Index Option Pricing via Nonparametric Regression

Ka Po Kung*♣

♣National University of Singapore, Singapore

Submitted: March 21, 2022 • Accepted: July 22, 2022

ABSTRACT: Investors typically use the Black-Scholes (B-S) parametric model to value financial options. However, there is extensive empirical evidence that the B-S model, assuming constant volatility of stock returns, is far from adequate to price options. This paper, using nonparametric regression, incorporates a volatility-adjusting mechanism into the B-S model and prices options on the S&P 500 Index. Specifically, the upgraded B-S model, referred to as the B-S nonparametric model, is equipped with such a mechanism whose function is to assign larger volatilities for larger log returns and smaller volatilities for smaller log returns to characterize volatility clustering, a phenomenon such that large/small log returns tend to be followed by large/small log returns. Using the B-S nonparametric models as a yardstick, our simulation results show that, across the board, the B-S parametric model considerably overprices both call and put options.

JEL classification: C14, C15, G13.

Keywords: Black-Scholes Parametric Model, Black-Scholes Nonparametric Models, Index Options, Volatility, Kernels.

1 Introduction

An index option is an option that gives the holder the right, but not the obligation, to buy or sell the value of a market index at a stated exercise price. Index options are cash-

*Corresponding Author. E-mail: kckung@u.nus.edu

DOI: 10.2478/erfin-2022-0004

settled and mostly European-style, meaning that they settle with an exchange of cash on the expiration date and have no provision for early exercise. A popular index option is on the S&P 500 Index (ticker SPX), which is actively traded on the Cboe Options Exchange[1] and has an expiration of up to 12 months.

This research is about pricing index options. Conventionally, when pricing options, financial economists express the dynamics of the underlying asset price (say z_t) in parametric form as an Ito stochastic differential equation with drift $\mu[\cdot]$ and volatility $\sigma[\cdot]$ as follows:

$$dz_t = \mu[\cdot]dt + \sigma[\cdot]dB_t \tag{1}$$

where B_t is a standard Brownian motion. Although the parametric formulation of the underlying asset price dynamics z_t has the advantage of analytic tractability, a potentially serious problem with such formulation is misspecification for the volatility. In practical applications, a misspecification for $\sigma[\cdot]$ can lead to systematic pricing errors for options driven by z_t.

As a case in point, the well-known Black-Scholes (B-S) (1973) parametric model has generally been accepted as the standard method for pricing options on stocks, market indices, and futures contracts. The B-S parametric model was developed on the assumption that the stock price dynamics P_t follow Equation (1) with $\mu[\cdot] = \mu P_t$ and $\sigma[\cdot] = \sigma P_t$. That is,

$$\frac{dP_t}{P_t} = \mu dt + \sigma dB_t \tag{2}$$

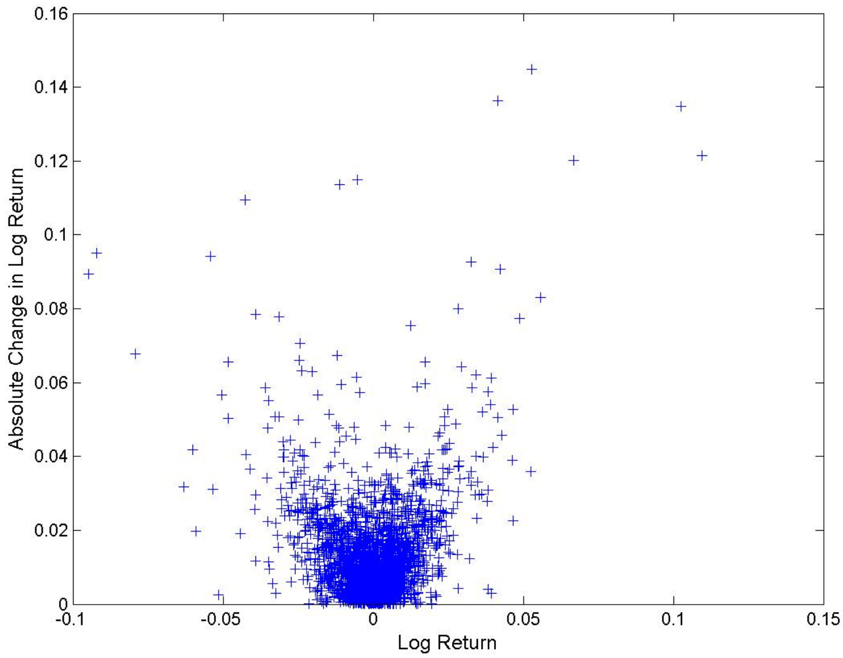
or

$$d\log P_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t \tag{3}$$

Commonly referred to as the geometric Brownian motion (GBM), Equation (2) or Equation (3) implies that the volatility σ of stock returns is constant over the life of the option. However, numerous empirical studies (see Blattberg and Gonedes, 1974; Castanias, 1979; Christie, 1982; MacBeth and Merville, 1979) have claimed that the volatility of stock returns is far from constant. To illustrate this claim, let daily log return s_t be defined as $s_t \equiv \log(P_t) - \log(P_{t-1})$ and let us employ $|s_t - s_{t-1}|$ as a proxy measure for volatility. Using 10-year daily log return data of the S&P 500 Index from 2012 to 2021, we plot a scatter diagram between $|s_t - s_{t-1}|$ and s_t. Somewhat in the shape of a funnel, the scatter diagram in Figure 1 shows that $|s_t - s_{t-1}|$ becomes larger as s_t moves both positively and negatively away from the mean of daily log returns – a phenomenon suggestive of volatility clustering. Volatility clustering (see Brooks, 2019; Cont, 2007) is that large returns (of either sign) are expected to follow large returns, and small returns (of

[1]The exchange, founded in 1973, was originally called the Chicago Board Options Exchange.

Figure 1: A scatter diagram between $|s_t - s_{t-1}|$ and s_t based on the S&P 500 Index



either sign) to follow small returns. Hence, the B-S parametric model, assuming constant volatility of stock returns as in Equation (3), is clearly inappropriate for pricing options on stocks or market indices.

To deal with this constant-volatility problem, this study, using nonparametric regression, incorporates a volatility-adjusting mechanism into the B-S model and values options on the S&P 500 Index. Specifically, the upgraded B-S model, referred to as the B-S nonparametric model, is equipped with such a mechanism whose function is to assign larger volatilities for larger $|s_t|$ and smaller volatilities for smaller $|s_t|$ to characterize volatility clustering. In Section 2, using daily log return data of the S&P 500 Index from 2012 to 2021, we will estimate nonparametrically these different values for the volatility σ in Equation (3) by means of the well-established Nadaraya-Watson (N-W) estimator[2].

Given the dependence of the volatility σ_t on the log return s_t (which in turn depends on time t) in our nonparametric formulation, Equation (3) can be expressed in discrete form under the risk-neutral probability[3] as follows:

$$\log P_t - \log P_{t-1} = \left(r_f - \frac{\sigma_t^2}{2}\right)\Delta t + \sigma_t \epsilon_t \sqrt{\Delta t} \tag{4}$$

where r_f is the short-term risk-free interest rate, ϵ_t is a standard normal variate, and Δt is the time change from $t-1$ to t.

[2]See Fan (2005); Ghosh (2018); Nadaraya (1964); Ullah and Pagan (1999); Watson (1964).

[3]The risk-neutral probability, not the original probability, is relevant for pricing options.

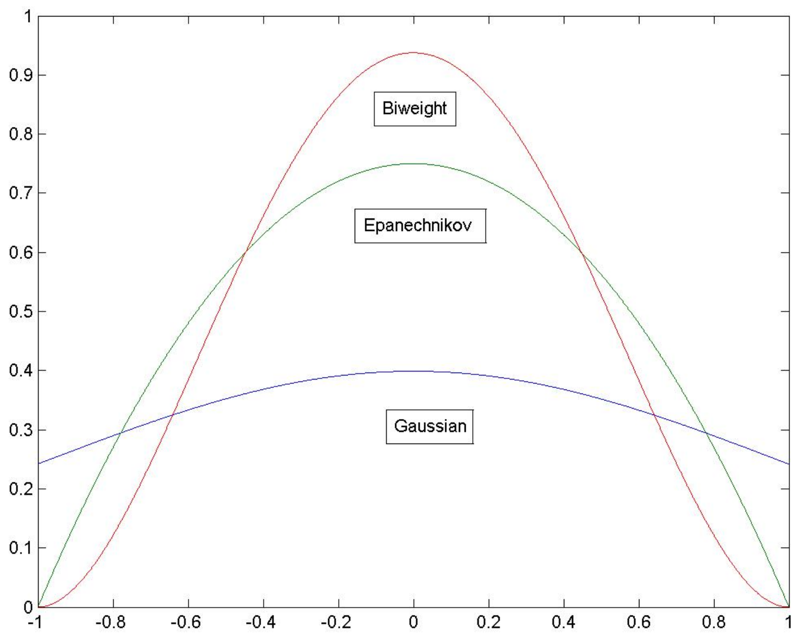
A major issue with nonparametric regression estimation is the choice of a kernel, which is a weight function incorporated into the regression estimator. Introduced first by Rosenblatt (1956), a kernel is a non-negative, real-valued, symmetric function K such that $\int_{-\infty}^{+\infty} K(v)dv = 1$. Although K is a probability density function, it is simply an effective way for calculating a weighted average and does not imply that the random variable v is distributed according to $K(v)$.

Table 1: Some commonly used kernel functions

Kernel	K(v)	Support
Biweight	$\frac{15}{16}\left(1-v^2\right)^2$	$\|v\|\leq 1$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}v\right)$	$\|v\|\leq 1$
Epanechnikov	$\frac{3}{4}\left(1-v^2\right)$	$\|v\|\leq 1$
Gaussian	$\frac{1}{h\sqrt{2\pi}}\exp\left(\frac{-v^2}{2h^2}\right)$	$-\infty < v < +\infty$
Logistic	$\frac{1}{2+e^v+e^{-v}}$	$-\infty < v < +\infty$
Triangular	$(1-\|v\|)$	$\|v\|\leq 1$
Tricule	$\frac{70}{81}\left(1-\|v\|^3\right)^3$	$\|v\|\leq 1$
Triweight	$\frac{35}{32}\left(1-v^2\right)^3$	$\|v\|\leq 1$
Uniform	$\frac{1}{2}$	$\|v\|\leq 1$

Note: exp stands for the exponential function.

Figure 2: The three kernel functions used: Gaussian, Epanechnikov, and Biweight



There are many kernel functions[4] proposed for nonparametric estimation, some of which are listed in Table 1. In this study, we try out three popular kernels (see Figure 2):

[4]See Fan and Yao (2003); Henderson and Parmeter (2015); Horowitz (2009); Li and Racine (2007).

the Gaussian, the Epanechnikov (1969), and the Biweight. Each of these kernels will be used to generate volatility estimates for the B-S nonparametric models. Given that a kernel plays an important part in nonparametric regression estimation, the goal of this study is to investigate, through simulation, the differences in index option prices between the B-S parametric model in Equation (3) and the B-S nonparametric models incorporated each with a kernel mechanism.

The rest of this paper is organized as follows. Section 2 introduces nonparametric regression and describes the Nadaraya-Watson estimators incorporated each with one of the three kernels. In Section 3, we lay out our simulation procedure in detail. Section 4 presents and discusses the option prices under the B-S parametric model and the three B-S nonparametric models incorporated each with a kernel mechanism. Section 5 concludes this research.

2 Nonparametric regression

2.1 An introduction

Suppose we wish to model a relation between two variables X and Y. In linear regression, having observed X, the average value of Y is obtained by the regression line. In many applications, however, we do not have adequate information to prespecify a linear relationship between X and Y. In these situations, we can resort to nonparametric regression to capture a wide variety of nonlinear relations between the two variables.

Specifically, in nonparametric regression, we want to determine a relation between X and Y such that

$$Y_t = f[X_t] + a_t \tag{5}$$

where a_t is a zero-mean error term process and $f[\cdot]$ is a smooth but unknown function. We wish to estimate $f[\cdot]$ at a specified time for which $X = x$. Suppose that at $X = x$ we can have m repeated independent observations $y_1 < y_2 < \cdots < y_m$. Then a rough estimator of $f[\cdot]$ at $X = x$ is

$$\widehat{f}[x] = \frac{1}{m}\sum_{t=1}^{m} y_t = \frac{1}{m}\sum_{t=1}^{m}\{f[x] + a_t\} = f[x] + \frac{1}{m}\sum_{t=1}^{m} a_t \tag{6}$$

By the law of large numbers, the average of $a_t's$ converges to 0 as m increases. Hence, the average $\widehat{f}[x] = \frac{1}{m}\sum_{t=1}^{m} y_t$ is a consistent estimator of $f[x]$.

For time series data, it is not possible to obtain repeated independent observations for which $X = x$. However, if the function $f[x]$ is adequately smooth, the values of Y for which X is near x will provide a close approximation of $f[x]$ and those for which X is further away from x will provide less close approximation of $f[x]$. Hence, if we incorporate a weight function $w(\cdot)$ into $\widehat{f}[x] = \frac{1}{m}\sum_{t=1}^{m} y_t$, then such a weighted average of the Y's is better than the simple average of the Y's in Equation (6) to estimate $f[\cdot]$. Hence, for any given x_i (where $i = 1, 2, \ldots, n$), a better estimator of $f[\cdot]$ can be expressed as

$$\widehat{f}[x_i] = \frac{1}{m} \sum_{t=1}^{m} w_t(x_i) y_t \tag{7}$$

Although nonparametric regression[5] requires few assumptions about the nature of the underlying data, it is highly data-dependent and is generally not effective for small sample sizes. Härdle (1990) pointed out four important reasons for the nonparametric approach to estimating a regression function.

First, it provides a versatile method of exploring a general relationship between two variables. Second, it gives predictions of observations yet to be made without reference to a fixed parametric model. Third, it provides a tool for finding spurious observations by studying the influence of isolated points. Fourth, it constitutes a flexible method of substituting for missing values or interpolating between adjacent values of the explanatory variable.

2.2 Nadaraya-Watson estimators with the three kernels

A key function of a nonparametric estimator is smoothing. By incorporating a kernel function into an estimator, nonparametric regression is a useful technique to deal with the amount of smoothing in a set of data. A kernel K has to satisfy two conditions:

$$K(v) \geq 0 \qquad \text{and} \qquad \int K(v) dv = 1 \tag{8}$$

By rescaling K using a variable $h > 0$, which is referred to as the bandwidth, the rescaled kernel becomes

$$K_h(v) \equiv \frac{1}{h} K\left(\frac{v}{h}\right) \qquad \text{and} \qquad \int K_h(v) dv = 1 \tag{9}$$

Given Equation (9), Nadaraya (1964) and Watson (1964) suggested the following weight function $w_t(x)$ in Equation (7):

$$w_t(x_i) = \frac{K_h[x_i - x_t]}{\frac{1}{m}\sum_{t=1}^{m} K_h[x_i - x_t]} \tag{10}$$

If the bandwidth h is small, the average or smoothing will be calculated over a small neighborhood around each of the $x_t's$. If h is large, the averaging will be calculated over a large neighborhood around each of the $x_t's$. Hence, adjusting the amount of smoothing is simply changing the bandwidth.

Substituting the weight function in Equation (10) into Equation (7), we obtain the following Nadaraya-Watson estimator $\widehat{g}[x_i]$ of $g[x_i] \equiv E[y_t|x_i]$ for an arbitrary x_i (where

[5]See Fan and Yao (2003); Henderson and Parmeter (2015); Horowitz (2009); Li and Racine (2007).

$i = 1, 2, \ldots, n$):

$$\widehat{g}[x_i] = \frac{\sum_{t=1}^{m} K_h [x_i - x_t] \, y_t}{\sum_{t=1}^{m} K_h [x_i - x_t]} \tag{11}$$

A feature of the N-W estimator in Equation (11) is that it is a weighted sum of those $y_t's$ that correspond to x_t in a neighborhood of x_i. The weights are small for $x_t's$ far away from x_i and large for $x_t's$ closer to x_i. Given the kernels in Table 1, we obtain the following N-W estimators incorporated with the Gaussian, the Epanechnikov, and the Biweight kernels, respectively.

$$\widehat{g}_G[x] = \frac{\sum_{t=1}^{m} \exp\left[\frac{-(x-x_t)^2}{2h^2}\right] y_t}{\sum_{t=1}^{m} \exp\left[\frac{-(x-x_t)^2}{2h^2}\right]} \tag{12}$$

$$\widehat{g}_E[x] = \frac{\sum_{t=1}^{m} \left[1 - \frac{(x-x_t)^2}{h^2}\right] y_t}{\sum_{t=1}^{m} \left[1 - \frac{(x-x_t)^2}{h^2}\right]} \tag{13}$$

$$\widehat{g}_B[x] = \frac{\sum_{t=1}^{m} \left[1 - \frac{(x-x_t)^2}{h^2}\right]^2 y_t}{\sum_{t=1}^{m} \left[1 - \frac{(x-x_t)^2}{h^2}\right]^2} \tag{14}$$

As suggested by many studies (see Henderson and Parmeter, 2015; Scott, 2015; Silverman, 1986), we use the following simple-but-popular rule for the bandwidth:

$$h = \widehat{\sigma} \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \tag{15}$$

where $\widehat{\sigma}$ is the overall standard deviation estimate of the variable, d is the dimension of the variable, and n is the number of observations. We use $s_t \equiv \log(P_t) - \log(P_{t-1}) =$ daily log return on the S&P 500 Index from 2012 to 2021 for estimation. With $d = 1$, $\widehat{\sigma} = 0.01046$, $n = 2{,}516$, the bandwidth based on Equation (15) is 0.00231.

Table 2 reports the three types of annualized N-W estimates for the volatility, obtained through estimation using the N-W estimators in Eqs. (12), (13), and (14), for different daily log returns from -0.05 to 0.05, a range wide enough to accommodate basically all possible values of daily log return s_t. In terms of magnitude, the differences between the volatilities generated by the three kernels are not large. For any s_t value, the Biweight kernel generates the largest volatility, followed by the Gaussian kernel and then the Epanechnikov kernel. For example, when $s_t = 0.002$, the annualized volatility is 0.0887 under the Biweight kernel, 0.0860 under the Gaussian kernel, and 0.0651 under the Epanechnikov kernel. Figure 3 gives a graphical exposition of the annualized N-W estimates for the volatility in Table 2 based on the three kernels. The three curves look rather similar — with larger volatility estimates for large $|s_t|$ and smaller volatility estimates for small $|s_t|$. This is a phenomenon of volatility clustering (see Brooks, 2019; Cont, 2007), a

Table 2: Annualized N-W estimates for the volatility of asset returns

s_t	$\widehat{\sigma}_g$	$\widehat{\sigma}_e$	$\widehat{\sigma}_b$	s_t	$\widehat{\sigma}_g$	$\widehat{\sigma}_e$	$\widehat{\sigma}_b$
−0.050	0.2594	0.2381	0.2828	0.002	0.0860	0.0651	0.0887
−0.048	0.2472	0.2300	0.2751	0.004	0.0842	0.0673	0.0852
−0.046	0.2350	0.2221	0.2680	0.006	0.0850	0.0735	0.0821
−0.044	0.2264	0.2135	0.2612	0.008	0.0886	0.0803	0.0779
−0.042	0.2186	0.2059	0.2553	0.010	0.0927	0.0858	0.0851
−0.040	0.2100	0.1978	0.2468	0.012	0.0970	0.0922	0.1009
−0.038	0.2021	0.1905	0.2367	0.014	0.1020	0.0983	0.1112
−0.036	0.1941	0.1825	0.2291	0.016	0.1061	0.1049	0.1238
−0.034	0.1867	0.1752	0.2239	0.018	0.1112	0.1119	0.1340
−0.032	0.1790	0.1679	0.2125	0.020	0.1155	0.1186	0.1464
−0.030	0.1722	0.1606	0.2011	0.022	0.1208	0.1252	0.1556
−0.028	0.1649	0.1499	0.1940	0.024	0.1258	0.1316	0.1660
−0.026	0.1572	0.1438	0.1805	0.026	0.1319	0.1389	0.1733
−0.024	0.1500	0.1374	0.1750	0.028	0.1375	0.1459	0.1805
−0.022	0.1429	0.1293	0.1643	0.030	0.1433	0.1526	0.1871
−0.020	0.1368	0.1249	0.1520	0.032	0.1493	0.1593	0.1950
−0.018	0.1325	0.1183	0.1449	0.034	0.1555	0.1663	0.2049
−0.016	0.1283	0.1113	0.1374	0.036	0.1617	0.1740	0.2125
−0.014	0.1238	0.1049	0.1290	0.038	0.1688	0.1813	0.2179
−0.012	0.1187	0.1020	0.1241	0.040	0.1752	0.1877	0.2238
−0.010	0.1133	0.0941	0.1219	0.042	0.1814	0.1944	0.2303
−0.008	0.1080	0.0884	0.1149	0.044	0.1901	0.2017	0.2368
−0.006	0.1029	0.0828	0.1084	0.046	0.2068	0.2094	0.2425
−0.004	0.0980	0.0781	0.1038	0.048	0.2163	0.2176	0.2489
−0.002	0.0933	0.0732	0.0978	0.050	0.2264	0.2253	0.2559
0.000	0.0892	0.0670	0.0916				

Note: $\widehat{\sigma}_g$, $\widehat{\sigma}_e$, and $\widehat{\sigma}_b$ are annualized volatility estimates from the Gaussian, the Epanechnikov, and the Biweight kernels, respectively.

tendency of large returns (both positive and negative) to follow large returns, and small returns (both positive and negative) to follow small returns.

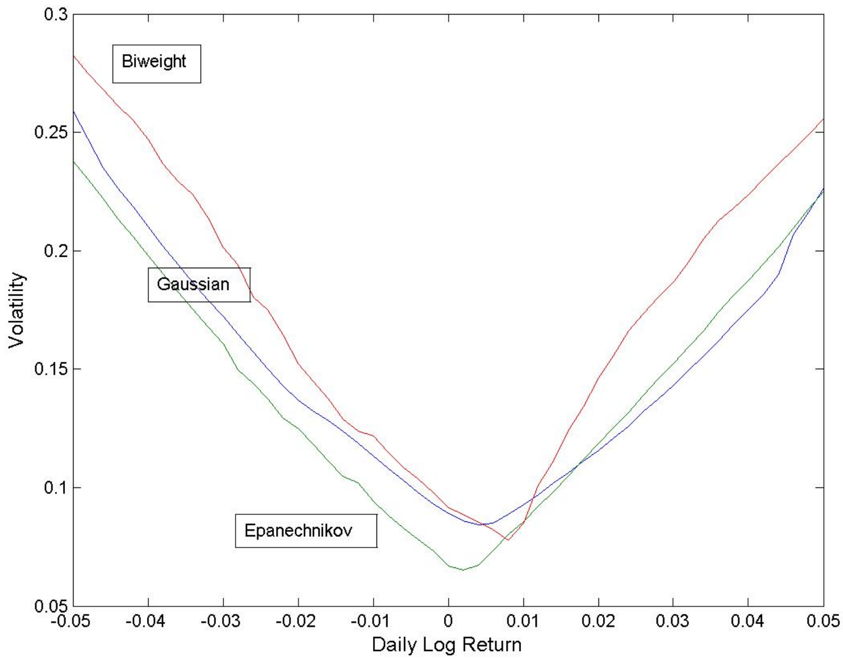
3 Simulation procedure

We price options on the S&P 500 Index (SPX) on 3 January 2022, the first trading day[6] of 2022. SPX options are of the European type and have an expiration of up to 12 months. We perform our computations through simulation and with different settings. Specifically, one setting is different from another by the values it assumes for the following four variables: the risk-free interest rate r_f, the dividend yield[7] d, the date of expiration T of the option, and the exercise price X. For our simulation, r_f is set to assume 0.02 and 0.05; d to assume 0.03; T to assume 3, 6, and 12 months; and P_0/X to range from 0.90 to 1.10 in increments of 0.01. For example, given $P_0 \equiv$ the S&P 500 Index on 3 January

[6]The day of 3 January 2022 is a Monday.

[7]The value of 0.03 for the dividend yield is based on an average estimate using the S&P 500 Index from 2012 to 2021.

Figure 3: A graphical exposition of the three annualized N-W estimates for the volatility



$2022 = 4{,}796.56$, X will be 4,568.15 when $P_0/X = 1.05$.

We will determine index option prices based on the B-S parametric model and the three B-S nonparametric models. The difference between the parametric model and the nonparametric models is that the former uses a constant value for the volatility σ in Equation (3) and the latter uses different values for σ_t. In the former, given that the volatility of daily log return $s_t \equiv \log(P_t) - \log(P_{t-1}) = 0.00826$ in the year 2021[8], the annualized volatility used is $0.00826x\sqrt{252} = 0.1311$. In the latter, the volatility σ_t used will be those in Table 2 based on the s_t value and the kernel incorporated. By the way, option prices based on the B-S parametric model can simply be computed using the standard B-S pricing formula. See Black and Scholes (1973) and Hull (2018).

Note that dividend payment will reduce the growth rate of the index. If the index grows from P_0 at time 0 to P_T at time T with an annualized dividend yield of d, then it will grow from P_0 at time 0 to $P_T e^{dt}$ at time T without the dividend yield. For simplicity, we replace P_0 by $P_0' = P_0 e^{-dt}$ and P_T by $P_T' = P_T e^{-dt}$, and then price index options as if the market index did not pay out any dividends. With ϵ_t to represent a standard normal variate, Equation (4) can be written in discrete form under the risk-neutral probability as

$$\log P_t' - \log P_{t-1}' = \left(r_f - \frac{\sigma_t^2}{2} \right) \Delta t + \sigma_t \epsilon_t \sqrt{\Delta t} \tag{16}$$

[8]Given that we price options with maturities of one year or less (3-month, 6-month, and 12-month) on the first trading day of 2022, it is appropriate to use the one-year data of 2021 to estimate the annualized volatility σ for the B-S parametric model.

For our simulation for the three B-S nonparametric models, consider a 6-month option to expire on 30 June 2022 with an exercise price of X. We denote 3 January 2022 by time 0 and 30 June 2022 by time T. We want to determine the price at time 0 of this option with expiration date T and exercise price X. At time 0, $P_0 \equiv$ the S&P 500 Index = 4,796.56. We partition the length of this 6-month option $[0, T]$ into 126 intervals[9] as $\{0 \equiv t_0 < \cdots < t_{126} \equiv T\}$. That is, each interval is of length $\Delta t = t_i - t_{i-1} = \frac{1}{126}$. Equation (16) now becomes

$$\log P'_T = \log P'_0 + \sum_{t=1}^{126} \left[\left(r_f - \frac{\sigma_t^2}{2} \right) \Delta t + \sigma_t \epsilon_t \sqrt{\Delta t} \right] \tag{17}$$

A typical trajectory of the index P_t from time 0 to time T is simulated as follows. We simulate 126 random draws of ϵ_t, each of which is used for each interval Δt, to determine P_T at time T. Note that at different points in time as we move forward through the life of the option, the volatility used over the next interval depends on the magnitude of the daily log return s_t of the market index over the previous interval. For example, at time 1, the volatility used is based on the daily log return from time 0 to time 1. Specifically, the S&P 500 Index was 4,796.56 on 3 January 2022 and 4,793.54 on 4 January 2022, which implies a daily log return of -0.00063. From Table 2, a daily log return of -0.00063 corresponds to an annualized volatility of 0.0892 for the Gaussian kernel, 0.0670 for the Epanechnikov kernel, and 0.0916 for the Biweight kernel. Hence, we use these values for σ_1 at time 1. The same procedure is used for choosing volatility at other time points.

Proceeding in similar fashion, we repeat the above procedure to simulate 100,000 independent trajectories for each setting. For each trajectory, we determine P'_T. Then the call and put for trajectory j are obtained as $c^j = \max\left[0, P'_T - X\right] e^{-r_f T}$ and $p^j = \max\left[0, X - P'_T\right] e^{-r_f T}$, where $j = 1, 2, \ldots, 100{,}000$. The prices of call and put at time 0 are

$$c = \frac{1}{100{,}000} \left[c^1 + c^2 + \cdots + c^{100{,}000}\right] \tag{18}$$

$$p = \frac{1}{100{,}000} \left[p^1 + p^2 + \cdots + p^{100{,}000}\right] \tag{19}$$

4 Results for option prices

For simplicity, we use Gauss, Epane, or Biwei to stand for the B-S nonparametric model equipped with the Gauss, the Epanechnikov, or the Biweight kernel, respectively. As shown in Section 2, the volatilities in Table 2 generated by the three kernels are tailor-made for the B-S nonparametric models and, as such, are better estimates than the "one-size-fits-all" volatility estimate of 0.1311 used by the B-S parametric model. With better

[9]Similarly, we partition the length of a 3-month option into 63 intervals and the length of a 12-month option into 252 intervals.

estimates for the volatility, the three B-S nonparametric models can correct a substantial amount of the systematic pricing biases in the B-S parametric model. Therefore, we use the three B-S nonparametric models as a yardstick against which the B-S parametric model is compared. In what follows, we first present and discuss the option prices under the B-S nonparametric models and the B-S parametric model, and then we illustrate, with concrete numbers, how much investors will overpay for options priced based on the B-S parametric model.

4.1 Option prices between nonparametric models and parametric model

Table 3: Prices of index options when expiration date $T = 3$ months and $r_f = 0.02$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	6.87	0.85	0.11	1.03	549.08	541.30	540.88	541.37
0.91	10.12	1.64	0.28	1.95	494.06	483.82	482.77	484.02
0.92	14.52	2.99	0.65	3.48	441.45	428.17	426.13	428.55
0.93	20.32	5.20	1.43	5.98	391.47	374.59	371.13	375.27
0.94	27.78	8.78	2.93	9.90	344.34	323.58	318.04	324.59
0.95	37.16	14.25	5.75	15.77	300.26	275.60	267.42	277.02
0.96	48.66	21.98	10.65	23.93	259.44	231.00	219.98	232.85
0.97	62.48	32.60	18.35	34.96	222.00	190.37	176.42	192.62
0.98	78.74	46.47	29.59	49.28	188.06	154.03	137.46	156.74
0.99	97.51	63.95	45.20	66.99	157.63	122.32	103.88	125.26
1.00	118.78	84.99	65.50	88.20	130.70	95.15	75.98	98.25
1.01	142.50	109.59	90.40	112.74	107.16	72.49	53.61	75.54
1.02	168.53	137.45	119.54	140.50	86.87	54.03	36.43	56.98
1.03	196.69	168.21	152.30	171.13	69.60	39.36	23.76	42.18
1.04	226.75	201.41	188.05	204.07	55.11	28.01	14.96	30.56
1.05	258.47	236.60	225.90	238.93	43.12	19.49	9.10	21.71
1.06	291.56	273.26	265.05	275.19	33.33	13.27	5.37	15.10
1.07	325.77	311.01	304.79	312.51	25.46	8.94	3.03	10.34
1.08	360.82	349.27	344.70	350.49	19.21	5.90	1.64	7.01
1.09	396.48	387.71	384.46	388.70	14.32	3.80	0.86	4.69
1.10	432.51	426.11	423.81	426.89	10.55	2.39	0.41	3.06

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

Tables 3 to 8 show the prices of index options with three different expiration dates T and with two different risk-free interest rates r_f. Regardless of the values[10] for T and r_f, the differences in option prices are not large between the three B-S nonparametric models. For both calls and puts, Biwei generates the largest option prices, followed by Gauss and then Epane. Suppose $r_f = 0.02$, and $P_0/X = 1.00$. When $T = 3$ months, the call and

[10]See Hull (2018) for how option prices change if there is a change in expiration date or in risk-free interest rate.

Table 4: Prices of index options when expiration date $T = 6$ months and $r_f = 0.02$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	26.48	7.02	1.80	8.08	577.81	556.93	551.67	557.87
0.91	33.42	10.14	3.11	11.49	526.77	502.06	495.00	503.30
0.92	41.64	14.31	5.13	15.99	478.27	449.52	440.30	451.07
0.93	51.27	19.86	8.09	21.94	432.39	399.56	387.75	401.52
0.94	62.39	27.03	12.38	29.52	389.19	352.41	337.73	354.77
0.95	75.08	36.06	18.46	38.93	348.71	308.26	290.62	311.01
0.96	89.42	47.16	26.71	50.47	310.97	267.29	246.80	270.48
0.97	105.43	60.63	37.51	64.32	275.99	229.77	206.60	233.33
0.98	123.14	76.62	51.30	80.59	243.74	195.79	170.43	199.65
0.99	142.53	94.99	68.25	99.17	214.18	165.22	138.44	169.28
1.00	163.57	115.82	88.53	120.20	187.25	138.08	110.76	142.35
1.01	186.20	139.17	112.01	143.53	162.87	114.42	87.22	118.65
1.02	210.35	164.67	138.40	169.03	140.92	93.81	67.51	98.05
1.03	235.93	192.21	167.34	196.43	121.30	76.16	51.24	80.26
1.04	262.82	221.58	198.58	225.58	103.86	61.20	38.16	65.07
1.05	290.90	252.48	231.81	256.21	88.45	48.61	27.90	52.22
1.06	320.06	284.69	266.61	288.14	74.94	38.15	20.03	41.48
1.07	350.14	318.09	302.50	321.12	63.15	29.68	14.06	32.59
1.08	381.02	352.31	339.24	355.02	52.94	22.81	9.70	25.39
1.09	412.57	387.13	376.43	389.55	44.15	17.29	6.55	19.59
1.10	444.65	422.43	413.82	424.53	36.62	12.98	4.33	14.96

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

put prices are 88.20 and 98.25 under Biwei, 84.99 and 95.15 under Gauss, and 65.50 and 75.98 under Epane. When $T = 6$ months, the call and put prices are 120.20 and 142.35 under Biwei, 115.82 and 138.08 under Gauss, and 88.53 and 110.76 under Epane. When $T = 12$ months, the call and put prices are 158.02 and 205.10 under Biwei, 152.02 and 199.45 under Gauss, and 114.99 and 161.77 under Epane.

On the other hand, the prices of both calls and puts are, across the board, larger based on the B-S parametric model than based on the three B-S nonparametric models, regardless of the values for T and r_f. A few numbers will suffice to bring out this phenomenon. Suppose $r_f = 0.05$, and $P_0/X = 1.05$. When $T = 3$ months, the call and put prices are 284.53 and 35.21 under the B-S parametric model, compared to 265.57 and 14.40 under Gauss, 256.66 and 5.98 under Epane, and 267.07 and 16.25 under Biwei. When $T = 6$ months, the call and put prices are 337.61 and 67.82 under the B-S parametric model, compared to 304.23 and 32.78 under Gauss, 287.32 and 16.23 under Epane, and 306.95 and 35.93 under Biwei. When $T = 12$ months, the call and put prices are 421.54 and 112.10 under the B-S parametric model, compared to 366.49 and 57.33 under Gauss, 340.96 and 31.49 under Epane, and 370.61 and 61.91 under Biwei. The B-S parametric model considerably overprices both calls and puts!

Table 5: Prices of index options when expiration date $T = 12$ months and $r_f = 0.02$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	67.29	23.63	8.82	26.21	636.47	593.46	578.00	595.69
0.91	77.81	30.00	12.39	33.00	589.58	542.42	524.16	545.08
0.92	89.40	37.58	17.02	41.05	545.01	493.85	472.63	496.97
0.93	102.08	46.47	22.96	50.50	502.74	447.78	423.63	451.47
0.94	115.87	56.86	30.39	61.30	462.75	404.39	377.27	408.49
0.95	130.78	68.77	39.55	73.60	425.01	363.65	333.79	368.13
0.96	146.81	82.18	50.53	87.40	389.49	325.51	293.21	330.38
0.97	163.96	97.21	63.50	102.71	356.15	290.05	255.69	295.21
0.98	182.20	113.85	78.54	119.62	324.93	257.23	221.27	262.66
0.99	201.51	132.13	95.68	138.05	295.78	227.05	189.96	232.62
1.00	221.86	152.02	114.99	158.02	268.64	199.45	161.77	205.10
1.01	243.21	173.47	136.42	179.60	243.44	174.34	136.65	180.13
1.02	265.51	196.47	159.83	202.54	220.11	151.71	114.42	157.44
1.03	288.71	220.93	185.10	226.88	198.56	131.42	94.94	137.03
1.04	312.76	246.72	212.23	252.48	178.71	113.32	78.18	118.73
1.05	337.60	273.70	240.91	279.24	160.49	97.25	63.81	102.44
1.06	363.15	301.66	270.92	307.01	143.81	82.97	51.57	87.96
1.07	389.37	330.64	302.26	335.77	128.57	70.49	41.47	75.28
1.08	416.18	360.52	334.61	365.37	114.69	59.68	33.12	64.19
1.09	443.51	391.06	367.65	395.61	102.09	50.29	26.23	54.49
1.10	471.31	422.17	401.22	426.42	90.67	42.18	20.59	46.09

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

4.2 How much investors overpay based on B-S parametric model

Options on the S&P 500 Index (SPX) are actively traded on the Cboe Options Exchange. One SPX option contract gives the holder the right to buy or sell 100 times the index at the specified exercise price. The following computations illustrate how much investors will overpay if the market prices the SPX options based on the B-S parametric model. To begin with, on 3 January 2022 when the risk-free interest rate r_f is 0.05, suppose a call buyer wishes to buy a 6-month call and a put buyer to buy a 6-month put at an exercise price of 4,796.56, the value of the S&P 500 Index on that day. This implies that $P_0/X = 4{,}796.56/4{,}796.56 = 1.00$. From Table 7, the call and put prices are 198.36 and 151.35 under the B-S parametric model, 151.88 and 103.19 under Gauss, 124.85 and 76.53 under Epane, and 155.85 and 107.59 under Biwei. Then the call buyer will overpay $(198.36 - 151.88) \times \$100 = \$4{,}648$ per option contract if the true underlying model at work is Gauss; $(198.36 - 124.85) \times \$100 = \$7{,}351$ if Epane; and $(198.36 - 155.85) \times \$100 = \$4{,}251$ if Biwei. On the other hand, the put buyer will overpay $(151.35 - 103.19) \times \$100 = \$4{,}816$ per option contract if the true model is Gauss; $(151.35 - 76.53) \times \$100 = \$7{,}482$ if Epane; and $(151.35 - 107.59) \times \$100 = \$4{,}376$ if Biwei.

Table 6: Prices of index options when expiration date $T = 3$ months and $r_f = 0.05$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	8.96	1.33	0.22	1.53	511.55	502.06	501.44	502.62
0.91	12.99	2.50	0.49	2.85	457.74	445.39	443.87	446.09
0.92	18.37	4.41	1.12	5.02	406.54	390.72	387.91	391.68
0.93	25.36	7.56	2.36	8.48	358.16	338.50	333.79	339.78
0.94	34.21	12.47	4.69	13.72	312.82	289.23	281.94	290.84
0.95	45.16	19.53	8.93	21.20	270.73	243.25	233.13	245.26
0.96	58.41	29.35	15.75	31.38	232.04	201.13	188.02	203.51
0.97	74.11	42.42	26.02	44.78	196.87	163.32	147.41	166.04
0.98	92.34	59.05	40.53	61.69	165.27	130.13	112.09	133.11
0.99	113.13	79.32	59.68	82.12	137.23	101.57	82.42	104.72
1.00	136.42	103.18	83.63	106.00	112.67	77.58	58.52	80.75
1.01	162.09	130.51	112.05	133.20	91.45	58.01	40.03	61.06
1.02	189.98	160.88	144.27	163.42	73.35	42.40	26.27	45.29
1.03	219.85	193.83	179.70	196.05	58.14	30.26	16.61	32.84
1.04	251.46	228.88	217.42	230.76	45.52	21.09	10.11	23.32
1.05	284.53	265.57	256.66	267.07	35.21	14.40	5.98	16.25
1.06	318.78	303.43	296.66	304.52	26.91	9.70	3.41	11.14
1.07	353.95	341.94	336.88	342.74	20.31	6.44	1.87	7.60
1.08	389.78	380.65	376.97	381.19	15.15	4.16	0.97	5.06
1.09	426.03	419.36	416.70	419.66	11.16	2.63	0.46	3.28
1.10	462.50	457.85	455.95	458.00	8.12	1.61	0.20	2.11

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

5 Conclusion

Financial economists and investors alike typically employ the B-S parametric model to price various kinds of financial options. However, there is extensive empirical evidence that the B-S model is inadequate to value options. A major problem with the B-S model is that it assumes constant volatility of stock returns. This erroneous assumption will cause the B-S model to price options incorrectly. By means of nonparametric regression, this study incorporates a volatility-adjusting mechanism into the B-S model and prices options on the S&P 500 Index. Specifically, the upgraded B-S model, referred to as the B-S nonparametric model, is incorporated with such a mechanism whose function is to assign larger volatilities for larger log returns and smaller volatilities for smaller log returns to characterize volatility clustering. In particular, we utilize three different kernels to formulate three different B-S nonparametric models. Using these three B-S nonparametric models as a yardstick, our simulation results show that, across the board, the B-S parametric model substantially overvalues both call and put options.

Table 7: Prices of index options when expiration date $T = 6$ months and $r_f = 0.05$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	36.23	11.41	3.71	12.89	509.00	482.52	475.18	484.43
0.91	45.09	16.15	6.06	18.00	460.74	430.13	420.42	432.41
0.92	55.42	22.38	9.52	24.62	415.20	380.48	368.00	383.15
0.93	67.34	30.39	14.55	32.99	372.44	333.82	318.35	336.85
0.94	80.91	40.37	21.54	43.44	332.50	290.29	271.82	293.79
0.95	96.19	52.69	30.94	56.08	295.39	250.22	228.84	254.04
0.96	113.21	67.44	43.16	71.14	261.11	213.68	189.76	217.80
0.97	131.96	84.74	58.59	88.60	229.63	180.74	154.95	185.03
0.98	152.45	104.60	77.38	108.59	200.90	151.39	124.53	155.81
0.99	174.60	127.00	99.51	131.05	174.84	125.57	98.45	130.05
1.00	198.36	151.88	124.85	155.85	151.35	103.19	76.53	107.59
1.01	223.64	178.89	153.05	182.77	130.31	83.89	58.42	88.20
1.02	250.33	207.91	183.77	211.49	111.59	67.49	43.73	71.51
1.03	278.31	238.59	216.61	241.98	95.04	53.65	32.04	57.46
1.04	307.45	270.76	251.29	273.86	80.51	42.15	23.04	45.67
1.05	337.61	304.23	287.32	306.95	67.82	32.78	16.23	35.93
1.06	368.64	338.68	324.33	341.04	56.83	25.19	11.22	27.98
1.07	400.42	373.89	361.96	375.88	47.36	19.16	7.59	21.58
1.08	432.80	409.61	399.88	411.29	39.25	14.40	5.03	16.51
1.09	465.65	445.64	437.86	447.01	32.36	10.68	3.27	12.49
1.10	498.84	481.81	475.68	482.90	26.54	7.84	2.08	9.35

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

Table 8: Prices of index options when expiration date $T = 12$ months and $r_f = 0.05$

	Call Price				Put Price			
$\frac{P_0}{X}$	B-S	Gauss	Epane	Biwei	B-S	Gauss	Epane	Biwei
0.90	98.69	43.76	21.18	47.64	513.47	458.84	435.94	463.17
0.91	112.62	54.10	28.39	58.41	471.69	413.46	387.44	418.23
0.92	127.74	66.07	37.42	70.65	432.32	370.93	341.97	375.98
0.93	144.05	79.67	48.37	84.52	395.30	331.21	299.60	336.52
0.94	161.54	94.94	61.36	100.03	360.60	294.29	260.40	299.83
0.95	180.20	111.92	76.52	117.22	328.17	260.18	224.46	265.93
0.96	200.00	130.59	93.93	136.08	297.94	228.81	191.84	234.77
0.97	220.91	150.99	113.71	156.56	269.85	200.22	162.62	206.24
0.98	242.87	173.04	135.77	178.69	243.82	174.27	136.69	180.38
0.99	265.85	196.77	159.98	202.38	219.77	150.97	113.87	157.04
1.00	289.79	222.00	186.13	227.51	197.62	130.12	93.93	136.08
1.01	314.62	248.62	214.18	253.97	177.28	111.56	76.80	117.37
1.02	340.28	276.50	243.87	281.64	158.65	95.15	62.21	100.75
1.03	366.70	305.49	275.06	310.33	141.64	80.72	49.97	86.02
1.04	393.81	335.52	307.51	340.01	126.16	68.15	39.83	73.10
1.05	421.54	366.49	340.96	370.61	112.10	57.33	31.49	61.91
1.06	449.81	398.15	375.13	401.91	99.38	48.00	24.67	52.22
1.07	478.56	430.40	409.86	503.90	87.90	40.02	19.17	29.64
1.08	507.71	463.08	444.95	466.08	77.57	33.22	14.77	36.69
1.09	537.20	496.02	480.22	498.70	68.30	27.41	11.29	30.54
1.10	566.96	529.17	515.50	531.54	60.00	22.51	8.52	25.33

Notes: B-S stands for Black-Scholes model, Gauss for model with Gaussian kernel, Epane for model with Epanechnikov kernel, and Biwei for model with Biweight kernel. X is exercise price, $d = 0.03$ is dividend yield, and $P_0 = 4{,}796.56$ is the value of the S&P Index on 3 January 2022.

References

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654.

Blattberg, R. C. and Gonedes, N. J. (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *Journal of Business*, 47:244–280.

Brooks, C. (2019). *Introductory Econometrics for Finance*. Cambridge University Press, Cambridge, 4th edition.

Castanias, R. P. (1979). Macroinformation and the Variability of Stock Market Prices. *The Journal of Finance*, 34(2):439–450.

Christie, A. A. (1982). The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics*, 10(4):407–432.

Cont, R. (2007). Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. In Teyssière, G. and Kirman, A. P., editors, *Long Memory in Economics*, pages 289–309. Springer.

Epanechnikov, V. A. (1969). Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & Its Applications*, 14(1):153–158.

Fan, J. (2005). A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science*, 20(4):317–337.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.

Ghosh, S. (2018). *Kernel Smoothing: Principles, Methods and Applications*. John Wiley & Sons.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.

Henderson, D. J. and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*. Cambridge University Press, Cambridge.

Horowitz, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer-Verlag, New York.

Hull, J. C. (2018). *Options, Futures, and Other Derivatives*. Prentice Hall, Upper Saddle River, 9 edition.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

MacBeth, J. D. and Merville, L. J. (1979). An Empirical Examination of the Black-Scholes Call Option Pricing Formula. *The Journal of Finance*, 34(5):1173–1186.

Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and Its Applications*, 9(1):141–142.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics*, 27(3):832–837.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Ullah, A. and Pagan, A. (1999). *Nonparametric Econometrics.* Cambridge University Press, Cambridge.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372.