Int. Journal of Math. Analysis, Vol. 4, 2010, no. 5, 201 - 209

The Constrained Option Pricing with Neural Networks and Isotonic Regression

Fatima Zahra Nqi

National School of Applied Sciences (E.N.S.A)
Faculty of Sciences, University Mohammed Premier
Oujda, Morocco
nqifz@yahoo.fr

Abdeluaab Lidouh

Department of Mathematics and Computer Sciences
Faculty of Sciences, University Mohammed Premier
Oujda, Morocco

Abstract

We estimate nonparametrically an option pricing function while imposing monotonicity and convexity constraints through isotonization. These two properties with respect to the price of the underlying asset appear as natural for option prices. To ensure monotonicity, an isotonic regression projects a nonparametric function on a set of monotonic functions. To ensure convexity, the same is done for the first derivative of the function. Although theoretical results state that nonparametric estimators approximate well both the function and its derivatives, it is often the case in practice that the assumed conditions are not met given a finite sample of data. With simulated data, we show that monotonicity and convexity constraints limit delta hedging errors.

Mathematics Subject Classification: financial mathematics

Keywords: Option Pricing, Nonparametric estimation, Neural networks, Isotonization

1 Introduction

A series of papers (Hutchinson, Lo and Poggio [11], Gourieroux, Monfort and Tenreiro [7], Aot-Sahalia and Lo [?], Garcia and Gencay [6], Broadie, Detemple, Ghysels and Torres [4]) have used nonparametric or semi-nonparametric methods to estimate a pricing formula for options. Since these methods do not rely on specific assumptions about the underlying asset price dynamics, they are robust to specification errors that might affect adversely parametric models. In most nonparametric studies, it is assumed that option pricing formula is homogeneous of degree one in the underling stock price. Dividing by the strike price reduces the number of inputs in learning the nonparametric pricing function. This parsimony is an advantage since the rate of convergence of nonparametric estimators slows down considerably as the number of inputs increases. Another reason invoked to justify homogeneity is the non-stationarity of option and stock prices.

The homogeneity of the option pricing function leads to its convexity. It appears natural that the option price at time t inherits the convexity property with respect to the underlying asset price of its terminal payoff.

When one estimates an option pricing function with nonparametric or semi-nonparametric methods, these shape restrictions are not automatically fulfilled. In general, imposing theoretical restrictions on an estimator of a function may be useful to reduce the variance of the estimator or to guarantee that the results are meaningful. For example, the derivative of the option pricing function with respect to the underlying asset price has the nature of a cumulative distribution function in the Black Scholes model and other models where the option pricing function is homogeneous. Estimating the function without any shape restrictions will not ensure that its derivative will obey the shape of a cumulative distribution function. Since a call option price is an increasing function of the stock price, monotonicity is related to pricing errors. The derivative of the option pricing function with respect to the stock price represents the hedging ratio. Ensuring convexity therefore controls the delta hedging error.

In this paper, we ensure the monotonicity and the convexity of the option pricing function through the method called isotonization. Following the lead of precursors such as Hildreth [8] and Brunk [3], isotonic regressions have been developed by Barlow [2] to ensure the monotonicity or the concavity of a regression functional estimator. Il consists in projecting a regression estimator on the class of monotone functions. To ensure monotonicity and convexity, we will apply the isotonization procedure twice, first to the function, second to its derivative. We will study a semi-nonparametric estimator, using feedforward neural networks, used by Hutchinson, Lo and Poggio [11] and Garcia and Gensay [6], for option pricing. For this estimator, authors usually specify the option price as a function of three variables : the underling asset price, the strike price and the time to maturity. The often made homogeneity assumptions reduces the number of independent variables to two, the asset price to strike price ratio and the time to maturity.

To simplify the estimation problem and assess the relevance of monotonicity and convexity restrictions, we will first study the option price solely as a function of the asset price, by fixing the maturity of the option. To be sure that the restrictions are satisfied by the data analyzed, we generate option prices that obey the Black-Scholes formula, which is a monotone and convex pricing function. We estimate the option pricing using a feed-forward neural network estimator and an isotonized estimator. Results show that imposing convexity restriction through isotonization helps recover the original function derivative and therefore reduces the hedging error.

2 A nonparametric approach to estimating an option pricing function

The option price is modeled as a function of variables that are deemed relevant for pricing the derivative security. For example, a natural nonparametric function for pricing a European call option on a non-dividend paying asset will relate the price of the option to the set of variables which characterize the option. i.e. the price of the underlying asset S_t, the strike price K, and the time to maturity τ. Therefore, the option pricing function can be written as:

$$C_t = f(S_t, K, \tau). \tag{1}$$

This approach is followed by Hutchinson, Lo and Poggio [11]. The function will also be valid to learn prices generated by a Black-Scholes model as the interest rate and volatility parameters present in the formula are constant and cannot be identified by a nonparametric estimator of the function f. However, when interest rate and volatility are stochastically changing over time, the nonparametric function should include these variables as well. Nonparametric estimation of such a general function will be marred by the curse of dimensionality. Therefore, no attempt has been made to estimate the function in such a general and unrestricted framework. On the contrary, researchers have been followed the reverse route. To reduce the number of variables, Hutchinson, Lo and Poggio [11] divide the function and its arguments by K and write the pricing function as follows:

$$\frac{C_t}{K} = f\left(\frac{S_t}{K}, 1, \tau\right). \tag{2}$$

While reducing the dimensionality of the problem, this division by the strike price assumes implicitly a property of the option pricing function f, namely its homogeneity of degree one in the asset price and the strike price. Garcia and Gensay [6] show that pricing accuracy gains can be made by exploiting the implications of this homogeneity property in terms of functional shape. They estimate a generalized option pricing formula that has functional shape similar to the usual Black-Scholes formula by a feed-forward neural network model.

While the option pricing function is of obvious economic interest, other quantities involving the derivatives of this function are of practical financial interest. The first derivative with respect to the stock price is the hedging ratio while the second derivative with respect to the stock price is proportional to the state price density or risk neural probability measure. Although theoretical results state that nonparametric estimators approximate well both the function and its derivatives, it is often the case in the practice that the assumed conditions are not met given a finite sample of data. If left unrestricted, a good nonparametric estimator of the function might not provide a good estimator of the hedging ratio or the state price density. Finally, the homogeneity of the option pricing function leads to its convexity. In the next section, we describe the methods that are used to impose monotonicity or convexity restrictions in nonparametric estimation.

3 Nonparametric Estimation Under Shape Restrictions

In the last section, we have motivated the use of nonparametric techniques in the estimation of option pricing functions. We have also indicated that monotonicity and convexity are weak and natural financial theoretical restrictions to be imposed on the option pricing function. Therefore, the question that has to be addressed first is to assess to what extent usual nonparametric estimation methods ensure that these restrictions are satisfied. Gallant and White [5] and Hornik, Stinchcombe and White [9] among others have demonstrated that under general regularity conditions, a sufficiently complex single hidden-layer feed-forward network can approximate a large class of functions and their derivatives to any desired degree of accuracy where the complexity of a single hidden layer feed-forward network is measured by the number of hidden units in the hidden layer. However, to learn the features of interest of an arbitrary function, the number of neural units must be permitted to go to infinity as the number of observations goes to infinity. In practice, with a finite number of observations, it might be difficult to reproduce well the function as well as its derivatives. Moreover, if some properties hold for the function, imposing them will reduce the variance of the estimators.

4 Estimating the Black-Scholes option pricing function and its derivative with Neural Networks Method

To illustrate the problem of estimating the function and its derivative with a finite sample of data, we simulate option prices from the Black-Scholes formula:

$$C_t = S_t \Phi(d_1(S_t)) - e^{-r(T-t)} K \Phi(d_2(S_t)), \tag{3}$$

where Φ denotes the normal distribution function, and where d_2 and d_2 are functions of S_t, more precisely:

$$d_1(S_t) = \frac{\ln(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \text{ and } d_2(S_t) = d_1(S_t) - \sigma\sqrt{T-t}, \tag{4}$$

where r is the risk free rate of interest, σ is the volatility and $T-t$ is the maturity. The hedging ratio Δ_t is defined as the first derivative of C_t with respect to S_t:

$$\Delta_t = \Phi(d_1(S_t)) + S_t \frac{\partial d_1(S_t)}{\partial S_t} \Phi'(d_1(S_t)) - K e^{r(T-t)} \frac{\partial d_2(S_t)}{\partial S_t} \Phi'(d_2(S_t)) \tag{5}$$

where Φ' denotes the derivative of Φ. Since :

$$\frac{\partial d_1(S_t)}{\partial S_t} = \frac{\partial d_2(S_t)}{\partial S_t} = \frac{1}{S_t \sigma \sqrt{T-t}}.$$

It can be shown that the second term in this expression cancels out and therefore, $\Delta_t = \Phi(d_1(S_t))$.

The Black-Scholes function is both monotonic and convex in the underling stock price and its first derivative is the normal distribution taking values from 0 to 1.

Given the Black-Scholes simulated prices, we estimate the option pricing function with neural network method. We start with the nonparametric regression

$$\frac{C_i}{K} = m\left(\frac{S_i}{K}\right) + \varepsilon_i \qquad i = 1, ..., N \tag{6}$$

If we denote C_i^{NN} and Δ_i^{NN} the option price and the hedging ratio estimated with the neural network method, then:

$$\frac{C_i^{NN}}{K} = \widehat{\beta}_0 + \sum_{j=1}^{d} \widehat{\beta}_j G\left(\widehat{\gamma}_j + \widehat{\theta}_j \frac{S_i}{K}\right), \text{ and } \Delta_i^{NN} = \sum_{j=1}^{d} \widehat{\beta}_j \widehat{\theta}_j G'\left(\widehat{\gamma}_j + \widehat{\theta}_j \frac{S_i}{K}\right), \tag{7}$$

where for all $j \in \{1, ..., d\}, \tilde{\beta}_0,\ \widehat{\beta}_j,\ \widehat{\gamma}_j$ and $\widehat{\theta}_j$ are solutions to the following optimization problem:

$$\min_{\beta_0,\beta_j,\gamma_j,\theta_j,j=1,..,d} \sum_{i=1}^{N} \left[\frac{C_i}{K} - \beta_0 - G(\gamma_j + \theta_j \frac{S_i}{K}) \right]^2, \tag{8}$$

d is the number of hidden units in the hidden layer and G the activation function, its given by: $G(x) = \frac{1}{1+e^{-\alpha x}}$, a logistic function which has the property of being a sigmoidal function. (G is sigmoidal function if $G : \mathbb{R} \to [0, 1], G(a) \to 0$ as $a \to -\infty$, $G(a) \to 1$ as $a \to +\infty$ and G is monotonic). G' is the first derivative of G.

5 Methods for estimating a function under shape restrictions

Shape restrictions on a functional parameter such as positivity, monotonicity or concavity arise often from theoretical considerations in economics or finance. For example, a cost function is known to be homogeneous of degree one, nondecreasing and concave in the practice in the price of inputs, while utility functions are assumed monotone and concave. Hildreth [8] considered the problem of estimating a production function with these theoretical restrictions, while Matzkin [10] estimated a monotone and concave utility function. The statistical nonparametric literature proposes several approaches to incorporate information about the shape of a function in an estimation procedure. The two main approaches are isotonized smoothers and constrained smoothers.

The general problem of estimating a function nonparametrically under restrictions can be formulated as follows. Let a set of data points $\{(X_i, Y_i)\}_{i=1}^{n}$, such as $X_1 \leq X_2 \leq ... \leq X_n$, and a functional parameter $m(x) = E[Y_i|X_i = x]$ to be estimated. Instead of estimating the regression function in an unconstrained manner as it is usually done, we impose restrictions on the estimate $\widehat{m}(x)$. For example, monotonicity would be expressed as: $\widehat{m}(X_1) \leq \widehat{m}(X_2) \leq ... \leq \widehat{m}(X_n)$, whereas concavity imposes the following constraint:

$$\frac{\widehat{m}(X_{i+1}) - \widehat{m}(X_i)}{X_{i+1} - X_i} \geq \frac{\widehat{m}(X_{i+2}) - \widehat{m}(X_{i+1})}{X_{i+2} - X_{i+1}},\ i = 1, ..., n-2$$

In the first case, we will perform a monotonic regression, a special case of an isotonic regression (where only a partial ordering is imposed on the X_i). These sets of constraints defined a cone of restrictions C and we look therefore for the solution $Y_i^\star$ to the minimization problem:

$$\min_{C} \sum_{i=1}^{n} (Y_i - m(X_i))^2.$$

In the case of monotonicity, the unique solution to this problem can be expressed by the two min-max formulas:

$$Y_i^\star = \max_{u\le i}\min_{v\ge i}\sum_{j=u}^{v} Y_j/(v-u+1) = \min_{u\ge i}\max_{v\le i}\sum_{j=u}^{v} Y_j/(u-v+1).$$

These expression are not optimal for the actual computation of the $Y_i^\star$. The most popular algorithm is the Pool Adjacent Violators (PAV) algorithm.

The PAV algorithm ensures that a monotonicity constraint is maintained for a nonparametric estimator of a function m known to be monotonic. More precisely, given a set of data points $\{(X_i, Y_i)\}_{i=1}^n$ such as $X_1 \le X_2 \le ... \le X_n$, we need to found estimated values $\{\tilde{m}(X_i)\}_{i=1}^n$ which minimize

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - m(X_i))^2$$

subject to the following monotonicity restriction:

$$\widehat{m}(X_1) \le \widehat{m}(X_2) \le ... \le \widehat{m}(X_n).$$

The solution of the this problem is found through the PAV algorithm. We will describe The PAV algorithm: given a series of data points $(X_i)_{i=1}^n$ such as $X_1 \le X_2 \le ... \le X_n$ and the series Y_i defined by $Y_i = m(X_i)\ \forall i = 1, n$. Let $Y = \{Y_1, Y_2, ..., Y_n\}$. A block will be defined as a subset of Y formed by consecutive elements of Y. Next, an order is defined over such blocks as follows: A block A is inferior to a block B if the highest element of A is smaller than the lowest element of B. We denote $Av(B)$ the average of the elements in block B. Initially, we subdivide the set Y in n blocks $\{B_i\}_{i=1}^n$ such as:

$$B_1 = \{Y_1\}, B_2 = \{Y_2\}, ..., B_n = \{Y_n\}.$$

If B_-, B, B_+ are three consecutive blocks in this order, we say that B is *up-satisfied* if $Av(B) < Av(B_+)$, and that B is *down-satisfied* if $Av(B_-) < Av(B)$. Extending this definition, we assume that any block containing Y_1 is *down-satisfied* and any block containing Y_n is *up-satisfied* . The PAV algorithm groups blocks together until all new blocks are both *up-satisfied* and *down-satisfied* . In this case, the isotonized series $Y_i^\star$ will be defined as follows:$\forall i$ such as $Y_i \in B, Y_i^\star = Av(B)$.

We start the algorithm with block $\{Y_i\}$. Assume that block B is active (the one on which the conditions *up-satisfied* and *down-satisfied* are tested). The computational procedure proceeds as follows:

• Step 1: Check if $Av(B) < Av(B_+$. If yes, go to the next block, i.e. B_+ becomes the active block and the procedure is repeated. If not, blocks B and B_+ are grouped together and the new active block becomes $B \cup B_+$.

• Step 2: Check if $Av(B_-) < Av(B)$. If yes, check if it *up-satisfied* (go back to step 1). If not, we step back and blocks B_- and B are grouped. The active block becomes $B_- \cup B$.

• Step 3: Step 2 is repeated until the active block is *down-satisfied* .

• Step 4: Once step 3 is finished, we go back to step 1 until the active block is *up-satisfied* .

• Step 5: Once the active block is both *up-satisfied* and *down-satisfied* , the next block becomes active and the procedure is repeated.

Then the isotonized neural network estimator is defined as m^{INN} and then:

$$\frac{C_i^{INN}}{K} = m^{INN}\left(\frac{S_i}{K}\right) = \min_{v \geq S_i/K} \max_{u \leq S_i/K} \frac{1}{v-u} \int_u^v m^{NN}(x)dx. \tag{9}$$

Using the expression of m^{NN}, we can write

$$m^{INN}\left(\frac{S_i}{K}\right) = \widehat{\beta}_0 + \sum_{j \in I_0} \widehat{\beta}_j G(\widehat{\theta}_j) + \min_{v \geq S_i/K} \max_{u \leq S_i/K} \frac{1}{v-u} \sum_{j \in I_1} \frac{\widehat{\beta}_j}{\widehat{\gamma}_j} \ln\left[\frac{1 - G(\widehat{\gamma}_j u + \widehat{\theta}_j)}{1 - G(\widehat{\gamma}_j v + \widehat{\theta}_j)}\right], \tag{10}$$

where

$$I_0 = \{j = 1, ..., d | \widehat{\gamma}_j = 0\} \text{ and } I_0 = \{j = 1, ..., d | \widehat{\gamma}_j \neq 0\}.$$

6 Conclusion

We look at the isotonized smoothers using the neural networks method. Since the data are generated by the Black-Scholes function, $\frac{C_t}{K}$ is a monotonic function of $\frac{S_t}{K}$. Therefore, there is no need to isotonize the data before estimating the function. Isotonization is done based on the PAV algorithm applied to the first derivative of the estimate function. Isotonization reduces roughly by half the mean integrable squared error (MISE) : With 2 units of hidden, we obtain a MISE-Hedging equal to $0.043\ 10^{-3}$ with NN method, and $0.023\,10^{-3}$ with INN method. With 3 units of hidden, we obtain a MISE-Hedging equal to $0.097\ 10^{-3}$ with NN method, and $0.060\,10^{-3}$ with INN method. With 4 units of hidden, we obtain a MISE-Hedging equal to $0.043\ 10^{-3}$ with NN method, and $0.024\,10^{-3}$ with INN method.

This paper proposed a method to incorporate shape restrictions, such as monotonicity and convexity, into neural networks estimator. The simulations results indicate that the isotonized estimator is monotone and convex and it is close to the Blach-Scholes estimator. One interesting extension of this methodology would be to do the same work for the kernel estimator and its derivative.

References

[1] Aot-Sahalia. Y. and A. Lo (1996), Nomparametric estimation of State-Price Densities Implicit in Finantial Asset Prices, *Journal of Finance*, 53, 499-547.

[2] Barlow, R, E. Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972),*Statistical Inference Under Order Restrictions, Wiley, New York.*

[3] Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities, *Annals of Mathematical Statistics*, 29,437-454.

[4] Broadie, M., Detemple, J., Ghysels, E., and Torres, O. (2000a). American options with stochastic dividends and volatility: A nonparametric investigation. *Journal of Economic Dynamics and Control*, 24, 1829-1857.

[5] Gallant, A. R. and H. White (1992), On learning the derivatives of an unknown mapping with multilayer feedforward networks,*Neural Networks*, 5, 129-138

[6] Garcia, R. and R. Gensay (2000), Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint. *Journal of Econometrics*, 94(1/2), 93-115

[7] Gourieroux, C., A., Monfort and C. Tenreiro (994). Kernel M-Estimators: Nonparametric Diagnostics for Structural Models.*Working Paper 9405, CEPREMAP, Paris.*

[8] Hildreth, C. (1954), Point Estimates of Ordinates of Concave Functions, *Journal of the American Statistical Association*, 49, 598-619.

[9] Hornik, K., M. Stinchcombe and H. White (1990), Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, 3,551-560.

[10] Matzkin, R. L. (1994), Restrictions of economic theory in nonparametric methods, in R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North Holland, Amsterdam. chapter 42, 2523-2558.

[11] Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 43(3), 851-889.

Received: May, 2009