volatility model, the underlying asset is assumed to follow a 1-factor continuous diffusion model:

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma(S_t, t)dW_t.$$
 (2)

Dupire (1994) showed that if all call option prices are available, then the local volatility

function in terms of European call option prices is given by:

$$\sigma(K,T) = \left(2\frac{\frac{\partial C}{\partial T} + qC(K,T) + K(r-q)\frac{\partial C}{\partial K}}{K^2\frac{\partial^2 C}{\partial K^2}}\right)^{\frac{1}{2}}$$
(3)

where r is the interest rate and q is the dividend yield.

However, as all call option prices are not usually available, the call option prices are usually obtained from the fitted IVS. Alternatively, an approximate local volatility surface can be obtained by minimizing the mean squares values:

$$\min_{\sigma(S,t)} \sum_{i=1}^{n} \left(C_i(\sigma(S,t)) - C_i \right)^2 \tag{4}$$

where the $C_{i^{-S}}$ are the market prices for options and the $C_i(\sigma(S,t))$ -s are the call option prices based on the model. However, to properly calibrate a spline-based local volatility model, additional regularization is required. Lagnado and Osher (1997) suggest a bicubic spline representation with smoothness penalty on the local volatility and minimizing the penalized objective.

$$\min_{\sigma(S,t)} \sum_{i=1}^{n} (C_i(\sigma(S,t)) - C_i)^2 + \lambda ||\nabla \sigma(S,t)||_2$$
(5)

where λ is a constant and is known as the regularization parameter. Another alternative is to introduce some sort of prior σ_{prior} , and penalizing the objective by deviation from this prior.

$$\min_{\sigma(S,t)} \sum_{i=1}^{n} (C_i(\sigma(S,t)) - C_i)^2 + \lambda ||\sigma(S,t) - \sigma_{prior}||_2^2$$
(6)



This is known as Tikhonov regularization. This approach is suggested by Achdou *et al* (2004) for calibrating local volatility for American-style options.

Figlewski (2009) employs a fourth order spline-based interpolant of the implied volatilities to extract the risk-neutral probability distribution. In addition, he observes that cubic splines produce bad results for the fitted risk-neutral probability density function, and smoothing splines (see Splines) are preferable to exact interpolants.

Employing splines in option pricing is not new; however, to our knowledge, no empirical tests of a spline-based IVS exist in the current literature. For example, Fengler (2009) introduces an arbitrage-free smoothing of the IVS, but he does not evaluate the model's empirical performance. Furthermore, our objective is not to build a model for the dynamic behavior of the IVS as considered Cont and Fonseca (2002). The primary purpose of the spline-based implied volatility model is to extract the price of an arbitrary call option from a set of liquid option prices.

SPLINES

The term spline is used to refer to a wide class of functions that are used in applications requiring data interpolation and/or smoothing. Splines may be used for interpolation and smoothing of either one-dimensional or multi-dimensional data.

A univariate natural polynomial spline of degree 2m-1, S(t) is a real-valued piecewise function that consists of polynomial pieces, $P_i(t)$. It is defined on an interval [a, b] with the aid of k given points t_i called knots.

$$-\infty \leqslant a < t_1 < t_2 \dots < t_{k-1} < t_k < b \leqslant \infty \tag{7}$$

It is also required that S(t) satisfies the following properties:

$$S(t) \in P^{m-1}, t \in [a, t_1], t \in [t_k, b]$$
 (8)

$$S(t) \in P^{2m-1}, t \in [t_i, t_{i+1}], i = 1, \dots, n-1$$
(9)

$$S(t) \in C^{2m-2}, t \in [-\infty, \infty]$$
 (10)

where P^k is the set of polynomials of degree k or less and C^k is the class of functions with k continuous derivatives.

Natural cubic splines (that is, for m = 2) are exact interpolating functions for n data points (x_i, y_i) , $t_i = x_i$, i = 1, ..., n. The second derivatives at the end points are zero (no bending at end points). It can be shown that the natural cubic spline is the smoothest, twice-continuously differentiable function in the Sobolev space of functions that matches the observations and minimizes

$$\int_{-a}^{b} (f^{(2)}(x))^2 dx \tag{11}$$

subject to $f(x_i) = y_i, i = 1, ..., n$.

It is also possible to provide a smoothing term; in this case, the interpolation is not exact.

A 1-dimensional cubic smoothing spline is the minimizer of:

$$\sum_{i=1}^{n} (\gamma_i - f(x_i))^2 + \lambda \int_{a}^{b} (f^{(2)}(x))^2 dx$$
 (12)

The smoothing parameter, λ , varies from zero to infinity. When $\lambda = 0$, the spline estimate interpolates the data and has a residual sum of

squares of zero; at $\lambda=\infty$, the spline estimate becomes the least squares straight line and as such it generally does not represent the data very well. Therefore, λ represents the rate of exchange between residual error and roughness of the curve. Thin-plate smoothing splines are the generalization of the one-dimensional cubic smoothing splines for arbitrarily spaced data (x_i, y_i, z_i) . The problem may be formulated as the minimization of the penalized sum of squares.

$$J[f] = \sum_{i=1}^{n} (z_i - f(x_i, y_i))^2 + \lambda \iint_{\mathbb{R}^2} f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 dx dy.$$
 (13)

For this variational problem, there exists a unique minimizer in an appropriately defined Sobolev space representing the set of all 'reasonable' candidate functions, and it can be shown that it has the form (Wahba, 1990)

$$f(x, y) = d_0 + d_1 u + d_2 v + \sum_{i=1}^{n} c_i E(u - x_i, v - y_i)$$
 (14)

$$E(u, v) = r^2 \log r \tag{15}$$

where $r = \sqrt{u^2 + v^2}$. This can also be written in matrix form, where $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^3$ are the minimizers of:

$$\frac{1}{n}||z - Td - \Phi \varepsilon||^2 + \lambda \varepsilon' \Phi \varepsilon \tag{16}$$

subject to

$$T'c = 0 \tag{17}$$

where *T* is the *n x* 3 matrix with rows $[1x_i y_i]_{i=1}^n$ and Φ is the *n x n* matrix with (*i*, *j*)-th

entry $E(x_i-x_j, y_i-y_j)$. As shown in Wahba (1990), the above can be written as equivalent to the linear system

$$Mc + Td = z \tag{18}$$

$$T'c = 0 \tag{19}$$

where

$$M = \Phi + n\lambda I. \tag{20}$$

Thin-plate splines with automatic optimal smoothing parameter selection via generalized cross-validation (GCV) have a number of desirable properties (Wahba, 1990). GCV provides an efficient and objective method to determine the correct degree of smoothing for optimally separating a smooth function from white noise. In addition, compared with other spline models such as bicubic splines, thin-plate splines have knots that are determined naturally by the data, and can model the data in a more natural, efficient manner.

Determining the smoothing constant

A suitable smoothing parameter can be efficiently calculated by minimizing the GCV statistic:

$$V(\lambda) = \frac{1}{n} \sum_{k=1}^{n} (\gamma_k - f_{\lambda}^k(x_k))^2 w_{kk}$$
$$= \frac{\frac{1}{n} ||(I - A(\lambda))\gamma||^2}{\left[\frac{1}{n} Tr(I - A(\lambda))\right]^2}$$
(21)

where f_{λ}^{k} is the spline fit with the k-th point removed and

$$w_{kk}(\lambda) = \frac{(1 - a_{kk}(\lambda))^2}{(1 - \mu(\lambda))^2}$$
 (22)



$$\mu(\lambda) = \frac{1}{n} \sum_{i=1}^{n} a_{ii}(\lambda)$$
 (23)

with $a_{ii}(\lambda) = 1/nTrA(\lambda)$, and $A(\lambda)$ (known as the influence matrix) is an $n \times n$ matrix defined such that it satisfies:

$$\begin{pmatrix} f_{\lambda}(x_1) \\ \vdots \\ f_{\lambda}(x_n) \end{pmatrix} = A(\lambda)\gamma. \tag{24}$$

The influence matrix for GCV

The influence matrix allows us to calculate the GCV statistic efficiently. Note that in the case of linear Tikhonov regularization problem, the influence matrix can be determined from the solution of the problem

$$x_{Tikhonov} = \min_{x} ||Bx - y||^2 + \lambda ||x||^2$$
$$= (B'B + \lambda I)^{-1} B'y$$
 (25)

and hence

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} = Bx_{Tikhonow} = B(B'B + \lambda I)^{-1}B'\gamma.$$

Therefore, the influence matrix $A(\lambda)$ is given by:

$$A(\lambda) = B(B'B + \lambda I)^{-1}B' \tag{27}$$

and the GCV statistic for linear Tikhonov regularization problem can be expressed as:

$$V(\lambda) = \frac{1}{n} \frac{||(I - A(\lambda))\gamma||^2}{\left[\frac{1}{n} Tr(I - A(\lambda))\right]^2}$$
$$= \frac{\frac{1}{n} ||(I - B(B'B + \lambda I)^{-1} B')\gamma||^2}{\left[\frac{1}{n} Tr(I - B(B'B + \lambda I)^{-1} B')\right]^2}. \quad (28)$$

By minimizing $V(\lambda)$, the optimal penalty constant λ can be determined.

GCV for smoothing splines

For smoothing splines, calculating the smoothing constant via GCV has been found to be reliable in numerous applications (Wahba, 1990). In addition, the GCV statistic can be efficiently determined from:

$$V(\lambda) = \frac{1}{n} \frac{||(I - A(\lambda))\gamma||^2}{\left[\frac{1}{n} Tr(I - A(\lambda))\right]^2}$$
(29)

and it was shown by Wahba (1990) that for smoothing splines

$$(I - A(\lambda)) = n\lambda Q_2 (Q_2' M Q_2)^{-1} Q_2'.$$
 (30)

Q₂ is obtained from the QR decomposition of T

$$T = QR = (Q_1 : Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$
 (31)

such that Q_1 is the first 3 rows of Q and Q_2 is given by the remaining n-3 rows. Therefore, Q_1 is $n \times 3$, Q_2 is $n \times (n-3)$, Q_1 and Q_2 are orthonormal, R is an $n \times 3$ upper triangular matrix, and R_1 is a 3×3 upper triangular matrix.

MODEL EVALUATION

Model selection is a topic of special relevance to our study. Option pricing models calibrated using a penalty function rely on a regularization parameter. To determine a suitable regularization parameter, a model selection criterion has to be applied. A model evaluation technique must also be used to evaluate the empirical performance of option pricing models. Our perspective coincides with a market maker's perspective, which is to determine the best interpolation method to price an illiquid option using all available option prices on a given day.

Cross-validation

One approach of model evaluation is to use the entire data set for model fitting, and then to use the model that provides the lowest error on the data set. The problem with this approach is that the final model might overfit the data. Therefore, it does not give an indication of how well the model will do when it is asked to make new predictions for data it has not seen before.

One way to overcome this problem is to remove some of the data before model calibration. Once the calibration is done, the data that was removed can be used to test the performance of the calibrated model. This is the basic idea for a whole class of model evaluation methods called cross-validation (Stone (1974)), which is one of the most commonly used model selection criteria.

The simplest kind of cross-validation is the holdout method. In this method, data set is separated into two disjoint subsets. One set is used for fitting each competing model and the other set is used to evaluate the model's performance, then the model with the best overall performance is selected. The advantage of this method is that it is usually preferable to the residual method and the computation does not take longer. However, the evaluation may depend heavily on which data points end up in the training set and which end up in the test set, thus the evaluation may be significantly different depending on how the division is made.

K-fold cross-validation is one way to improve on the holdout method. Under *K*-fold

cross-validation, the available data are first divided into k disjoint sets. Then K models are fitted, each on a different combination of K-1 partitions, and each of these models are tested on the remaining partition. The advantage of this method is that it matters less how the data gets divided. Every data point get to be in a test set exactly once, and get to be in a training set k-1 times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times so that it takes k times as much computation to make an evaluation. A variant of this method is to randomly divide the data into a test and training set k different times. The advantage of doing this is that you can independently choose how large each test set is, and how many trials you average over.

The most extreme form of cross-validation, where k is equal to the number of data points, is known as leave-one-out cross-validation (LOOCV). The LOOCV statistic is highly attractive for the purpose of model selection because it gives an almost unbiased estimator for the generalization ability of the model (Yang (2007)).

Evaluating the performance of option pricing models

Measuring in-sample and out-of-sample errors after daily calibration has been a common approach to evaluate the performance of option pricing models. For example, Bakshi *et al* (1997); Dumas *et al* (1997); Christoffersen *et al* (2009), and Carr and Wu (2004) base some of their conclusions on this method. Note that this test is a version of the holdout method. Although the test is informative, Bates (2003) points out that it



has several drawbacks. For example, the best predictor of future IVS is today's IVS. Therefore, any model with multiple free parameters and a good in-sample fit will perform roughly as well out-of-sample. This drawback is especially relevant when one attempts to evaluate the performance of nonparametric models that near-perfectly reproduce the IVS. In order to conclude that a nonparametric model is indeed preferable to other models, additional tests have to be carried out.

Bates (2003) suggests that measuring in-sample and out-of-sample hedging errors are probably more informative than measuring in-sample and out-of-sample pricing errors. Although this test would reveal more easily whether a model overfits, the results would not be conclusive when the primary objective is to minimize pricing errors. Moreover, hedging errors are especially unreliable when one attempts to evaluate the empirical performance of *ad hoc* models that are known to produce incorrect hedge ratios.

Several authors also measure Akaike Information Criterion (Dumas *et al* (1997)) or Schwartz Information Criterion (Carr and Wu (2003)). These tests are supposed to prevent overfitting by penalizing the number of free parameters used. However, there is no theoretical basis for their use when evaluating the performance of nonlinear option pricing models.

A statistical test that is much more informative than the previously described tests is the LOOCV statistic. This test is not only assumption free and statistically meaningful, it also coincides with a market maker's objective. To our knowledge, we are the first ones to evaluate the performance of option pricing models based on the LOOCV statistic. Unfortunately, the statistic is very expensive to

compute for models that are calibrated by minimizing mean square option pricing errors such as the Heston model or PBS model calibrated by the NLS objective. However, it is computationally feasible to measure the LOOCV statistic for certain spline-based implied volatility models.

METHODOLOGY

Data

The empirical tests are based on S&P500 index call options obtained from the OptionMetrics¹ database. S&P500 index options are well suited as test cases because extensive empirical studies have been carried out on them and are the most liquid European options available on the Chicago Board of Exchange. The period covered ranges from 3 January 2000 to 30 December 2005.

To filter the data for possible biases, the following criteria were applied. First, following Bakshi *et al* (1997), we exclude all options that violate at least one of a number of basic, noarbitrage conditions and contracts that cost less than US\$3/8. Furthermore, we consider only data for contracts with more than 6 trading days to maturity and less than a year to maturity. Finally, following Dumas *et al* (1997), we exclude options with absolute moneyness in excess of 10 per cent, where absolute moneyness is given by the following: M = K/S - 1, with S the stock price and K the strike price. Table 1 summarizes the remaining data set consisting of 24 039 contracts.

Empirical method

To evaluate the performance of the spline-based local volatility models, empirical tests similar to

Table 1: S&P500 Index call option data by absolute moneyness (M) and days to maturity (D)

	D<30	30 <d<90< th=""><th>D > 90</th><th>D>180</th><th>All</th></d<90<>	D > 90	D>180	All
-0.05 > M > -0.1	435	2053	862	1473	4823
-0.025 > M > -0.05	732	1932	532	862	4058
0 > M > -0.025	907	2135	500	797	4339
0.025 > M > 0	789	1819	425	695	3728
0.05 > M > 0.025	529	1855	701	1200	2806
0.1 > M > 0.05	507	1231	400	668	4285
All	3899	11025	3420	5695	24039

Christoffersen *et al* (2009) were performed. For each of the 309 available Wednesdays in the 2000–2005 sample, the previously described spline-based implied volatility model and three additional benchmark models were calibrated by minimizing the non-linear least-squares objective:

$$\sum_{i=1}^{n} (C_i(\theta) - C_i)^2$$
 (32)

where the C_{i} -s are the market prices for options, the $C_{i}(\theta)$ -s are the call option prices based on the model and θ is the set of parameters of the model. After calibration, for each of the 309 available Wednesdays in the 2000–2005 sample, the in-sample option pricing errors are measured using root mean square pricing errors (RMSE).

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(C_{ij}(\theta_j) - C_{ij} \right)^2}$$
 (33)

where θ_j represents the parameters for the *j*th given trading day, the $\{C_{ij}\}_{i=1}^n$ -s are the market prices for options for the *j*th trading day (for all strikes and expiries) and the $C_{ij}(\theta_j)$ -s are the option prices based on the model. The average

in-sample error (we will refer to this as the in-sample error) of an option pricing model is obtained by measuring the average in-sample option pricing error over all 309 trading days in the sample.

The 1-day and 5-day out-of-sample errors are computed using option prices on the first day and the fifth day following the calibration period, respectively. Thus, the 1-day out-of-sample pricing error for the *j*th trading day is given by:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_{ij}(\theta_{j-1})-C_{ij})^2},$$

and the 5-day out-of-sample pricing error is given by:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(C_{ij}(\theta_{j-5})-C_{ij}\right)^{2}}.$$

We will refer to the average of these (over all trading days) as the out-of-sample pricing errors. Note that the in-sample pricing error indicates how well a model captures the characteristics of the observed option prices on average.



The first benchmark model is the NLS PBS model and the second is the stochastic volatility model of Heston (1993). On the basis of the 1-day and 5-day out-of-sample error criteria, Christoffersen *et al* (2009) consider the performance of the NLS PBS model to be the best in the current literature. However, it is possible that a model has a relatively poor 1-day and 5-day out-of-sample performance, despite being an effective interpolant. Therefore, we introduce a more complex implied volatility-based model by adding three more parameters to the NLS PBS model:

$$\sigma(K, T) = \beta_1 + \beta_2 K + \beta_3 K^2 + \beta_4 K T + \beta_5 T + \beta_6 K^3 + \beta_7 K^2 T + \beta_8 T^2.$$
 (34)

We will refer to this benchmark model as NLS PBS 2. We have also experimented with other representations of the IVS by including more parameters, but these models resulted in inferior performance.

As we have noted earlier, nonparametric models can exhibit a high degree of flexibility that may ultimately result in overfitting.

Although such models can achieve zero error on the training data, they might not model the underlying function well, thus performing poorly when presented with a new data set.

Therefore, it is necessary to measure the LOOCV statistic of the spline-based model to make sure that it does not overfit. The LOOCV statistics of the spline-based model were also determined for each of the 309 trading days in the sample.

Determining the LOOCV statistic is extremely computationally intensive. To save significant computational time, we have used a smoothing constant with a preset value instead of minimizing the GCV statistic. Setting the smoothing constant to zero is reasonable because the results presented in Table 3 (see Optimal Smoothing) suggest that the ideal smoothing constant is very close to zero. Therefore, all LOOCV statistics presented in our study are based on a thin-plate spline model that is an exact interpolant.

Implementation

The empirical tests were carried out by using the Matlab programming language. To calculate the parameters for the benchmark model, the builtin *fminsearch* function was used. The thin-plate spline model was implemented directly in Matlab instead of using the built-in *tpaps* function in the Matlab spline toolbox. This function calculates the optimal smoothing constant based on an ad hoc method that is inferior to determining the smoothing constant by GCV. To verify that our implementation of the thin-plate spline is correct, we compared its output with the built-in tpaps function. Finally, the smoothing constant was determined by minimizing the GCV function using golden section search. For the golden section search, the lower bound was set to 10^{-9} , the upper bound was set to 10^{-1} and 30 iterations were used. In addition, in the case of splines, a linear transformation of K' = K/1000, is used. This transformation did not improve the benchmark models.

RESULTS AND DISCUSSION

Performance compared with benchmark models

The in-sample and out-of-sample RMSE are summarized in Table 2. The relative in-sample

and out-of-sample results for the NLS PBS and Heston benchmark models are comparable with Christoffersen and Jacobs (2004), even though the data sets are different. This confirms that our data set are appropriate for the study. Note that the larger change in absolute value in our data set can be explained by the significantly larger index value.

It can be observed that the thin-plate splinebased volatility surface model leads to an improvement over all benchmark models both in-sample and out-of-sample. The 5-day out-ofsample performance of the thin-plate splinebased volatility surface model is comparable with the NLS PBS model; however, by examining the LOOCV errors of the spline-based model, we demonstrate that this is not the result of overfitting. Although the NLS PBS 2 model significantly outperforms the NLS PBS model in-sample, it does not outperform the NLS PBS model out-of-sample. Therefore, increasing the number of parameters to model the IVS does not necessarily lead to smaller out-of-sample pricing errors.

Optimal smoothing

For each of the 309 trading days, the value of the smoothing constant λ was recorded and simple statistics were calculated to study its distribution. These are shown in Table 3. It can be observed that, when using GCV, very little smoothing is applied.

To better study the distribution of the smoothing constant, (17) is expressed in terms of two weights summing to one.

$$w_1||\gamma - Td - Kc||^2 + w_2c'Kc$$
 (35)

so that $w_1 = 1/1 + n\lambda$ and $w_2 = n\lambda/1 + n\lambda$. Statistics for w_1 are shown in Table 3. Again, it is evident that, based on GCV, very little smoothing is required.

By representing the IVS as a thin-plate spline model, the regularization constant is determined from the implied volatilities. The advantage of this approach is that λ can be determined efficiently by minimizing (24). However, GCV or cross-validation should be applied to the nonlinear option-pricing problem shown in (6).

Table 2: In- and out-of-sample RMSE results

	NLS PBS	NLS PBS 2	Heston	Spline
In-sample	0.7603	0.5739	0.961	0.085
1-day out-of-sample	1.736	1.8252	2.158	1.580
5-day out-of-sample	2.0571	2.1336	2.716	1.968

Table 3: Distribution of the smoothing constant

w_1	Mean	Maximum	Minimum
Smoothing constant	5.46*10 ⁻⁶	9.92*10 ⁻⁵	1.00*10 ⁻⁹
Smoothing constant (weight based)	0.9994	1.0000	0.9912



Table 4: In- and out-of-sample RMSE for various smoothing constants

Smoothing constant	0	10^{-6}	10^{-5}	10 ⁻⁴	10^{-3}	10^{-2}
In-sample	0	0.039	0.079	0.142	0.346	0.965
1-day out-of-sample	1.584	1.582	1.579	1.578	1.624	1.946
5-day out-of-sample	1.975	1.972	1.967	1.964	1.997	2.293

To confirm that GCV is appropriate, the performance of the thin-plate spline model is evaluated for several constant λ -s and the results are shown in Table 4. From this, it can be concluded that GCV is a reliable method to determine the value of the smoothing constant. Moreover, if reducing computational time is crucial, then it is suggested that instead of performing GCV a small constant value of λ should be used.

Therefore, increasing the smoothing constant does not provide additional regularization in the case of S&P500 index options. This indicates that the observed structure of the IVS is representative of the actual structure.

LOOCV option pricing error

In order to properly assess the generalization ability of the thin-plate spline model, we have calculated various error metrics. We have measured the LOOCV RMSE, mean absolute percentage option pricing error (MAPE), and the mean absolute option pricing error (LOOCV MAE). The results are presented in Table 5.

Table 5 also shows the in-sample RMSE, MAPE and MAE of the PBS and Heston benchmark models. Ideally, we would like to measure the LOOCV statistics of the three models instead of comparing the LOOCV

statistics of the spline-based model with the in-sample values of the benchmark models. However, as we have noted earlier, it is extremely computationally expensive to measure the LOOCV statistics for the benchmark models. Therefore, we employ the in-sample values of the benchmark models as a proxy for their LOOCV values. Note that this is reasonable because the LOOCV values can be expected to be higher than the in-sample values.

It can be observed that LOOCV statistics of the thin-plate spline model is significantly lower than the corresponding in-sample errors of the benchmark models. The LOOCV RMSE is 60.9 per cent lower than the RMSE of the best performing benchmark model (NLS PBS 2).

Particularly informative are the in-sample MAPE. These are 5.30 per cent for the Heston model, 4.33 per cent for the NLS PBS model and 3.16 per cent for the NLS PBS 2 model. These clearly show that the benchmark models are not sufficient to capture all the characteristics of the IVS, making these models undesirable for certain applications. The corresponding LOOCV MAPE for the spline-based model is 0.98 per cent. These results suggest that a nonparametric representation is necessary and the superior performance of the model is not a result of overfitting.

Note that the benchmark models were calibrated by minimizing the RMSE option

Table 5: LOOCV statistics of the thin-plate spline-based model and the corresponding insample statistics of the benchmark models

	Spline (LOOCV)	Heston (In-sample)	NLS PBS (In-sample)	NLS PBS 2 (In-sample)
RMSE	0.224	0.961	0.7603	0.5739
MAPE	0.0098	0.0530	0.0433	0.0316
MAE	0.147	0.734	0.635	0.4907

pricing loss function and the thin-plate spline was calibrated by minimizing the MSE implied volatility loss function. Although Christoffersen and Jacobs (2004) point out the critical importance of matching up in-sample and out-of-sample loss functions, in this case it is reasonable because all three LOOCV values of the spline-based model are significantly lower than the in-sample values of the benchmark models.

Although the LOOCV pricing error of the spline-based implied volatility model is significantly lower than the in-sample pricing error of the PBS model, in the presence of a large bid-ask spread a low pricing error would not necessarily lead to a superior performance. To determine the effect of the bid-ask spread, for each of the 309 trading days in the sample, we have measured how often the LOOCV option pricing error of the thin-plate spline-based model is less than half the bid-ask spread. The average values over all trading days are reported in Table 6. Note that this value equals the proportion of option prices (based on the thinplate spline-based model and averaged over all trading days) that lies inside the bid-ask spread. Similarly, the proportion of option prices (averaged over all trading days) that lie inside the bid-ask spread based on the PBS model are

Table 6: Statistics showing how often the option prices based on the models lie inside the bid-ask spread

	Mean per cent	Standard deviation per cent
NLS PBS	69.95	17.20
NLS PBS 2	91.73	26.29
Spline	96.44	5.12

reported in Table 6. It can be observed that the spline-based model outperforms all the benchmark models based on this criteria.

Performance compared with other spline-based models

Orosi (2010) has previously shown that local volatility spline-based models have superior empirical performance compared with the benchmark models considered in this study. In our experience, implied volatility and local volatility thin-plate spline-based models have similar performance for the purpose of interpolation if the number of parameters of both models is comparable. Moreover, if retrieving the local volatility surface is not necessary, the implied volatility model has several advantages over the local volatility model.



First, the implied volatility-based model has substantially smaller 1-day and 5-day out-of-sample errors. Goncalves and Guidolin (2006) point out that minimizing the 1-day and 5-day out-of-sample errors is important for constructing dynamic models of IVS. Moreover, the calibration time of the spline-based implied volatility model is significantly less because minimizing a non-linear least-squares objective is not necessary. Therefore, the LOOCV pricing error of the model can be easily measured.

Furthermore, we would like to point out that not all exact interpolants of option prices have the same performance. If the scaling K' = K/1000 is not included in the spline-based implied volatility model, the model based on LOOCV option pricing error does not outperform neither of the NLS PBS and NLS PBS 2 models (results not included). Therefore, models with small in-sample option pricing errors are not always effective interpolants.

Finally, it is well known that thin-plate splines do not extrapolate well. To overcome this drawback, the model presented in this article can be combined with an arbitrage-free interpolant (such as the one considered by Fengler (2009)) when option prices outside the interpolating region have to be determined.

CONCLUSIONS

This article considers a nonparametric spline representation of the IVS and compares its performance with the best performing implied volatility-based model in the current literature. Besides outperforming the benchmark model, the proposed spline model has additional advantages. It requires significantly less computational time and regularization can be easily implemented.

Although these results are primarily of practical importance, some of our findings have theoretical implications. We also find that the optimal smoothing parameter of the volatility surface should be fairly close to zero. These results indicate that option pricing models should replicate the observed IVS fairly closely.

In addition, our findings should motivate further research in the use of nonparametric methods for option pricing. In particular, other nonparametric representation of the IVS, such as other types of radial basis functions, could be explored.

NOTE

1. http://www.optionmetrics.com

REFERENCES

- Achdou, Y., Indragoby, G. and Pironneau, O. (2004) Volatility calibration with American options. *Methods* and Applications of Analysis 11(3): 001–024.
- Bakshi, G., Cao, C. and Chen, Z. (1997) Empirical performance of alternative option pricing models. *Journal of Finance* 52(5): 2003–2050.
- Bates, D.S. (2003) Empirical option pricing: A retrospection. *Journal of Econometrics* 116(1–2): 387–404.
- Black, F. and Scholes, M.S. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81(3): 637–659.
- Carr, P. and Wu, L. (2003) Finite moment log stable process and option pricing. *Journal of Finance* 58(2): 753–777.
- Carr, P. and Wu, L. (2004) Time-changed levy processes and option pricing. *Journal of Financial Economics* 17(1): 113–141.
- Christoffersen, P., Heston, S. and Jacobs, K. (2009) The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science* 55(12): 1914–1932.
- Christoffersen, P. and Jacobs, K. (2004) The importance of the loss function in option valuation. *Journal of Financial Economics* 72(2): 291–318.
- Cont, R. and Fonseca, J. (2002) Dynamics of implied volatility surfaces. *Quantitative Finance* 2(2): 45–60.



- Crépey, S. (2003) Calibration of the local volatility in a trinomial tree using tikhonov regularization. *Inverse Problems* 19(1): 91–127.
- Derman, E. and Kani, I. (1994) Riding on a smile. *Risk* 7(2): 32–39.
- Duan, J. and Wei, J. (2009) Systematic risk and the price structure of individual equity options. *Review of Financial* Studies 22(5): 1981–2006.
- Dumas, B., Fleming, J. and Whaley, R.E. (1997) Implied volatility functions: Empirical tests. *Journal of Finance* 53(6): 2059–2106.
- Dupire, B. (1994) Pricing with a smile. Risk 7(1): 18–20.
 Egger, H. and Engl, H. (2005) Tikhonov regularization applied to the inverse problem of option pricing:
 Convergence analysis and rates. Inverse Problems 21(3): 1027–1045.
- Fengler, M.R. (2009) Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance* 9(4): 417–428.
- Figlewski, S. (2009) Estimating the Implied Risk-Neutral Density for the US Market Portfolio, Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle, Oxford University Press.
- Goncalves, S. and Guidolin, M. (2006) Predictable dynamics in the S&P 500 index options volatility vurface. *Journal of Business* 79(3): 1591–1635.

- Heston, S.L. (1993) A closed-form solution for options with stochastic volatility applications to bond and currency options. *Review of Financial Studies* 6(2): 327–343.
- Jiang, G. and Tian, Y. (2007) Extracting model-free volatility from option prices: An examination of the VIX index. *Journal of Derivatives* 14(3): 35–60.
- Lagnado, R. and Osher, S. (1997) A technique for calibrating derivative security pricing models:
 Numerical solution of an inverse problem. *Journal of Computational Finance* 1(1): 13–25.
- Merton, R. (1976) Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3(1–2): 125–144.
- Orosi, G. (2010) Improved implementation of local volatility and its application to S&P 500 index options. *Journal of Derivatives* 17(3): 53–64.
- Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. B 36(1): 111–147.
- Wahba, G. (1990) Splines Models for Observational Data, Series in Applied Mathematics. CBMS-NSF Regional Conference Series in Applied Mathematics, p. 56.
- Yang, Y. (2007) Consistency of cross validation for comparing regression procedures. *Annals of Statistics* 35(6): 1564–1599.

After initial online publication this article was temporarily withdrawn until a permission issue was resolved. This reinstated and final version is completely unchanged from the originally published version.