

# Option Prices and Pricing Theory: Combining Financial Mathematics with Statistical Modeling

Ling Chen

Securities Division, Goldman Sachs

Tze Leung Lai

Department of Statistics, Stanford University

Tiong Wee Lim

Department of Statistics and Applied Probability, National University of Singapore

## Keywords

Option pricing, Substantive models, Nonparametric regression, Semiparametric regression, Time series modeling

## Abstract

After an overview of important developments of option pricing theory, this article describes statistical approaches to modeling the difference between the theoretical and actual prices. An empirical study is given to compare various approaches.

A cornerstone of Financial Mathematics is option pricing theory, which Ross<sup>1</sup> has described as “the most successful theory not only in finance, but in all of economics.” A *call (put)* option gives the holder the right to buy (sell) the underlying asset by a certain expiration date  $T$  at a certain price  $K$ , which is called the strike price. If the option can be exercised only at  $T$ , it is termed *European* or European-style, whereas if early exercise is also allowed at any time before  $T$ , the option is termed *American* or American-style. In the seminal work of Black and Scholes,<sup>2</sup> a dynamic hedging argument was used in conjunction with a no-arbitrage condition to derive closed-form pricing formulas for European options written on an asset whose price  $S_t$  at time  $t$  is modeled by a geometric Brownian motion

$$S_t = S_0 \exp\{(\mu - \sigma^2/2)t + \sigma B_t\}, \quad (1)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of the asset's return  $dS_t/S_t$  and  $\{B_t, t \geq 0\}$  is a standard Brownian motion;  $\sigma$  is commonly known as the asset's volatility. Merton<sup>3</sup> extended the Black-Scholes theory to American options, showing that optimal exercise of the option occurs when the asset price exceeds (or falls below) an exercise boundary for a call (or put) option. The Black-Scholes-Merton theory for pricing and hedging options has played a fundamental role in the development of financial derivatives; a derivative is a financial instrument having a value derived from or contingent on the values of more basic underlying variables. The European and American call and put options are “plain vanilla” products that are actively traded on many exchanges throughout the world, e.g., the Chicago Board Options Exchange that started trading options contracts in 1973. One can therefore compare the actual option prices with those given by the Black-Scholes-Merton formula that involves the price of the underlying asset and the risk-free interest rate, which are directly observable, and the volatility  $\sigma$  of the asset's return, which has to be estimated from past data. Discrepancies between the theoretical and actual prices suggest possible misspecification of the asset price dynamics via (1). One way to address the issue is to develop more flexible (and increasingly complex) stochastic models which incorporate stochastic volatility, stochastic interest rate, random jumps, and models incorporating both stochastic volatility and contemporaneous jumps (SVCJ) in prices and volatilities. This approach has an extensive literature and is reviewed in the next section, which is followed by a section that reviews the nonparametric approach of Hutchinson et al.<sup>4</sup> for European options and Broadie et al.<sup>5</sup> for American options.

A semiparametric approach, which combines the parametric pricing formulas given by the Black-Scholes-Merton theory with nonparametric regression applied to the discrepancies between the theoretical and

observed option prices, was introduced by Lai and Wong,<sup>6</sup> who also performed simulation studies showing that this semiparametric approach provides good approximations to the true underlying option prices. An empirical study of its performance is given in this article, and the findings show that its improvement in predictive (out-of-sample) performance over the Black-Scholes pricing formula is limited to stable periods in which the future is similar to the past. The success of a statistical learning approach relies critically on the stationarity of the data to which it is applied. A nonparametric or semiparametric approach works best when the underlying distribution and characteristics of the data in the training period can be well extended to the data in the test period. While this is the case in typical simulation studies, the market environment may change substantially from the start of the training period till the end of the test period for an empirical study in which six-month training and test subperiods are used. We describe herein a simple modification of the semiparametric approach by using time series modeling, instead of nonparametric regression, of the discrepancies between the actual and theoretical option prices.

## From Black-Scholes to SVCJ models for option pricing

Assuming that the asset price  $S_t$  follows a geometric Brownian motion, that the market has a risk-free asset with constant interest rate  $r$ , that short selling is allowed, and that continuous trading of the perfectly divisible asset can occur with no transaction costs, Black and Scholes<sup>2</sup> use the absence of arbitrage to show that the price of a call or put option satisfies a partial differential equation (PDE) whose solution is a function of the time  $u = T - t$  to expiration and  $R_t = S_t/K$ :

$$P^{\text{bs}}(u, R) = \begin{cases} K \{ Re^{-qu} \Phi(d(u, R)) - e^{-ru} \Phi(d(u, R) - \sigma\sqrt{u}) \} & \text{for call,} \\ K \{ e^{-ru} \Phi(\sigma\sqrt{u} - d(u, R)) - Re^{-qu} \Phi(-d(u, R)) \} & \text{for put,} \end{cases} \quad (2)$$

where  $q$  is the dividend rate paid by the underlying asset,  $\Phi$  is the standard normal distribution function and

$$d(u, R) = \{\log R + (r - q)u\}/(\sigma\sqrt{u}) + \sigma\sqrt{u}/2. \quad (3)$$

Merton<sup>3</sup> extended the Black-Scholes theory for pricing European options to American options. Optimal exercise of the option occurs when the asset price exceeds or falls below an exercise boundary  $\partial\mathcal{C}$  for a call or put option, respectively. The Black-Scholes PDE still holds in the continuation region  $\mathcal{C}$  of  $(t, S_t)$  before exercise, and  $\partial\mathcal{C}$  is determined by the *free boundary condition*  $\partial P/\partial S = 1$  (or  $-1$ ) for a call (or put) option. Unlike the explicit formula (2) for European options, there is no closed-form solution of the free-boundary PDE, and numerical methods such as finite differences are needed to compute American option prices under this theory. The free-boundary PDE can also be represented probabilistically as the value function of the optimal stopping problem

$$P^{\text{a}}(u, R) = K \sup_{\tau \in \mathcal{T}_{T-u, T}} E[e^{-r(\tau - (T-u))} g(R_\tau) | R_t = R], \quad (4)$$

where  $g(R) = (R - 1)_+$  or  $(1 - R)_+$  for the call or put, with  $x_+ = \max\{0, x\}$ ,  $\mathcal{T}_{t, T}$  is the set of stopping times whose values are between  $t$  and  $T$ , and  $E$  is expectation with respect to the *risk-neutral measure* under which  $\mu = r - q$  in (1).

The interest rate  $r$  in the Black-Scholes formula (2) for the price of a European option is usually taken to be the yield of a short-maturity Treasury bill at the time when the contract is initiated. The parameter in (2) that cannot be directly observed is  $\sigma$ . Equating (2) to the actual price of the option yields a nonlinear equation in  $\sigma$  whose solution is called the “implied volatility” of the underlying asset. The implied volatilities computed from call and put options with the same strike price  $K$  and time to maturity  $u$  should be equal because the put-call parity relationship (call price  $-$  put price  $= S_t e^{-qu} - K e^{-ru}$ ) holds for both the Black-Scholes price pair in (2) and the corresponding market price pair. A call option, whose payoff function is  $(S - K)_+$ , is said to be *in the money*, *at the money*, or *out of the money* according to whether  $S_t > K$ ,  $S_t = K$ , or  $S_t < K$ , respectively. Puts have the reverse terminology since the payoff function is  $(K - S)_+$ . According to the Black-Scholes theory, the  $\sigma$  in (2) is the volatility of the underlying asset and therefore does not vary with  $K$  and  $T$ . However, for some equity options, a “volatility skew” is observed (i.e., the implied volatility is a decreasing function of the strike price  $K$ ). The “volatility smile” is common in foreign

currency options, for which the implied volatility is relatively low for at-the-money options and becomes higher as the option moves into the money or out of the money, giving the smile shape of the implied volatility curve as a function of  $K$  (with minimum around  $K = S_t$ ). Moreover, implied volatilities also tend to vary with time to maturity. These volatility smiles and skews show that there are marked discrepancies between theoretical and observed prices, and an important area of financial mathematics consists of more realistic models of asset price dynamics than (1) that still have tractable option pricing formulas.

In particular, the *implied volatility* (also called the *implied tree*) model provides an exact fit to all European option prices on any given day. It assumes that the risk-neutral process of the asset price has the more general form

$$dS_t = (r_t - q_t)S_t dt + \sigma(t, S_t)S_t dB_t \quad (5)$$

rather than that in (1) with  $r_t = r$ ,  $q_t = q$  and  $\sigma(t, S_t) = \sigma$ . Dupire<sup>7</sup> has shown that the function  $\sigma(t, S)$  is given analytically by

$$\frac{\sigma^2(T, K)}{2} = \left\{ \frac{\partial c_T}{\partial T} + q_T c_T + K(r_T - q_T) \frac{\partial c_T}{\partial K} \right\} / K^2 \frac{\partial^2 c_T}{\partial K^2}, \quad (6)$$

where  $c_T$  is the market price of a European call option with strike price  $K$  and maturity  $T$ . To price a European option under the risk-neutral model (5), Andersen and Brotherton-Ratcliffe<sup>8</sup> use finite difference approximations of (6) to recalibrate the model (5) daily to the market prices of standard European options. An alternative approach, proposed by Derman and Kani<sup>9</sup> and Rubinstein,<sup>10</sup> approximates (5) by an implied tree, which is a discrete-time Markov chain approximation to (5) in the form of a binomial tree that is recalibrated daily to the market prices of standard options.

Whereas the implied volatility model changes the constant  $\sigma$  in the Black-Scholes theory by a function  $\sigma(t, S_t)$ , the constant elasticity of variance (CEV) model replaces  $\sigma$  by  $\sigma S_t^\alpha$ , imposing an additional parameter for the Black-Scholes model. Specifically, the risk-neutral process of the asset price follows the CEV model  $dS_t = (r - q)S_t dt + \sigma S_t^\alpha dB_t$  introduced by Cox and Ross,<sup>11</sup> who show that the formula (2) can be modified by replacing the standard normal distribution functions by the distribution functions of certain noncentral  $\chi^2$ -distributions. Another approach is to replace  $\sigma$  by a stochastic process  $\sigma_t$ . In particular, the *stochastic volatility* (SV) model under the risk-neutral measure is  $dS_t/S_t = (r - q) dt + \sigma_t dB_t$ , in which  $v_t := \sigma_t^2$  is modeled by

$$dv_t = \alpha(v^* - v_t) dt + \beta v_t^\xi d\tilde{B}_t, \quad (7)$$

where  $\tilde{B}_t$  is Brownian motion that is independent of  $B_t$ . For this SV model, Hull and White<sup>12</sup> have shown that the price of a European option is given by  $\int_0^\infty b(w) f(w) dw$ , where  $b(w)$  is the Black-Scholes price in which  $\sigma$  is replaced by  $w$ , and  $w$  is the average variance rate during the life of the option, which is a random variable with density function  $f$  determined by the stochastic dynamics (7) for  $v_t$ . Although there is no analytic formula for  $f$ , Hull and White<sup>12</sup> have used this representation of the option price to develop closed-form approximations to the model price. The parameters  $\alpha$ ,  $\beta$ ,  $v^*$ , and  $\xi$  in (7) can be estimated by minimizing the sum of squared differences between the model prices and the market prices. The SV model has been used to account for the volatility smile associated with the Black-Scholes prices.

Stochastic volatility has paved the way for increasingly complex stochastic models which incorporate stochastic volatility, stochastic interest rate (SI), and random jumps (J). A representative list of such SVSI-J models is provided by Bakshi et al.,<sup>13</sup> who compare these alternative parametric models on the basis of internal consistency of implied parameters/volatility with relevant time-series data, out-of-sample pricing, and hedging. More recently, Broadie et al.<sup>14</sup> have developed a test to detect jumps in volatility and find that whereas the tests reject the stochastic volatility model and an extension with jumps in prices, models that admit stochastic volatility with *contemporaneous jumps* in volatility and prices, the so-called SVCJ models introduced by Duffie et al.,<sup>15</sup> Eraker et al.,<sup>16</sup> and Chernov et al.,<sup>17</sup> easily pass these tests.

## Nonparametric methods for option pricing

The methods reviewed in the preceding section basically replace the geometric Brownian motion (1) in the Black-Scholes theory by other stochastic models for the asset price  $S_t$  under the risk-neutral measure to account for the

observed implied volatility smile, which is incompatible with the Black-Scholes theory. Instead of specifying a particular model for  $S_t$ , Hutchinson et al.<sup>4</sup> propose to use a nonparametric model that only requires  $S_t$  to have independent increments. Noting that the option price  $P_t$  is a function of  $u = T - t$  and  $R_t = S_t/K$  with  $r$  and  $\sigma$  being constant, they assume  $P_t = Kf(u, R_t)$  and approximate  $f$  by taking the predictor variables to be the vector  $(u, R_t)^T$  in one of the following nonparametric regression models: (i) projection pursuit regression, (ii) neural networks, and (iii) radial basis networks. While the transformation of  $S_t$  to  $R_t$  above can be motivated from the assumption on  $S_t$ , choosing a regression function of the form  $f(u, R)$  has the advantage of circumventing the problem associated with sparsity of data in the space of  $(u, S_t, K_t)$  and the consequent difficulty in making good predictions by making use of the fact that all observed and future  $S_t/K_t$  are close to 1. Another point highlighted by Hutchinson et al.<sup>4</sup> is the measure of performance of the estimated pricing formula. According to their simulation study, even a linear function  $f(u, R)$  can give  $R^2 \approx 90\%$  (see their Table I). However, such a linear  $f$  implies a constant delta hedging scheme, which gives poor hedging results. Since the primary function of options is to hedge the risk created by changes in the price of the underlying asset, Hutchinson et al.<sup>4</sup> suggest using, instead of  $R^2$ , the hedging error measures  $\xi = e^{-rT}E[|V(T)|]$  and  $\eta = e^{-rT}[EV^2(T)]^{1/2}$ , where  $V(T)$  is the value of the hedged portfolio at the expiration date  $T$ . If all the Black-Scholes assumptions hold,  $V(T)$  should be 0 when one uses the Black-Scholes formula. Hutchinson et al.<sup>4</sup> report that the nonparametric delta gives values for the hedging error measures comparable to those for the Black-Scholes delta in their simulation study, which makes use of the Black-Scholes assumptions with the exception that time is discrete (for daily closing prices) so that daily rebalancing with the Black-Scholes delta still gives  $\xi > 0$  and  $\eta > 0$ . They have also carried out an empirical study of out-of-sample hedging performance for S&P 500 futures options from January 1987 to December 1991 and report that the nonparametric delta performs better than the Black-Scholes delta.

For American vanilla options, Broadie et al.<sup>5</sup> used kernel smoothers to estimate the option pricing formula of an American option. Using a training sample of daily closing prices of American calls on the S&P 100 index from January 3, 1984 to March 30, 1990, they compared the nonparametric estimates of American call option prices at a set of  $(S_t/K, T - t)$  values with those obtained by approximations to (4) due to Broadie and Detemple<sup>18</sup> and reported significant differences between the parametric and nonparametric estimates.

## A semiparametric regression approach

For European call or put options, instead of using nonparametric modeling of  $f(u, R)$  as in Hutchinson et al.,<sup>4</sup> an alternative approach is to express the option price as  $P^{bs} + Ke^{-ru}f(u, R)$ , where  $P^{bs}$  is the Black-Scholes price (2), because the Black-Scholes formula has been widely used by option traders. This is tantamount to including  $P^{bs}(u, R)$  as one of the basis functions (with prescribed weight 1) to come up with a more parsimonious approximation to the actual option price. The usefulness of this idea is even more apparent in the case of American options, which do not have explicit pricing and delta-hedging formulas even under the Black-Scholes assumptions. Because one can express an American call (when  $q > 0$ ) or put option price  $P^a$  as the sum of a European call or put price  $P^{bs}$  and the early exercise premium (“decomposition formula”), which is typically small relative to  $P^{bs}$ , this suggests that  $P^{bs}$  should be included as one of the basis functions (with prescribed weight 1). Lai and Wong<sup>6</sup> propose to use additive regression splines after the change of variables  $s = -\sigma^2 u$  and  $x = \log R$ . Specifically, for small  $u$  (say within 5 trading days prior to expiration; i.e.,  $u = T - t = 5/253$  under the assumption of 253 trading days per year), they approximate  $P^a$  by  $P^{bs}$ . For  $u > 5/253$  (or equivalently  $s < -5\sigma^2/253$ ), they approximate  $P^a$  by

$$P^a = P^{bs} + Ke^{\rho s} \left\{ \alpha + \alpha_1 s + \sum_{j=1}^{J_s} \alpha_{1+j} (s - s^{(j)})_+ + \beta_1 x + \beta_2 x^2 + \sum_{j=1}^{J_x} \beta_{2+j} (x - x^{(j)})_+^2 + \gamma_1 w + \gamma_2 w^2 + \sum_{j=1}^{J_w} \gamma_{2+j} (w - w^{(j)})_+^2 \right\}, \quad (8)$$

where  $\rho = r/\sigma^2$ ,  $w = |s|^{-1/2}x - (\rho - \theta\rho - 1/2)s$  (with  $\theta = q/r$ ) is an “interaction” variable derived from  $s$  and  $x$ , and  $\alpha$ ,  $\alpha_j$ ,  $\beta_j$ , and  $\gamma_j$  are regression parameters to be estimated by least squares from the training sample. The motivation behind the centering term  $(\rho - \theta\rho - 1/2)s$  comes from the transformation of geometric Brownian motion into Brownian motion, whereas that behind the normalization  $|s|^{-1/2}$  comes from the formula for  $d$  in (3). The knots

$s^{(j)}$  (resp.  $x^{(j)}$  or  $w^{(j)}$ ) of the linear (resp. quadratic) spline in (8) are the  $j/J_s$ -th (resp.  $j/J_x$ -th or  $j/J_w$ -th) quantiles of  $s_1, \dots, s_n$  (resp.  $x_1, \dots, x_n$  or  $w_1, \dots, w_n$ ). The choice of  $J_s$ ,  $J_x$  and  $J_w$  is over all possible integers between 1 and 10 to minimize the generalized cross-validation (GCV) criterion, which can be expressed as

$$\text{GCV}(J_s, J_x, J_w) = \sum_{i=1}^n (P_i^{\text{obs}} - \hat{P}_i)^2 \Big/ \left\{ n \left( 1 - \frac{J_s + J_x + J_w + 6}{n} \right)^2 \right\},$$

where the  $P_i^{\text{obs}}$  are the observed American option prices in the past  $n$  periods and the  $\hat{P}_i$  are the corresponding fitted values given by (8), in which the regression coefficients are estimated by least squares.

In the preceding, we have assumed prescribed constants  $r$  and  $\sigma$  as in the Black-Scholes model; these parameters appear in (8) as  $\rho = r/\sigma^2$ . In practice,  $\sigma$  is unknown and may also vary with time. Lai and Wong<sup>6</sup> replace it in (8) by the standard deviation  $\hat{\sigma}_t$  of the most recent asset prices, say, during the past 60 trading days prior to  $t$  as in Hutchinson et al.<sup>4</sup> (p. 881). This is tantamount to incorporating the asset prices  $S_{t-1}, \dots, S_{t-60}$  in the formula  $P^a(t, S; S_{t-1}, \dots, S_{t-60})$  with  $S_t = S$ . Moreover, the risk-free rate  $r$  may also change with time and can be replaced by the yield  $\hat{r}_t$  of a short-maturity Treasury bill on the close of the month before  $t$ . The same remark also applies to the dividend rate. Lai and Wong<sup>6</sup> report a simulation study of the performance, measured by  $E|\hat{P} - P|$  and  $E\{e^{-r\tau}|V(\tau)|\}$  (which corresponds to changing  $T$  to the early exercise time  $\tau$  in the measure proposed by Hutchinson et al.<sup>4</sup> for European options), of their pricing formula  $\hat{P}$  and the Black-Scholes-Merton decomposition formula when the portfolio is rebalanced daily, where  $P$  is the true option price. The simulation study shows that whereas the decomposition formula can perform quite poorly when the Black-Scholes assumptions are violated,  $\hat{P}$  does not encounter such difficulties.

## An empirical study

The semiparametric approach of Lai and Wong<sup>6</sup> has not been tested on empirical data. Here we evaluate its performance by using daily settlement prices of S&P 500 futures and futures options (obtained from the Chicago Mercantile Exchange) for the 22-year period from January 1987 to December 2008. We use the settlement prices rather than closing prices as the former are used to calculate gains and losses in market accounts. S&P 500 futures options are among the most actively traded options available in the market. They have been used in empirical studies of option prices and pricing formulas by Hutchinson et al.,<sup>4</sup> Bates,<sup>19</sup> Broadie et al.,<sup>14</sup> and a number of other authors. Since both the futures and futures options are traded at the same location, the issue of nonsynchronous underlying asset price and option price does not arise here as in the case of S&P 500 index options, which are used by Bakshi et al.<sup>13</sup> and others. For options that expire in the March quarterly cycle (i.e., March, June, September, and December), the underlying futures contract is the futures contract for the month in which the option expires. For options (first introduced in September 1987) that expire in months other than those in the March quarterly cycle, the underlying futures contract is the next futures contract in the March quarterly cycle that is nearest the expiration of the option. Figure 1 displays the S&P 500 futures price history for the 22-year period from December 19, 1986, to December 18, 2008, the endpoints being option expiration dates. For each day, the futures contract that is closest to its expiration is chosen as the representative futures contract, whose price is then plotted. The vertical axis is in logarithmic scale, so that the vertical distance between any two points on the path reflects the percentage change in futures price. As long-dated, deep in-the-money and deep out-of-the-money options are often thinly traded and their price quotes are generally not supported by actual trades, we apply two exclusionary filters to the data. First, we eliminate data points for which the time to expiration is more than half a year; second, we also eliminate data points for which the absolute moneyness  $|S/K - 1|$  is greater than 10%.

The S&P 500 data are divided into 44 nonoverlapping six-month subperiods. After applying the two exclusionary filters, the number of options per subperiod ranges from 38 to 467, with an average of 242; the total number of data points for each subperiod ranges from 2,573 to 21,170, with an average of 11,280. Each six-month subperiod, except for the last one, is used as a training sample to estimate the parameters of the pricing formula to be used in the next six-month subperiod as a test sample; there are therefore 43 test samples.

INSERT FIGURE 1 ABOUT HERE

## Deviations from Black-Scholes prices and their time series analysis

Whereas the semiparametric approach uses regression splines in  $u$  and  $R$  to fit the deviations of the observed option prices from the theoretical values given by the Black-Scholes formula, an alternative approach is to model the time series of these deviations  $\epsilon_t = P_t^{\text{obs}} - P_t^{\text{bs}}$  to predict future deviations from the past ones, where  $P_t^{\text{obs}}$  is the actual option price at time  $t$ .

The two Black-Scholes parameters, namely risk-free interest rate and volatility, need to be estimated and substituted into the Black-Scholes formula to calculate option prices. We approximate the risk-free rate  $r$  by the yield of the 3-month Treasury bill (obtainable as a secondary market rate in the Federal Reserve Board H.15 publication). We distinguish between three approaches for estimating the volatility  $\sigma$  of a given S&P 500 futures contract. Hutchinson et al.<sup>4</sup> and Lai and Wong<sup>6</sup> propose to estimate  $\sigma$  using only a window of the most recent data via the historic volatility given by  $\sigma_H = s/\sqrt{60}$ , where  $s$  is the standard deviation of the 60 most recent continuously compounded daily returns of the contract. The solid green curve in Figure 2 displays the historic volatility for the 22-year period in our study. Whereas  $\sigma_H$  is a backward-looking estimate of  $\sigma$ , practitioners sometimes prefer a forward-looking estimate which incorporates the expectation of market participants about future volatility. A commonly used forward-looking estimate is the option's implied volatility  $\sigma_I$ , which is defined as the value of volatility that equates the Black-Scholes option price to the actual market option price, i.e.,  $\sigma_I$  satisfies  $P_t^{\text{obs}} = P^{\text{bs}}(u, R; \sigma_I)$ . Another forward-looking estimate of  $\sigma$  is the one that calibrates to the daily cross-section of observed option prices, instead of to each individual option as the implied volatility does. This ‘‘calibrated volatility’’  $\sigma_C$  is defined as the minimizer of the sum of squared differences between the market option prices and the corresponding Black-Scholes prices:

$$\sigma_C = \arg \min_{\sigma \geq 0.05} \sum_{i=1}^n [P_i^{\text{obs}} - P^{\text{bs}}(u_i, R_i; \sigma)]^2,$$

where the summation is taken over all options in the cross-section, indexed by  $i$ , and a minimum value of 0.05 is imposed to prevent negative values in the estimated volatility; see Dumas et al.<sup>20</sup> For both  $\sigma_I$  and  $\sigma_C$ , the value from the previous day is used to estimate  $\sigma$  for the present day to be substituted into the Black-Scholes formula. In Figure 2, the dashed red curve and the solid black curve correspond to the implied volatility and calibrated volatility, respectively, for S&P 500 futures options over the 22-year period in our study. These two curves lie quite close to each other; a large proportion of the two curves are not even distinguishable under the displayed resolution. However, the implied volatility may not exist and is plotted only where it exists, and appears more volatile when large price changes occur. For example,  $\sigma_I$  shoots up to around 90% during the stock market crash in October 1987, while the peak of  $\sigma_C$  is only around 60% at that time. This is not surprising as the latter is smoothed out by calibrating to the daily cross-section of call option prices instead of each individual option price. Both  $\sigma_I$  and  $\sigma_C$  show significant deviation from the historic volatility curve.

INSERT FIGURE 2 ABOUT HERE

To study deviations of the observed options prices from  $P_t^{\text{bs}}$ , we begin by substituting  $\sigma$  in  $P_t^{\text{bs}}$  by the historic volatility  $\sigma_H$ . Denote the set of trading days in the  $j$ th six-month subperiod by  $T_j$ ,  $1 \leq j \leq 44$ , and let  $\rho_j$  be the lag-1 autocorrelation (i.e., correlation coefficient between  $\epsilon_t$  and  $\epsilon_{t-1}$ ) of the time series  $\{\epsilon_t, t \in T_j\}$ . The lag-1 autocorrelations  $\rho_1, \dots, \rho_{44}$  range between 0.8 and 0.99, with median 0.97 and first quartile 0.96. These high autocorrelations suggest the forecast  $\epsilon_{t-1}$  of  $\epsilon_t$  at time  $t - 1$  before  $P_t$  is observed, leading to the lag-1 correction of the Black-Scholes price if  $\sigma = \sigma_H$  is used in  $P_t^{\text{bs}}$ :

$$\hat{P}_t = P_t^{\text{bs}} + \epsilon_{t-1} = P_t^{\text{bs}} + (P_{t-1}^{\text{obs}} - P_{t-1}^{\text{bs}}). \quad (9)$$

When  $\sigma_C$  is used in  $P_t^{\text{bs}}$ , the corresponding autocorrelations are considerably lower, ranging from 0.13 to 0.99. Instead of using the simple lag-1 forecast  $\epsilon_{t-1}$  of  $\epsilon_t$ , we fit an AR(1) model to  $\{\epsilon_t, t \in T_j\}$  for the  $j$ th subperiod and use the forecast  $\hat{\epsilon}_t = \hat{\alpha} + \hat{\beta}\epsilon_{t-1}$ , in which  $\hat{\alpha}$  and  $\hat{\beta}$  are estimated from the time series  $\{\epsilon_t, t \in T_j\}$ , leading to the AR(1) correction of the Black-Scholes price if  $\sigma = \sigma_C$  is used in  $P_t^{\text{bs}}$  for  $t \in T_{j+1}$ :

$$\hat{P}_t = P_t^{\text{bs}} + \hat{\alpha} + \hat{\beta}\epsilon_{t-1}. \quad (10)$$

In this way,  $T_j$  is used as a training sample while  $T_{j+1}$  as a test sample. As we have already noted, the equation  $P_t^{\text{obs}} = P^{\text{bs}}(u, R; \sigma)$  does not always have a solution  $\sigma = \sigma_1$  for the S&P 500 futures option data. This is in theory impossible. The Black-Scholes formula (2) is an increasing continuous function of  $\sigma$ , whose lower limit is the no-arbitrage lower bound of the option price implied from the put-call parity as  $\sigma$  approaches 0 and whose upper limit is the underlying asset price as  $\sigma$  approaches infinity. Therefore, in the absence of arbitrage opportunities, the implied volatility  $\sigma_1$  should always exist. However, several factors may cause the implied volatility not to exist in practice. First, the observed option price is not supported by an actual trade so that the price quote is out-of-date; this can happen for thinly traded long-dated options or options far from the money. Moreover, the option price and the underlying asset price are not quoted concurrently, which actually happens all the time, as the two instruments are traded independently in different markets. Furthermore, the risk-free rate  $r$  and hence the lower limit of the Black-Scholes price,  $S - e^{-ru}K$ , can be overestimated, particularly for options with less than three months to expire as the three-month Treasury bill rate is used as a proxy for  $r$ . When the implied volatility  $\sigma_1$  does not exist for an option, we use the calibrated volatility  $\sigma_C$  of the same day as a proxy for  $\sigma_1$ . When  $\sigma = \sigma_1$  (or its proxy  $\sigma_C$  if  $\sigma_1$  does not exist) is used in  $P_t^{\text{bs}}$ , we again use the AR(1) correction (10), in which  $P_t^{\text{bs}}$  and therefore  $\epsilon_{t-1}$  also are based on  $\sigma = \sigma_1$ .

## Out-of-sample pricing errors

As we have already noted, each of the 44 six-month subperiods from 1987 to 2008, except the last one, is used as a training sample to determine the semiparametric pricing formula for performance evaluation in the subsequent period. Two semiparametric pricing formulas are considered. The first is that introduced in the preceding section and uses time series modeling of the deviations  $\epsilon_t$  from the Black-Scholes prices  $P_t^{\text{bs}}$ . The second, proposed by Lai and Wong,<sup>6</sup> uses nonparametric regression of  $\epsilon_t$  on moneyness and time to expiration. For each test period  $j = 2, \dots, 44$ , we evaluate the root-mean squared error

$$\text{RMSE} = \left\{ \frac{1}{\#_j} \sum_{m \in \mathcal{M}_j} (P_m^{\text{obs}} - \hat{P}_m)^2 \right\}^{1/2} \quad (11)$$

of the pricing formula that gives  $\hat{P}_m$ , in which  $\#_j$  denotes the total number of options summed over the trading days in the  $j$ th period and  $\mathcal{M}_j$  denotes the corresponding set of option-day pairs  $m$ . We have modified somewhat the original proposal (8) of Lai and Wong<sup>6</sup> for American options and use the pricing formula

$$\tilde{P}(v, z) = P^{\text{bs}}(v/\sigma^2, e^{z+\rho v}) + Ke^{-\rho v} \{a_0 + f_1(v) + f_2(z)\}, \quad (12)$$

where  $v = -s$ ,  $z = x - \rho v$ ,  $f_1(v) = a_1 v + a_2 v^2 + a_3 v^3 + \sum_{i=1}^{I_v} a_{3+i} (v - v^{(i)})_+^3$  and  $f_2(z) = b_1 z + b_2 z^2 + b_3 z^3 + \sum_{i=1}^{I_z} b_{3+i} (z - z^{(i)})_+^3$  are cubic splines. Here  $a_k$  and  $b_k$  are regression parameters that can be estimated from the training sample by ordinary least squares. The knots  $v^{(i)}$  (or  $z^{(j)}$ ) are chosen to be equally spaced percentiles of the observations  $v_t$  (or  $z_t$ ) in the training sample. For the  $j$ th training sample, the number of knots,  $I_v$  and  $I_z$  respectively, is chosen to minimize the generalized cross-validation criterion

$$\text{GCV}(I_v, I_z) = \frac{1}{\#_j} \sum_{m \in \mathcal{M}_j} \left( \frac{P_m^{\text{obs}} - \tilde{P}_m}{1 - (I_v + I_z + 7)/\#_j} \right)^2$$

under the constraint  $\max\{I_v, I_z\} \leq 10$ .

Besides the much lower computational complexity of using ordinary least squares and the simple GCV criterion, another advantage of this semiparametric approach is that the time variation of interest rate  $r$  and volatility  $\sigma$  is incorporated by the inclusion of the Black-Scholes price as a basis function and the choice of the predictor variables  $v$  and  $z$ . Strictly speaking, the estimated pricing formula can only be used inside the convex hull  $\mathcal{H} = \{(v, z) : 0 \leq v \leq v_{\max}, \underline{z}(v) \leq z \leq \bar{z}(v)\}$  of the training sample  $\mathcal{M}$ , where  $v_{\max} = \max\{v : v \in \mathcal{M}\}$ , and  $\underline{z}(\cdot)$  and  $\bar{z}(\cdot)$  are the lower and upper boundaries of the convex hull, respectively. Thus, we set  $\hat{P} = \tilde{P}$  for  $(v, z) \in \mathcal{H}$ . For  $(v, z)$  outside  $\mathcal{H}$ , we use the following method to do safer extrapolation and correct the potential erratic behavior of

polynomial basis functions near or outside the data boundaries. For a new data point  $(v, z)$  with  $v > v_{\max}$ , we simply use  $\hat{P}(v, z) = P^{\text{bs}}(v/\sigma^2, e^{z+\rho v})$ . For  $0 \leq v \leq v_{\max}$  and  $z \notin [\underline{z}(v), \bar{z}(v)]$ , we use

$$\hat{P}(v, z) = \begin{cases} \underline{w}(v)\tilde{P}(v, z) + (1 - \underline{w}(v))P^{\text{bs}}(v/\sigma^2, e^{z+\rho v}) & \text{if } z < \underline{z}(v), \\ \bar{w}(v)\tilde{P}(v, z) + (1 - \bar{w}(v))P^{\text{bs}}(v/\sigma^2, e^{z+\rho v}) & \text{if } z > \bar{z}(v), \end{cases} \quad (13)$$

where  $\tilde{P}(\cdot, \cdot)$  is the given by (12),  $\underline{w}(v) = e^{z-\underline{z}(v)}$  and  $\bar{w}(v) = e^{\bar{z}(v)-z}$ .

Table 1 gives for each of the 43 test periods the RMSE, defined in (11), of the estimated option prices given by the lag-1 correction (9) or the AR(1) correction (10), or by the semiparametric (SP) regression correction (13). Also given for comparison is the RMSE of the Black-Scholes (BS) pricing formula. As noted in the preceding section, there are three choices of  $\sigma$  in the Black-Scholes formula:  $\sigma = \sigma_H, \sigma_C$ , or  $\sigma_I$ . Table 1 considers each of these choices.

INSERT TABLE 1 ABOUT HERE.

The semiparametric approach proposed by Lai and Wong<sup>6</sup> only considers the case  $\sigma = \sigma_H$ , and we first focus on this case and its comparison to the Black-Scholes formula. Table 1 shows that out of the 43 test periods, SP gives a smaller RMSE than BS for 29 subperiods, among which only a few show substantially reduced RMSEs. An interesting pattern shown by the table is that large values of RMSE tend to cluster in time. This is consistent with the pattern of clustering of large jumps in the S&P 500 futures price and of large and abruptly changing volatilities, which can be observed in Figures 1 and 2. Making use of the events identified by Broadie et al.<sup>14</sup> in their Figure 1, we can distinguish the “unstable” periods from the stable ones in our empirical study:

1. July 1987 to June 1988, associated with the stock market crash of 1987, when the largest single-day stock market decline (the S&P 500 index lost 20.5%) in history occurred on October 19, 1987;
2. January 1998 to December 2002, associated with a series of events including the Asian currency crises, the Russian default/LTCM, the bursting of the dot-com bubble, 9-11 and the Gulf War II;
3. July 2007 to December 2008, associated with the (recent) subprime mortgage crisis.

Thus, there are 28 stable subperiods and 15 unstable subperiods, which are characterized by sustained large volatilities exceeding the threshold of 25% in Figure 2 (or 30% in Figure 1 of Broadie et al.,<sup>14</sup> where the VIX index was used). Table 2 summarizes the extent to which the RMSE of option prices is higher in the unstable subperiods than in the stable subperiods; the definition (11) of RMSE is now modified to average over all squared errors in the specified period. We note that the improvement in pricing performance of SP over BS is limited to the stable subperiods; both approaches perform comparably in the unstable subperiods. In fact, 21 of the 29 six-month subperiods in which SP outperforms BS belong to the stable regime.

INSERT TABLE 2 ABOUT HERE.

For  $\sigma = \sigma_H$ , unlike SP that only has limited improvement over BS, Table 1 shows that lag-1 correction (9) has marked improvement over BS for every six-month test subperiod. The percentage reduction in RMSE of predicted option prices exceeds 64.7% for every test subperiod except the two in 1999, which turn out to have the smallest lag-1 autocorrelations of  $\epsilon_t$  among the 43 test subperiods; the percentage reduction in RMSE for these two subperiods are 38.5% and 50.5%, respectively. The maximum and mean of the percentage reduction in RMSE are 88.4% and 77.4%, respectively, and the overall percentage reduction in RMSE for all 43 test subperiods combined is 76.8%. In Table 2 which groups the 44 six-month subperiods into two stable periods and three unstable ones, the percentage reduction in RMSE from the uncorrected Black-Scholes price are 80.1% and 80.4% for the two stable periods while for the three unstable periods, the percentage reduction in RMSE are 83.9%, 76.8% and 72.9%.

It is clear from Table 1 that BS with  $\sigma = \sigma_C$  outperforms BS with  $\sigma = \sigma_H$  for every six-month test subperiod and BS with  $\sigma = \sigma_I$  further dominates the former in its pricing performance. For  $\sigma = \sigma_C$ , except for six subperiods with improvements below 20.6%, the AR(1) correction (10) has marked improvement over BS for the remaining



37 subperiods, with percentage reduction in RMSE of at least 29.5%. The maximum and mean of the percentage reduction in RMSE are 83.7% and 45.2%, respectively, and the overall percentage reduction in RMSE for all 43 test subperiods combined is 34.0%. For  $\sigma = \sigma_1$ , the AR(1) correction still improves the uncorrected Black-Scholes price, even though the latter already has good pricing performance. The mean percentage reduction in RMSE is 21.6% whereas the overall percentage reduction in RMSE for all 43 test subperiods combined is 19.0%.

## Conclusion

Cox<sup>21</sup> and Lehmann<sup>22</sup> have described two major types of stochastic models in statistical analysis, namely, *empirical* and *substantive*. Substantive models are explanatory and related to the subject-matter theory on the data generating mechanisms. Empirical models are interpolating and aim at representing the observed data as a realization of a statistical model which is chosen largely for its flexibility and statistical tractability. A combined *substantive-empirical* approach has been introduced by Lai and Wong,<sup>23</sup> in which the substantive component is determined by the underlying theory and the empirical component uses computationally convenient basis functions, such as regression splines, to correct the error in using the substantive model as a first approximation. Lai and Wong<sup>23</sup> have used this approach to re-analyze the extensively studied time series of annual numbers of the Canadian lynx trapped for the period 1821–1934, and have noted that their earlier work on pricing American options is an example of this approach. The empirical study, given in the preceding section, of Lai and Wong's implementation of the substantive-empirical approach by semiparametric regression shows that it should not be used during unstable periods featuring a highly nonstationary market. On the other hand, time series modeling of the empirical component gives excellent pricing performance for S&P 500 futures options, even during unstable periods.

Whereas financial mathematics has developed increasingly complex models of asset price dynamics to address discrepancies between option pricing formulas and option prices, the approach initiated by Lai and Wong<sup>6</sup> addresses these discrepancies by using versatile statistical methods that can greatly reduce these discrepancies. Financial mathematics often focuses on the development of stochastic dynamical models without considering the statistical issues in estimating the model parameters, resulting in increasingly complex models that are increasingly difficult to estimate with time series data. Combining financial mathematics with the principles and advances of statistical modeling circumvents these difficulties and has the advantage of providing simple and yet highly reliable corrections to the widely used pricing formulas of conventional models.

## References

1. Ross SA. Finance. In Eatwell J, Milgate M, and Newman P, editors, *The New Palgrave: A Dictionary of Economics*, volume 2, pages 322–336. New York: Stockton Press, 1987.
2. Black F and Scholes M. The pricing of options and corporate liabilities. *J Political Econom*, 81:637–654, 1973.
3. Merton RC. Theory of rational option pricing. *Bell J Econo Manage Sci*, 4:141–181, 1973.
4. Hutchinson JM, Lo AW, and Poggio T. A nonparametric approach to pricing and hedging derivative securities via learning networks. *J Finance*, 49:851–889, 1994.
5. Broadie M, Detemple J, Ghysels E, and Torrès O. Nonparametric estimation of american options' exercise boundaries and call prices. *J Econom Dynamics Control*, 24:1829–1857, 2000.
6. Lai TL and Wong SPS. Valuation of American options via basis functions. *IEEE Trans Automat Contr*, 49:374–385, 2004.
7. Dupire B. Pricing with a smile. *RISK*, 7:18–20, January 1994.
8. Andersen L and Brotherton-Ratcliffe R. The equity option volatility smile: An implicit finite-difference approach. *J Comput Finance*, 1:5–37, 1997.
9. Derman E and Kani I. Riding on the smile. *RISK*, 7:32–39, February 1994.

10. Rubinstein M. Implied binomial trees. *J Finance*, 49:771–818, 1994.
11. Cox JC and Ross SA. The valuation of options for alternative stochastic processes. *J Financial Econom*, 3:145–166, 1976.
12. Hull J and White A. The pricing of options on assets with stochastic volatilities. *J Finance*, 42:281–300, 1987.
13. Bakshi G, Cao C, and Chen Z. Empirical performance of alternative option pricing models. *J Finance*, 52:2003–2049, 1997.
14. Broadie M, Chernov M, and Johannes M. Model specification and risk premia: Evidence from futures options. *J Finance*, 62:1453–1490, 2007.
15. Duffie D, Pan J, and Singleton K. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376, 2000.
16. Eraker B, Johannes M, and Polson N. The impact of jumps in volatility and returns. *J Finance*, 58:1269–1300, 2003.
17. Chernov M, Gallant AR, Ghysels E, and Tauchen G. Alternative models for stock price dynamics. *J Econometrics*, 116:225–257, 2003.
18. Broadie M and Detemple J. American option valuation: New bounds, approximations, and a comparison of existing methods. *Rev Financial Stud*, 9:1211–1250, 1996.
19. Bates DS. Post-'87 crash fears in the S&P 500 futures option market. *J Econometrics*, 94:181–238, 2000.
20. Dumas B, Fleming J, and Whaley RE. Implied volatility functions: Empirical tests. *J Finance*, 53:2059–2106, 1998.
21. Cox DR. Role of models in statistical analysis. *Statist Sci*, 5:169–174, 1990.
22. Lehmann EL. Model specification: The views of Fisher and Neyman. *Statist Sci*, 5:160–168, 1990.
23. Lai TL and Wong SPS. Combining domain knowledge and statistical models in time series analysis. In Ho HC, Ing CK, and Lai TL, editors, *Time Series and Related Topics*, pages 193–209. Institute of Mathematical Statistics, 2006.

## Further Reading

Hull J. *Options, Futures, and Other Derivatives*. 8th ed. New Jersey: Prentice Hall; 2011.

Cox JC, Ross SA, and Rubinstein M. Option pricing: A simplified approach. *J Financial Econom*, 7:229–263, 1979.

Ju N. Pricing an American option by approximating its early exercise boundary as a multipiece exponential function. *Rev Financial Stud*, 11:627–646, 1998.

AitSahlia F and Lai TL. Exercise boundaries and efficient approximations to American options prices and hedge parameters. *J Comput Finance*, 4:85–103, 2001.

Lai TL and Xing H. *Statistical Models and Methods for Financial Markets*. New York: Springer; 2008.

Table 1 Out-of-sample RMSE of predicted S&P 500 futures option prices given by various pricing methods for the six-month subperiods from June 1987 to December 2008. BS stands for Black-Scholes and SP stands for semiparametric. The mean, standard deviation, minimum and maximum of the out-of-sample RMSE values are taken over all the 43 subperiods.

Subperiod	$\sigma = \sigma_H$			$\sigma = \sigma_C$		$\sigma = \sigma_I$	
	BS	SP	Lag-1	BS	AR(1)	BS	AR(1)
Jun 87 – Dec 87	9.626	9.639	1.556	1.757	1.239	1.445	1.342
Dec 87 – Jun 88	9.253	6.387	1.467	0.758	0.654	0.544	0.610
Jun 88 – Dec 88	1.501	0.973	0.301	0.925	0.524	0.798	0.796
Dec 88 – Jun 89	2.376	1.534	0.276	1.887	0.395	1.805	0.493
Jun 89 – Dec 89	3.547	2.794	0.422	1.897	0.692	1.803	0.721
Dec 89 – Jun 90	2.952	2.142	0.586	1.228	0.558	1.090	0.601
Jun 90 – Dec 90	1.978	2.445	0.413	1.025	0.519	0.799	0.520
Dec 90 – Jun 91	2.601	1.437	0.409	1.267	0.473	1.127	0.539
Jun 91 – Dec 91	1.121	0.951	0.356	0.723	0.456	0.643	0.474
Dec 91 – Jun 92	1.139	0.893	0.281	0.621	0.298	0.514	0.314
Jun 92 – Dec 92	1.069	1.130	0.261	0.412	0.259	0.267	0.249
Dec 92 – Jun 93	0.864	1.111	0.245	0.480	0.264	0.418	0.297
Jun 93 – Dec 93	1.033	0.901	0.235	0.589	0.244	0.530	0.261
Dec 93 – Jun 94	1.208	1.191	0.318	0.597	0.414	0.450	0.392
Jun 94 – Dec 94	1.054	0.976	0.334	0.706	0.403	0.604	0.413
Dec 94 – Jun 95	2.717	1.929	0.334	2.175	0.526	2.137	0.683
Jun 95 – Dec 95	2.466	1.470	0.320	2.190	0.357	2.129	0.416
Dec 95 – Jun 96	3.375	3.291	0.515	1.848	0.710	1.709	0.794
Jun 96 – Dec 96	2.966	2.306	0.644	1.451	0.702	1.289	0.754
Dec 96 – Jun 97	3.816	3.560	0.663	1.968	0.825	1.640	0.934
Jun 97 – Dec 97	5.195	5.287	1.225	2.203	1.448	1.432	1.467
Dec 97 – Jun 98	7.212	6.634	1.432	2.174	1.275	1.715	1.326
Jun 98 – Dec 98	10.182	11.198	2.070	3.637	2.353	1.780	1.865
Dec 98 – Jun 99	4.976	6.915	3.058	3.794	4.057	2.865	2.976
Jun 99 – Dec 99	5.280	5.745	2.615	3.301	3.013	2.483	2.203
Dec 99 – Jun 00	13.087	13.051	2.134	3.202	2.913	1.971	2.191
Jun 00 – Dec 00	10.084	5.709	1.568	2.704	1.505	1.766	1.579
Dec 00 – Jun 01	10.110	5.583	1.560	2.622	1.409	1.272	1.254
Jun 01 – Dec 01	4.100	7.038	1.078	2.384	1.184	0.850	0.850
Dec 01 – Jun 02	2.806	2.459	0.765	1.620	0.830	0.526	0.526
Jun 02 – Dec 02	6.814	6.646	1.478	2.857	1.342	0.916	0.899
Dec 02 – Jun 03	4.149	5.405	0.853	1.894	0.820	0.588	0.587
Jun 03 – Dec 03	4.688	3.643	0.627	1.774	0.665	0.456	0.453
Dec 03 – Jun 04	5.276	2.883	0.725	1.985	0.851	0.559	0.550
Jun 04 – Dec 04	3.178	3.213	0.537	1.534	0.682	0.517	0.501
Dec 04 – Jun 05	1.875	3.511	0.661	0.936	0.851	0.722	0.697
Jun 05 – Dec 05	2.911	2.600	0.709	1.544	0.956	1.430	1.183
Dec 05 – Jun 06	3.172	2.007	0.876	2.637	1.075	2.509	1.141
Jun 06 – Dec 06	5.346	3.978	1.038	3.084	1.135	2.951	1.167
Dec 06 – Jun 07	6.559	3.422	1.303	3.344	1.705	3.202	1.711
Jun 07 – Dec 07	6.098	8.694	1.967	3.706	2.941	1.961	2.598
Dec 07 – Jun 08	6.142	5.790	1.467	3.375	1.989	1.074	1.047
Jun 08 – Dec 08	13.136	13.817	3.484	5.182	3.336	1.979	1.938
Mean	4.629	4.239	1.004	2.000	1.136	1.332	0.984
SD	3.284	3.275	0.790	1.093	0.920	0.777	0.662
Minimum	0.864	0.893	0.235	0.412	0.244	0.267	0.249
Maximum	13.136	13.817	3.484	5.182	4.057	3.202	2.976

Table 2 Out-of-sample RMSE of predicted S&P 500 futures option prices given by various pricing methods for periods separated by the 25% volatility cutoff point in Figure 2. The shaded rows correspond to unstable periods with large volatilities. BS stands for Black-Scholes and SP stands for semiparametric.

Subperiod	$\sigma = \sigma_H$			$\sigma = \sigma_C$		$\sigma = \sigma_I$	
	BS	SP	Lag-1	BS	AR(1)	BS	AR(1)
Jun 87 – Jun 88	9.484	8.532	1.522	1.455	1.053	1.182	1.118
Jun 88 – Dec 97	2.865	2.608	0.569	1.539	0.690	1.331	0.742
Dec 97 – Dec 02	8.280	7.838	1.923	2.937	2.259	1.791	1.753
Dec 02 – Jun 07	4.428	3.433	0.867	2.278	1.058	1.916	1.032
Jun 07 – Dec 08	9.086	10.034	2.466	4.167	2.826	1.735	1.987

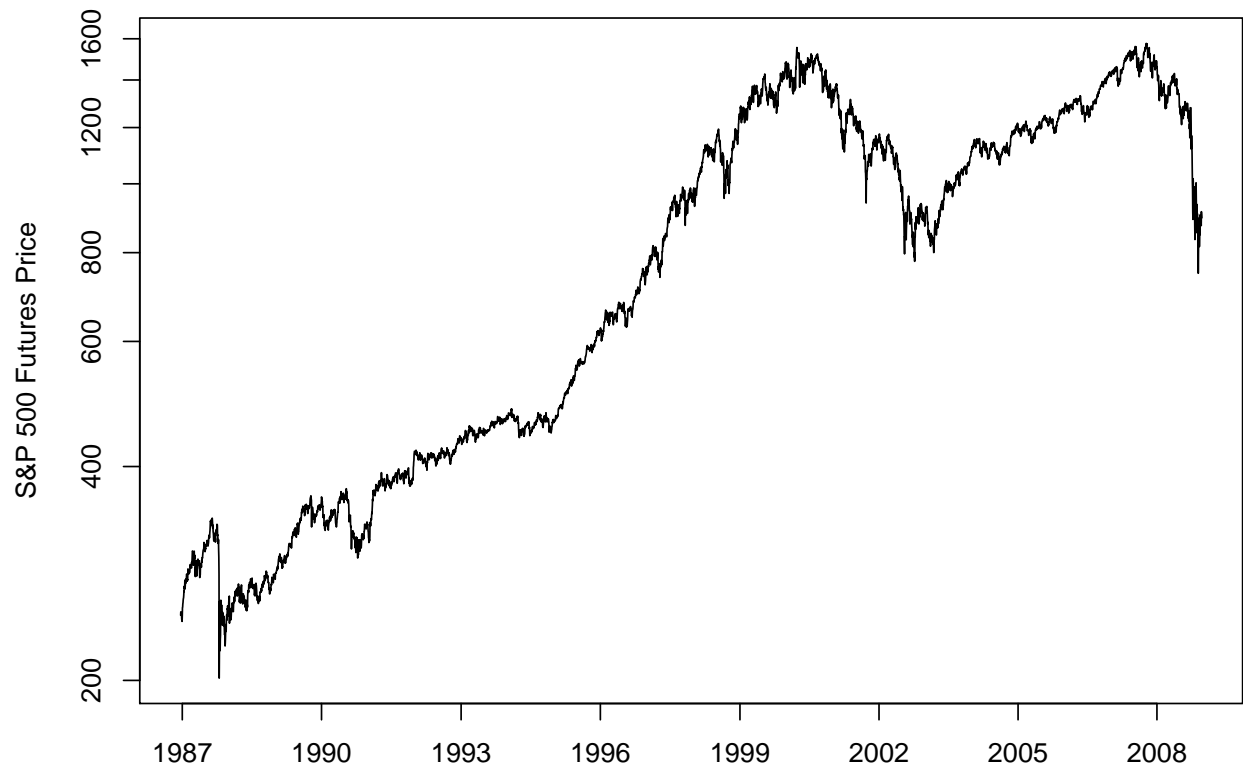


Figure 1: S&P 500 futures prices from December 19, 1987, to December 18, 2008. Only the price of the futures contract that is closest to its expiration is plotted. The vertical axis is in the logarithmic scale.

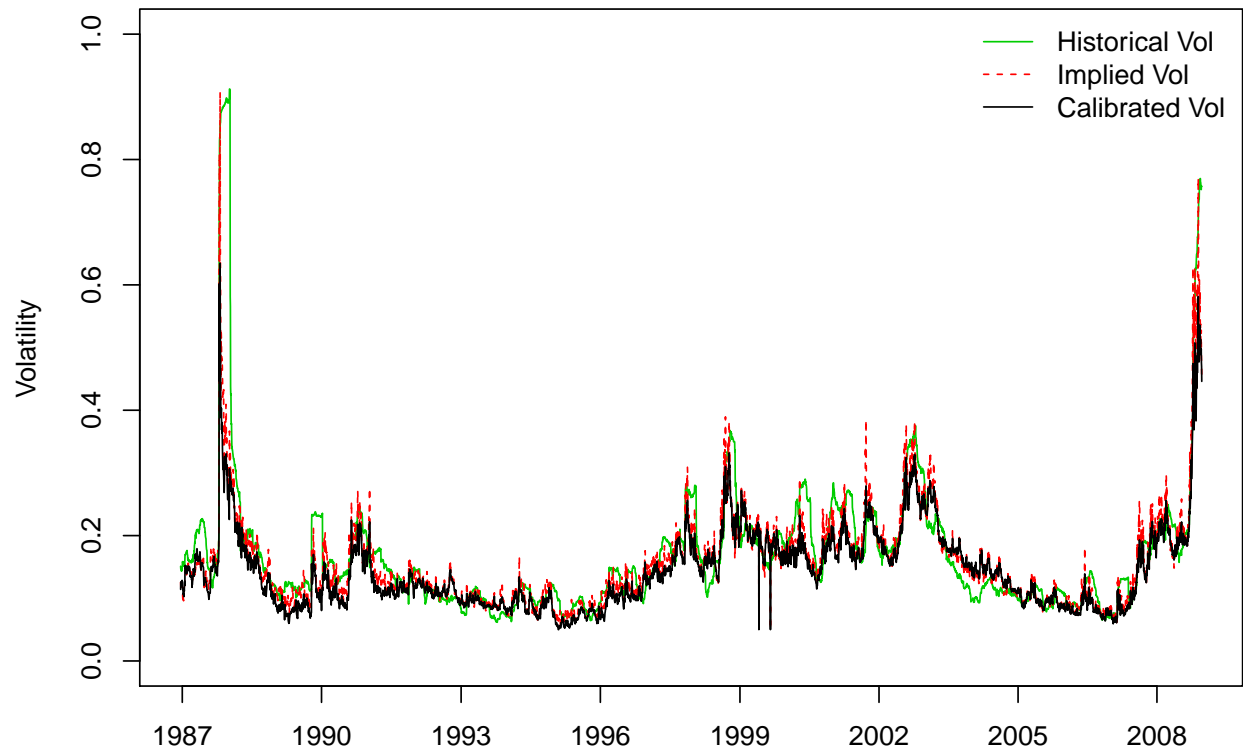


Figure 2: Historic volatility, at-the-money implied volatility and calibrated volatility for the S&P 500 futures and futures option data from December 19, 1987, to December 18, 2008. Implied volatility is calculated for the option that has the smallest log moneyness among the options that are closest to their expiration dates.