Expert Systems with Applications 38 (2011) 1702–1708

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Predicting a distribution of implied volatilities for option pricing

Seung-Ho Yang, Jaewook Lee *

Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Kyungbuk 790-784, Republic of Korea

ARTICLE INFO

Keywords:
Option pricing
Implied volatility
Bayesian approaches
Kernel methods
Gaussian processes
Black–Scholes model

ABSTRACT

In this paper, we propose a method that predicts a distribution of the implied volatility functions and that provides confidence intervals for the option prices from it. The proposed method, based on a Bayesian approach, employs a Bayesian kernel machine, so-called Gaussian process regression. To verify the performance of the proposed method, we conducted simulations on some model-generated option prices data and real option market data. The simulation results show that the proposed method performs well with practically meaningful option ranges as well as overcomes the problem of containing negative prices in their predicted confidence intervals by the previous works.

1. Introduction

Since the appearance of the Black and Scholes model in 1973, many option pricing formulas have been developed to overcome the restrictive assumptions of Black and Scholes models and to give more accurate prices (Lajbcygier, 1999). Most of the methods are focused on a point prediction of option price. Since there is always discrepancies between the option prices predicted by the models and the real option market prices due to some market frictions such as bid-ask spreads, noisy information, and etc., it is desirable to have a predictive distribution of option prices. To give a distribution of option prices, the use of a neural network kernel-based Bayesian method is addressed in Jung, Kim, and Lee (2006). Han and Lee (2008) suggested the uses of mixed kernels to give more accurate ranges of option prices than other neural networks models in Choi, Lee, Han, and Lee (2004), Gencay and Qi (2001), Hutchinson, Lo, and Poggio (1994) and applied it to pricing one-sided equity-linked warrants (ELWs, only the buy-side is allowed but the sell-side is limited for the investors). However, these previous approaches have serious problems in their predicted confidence intervals (CIs) of option prices. First, some option prices in their ranges can take negative values in the out of the money (OTM) regions. Second, when the moneyness goes to OTM or in-the-money (ITM) regions, the predicted CIs becomes too broad. Finally, the predicted option ranges around at-the-money (ATM) regions becomes too narrow to encompass the high volume of traded options.

To overcome such problems, in this paper, we propose a method that estimates option prices with more practically useful range of options. The proposed method consists of two phases: first phase for constructing a predictive distribution of implied volatility functions and second phase for estimating a confidence intervals of option prices using the predicted volatility distributions. The proposed method will, through simulation results, be shown to contain only positive option prices in their predicted confidence intervals as well as provides more tight option ranges for ITM and OTM regions and more broad option ranges near ATM.

The paper is structured as follows. In Section 2, we give some preliminary notions and terminologies that are needed for the subsequent sections. In Section 3, we give a short review of Bayesian linear basis function models and present a method to constructing a distribution of implied volatility functions using a recently developed Gaussian processes regression models. In Section 4, we then provide a way to compute a confidence interval (CI) for option prices using the predicted distribution of implied volatilities. We show some simulation results in Section 5 and conclude the paper in Section 6.

2. Preliminaries

2.1. Black–Scholes formula

Black–Scholes formula is one of the most well-known and widely used method to compute European option prices. Black–Scholes formula is derived under some restrictive assumptions, say, stock price follows a geometric Brownian motion with constant drift, volatility, and interest rate is fixed. It also assumes no transaction costs. In the case of paying continuous yield dividends the Black–Scholes formula for the European call and put option prices, denoted by C_{BS} and P_{BS}, are given by

$$C_{BS}(S_0, K, T, \Sigma, r_f, d) = \exp(-r_f T)(F_T \mathcal{N}(d_1) - K\mathcal{N}(d_2)), \quad (1)$$

$$P_{BS}(S_0, K, T, \Sigma, r_f, d) = \exp(-r_f T)(K\mathcal{N}(-d_2) - F_T \mathcal{N}(-d_1)), \quad (2)$$

$$d_1 = \frac{\ln(F_T/K) + (\Sigma^2/2)T}{\Sigma\sqrt{T}}, \quad d_2 = d_1 - \Sigma\sqrt{T},$$

* Corresponding author.
E-mail address: jaewookl@postech.ac.kr (J. Lee).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.095

where S_0 is the stock price, K is the strike price, T is the expiry date, Σ is the volatility of the stock, r_f is the risk-free interest rate, d is the dividend rate and $F_T = S_0 e^{(r_f - d)T}$ is the forward price. Here $\mathcal{N}$ stands for the probability of a cumulative standard normal distribution.

2.2. Implied volatility

In the Black–Scholes formula, every parameters except the volatility of the stock are given in the market. Although the volatility parameter can be estimated from the time series of stock prices data, it gives usually mismatched Black–Scholes option prices with the market option prices since the hypotheses of the Black–Sholes model do not hold in real markets. Instead, option traders use so-called implied volatility to match the market prices of options with Black–Sholes formula. Specifically, they represent the market prices of options in terms of their Black–Sholes (BS) implied volatilities:

$$C^{mkt} = C_{BS}(S_0, K, T, \Sigma_{im}(K/S_0, T), r_f, d), \tag{3}$$

$$P^{mkt} = P_{BS}(S_0, K, T, \Sigma_{im}(K/S_0, T), r_f, d), \tag{4}$$

where C^{mkt} and P^{mkt} denote the observed market European call and put option prices and $\Sigma_{im}(K/S_0,T)$ stands for the BS implied volatility depending on $(K/S_0, T)$. Here K/S_0 is called a moneyness.

3. Constructing a distribution of implied volatility functions

The main goal of this section is to construct a probability distribution of implied volatility functions under the Gaussian process framework. To accomplish this task, we consider the following problem of constructing a implied volatility function that match the observed market BS implied volatilities as closely as possible:

$$y = f(\mathbf{x}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \tag{5}$$

where the output variable y is the observed BS implied volatility Σ_{im}, f is the implied volatility function to be estimated, and $\mathbf{x}$ is an input vector, for example, representing two-dimensional variables $(K/S_0, T)$ or one-dimensional variable moneyness for a given maturity. The noise term ε is due to the presence of observation errors in the market data such as bid-ask spreads or other market frictions and is assumed to follows a normal distribution.

3.1. Review of Bayesian linear basis function models

We first review Bayesian linear basis function methods for regression. More detailed expositions can be found in Bishop (2006) or references therein. A general linear model for the function f in (5) is one that involves the following linear combination of fixed nonlinear basis functions of the input variables

$$f(\mathbf{x}; \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}), \tag{6}$$

where $\mathbf{w} = (w_0, \ldots, w_M)^T$ is the parameters and $\phi = (\phi_0, \ldots, \phi_{M-1})^T$ is the basis functions. We set $\phi_0(\mathbf{x}) = 1$, so that the corresponding parameter w_0 plays the role of a bias.

The noise assumption together with the model directly gives rise to the *likelihood*, the probability density of the observations $\mathbf{y} = (y_1, \ldots, y_N)^T$ given the parameters, which is factored over cases in the training set (because of the independence assumption) to give (conditioned on the input values $\mathbf{x}_1, \ldots, \mathbf{x}_N$)

$$p(\mathbf{y}|\mathbf{w}) = \mathcal{N}(\mathbf{\Phi w}, \sigma^2 \mathbf{I}),$$

where $\mathbf{\Phi}$ is the design matrix with elements $\mathbf{\Phi}_{nk} = \phi_k(\mathbf{x}_n)$. Now we assume that a prior distribution over $\mathbf{w}$ is given by an isotropic Gaussian of the form

$$p(\mathbf{w}) = \mathcal{N}(0, \mathbf{\Sigma}_0).$$

Then the posterior distribution, conditioned on the input values $\mathbf{x}_1, \ldots, \mathbf{x}_N$, becomes

$$p(\mathbf{w}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_N, \mathbf{\Sigma}_N),$$

where

$$\boldsymbol{\mu}_N = \frac{1}{\sigma^2} \mathbf{\Sigma}_N \mathbf{\Phi}^T \mathbf{y}, \quad \mathbf{\Sigma}_N^{-1} = \mathbf{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{\Phi}^T \mathbf{\Phi}.$$

The *predictive distribution* for $f^* \triangleq f(\mathbf{x}_*)$, conditioned on the input values $\mathbf{x}_1, \ldots, \mathbf{x}_N$, is given by averaging the output of all possible linear models with respect to the Gaussian posterior

$$p(f_*|\mathbf{x}_*, \mathbf{y}) = \int p(f_*|\mathbf{x}_*, \mathbf{w}) p(\mathbf{w}|\mathbf{y}) d\mathbf{w} = p(\phi(\mathbf{x}_*)^T \mathbf{w}|\mathbf{x}_*, \mathbf{y})$$
$$= \mathcal{N}(\phi(\mathbf{x}_*)^T \boldsymbol{\mu}_N, \phi(\mathbf{x}_*)^T \mathbf{\Sigma}_N \phi(\mathbf{x}_*)^T). \tag{7}$$

Using the shorthand $\phi(\mathbf{x}^*) = \phi^*$, $\mathbf{K} = \mathbf{\Phi \Sigma}_0 \mathbf{\Phi}^T$, we can rewrite the Eq. (7) in the following way

$$p(f_*|\mathbf{x}_*, \mathbf{y}) = \mathcal{N}(m(\mathbf{x}_*), \sigma^2(\mathbf{x}_*)) \tag{8}$$

with the mean and the variance functions

$$m(\mathbf{x}_*) = \phi_*^T \mathbf{\Sigma}_0 \mathbf{\Phi}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} = \mathbf{k}^T \mathbf{K}_y^{-1} \mathbf{y},$$

$$\sigma^2(\mathbf{x}_*) = \phi_*^T \mathbf{\Sigma}_0 \phi_* - \phi_*^T \mathbf{\Sigma}_0 \mathbf{\Phi}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{\Phi \Sigma}_0 \phi_* = k_* - \mathbf{k}^T \mathbf{K}_y^{-1} \mathbf{k},$$

where $\mathbf{k} = (k(\mathbf{x}_1, \mathbf{x}^*), \ldots, k(\mathbf{x}_N, \mathbf{x}^*))^T$, $k^* = k(\mathbf{x}^*, \mathbf{x}^*)$, and $\mathbf{K}_y = \mathbf{K} + \sigma^2 \mathbf{I}$.

To show this for the mean, note that $\frac{1}{\sigma^2} \mathbf{\Phi}^T (\mathbf{K} + \sigma^2 \mathbf{I}) = \frac{1}{\sigma^2} \mathbf{\Phi}^T (\mathbf{\Phi \Sigma}_0 \mathbf{\Phi}^T + \sigma^2 \mathbf{I}) = (\mathbf{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{\Phi}^T \mathbf{\Phi}) \mathbf{\Sigma}_0 \mathbf{\Phi}^T$. Hence we have $\frac{1}{\sigma^2} (\mathbf{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T = \mathbf{\Sigma}_0 \mathbf{\Phi}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1}$. For the variance, we can prove it by using the *Sherman–Woodbury–Morrison* formula.

Note that $\mathbf{K} = \mathbf{\Phi \Sigma}_0 \mathbf{\Phi}^T$ is called the Gram matrix with kernel elements

$$K_{nm} = k(\mathbf{x}_n, \mathbf{x}_m) = \phi(\mathbf{x}_n)^T \mathbf{\Sigma}_0 \phi(\mathbf{x}_m).$$

The kernel $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \mathbf{\Sigma}_0 \phi(\mathbf{x}')$ can be written as a simple dot product representation $k(\mathbf{x}, \mathbf{x}') = \psi(\mathbf{x})^T \psi(\mathbf{x}')$ where $\psi(\mathbf{x}) = \mathbf{\Sigma}_0^{1/2} \phi(\mathbf{x})$.

3.2. Gaussian process models

Instead of dealing with the parametric model in (6), we can directly define a prior probability distribution over functions. Specifically, let $\mathbf{f}$ denote the vector with elements $f_n = f(\mathbf{x}_n; \mathbf{w})$ for $n = 1, \ldots, N$. Then $\mathbf{f} = \mathbf{\Phi w}$ has the mean and covariance

$$\mathbb{E}[\mathbf{f}] = \mathbf{\Phi} \mathbb{E}[\mathbf{w}] = 0,$$

$$\text{cov}[\mathbf{f}] = \mathbb{E}[\mathbf{f}\mathbf{f}^T] = \mathbf{\Phi} \mathbb{E}[\mathbf{w}\mathbf{w}^T] \mathbf{\Phi}^T = \mathbf{\Phi}^T \mathbf{\Sigma}_0 \mathbf{\Phi} = \mathbf{K}$$

which shows that if we explicitly specify $\mathbf{K}$, then the mean and covariance of $\mathbf{f}$ does not explicitly depend on the parameter vector $\mathbf{w}$.

A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution (Williams & Rasmussen, 2006). In our formulation, it specifies the marginal distribution of $\mathbf{f}$ by

$$p(\mathbf{f}) = \mathcal{N}(\mathbb{E}[\mathbf{f}], \text{cov}[\mathbf{f}]) = \mathcal{N}(\mathbf{0}, \mathbf{K}).$$

The marginal distribution $p(\mathbf{y})$, conditioned on the input values $\mathbf{x}_1, \ldots, \mathbf{x}_N$, is then given by

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f}) p(\mathbf{f}) d\mathbf{f} = \mathcal{N}(\mathbf{0}, \mathbf{K}_y), \tag{9}$$

where $\mathbf{K}_y = \mathbf{K} + \sigma^2 \mathbf{I}$ is the covariance matrix. Since the joint distribution of the observed target values, $\mathbf{y}$, and the test output, $f^* = f(\mathbf{x}^*)$, is given by