A Non-Parametric Technique of Option Pricing

In our quest for a proper option-pricing model, we have so far relied on making assumptions regarding the dynamics of the underlying asset (more or less realistic) and attempted to derive an option pricing formula for an option written on an asset following these – assumed – dynamics. One important drawback is the restrictive aspect of the assumptions on the dynamics of the underlying asset.

Another approach to option pricing consists of inferring the transition probability density function from the cross-section of option prices, instead of relying on assumed underlying asset. The density function can then be used to price other derivatives written on the same asset. Note that by using historical data on that asset a density function could also be obtained, but it would be the objective density function, not the risk-neutral, and we would have to determine the appropriate discount rate as well – a difficult task. Moreover, the cross-section of option prices gives us a forward-looking view for the density, as opposed to a (possibly) outdated historical density function.
A. How to infer the risk-neutral density from a cross-section of option prices:

a. Using risk-neutral pricing, the price of a call is:

\[
C(S,K,r,t,T) = e^{-r(T-t)} E^Q \left[ g(S_T) \mid S_t = S \right] \\
= e^{-r(T-t)} E^Q \left[ \max(0, S_T - K) \mid S_t = S \right] \\
= e^{-r(T-t)} \int_0^{+\infty} \max(0, S_T - K) \rho^Q(S_T \mid S_t = S) dS_T \\
= e^{-r(T-t)} \int_K^{+\infty} (S_T - K) \rho^Q(S_T \mid S_t = S) dS_T
\]

b. Since we need to obtain the risk-neutral density function \( \rho^Q \) from the observed market prices of call options, we can differentiate the expression above twice with respect to \( K \) and get:

\[
\frac{\partial^2 C(S,K,r,t,T)}{\partial^2 K} = e^{-r(T-t)} \rho^Q(K \mid S_t = S)
\]

Note that the options must be very liquid in order to use this method, and that you can also use put options as they carry the same information (as a result of put-call parity).
B. Non-Parametric Regressions

1. A Review of Parametric Regressions First

a. Linear Model: \( y = x'\beta + \epsilon \) with \( E[y \mid x] = x'\beta \), can be estimated by least squares.

b. Nonlinear Model: \( y = \varphi(x,\beta) + \epsilon \) with \( E[y \mid x] = \varphi(x,\beta) \), can be estimated by least squares:
\[
\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{n} \left\{ y_i - \varphi(x_i,\beta) \right\}^2
\]

c. The key point here is that both models require the functional form of \( E[y \mid x] \), indicating that a process has been specified for the underlying asset. In the case of Black-Scholes, for instance, we have \( y=C, x=(S,K,r,t,T) \) and \( \beta=\sigma \):
\[
E[C \mid S, K, r, t, T] = BS(S, K, r, t, T; \sigma)
\]

2. Non-Parametric Regression

a. What we need is an estimate of \( E[y \mid x] \) without any specific assumption about the dynamics of the process.

b. Various methods can be used: Kernel functions, splines, or neural networks.

c. Kernel function: \( k(u) = \left(1/\sqrt{2\pi}\right)\exp\left\{-u^2/2\right\} \)

- Need to estimate \( E[C \mid S, K, r, \tau] \), where \( \tau=T-t \).
d. Finally, compute an estimate of the risk-neutral density:

\[ \hat{\rho}^Q(K \mid S_t = S) = e^{r(T-t)} \frac{\partial^2 \hat{E}[C \mid S, K, r, t, T]}{\partial K^2} \]

This is achieved by taking derivatives of the non-parametric call option pricing function (with respect to the exercise price) that we estimated through prior regression.

C. Monotonicity and Convexity of Option Pricing Functions

- The call pricing function at time \( t \) is then given by:

\[
C(S_t, X, \tau, r_t, \delta_t, \tau) = e^{-r_t \tau} \int_0^{+\infty} \max(S_T - X, 0) \hat{p}^*(S_T \mid S_t, \tau, r_t, \delta_t, \tau) dS_T
\]

where \( C \) is the call option price, \( X \) the exercise price, \( \tau \) the time to maturity, \( r \) the risk-free rate and \( \delta \) the dividend yield.

\( C \) must be a decreasing function of \( X \) and the first derivative of \( C \) with respect to \( X \) must be greater than \(-\exp(-r\tau)\) because:

\[
\frac{\partial C(S_t, X, \tau, r_t, \delta_t, \tau)}{\partial X} = -e^{-r_t \tau} \int_X^{+\infty} \hat{p}^*(S_T) dS_T
\]

( since we know that the density function is positive)

Thus we must have:

\[-e^{-r_t \tau} \leq \frac{\partial C(S_t, X, \tau, r_t, \delta_t, \tau)}{\partial X} \leq 0\]
By differentiating the call price twice with respect to the exercise price as we previously did, we obtain:

$$\frac{\partial^2 C(S_t, X, \tau, r_{t, \tau}, \delta_{t, \tau})}{\partial X^2} = e^{-r_{t, \tau} \tau} p^*(X) \geq 0$$

(Breeden and Litzenberger (1978) and Banz and Miller (1978))

This indicates that the second derivative must be convex. We thus have two conditions that must be satisfied:

$$\begin{cases} 
-e^{-r_{t, \tau} \tau} \leq C'_{t, \tau}(X) \leq 0 \\
C''_{t, \tau}(X) \geq 0.
\end{cases}$$

Ait-Sahalia and Lo (1998) proposed to estimate the state price density function by estimating an option pricing formula $\hat{C}(X)$ non-parametrically and by then differentiating the function with respect to $X$ in order to obtain $\partial^2 \hat{C} / \partial X^2$.

When the sample is large enough, non-parametric estimators of the call pricing function will satisfy the restrictions given above, and the technique described in Ait-Sahalia and Lo (1998) works well. However, the methodology requires a large amount of cross-sectional (over $X$, the strike price) information that is not always available (We usually only have about 20 to 50 options with different strike prices). Moreover, attempting to increase the sample size by gathering data from different dates seems like a good idea to improve the estimation of the SPD but can create problems: the underlying asset price, its volatility or the interest rate can be volatile enough to bias severely the estimation as the changes over even a few days could still be significant.

Therefore, since with a limited sample the estimated SPD can violate the constraints, we need to include these constraints by conducting a constrained regression (when estimating the call pricing function).
D. Constrained Non-Parametric Estimation

In order to obtain a non-parametric estimator that satisfies the constraints on the shape of the SPD, it is necessary to use a combination of constrained least squares regression and smoothing.

1. Constrained Least Squares Regression

Given \( n \) observations \( y_1, y_2, \ldots, y_n \) of call option prices, we first need to estimate a call pricing function satisfying a set of constraints. The explanatory variables (exercise or strike prices) are denoted by \( x_1, x_2, \ldots, x_n \) and assume that we have ordered the strike prices so that \( x_i > x_j \) for \( i > j \) and with \( i, j \in \{1, 2, \ldots, n\} \).

The constrained least squares regression consists in finding a vector \( m \) solving:

\[
\min_{m \in \mathbb{R}^n} \sum_{i=1}^{n} (m_i - y_i)^2 = \min_{m \in \mathbb{R}^n} \|m - y\|^2
\]

Subject to the slope and convexity constraints:

\[
\begin{cases}
-e^{-r_t} \leq \frac{m_{i+1} - m_i}{x_{i+1} - x_i} \leq 0 & \text{for all } i = 1, \ldots, n - 1 \\
\frac{m_{i+2} - m_{i+1}}{x_{i+2} - x_{i+1}} \geq \frac{m_{i+1} - m_i}{x_{i+1} - x_i} & \text{for all } i = 1, \ldots, n - 2
\end{cases}
\]
2. Local Polynomial Kernel Smoothing

The goal is to approximate smoothly the call option pricing function (as a function of $X$, the exercise price) for each quoted value of $X$. This way, we can then take the second derivative of that smooth function with respect to $X$ in order to derive the SPD, using the relationship between the second derivative of the call price and the state price density (Breeden and Litzenberger [1978]).

Let us approximate the pricing function $m$ in the neighborhood $z$ of a given state value $x$ (the exercise price) using a Taylor expansion or order $p$:

$$m(z) \approx \sum_{k=0}^{p} \beta_k(x) \times (z-x)^k$$

with $\beta_k(x) = m^{(k)}(x)/k!$ and where $m(z)=E[Y|Z=z]$.

To make sure that the estimated coefficients $\beta_k(x)$ reflect the local nature of the representation, it intuitively makes sense to assign more weight on data points that are closer to the value $x$. A simple way to achieve this is to introduce a kernel function $K(.)$ with a bandwidth $h$ and to use as weights $K_h(x_i-x) = K((x_i-x)/h)/h$. A typical kernel density is:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

(normal density)
The bandwidth controls the smoothness of the estimated function. Intuitively, if h is small, the weights assigned to data points away from x decrease very fast, so strong emphasis is placed on getting accurate pricing in the area very close to x, thus making the function not so smooth. If h is very large, however, significant weights are placed on data points away from x, so emphasis is placed on spreading the pricing accuracy over a “larger area” or “more points”, thus making the function smoother.

For each fixed point x we thus need to estimate the coefficients $\beta_k(x)$ as minimizers of:

$$
\sum_{i=1}^{n} \left\{ m_i - \sum_{k=0}^{p} \beta_{k,p}(x) \times (x_i - x)^k \right\}^2 K_h(x_i - x)
$$

This is a Generalized Least Squares regression of the $m_i$ (observed option prices) on various powers of $(x_i-x)$ with Kernel weights assigned. Each regression is “local” in the sense that the $\beta_k(x)$ obtained are valid in the neighborhood of x only. The regression must thus be repeated for each value of x (each available exercise price).
3. Estimating the Derivative

In order to estimate \( \frac{\partial^2 C}{\partial X^2} \) for each value of \( x \) observed, we can simply set \( p=k+1 \) with \( k=2 \) and use the estimator \( \tilde{m}^{(k)}(x) = k! \tilde{\beta}_{k,p}(x) \) with \( \tilde{\beta}_{2,3}(x) \) obtained from the GLS regression for that particular value of \( x \).

WHY?

We basically need to expand up to the third order:

\[
m(z) = \beta_{0,3}(x) + \beta_{1,3}(x)(z - x) + \beta_{2,3}(x)(z - x)^2 + \beta_{3,3}(x)(z - x)^3
\]

so that the second order estimate \( \tilde{\beta}_{2,3}(x) \) does not contain (or is “clean of”) all the remaining terms of higher order. The terms of higher order will be accounted for within the \( \tilde{\beta}_{3,3}(x) \) estimate.

Notice that \( \hat{\rho}^Q(X | S_1 = S) = e^{r(T-t)} \frac{\partial^2 \hat{E}[C | S, X, r, t, T]}{\partial X^2} \) and that \( \frac{\partial^2 \hat{E}[C | S, X, r, t, T]}{\partial X^2} = \tilde{m}^{(2)}(x) \)

So we finally need to back out \( \tilde{m}^{(2)}(x) = 2! \tilde{\beta}_{2,3}(x) \)

(since \( \beta_k(x) = m^{(k)}(x)/k! \))