Hedging and Pricing Options – using Machine Learning –

Jacob Michelsen Kolind, Jon Harris
and Karol Przybytkowski

December 10, 2009

Introduction

Options hedging has important applications in risk management. In its most simple form, options hedging is a trading strategy in a security and a risk-free bank account. An option written on the security is hedged by this strategy if the strategy is self-financing, and replicates the price of the option at all times and in all states of the world. In the simple Black Scholes model, where only one source of uncertainty is present, it can be shown that such strategies do exist and an analytical expression can be found for the proportion of wealth that should be invested in the underlying security. For an options seller as well as an options buyer, this trading strategy is important to know, since being short the option but long the hedging strategy or vice versa allows the seller/buyer to eliminate the risk associated with selling/buying the option.

In a real life setting, many of the Black-Scholes assumptions are violated. There are many sources of risk and agents incur transaction costs, so a perfect hedging strategy cannot be expected to exist. A more realistic approach for an options trader is to minimize, not eliminate, their risk by rebalancing their portfolio discretely. The size of the time steps between adjustments would then depend on the volatility of the market, the size of the portfolio, and the size of the transaction costs incurred at each trading round. A discretely rebalanced trading strategy will only approximately hedge the option, and it is not immediately clear how to choose the proportion of wealth to invest in the underlying security at each trade. If we abstract away from choosing the size of the time steps between portfolio adjustments, we are left with an interesting machine learning problem: How do we optimally choose the proportion of wealth to invest in the security, the so called delta, Δ, so that the value of our portfolio at the next readjustment point in time is as close as possible to the actual options price.

Data

We consider options data on the S&P500 index from the period 2009-01-02 09:30:01.052 to 2009-05-29 16:12:15.264. Our data set contains tick by tick data of every trade made in call and put options on the index. Each observation contains the following variables:

1. The time at which the trade took place, t.
2. The type of the option (call or put).
3. The strike price of the option, K.
4. The maturity date of the option, T.
5. The S&P 500 index price, S.
6. The price of the option, P.
7. An implied volatility proxy, σ.

Since the numeraire in price data can be arbitrarily chosen, there is some redundancy in our data set. We choose to use the strike

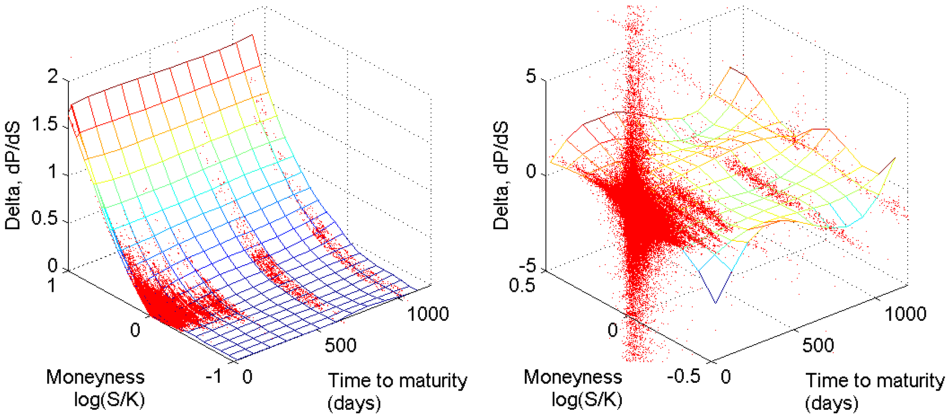

Figure 1: *Left panel:* A 3D scatter plot of the call-options data including a fitted mesh. *Right panel:* A 3D scatter plot of the observed deltas including a fitted mesh.

price as the numeraire and thus henceforth only look at 'moneyness', $\log(S/K)$, and 'calliness'/'puttiness', P/K. Ignoring the time aspect of the data we have plotted (the dots)

$$P/K \sim (\log(S/K), (T-t))$$

in the left panel of Figure 1. Qualitatively the data behaves very much as would be expected from the Black Scholes model. Close to maturity, P/K as a function of S/K resembles a hockey stick, whereas further away from maturity this hockey stick has been smoothed out. The data contains a high degree of time variation. This cannot be seen from left panel of Figure 1, but a closer inspection of the data reveals that the surface has a positive thickness. If market conditions were stable arbitrage would force the surface to have zero thickness.

By identifying options with the same strike and maturity date, we extracted one-day movements in option prices and included that as an extra column in our data, (dP). Corresponding to these one day movements, we also extracted the movements in the S&P500 index, (dS). Ignoring interest payments, the gain/loss of a strategy that buys one option and sells Δ units of the index at time t and then reverses these trades at time $t+1$ is

$$\begin{aligned}(P(t+1) - \Delta S(t+1)) - (P(t) - \Delta S(t))\\ = dP(t) - \Delta dS(t)\end{aligned}$$

In the right panel of Figure 1 we have plotted dP/dS for those observations where $dS \neq 0$. dP/dS corresponds to the value of Δ that would make the loss/gain of the hedging portfolio equal to zero.

Methodology

We explore two ways to model the price and the optimal hedge: parametrically and non-parametrically. The goal of both models is first to be able to fit the surfaces like the one displayed in Figure 1. For the price surface we use a weighted squared error loss function

$$\hat{P}_i \mapsto w_i(P_i - \hat{P}_i)^2$$

where w_i is a weight that will be specified shortly. For the delta surface we use the loss function

$$\begin{aligned}\hat{\Delta}_i &\mapsto w_i(dP_i - \hat{\Delta} dS_i)^2\\ &= w_i dS_i^2 (dP_i/dS_i - \hat{\Delta})^2\end{aligned}$$

i.e. a weighted version of the squared loss/gain of the one-day hedge. Minimizing both loss functions results in an WLS estimator of $\hat{P}$ and $\hat{\Delta}$ respectively when these are considered as being functions of the covariates log-moneyness $\log(S/K)$ and time to maturity $T-t$. Since neither surface looks like a hyperplane we approximate them using fourth order tensor spline by a B-spline basis expansion of the standard linear design matrix. Our parametric

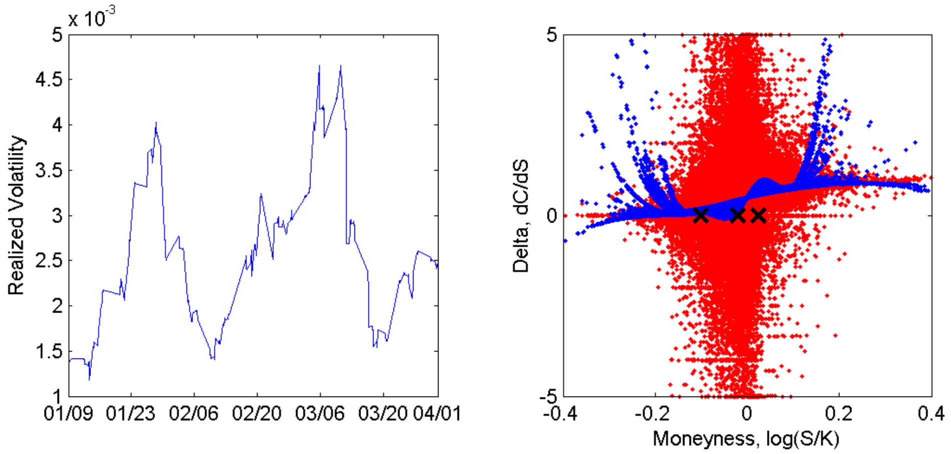

Figure 2: *Left panel:* Seven days trailing realized volatility calculated from the spot S&P500 index price. *Right panel:* A plot of the cross-section at $0 < T-t < 50$ of the data plotted in the right panel of Figure 1. The blue points are fitted values.

model also includes a third covariate, trailing realized volatility, ($\tilde{\sigma}$), a measure of the observed market volatility during the last seven days before the data point. Trailing realized volatility, depicted in the left panel of Figure 2, is added to the parametric model as a linear factor independent of the tensor spline basis in $\log(S/K)$ and $T-t$. Our calculations of realized volatility follows, to some extent, the papers by Andersen et. al., 2003 and Barndorff-Nielsen et. al. 2002.

The trailing realized volatility is included in the parametric model to accomodate the time variation inherent in the data. That is, we use the trailing realized volatility as a proxy for the time variation in the data, and hope that the observed call prices/deltas can be described as noisy observations of a function of the vector $(\log(S/K), T-t, \tilde{\sigma})$. In the parametric model $w_i = 1$.

Another way of incorporating time-variation into our model, the non-parametric approach, is to let the weights (w_i) account for the time variation. Our approach is to let

$$w_i = e^{-\lambda(t-t_i)} 1\{t_i + \delta > t > t_i\}$$

where t_i is the time corresponding to observation i, and t is the time at which the model is used for prediction. The non-parametric model weighs data points observed just prior to a point of prediction more heavily than data points observed further away in the past[1].

Results

Examples of how our models fit to the data are displayed in the panels of Figure 1. The plotted grids represent the fitted spline surfaces.

It is most interesting to focus on the performance of our delta estimates. In the left panel of Figure 3, a histogram of the distribution of gains/losses using predicted deltas for the month of April 2009 is displayed. In the right panel, the corresponding distribution of gains/losses using Black Scholes deltas is displayed. The Black Scholes deltas are simply calculated using the formula

$$\Delta = N\left(\frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right)$$

where σ is the implied volatility and r is the the risk-free interest rate. The Black Scholes delta should result in a perfect hedge

[1] δ is a fixed cutoff that sets weights that would otherwise have been very small to zero. This mitigates some of the computational burden involved in fitting the non-parametric model.

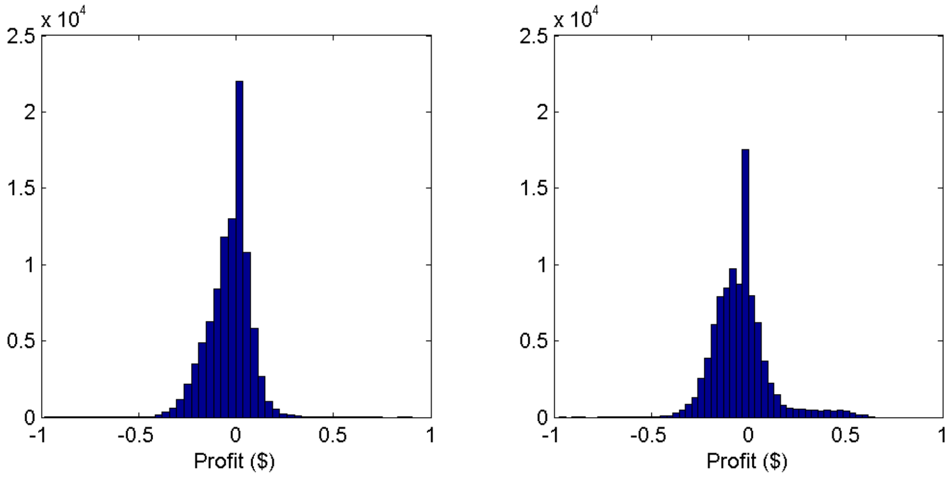

Figure 3: *Left panel:* A histogram of the distribution of gains/losses using predicted deltas from the non-parametric model with $\lambda = 1$/day for the month of April 2009. *Right panel:* The corresponding plot for Black Scholes deltas.

with continuous rebalancing of the hedging portfolio, but with discrete rebalancing the BS delta can be seen as a first order approximation to an optimal delta. Our deltas perform slightly better than the Black-Scholes deltas. The average absolute loss/gains are, respectively, \$0.156112 on the training data, \$0.107760 on the test set with Black Scholes deltas, and \$0.082342 on the test set with the non-parametric spline and $\lambda = 1$.

It seems interesting to investigate whether the better performance of our delta estimates is due to the fact that the data was not generated from the Black Scholes model. We generated a simulated data set using the Black Scholes model, using the strikes, the times and the maturities from our original data. We then applied the two ways of predicting deltas to the simulated data. The result can be seen in Figure 4. Again our model outperforms the Black-Scholes formula: The average absolute loss/gains are, respectively,

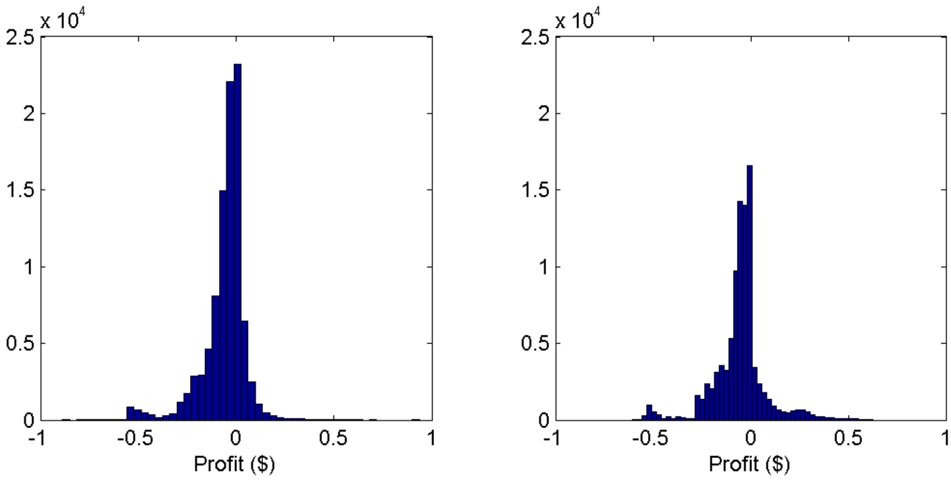

Figure 4: *Left panel:* A histogram of the distribution of gains/losses using predicted deltas from the non-parametric model with $\lambda = 1$/day for the month of April 2009 on the simulated data. *Right panel:* The corresponding plot for Black Scholes deltas.

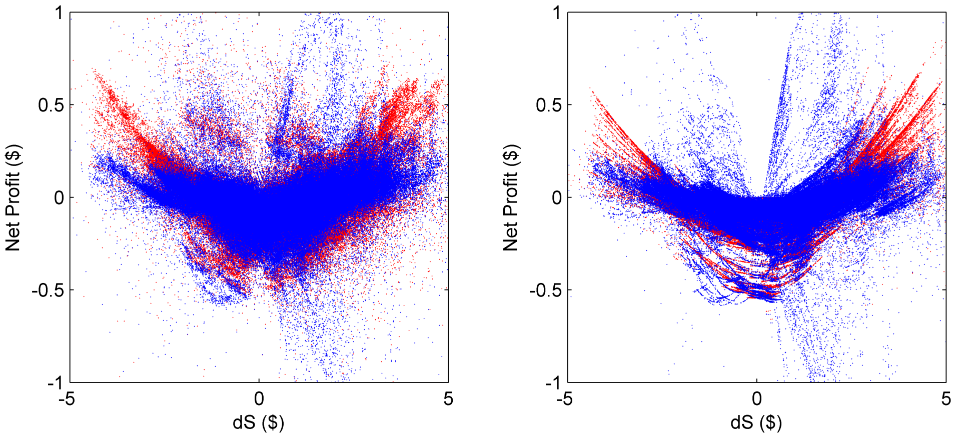

Figure 5: *Left panel:* Real data *Right panel:* Simulated data. The red dots are calculated using Black-Scholes deltas. Blue dots are spline fitted deltas.

$0.138111 on the training data, $0.096381 on the test set with Black Scholes deltas, and $0.078296 on the test set with the non-parametric spline and $\lambda = 1$.

Figure 5 gives some intuition behind why our method outperforms the Black-Scholes delta. There is a convexity effect inherent in the Black Scholes deltas that creates large gains when the underlying stock moves a lot. Our method specifically targets these large movements, using dS^2 as a weight in the fitting, to try to minimize the impact of these movements on the loss/gain in the hedging portfolio. The point cloud representing gains/losses using our models thus curves less than the point cloud corresponding to the analytical Black-Scholes deltas. We are aware that some trading strategies exploit the convexity effect by e.g. hedging straddles, but these kinds of strategies requires insight/an opinon about future volatility.

We do not present any results from the parametric model, since it, in its current form, performed worse than the Black Scholes deltas.

Conclusion

We have presented a new way of estimating good delta hedges using tensor splines. The methods we developed seem to work well on simulated as well as real data, and generally outperform the generic choice, the analytical Black-Scholes delta hedging formula. Related to our work are the papers by Hutchinson et. al., 1994, Lai et. al., 2004 and Bennell et. al., 2005.

References

Andersen T. G., Bollerslev T., Diebold F. X. and Labys P., *Modeling and Forecasting Realized Volatility*, Econometrica, Vol. 71, No. 2 (Mar., 2003), pp. 579-625.

Barndorff-Nielsen O. E. and Shephard N., *Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models*, Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 64, No. 2 (2002), pp. 253-280.

Bennell J. and Sutcliffe C., *Black-Scholes versus artificial neural networks in pricing FTSE 100 options*, Intelligent Systems in Accounting, Finance & Management, Vol. 12 Issue 4 2005 pp. 243-260.

Hutchinson J., Lo A. and Poggio T., *A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks*, Journal of Finance, Vol. 49 1994 pp. 851-889.

Lai T.L. and Wong S., *Valuation of American options via basis functions*, IEEE Transactions on Automatic Control, Vol. 49 2004 pp. 374-385.