Practical Option Pricing with
Support Vector Regression and MART

by

Ian I-En Choo

Stanford University

1. Introduction The Black-Scholes [BS73] approach to option pricing is arguably one of the most important ideas in all of finance today. From the assumption that the price of a stock follows a geometric Brownian motion with constant volatility, Black and Scholes derived a formula that gives the price of a European call option on the stock, C, as a function of six variables – the stock price S, the strike price K, the time to the expiration of the option T, the risk-free interest rate r, the dividend rate paid by the stock, q, and the volatility of the stock's return, σ :

$$C(\cdot) = Se^{-qT}\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$
$$d_1 = \frac{\log(S \, / \, K) + (r - q + \sigma^2 \, / \, 2)T}{\sigma\sqrt{T}}, d_2 = d_1 - \sigma\sqrt{T}$$

where $\Phi(\cdot)$ is the standard normal CDF.

In general, an estimate of the volatility of the stock return σ can be obtained by estimating the standard error of the stock returns from past data. However, the practice most commonly used by option traders is to assume that the Black Scholes formula is correct and to solve for σ by inverting the formula given above. This practice of calculating σ (which are called the implied volatilities) is curious because the Black Scholes equation implies the volatilities obtained from options on the same stock are constant across different strikes K and maturities T. This empirical prediction is frequently violated in practice – implied volatilities plotted against strikes for most stocks typically exhibit a smile or skew effect.

To address this shortcoming, numerous attempts have been made to adapt the Black-Scholes model to make it consistent with this empirical observation. One approach is to directly model σ as a deterministic function of the K and T. Some notable attempts include implied binomial trees by Rubenstein [R94], stochastic volatility models by Heston [H93] and discrete-time GARCH models by Duan [D95].

One of the most widely used option-valuation techniques used in practice embodies this approach and is what Christofersen and Jocobs [CJ04] term the Practitioner Black-Scholes (PBS) pricing scheme. The implementation of the PBS method is straightforward and can be summarized in four steps as follows:

1. Use a cross section of European call options on the same stock with differing S, C, K, T, r and q to obtain the set of implied volatilities σ by inverting the Black Scholes formula.

2. Choose a linear functional form for the volatilities, $\hat{\sigma}(K,T)$ and estimate it by regressing the implied volatilities obtained in step 1 on powers (usually up to 2) of K, T and their cross terms using Ordinary Least Squares (OLS) .

3. For a new option we wish to price, use its values of K and T to obtain an estimate of its implied volatility through the function $\hat{\sigma}(K,T)$ estimated in step 2.

4. Obtain the estimated option price using the Black Scholes formula by using $\hat{\sigma}(K,T)$ as an argument (ie. calculating $C(S,K,T,r,q,\hat{\sigma}(K,T))$.

Although the PBS model using OLS to estimate implied volatility is remarkably simple to implement, Berkowitz [B04] notes that it surprisingly dominates the performance of other more complex, and theoretically sound approaches. Most notably, the PBS pricing scheme has been found to outperform pricing methods based on the deterministic local volatility function of Rubenstein [R94] and Heston's stochastic volatility model [DFW00]. Berkowitz [B04] offers some justification for the excellent performance of the PBS by proving that the pricing scheme can be made arbitrarily accurate as the frequency at which we re-estimate $\hat{\sigma}(K,T)$ goes to infinity.

Our aim is largely grounded in empirics - given the widespread use of the PBS model by traders and the market in practice, we consider the possibility of enhancing the performance of the PBS model through estimating $\hat{\sigma}(K,T)$ using machine learning techniques.

We estimate the implied volatility function (for data consisting of the daily prices for European call options on the S&P 500 index from February 2003 to August 2004) using Support Vector Regression (SVR) and the Multiple Additive Regression Tress (MART) algorithm, and compare the results with those obtained from an Ordinary Least Squares (OLS) regression.

2. Support vector regression Given a data set $D = \left\{ x_i, y_i \right\}_{i=1}^n$ of N points, The method of ε-Support Vector Regression [V98] (from henceforth denoted SVR) fits a function f to the data D of the following form:

$$f(x) = w^T \phi(x) + b$$

where ϕ is a mapping from the lower dimensional predictor space (or x-space), to a higher dimensional feature space, and w and b are coefficients to be estimated. The SVR stated as constrained optimization problem is:

$$\min_{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \left(\xi_i + \xi_i^*\right)$$
$$\text{subject to} \begin{cases} y_i - w^T \phi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

The dual of the SVR primal problem is

$$\max \quad -\frac{1}{2}\sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i,x_j) - \varepsilon\sum_{i=1}^n (\alpha_i - \alpha_i^*) + \sum_{i=1}^n y_i(\alpha_i - \alpha_i^*)$$
$$\text{subject to} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in \left[0, C\right] \end{cases}$$

where $k(x_i,x_j)$ is the **kernel** function that is known to correspond to the inner products between the vectors $\phi(x_i)$ and $\phi(x_j)$ in the high dimensional feature space (ie. $k(x_i,x_j) = \phi(x_i)^T \phi(x_j)$). The Radial Basis Function (RBF) kernel [SS98], $k(x_i,x_j) = \exp(-\gamma\|x_i - x_j\|)$, is known to correspond to a mapping to an infinite dimensional feature space and is adopted in this paper.

3. MART (Multiple Additive Regression Trees) The MART algorithm is an ensemble learning method where the base learners are binary decision trees with a small number of terminal nodes, m (m is usually between 4 and 8). The structural model behind MART, which belongs to a class of learning algorithms called booted tree algorithms, is that the underlying function to be estimated f, is an additive expansion in all the N possible trees with m terminal nodes than can be created from the training data, i.e. $f(x) = \sum_{i=1}^N \beta_i h_i(x)$ where h_i is the i tree whose "opinion" is weighted by β_i. Since N is typically a enormous number, tractability is an immediate concern; thus, the goal is to find sparse solutions for the β_i's.

MART adds trees in sequence starting from an initial estimate for the underlying function f. At the n iteration of the algorithm, the tree added best fits what Friedman [F01] calls the "pseudo-residuals" from the $n-1$ previous fits to the training data. This procedure is known as gradient boosting because the pseudo-residual of the i training point from the n run of the algorithm turns out to be the gradient of the squared error loss function $L(\cdot)$, $\delta L(y_i, f_n(x_i))/\delta f_n(x_i)$. The generalization properties of the MART are enhanced by the introduction of a stochastic element in the following way: at each iteration of MART, a tree is fit to the pseudo-residuals from a randomly chosen, strict subset of the training data. This mitigates the problem of "overfitting" as a "different" data set enters at each training stage. Also, boosted trees are robust to outliers as this randomization process prevents them from entering every stage of the fitting procedure. Successive trees are trained on the training set while updated estimates for f are validated on a test set;

the procedure is iterated until there is no appreciable decrease in the test error. For specific details on the MART, the reader is directed to Friedman [F01,F02].

4. Experimental Settings The data set consisting of the daily prices for European call options on the S&P 500 index (obtained from http:// www.marketexpressdata.com) from February 2003 to August 2004 was selected for this experiment. There were 398 trading days giving 252493 unique price quotes in the data. We chose the same data used in Panayiotis et al. [PCS08] where SVR was used to directly predict option prices, so that the results between our approach and theirs could be compared. As such, we have applied most of their editing and filtering rules to the data as follows: all observations that have zero trading volume were eliminated, since they do not represent actual trades. Next, we eliminated all options with less than 6 days or more than 260 days to expiration to avoid extreme option prices that are observed due to potential liquidity problems. T was calculated on the assumption that there are 252 trading days in a year, while the daily 90-day T-bill rates obtained from the Federal Reserve Statistical Release (http:// www.federalreserve.gov/releases/h15/update/) were used as an approximation for r. The annual dividend rate for the period was q=0.01587. The implied volatilities vol were then calculated using the Financial toolbox in MATLAB. This yielded a final data set of 22867 points.

This data was then randomly partitioned into an In-Sample data set consisting of 80% of the data (18293 data points) and an Out-of-Sample set consisting of the remaining data. The In-Sample data was then used train models in each of the three competing function classes – OLS, SVR and MART.

The Out-of-Sample set of 4574 data points was then used to gauge the out-of-sample performance of the three competing models. Out-of-sample estimates of the implied volatilities $\hat{\sigma}_{OLS}(\cdot)$, $\hat{\sigma}_{SVR}(\cdot)$ and $\hat{\sigma}_{MART}(\cdot)$ were obtained and subsequently plugged into the Black-Scholes formula to obtain the Practitioner Black Scholes (PBS)-estimated option prices $\hat{C}_{OLS}(S,K,T,r,\hat{\sigma}_{OLS}(\cdot))$, $\hat{C}_{SVR}(S,K,T,r,\hat{\sigma}_{SVR}(\cdot))$ and $\hat{C}_{MART}(S,K,T,r,\hat{\sigma}_{MART}(\cdot))$. These estimated values were then combined with their observed values to construct statistical metrics that we used to compare the competing methodologies. These metrics are the Root Mean Squared Error (RMSE) and the Average Absolute Error (AAE) of the PBS-predicted option prices, and the Root Mean Squared Error (IV-RMSE) and the Average Absolute Error (IV-AAE) of the predicted implied volatilities (we accessed the quality of the predictions of the volatilities as σ is frequently used in hedging applications). The definitions of these metrics are (for N=4574):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_i - \hat{C}_i)^2} \quad AAE = \frac{1}{N}\sum_{i=1}^{N}\left|C_i - \hat{C}_i\right| \quad IV-RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\sigma_i - \hat{\sigma}_i)^2} \quad IV-AAE = \frac{1}{N}\sum_{i=1}^{N}\left|\sigma_i - \hat{\sigma}_i\right|$$

5. Experimental Procedure and Results Fitting an OLS linear regression to any data set is a fairly standard procedure and there exist a large number of numerical routines and software packages that can execute the task well. In our case, we used the built-in function capabilities of the R statistical programming language to estimate the parameters of the most general specification of the quadratic deterministic volatility function in the Dumas et al [DFW00] study:

$$vol = \alpha_0 + \alpha_1 K + \alpha_2 K^2 + \alpha_3 T + \alpha_4 T^2 + \alpha_5 KT$$

The results of the fit are as follows:

	estimate	std.error	t value	P(>|t|)
α_0	1.51	9.79e-03	154.74	<2e-16
α_1	-2.17e-03	1.81e-05	-119.86	<2e-16

α_2	8.51e-03	8.64e-09	98.52	<2e-16
α_3	-4.94e-01	1.39e-02	-35.63	<2e-16
α_4	6.77e-02	6.70e-03	10.10	<2e-16
α_5	4.24e-04	1.21e-05	35.00	<2e-16

R^2=0.6079,   F=5670, $P(>|F|) < 2e\text{-}16$

As we can see, the F-statistic and t-statistics on all of the estimated coefficients are extremely significant. These results, coupled with the relatively high R^2 value indicates a good fit of the model to the data.

To fit the SVR model, we used the R implementation of the LIBSVM library of Support Vector Machine algorithms developed by Chang and Lin [CL09]. The inputs to the SVR were the same as those used for OLS. The optimal values of the free parameters C, ε, and γ (the parameter in the RBF kernel) were found by conducting a full grid search over a range of specified values for each parameter using 5-fold cross validation. The values of the parameters we used for the grid search were $\gamma \in \{0.01, 0.001, 0.0001, 0.00001\}$, $\varepsilon \in \{0.1, 0.01, 0.001\}$ and $C \in \{10, 100, 1000\}$. The best set of parameters thus obtained was C=10, $\gamma = 0.01$ and $\varepsilon = 0.001$. The SVR was then trained on the entire In-Sample data set with these parameter values.

The R implementation of the MART algorithm that was developed by Friedman [F02] was used to fit the boosted tree model. The inputs to the MART algorithm were the same as that in OLS and the SVR. The number of terminal nodes of each base learner was set to 6. 80% of the In-Sample data was used to train the model while the remaining 20% was used as a test set. A plot of the training and test errors is displayed below. The smooth plot represents the test errors and we observe that they start to level off after about 400 iterations.

Training and test absolute error
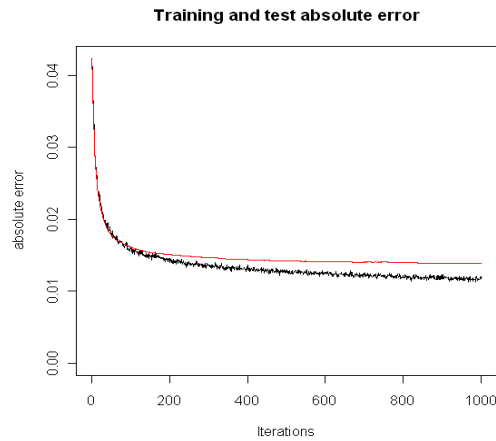
After all three estimated functions were obtained, the metrics AAE, RMSE, IV-AAE and IV-RMSE were calculated for each estimate as detailed in section 4. These results are summarized in the table below:

	RMSE	AAE	IV-RMSE	IV-AAE
OLS	2.616804	1.848029	0.04562832	0.02536823
SVR	1.743571	1.23512	0.03749572	0.01744673
MART	**1.434732**	**1.009406**	**0.03231285**	**0.01358134**

The results of the experiment are evident: the MART procedure clearly performed the best (gives the lowest Out-of-Sample error), followed by the SVR procedure across all metrics. As the metrics were calculated using Out-of-Sample data that was not used in the training of any of the models, they make a clear statement about the superior generalization properties of the machine learning based PBS-pricing models compared to the OLS based PBS-pricing model, offering the tantalizing prospect of significantly enhancing the performance of the PBS pricing model in practice.

What is perhaps most startling about these results is the size of the reduction in the errors of the option price predictions when the PBS pricing model is used with the SVR or MART instead of the OLS. Use of MART in the PBS pricing model reduced both the Average Absolute Error (AAE) and the Root Mean Squared Error (RMSE) of the call price predictions by more than 45% on the Out-of-Sample data, compared to the OLS-based PBS model. The SVR-based PBS pricing model offered more modest, but still significant improvements over the OLS-based PBS pricing model. Furthermore, our ad-hoc approach to selecting the free parameters in the machine learning algorithms used in this paper suggest that the performance of both machine learning algorithms can be significantly enhanced through a more careful and systematic parameter selection procedure. Bi et al.[BH05], and Lendasse et al. [LW03] prescribe novel ways of selecting the gamma parameter in the Gaussian kernel of the SVR, but these are by no means standard and the matter of parameter selection in SVM remains an active area of research.

References

[B04] Berkowitz J. 2004. On Justifications for the ad hoc Black-Scholes Method of
Option Pricing. Working Paper, Department of Finance, University of Houston.

[B06] Bishop C. 2006. Machine Learning and Pattern Recognition. Springer, New York.

[BH05] Bi L., Huang H., Zheng Z., Song H. 2005. New heuristic for determining Gaussian kernel's parameter. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21, August.

[BS73] Black, F., Scholes, M. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*.81: 637-659

[CL09] Chang C. and Lin C. 2001. LIBSVM : a library for support vector machines. Software available at http:// www.csie.ntu.edu.tw/~cjlin/libsvm

[CJ04] Christoffersen P., Jacobs K. 2004. The importance of the loss function in option valuation. *Journal of Financial Economics*, Vol 72 Issue 2 291-318.

[D95] Duan, J. 1995. The GARCH option pricing model. *Mathematical Finance*. 5: 13-32

[DFW00] Dumas B., Fleming J., Whaley R. 2000. Implied Volatility Functions: Empirical Tests. NBER Working Paper No. W5500.

[F01] Friedman J. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. Vol 29 No. 5: 1189-1232

[F02] Friedman J. 2002. Stochastic Gradient Boosting. *Computational Statistics and Data Analysis*. Vol 38 Issue 4: 367-378.

[F02A] Friedman J. 2002. Getting Started with MART in R. Software available at http:// www- stat. stanford.edu/~jhf/R_MART.html

[H93] Heston, S. 1993. A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*. 6: 327-343

[H073] Huber P. 1973. Robust Regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*. 1: 799-821

[H06] Huang, S. 2006. Combining Extended Kalman Filters and Support Vector Machines for Online Option Price Forecasting. Lectures in Computer Science. Vol 4259/2006. Springer. Berlin.

[HN00] Heston, S., Nandi, S. 2000. A Methodology for Assessing Model Risk and its application to the Implied Volatility Function Model. Manuscript, Rotman School of Management, University of Toronto.

[HTF01] Hastie, T., Tibshirani, R. and Friedman, J. H. 2001. *Elements of Statistical Learning*. Springer.

[LW03] Lendasse A., Wertz V. Verleysen M. 2003. Model selection with cross-validations and bootstraps – application to time series prediction with RBFN Models. ICANN/ICONIP 2003, LNCS 2714 pp. 573-580.

[LX08] Lai T., Xing H. 2008. Statistical Models and Methods for Financial Markets. Springer, New York.

[MP04] Marwala, T. Pires, M. 2004. American Option Pricing Using Multi-Layer Perceptron and Support Vector Machines. *IEEE Internation Conference on Systems, Man and Cybernetics*. Vol. 2: 1279-1285

[P04] Primbs, J. 2004. FinGroup MATLAB toolbox. Software available at http//www. stanford.edu/ ~japrimbs/fingroup.htm

[PA03] Perez-Cruz F., Afonso-Rodriguez J.A., Giner J. 2003. Estimating GARCH models using support vector machines. *Quantitative Finance*. Vol 3: 1-10.

[PCS08] Panayiotis A., Chris C., Spiros M. 2008. Options Pricing Via Statistical Learning Techniques: The Support Vector Regression Approach. Working Paper.

[R94] Rubenstein, M. 1994. Implied Binomial Trees. Journal of Finance. 49: 771-818

[RV07] Rouah F., Vainberg G. 2007. Option pricing models and volatility using Excel-VBA. John Wiley, New Jersey.

[SS98] Smola AJ, Scholkopf B. 1998 A tutorial on support vector regression. NeuroCOLT Technical Report TR, Royal Holloway College, London, UK.

[SS00] Scholkopf B., Smola A. Williamson R.C. and Barlett P.L. 2000. New support vector algorithms. *Neural Computation*, 12: 1207-1245.

[V98] Vapnik V. 1998. Statistical Learning Theory. John Wiley, New York.