PRICING THE IMPLIED VOLATILITY SURFACE

by

Ionut Florescu
Purdue University

Technical Report #05-02

Department of Statistics
Purdue University

April 2005

Pricing the implied volatility surface

Ionut Florescu

Department of Statistics
Purdue University, 150 N. University St.
West Lafayette, IN 47907-2067, USA
ifloresc@purdue.edu

Abstract: Today's practitioners model the evolution of the price of companies' stock using the well-known Black-Scholes-Merton model. To this end they need to estimate the volatility of the model. There exists a direct relationship between this volatility and the price of the options on the stocks. We will concentrate on this relationship and come up with a new method to compute the volatility from these prices available on the market.

1. Introduction

Expected future volatility plays a central role in finance theory. Consequently, accurate estimation of this parameter is crucial to meaningful financial decision-making. Researchers rely generally on the past behavior of asset prices to estimate volatility, relating movements in volatility value with prior volatility and/or variables in the investors' information set. These procedures are by nature backward looking using past behavior to predict the future. An alternative approach is to use reported option prices to infer volatility expectations. Since option value depends critically on the expected future volatility, the volatility expectation of market participants can be recovered by inverting the option valuation formula when such a formula exists.

The volatility expectation derived from reported option prices depends on the assumptions used to valuate the option. For example, the Black-Scholes model assumes the asset price follows a Geometric Brownian motion with constant volatility. Consequently the options on the same asset, but with different strike prices and maturity dates should provide the same implied volatility. In practice, however, the implied volatility tens to differ across exercise prices and time to expiration producing the so called "volatility smile".

The failure of the Black- Scholes model to describe the structure of reported option prices is thought to arise from its constant volatility assumption. It has been observed that when stock prices go up volatility goes down and vice versa. If the volatility is allowed to depend on the stock price it becomes stochastic and accounting for stochastic volatility within an option valuation formula is not an easy task. In the special case when the volatility is a deterministic function of asset price and/or time, it is possible to valuate options by means of the Black - Scholes partial differential equation.

Derman and Kani (1994), Dupire (1994), and Rubinstein (1994) develop variations of this deterministic case approach. Their methods attempt to fit a cross section of option prices and deduce the future behavior of volatility as anticipated by market participants. Rather than give a formula or a structural form for the volatility function, they search for a binomial or trinomial lattice that achieves an exact cross-sectional fit of reported option prices.

Bellow I will try to give an approximation algorithm for a stochastic volatility function. The method is based on cubic splines interpolation and it is an inverse function approximation problem since from the market we observe the option price not the volatility itself.

2. Formulation of the problem

Assume that the underlying asset follows a continuous 1-factor diffusion process with the initial value S_0:

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma^*(S_t, t)dW_t \qquad , t \in [0, T] \qquad (1)$$

for some finite horizon T, where W_t is a finite Brownian motion and $\mu(s,t)$ and $\sigma^*(s,t)$ are deterministic functions sufficiently well behaved to guarantee a unique solution to the above equation. We assume for simplicity that the instantaneous interest rate is a constant r>0 and the dividend rate is a constant q>0. Given S_0, r, and q under the no arbitrage assumption, an option with volatility σ(s,t), strike price K, and maturity T has a unique price v(σ(s,t), K, T).

Assume that we are given m market option bid-ask pairs, $\{(bid_j, ask_j)\}_{j=1,\ldots,m}$ corresponding to strike prices / expiration times $\{(K_j, T_j)\}_{j=1,\ldots,m}$. Let

$$v_j(\sigma(s,t)) \overset{def}{=} v(\sigma(s,t), K_j, T_j) \quad , \quad j=1,\ldots,m$$

We want to approximate as accurately as possible, the local volatility function $\sigma^*(s,t): \Re^+ \times [0,\tau] \rightarrow \Re^+$, from the requirement: $bid_j \leq v_j(\sigma(s,t)) \leq ask_j$, for all j's. Since the observation data is finite and the restriction is on the option values, not on the volatility itself, the problem is an inverse approximation problem, from finite data. The problem can be written as an optimization problem that is finding:

$$\min_{\sigma(s,t)} \sum_{j=1}^{m} \left| bid_j - v_j(\sigma(s,t)) \right| + \sum_{j=1}^{m} \left| v_j(\sigma(s,t)) - ask_j \right| \qquad (2)$$

Unfortunately the function to be minimized is not differentiable. To overcome non-differentiability we can solve the least squares minimization problem:

$$\min_{\sigma(s,t)} \sum_{j=1}^{m} \left(v_j(\sigma(s,t)) - \overline{v}_j \right)^2 \qquad (3)$$

where $\overline{v}_j = (bid_j + ask_j)/2$. Since the observation data is finite this problem typically has an infinite number of solutions. Notice that the option price as a functional of the volatility is non-linear.

There is an obvious solution of the above problem, that is: match the option price exactly by finding the local volatility such that $v_j(\sigma(s,t)) = \overline{v}_j$ for all j. This method has been emphasized in the majority of the papers I read. However, since the problem has an infinite number of solutions, a function which matches the finite set of data points will tend to over-fit the data and can be very different from the local volatility. Moreover, the price $\bar{v}_j$ generally has errors (for example when a bid-ask spread doesn't exist). In addition, in general, the option value $v_j(\sigma(s,t))$ can only be computed numerically using a binomial tree or a PDE approach (the general diffusion equation (1) doesn't have closed form solution). Hence it might not be desirable to insist that the $\bar{v}_j$'s are matched exactly. For pricing and hedging of exotic options it is more important to compute a local volatility function which fits the market as closely as possible but also approximates well the "real" local volatility function $\sigma^*(s,t)$ of model (1).

Smoothness has long been used as a regularization condition for a function approximation problem with limited observation data. A smoothing spline minimizes a compromise between goodness of fit and the degree of goodness, that is minimizing the functional:

$$\min_{\sigma(s,t)} \frac{1}{m}\sum_{j=1}^{m}\left[\sigma_j(s,t)-\bar{\sigma}_j\right]^2 + \lambda \iint \left[\sigma''(s,t)\right]^2 ds dt. \tag{4}$$

The choice of the regularization parameter λ is crucial here.

2.1. Choosing the right λ

Splines have long been used in approximating smooth curves and surfaces. In a typical 1 dimensional spline interpolating settings, m-pairs (x_i, Y_i) are given according to the model: $Y_i=f(x_i)+\varepsilon_i$, $i=1,\dots,m$ with $\varepsilon_i \sim N(0,\sigma^2)$. They represent a finite sample of the predictor x and of the observable dependent variable $f(x)$. Using this model one estimates $f(x)$, using the minimizer of:

$$J_\lambda(f)=\frac{1}{m}\sum_{j=1}^{m}\left[Y_j - f(x_j)\right]^2 + \lambda \int \left[f''(x)\right]^2 dx.$$

The first term in $J_\lambda(f)$ measures the closeness of $f(x)$ to y; the second one adds a penalty for the curvature of $f(x)$. The smoothing parameter $\lambda \geq 0$ controls the tradeoff between the two conflicting goals. Small values of λ favor jagged curves that follow the data points closely, while when $\lambda \to \infty$ one has the simple linear regression model $f(x)=\beta_0+\beta_1 x$.

The freedom of the spline is determined by the degree of the polynomial used for approximation, on each segment. Craftsmen and engineers have long used the cubic spline as the mechanical spline. It is the smoothest twice-continuously differentiable function that matches the observations and the second derivative at fixed points known as the knots. It can be shown that for any fixed value of λ, the minimizer of $J_\lambda(f)$ is a cubic spline.

What value of λ is best for our data? Denote $\hat{f}_\lambda(x)$ the function estimate based on our data set and a fixed value of λ. If we had new data (x', Y') it would be reasonable to choose λ to minimize the expected prediction error:

$$pse(\lambda) = E\left[Y' - \hat{f}_\lambda(x')\right]^2$$

How do we compute λ using this principle since in general we don't have a new data set? Fortunately there are methods developed especially for this purpose, namely Bootstrap, Cross-validation, and generalized Cross-validation. We will describe these methods since we are going to use them in our algorithm. Say our original sample is made of n pairs (x_i, Y_i).

The Bootstrap method generates bootstrap samples, which I will denote (x^*, Y^*). A Bootstrap sample is made of n pairs (x_i, Y_i) drawn from the original data set with replacement. Then the method proceeds by computing the curve estimate $\hat{f}_\lambda^*(x)$ based on each Bootstrap sample and a fixed value of λ. For each sample, the error that $\hat{f}_\lambda^*(x)$ makes in predicting the original sample is:

$$pse^*(\lambda) = \frac{1}{n} \sum_{\substack{original \\ data \\ points}} \left[Y_i - \hat{f}_\lambda^*(x_i)\right]^2$$

Averaging this quantity over B bootstrap samples provides an estimator of the prediction error $pse(\lambda)$.

The Generalized Cross-Validation Method divides the data into K roughly equal-sized parts. For each part k, finds $\hat{f}_\lambda(x)$ based on the other K-1 parts of the data, and then calculates the prediction error of the fitted model when predicting the k^{th} part of the data. Then the Generalized Cross-Validation estimator is the average prediction error over all K parts of the data.

Cross Validation is an extreme case of the Generalized Cross-Validation Method, here the data is divided into n parts, each containing one data point, the so called "leave-one-out" cross-validation.

2.2. How all this apply to our problem.

Our problem is slightly more complicated since the function to be approximated is two dimensional, and furthermore, is given in implicit form. We overcome these difficulties by using a 2 dimensional spline, a so-called bicubic spline, and estimating the volatility at given points using the diffusion equation (1), and the observed option values from the market. Specifically given $\bar{v}_j$ for the respective pair (K_j, T_j), we determine the local volatility values $\bar{\sigma}_j$ by calibrating the market observable prices $\bar{v}_j$. Then we determine the best λ for the data given and fit the respective bicubic spline, which will be our estimator of the true local volatility function.

This procedure will give us a way to estimate the local volatility function for any option with strike price and maturity date set. Finding the local volatility will allow us to estimate a fair price for the premium of an option and then compare this estimate with the premium prices currently on the market.

2.3. Assumptions and Applications

Approximating the local volatility function by a spline is particularly reasonable if the local volatility function is smooth. Is this a reasonable expectation for the local volatility function? Assume that the underlying follows the 1-factor diffusion process (1). Let *v(K,T)* be observable arbitrage-free market European Call prices, for all strikes $K \in [0,\infty)$ and all maturities $T \in [0,\tau]$. From the diffusion equation of the European Call with dividends (see for example Hull 1997) the local volatility function is given uniquely by:

$$[\sigma^*(K,T)]^2 = 2\frac{\frac{\partial v}{\partial T} + qv(K,T) + K(r-q)\frac{\partial v}{\partial K}}{K^2\frac{\partial^2 v}{\partial K^2}}$$

This formula suggests that, assuming *v(K,T)* is sufficiently smooth (note that $\frac{\partial^2 v}{\partial K^2}$ and $\frac{\partial v}{\partial T}$ already exist) and $\frac{\partial^2 v}{\partial K^2} \neq 0$, the $[\sigma^*(K,T)]^2$ is sufficiently smooth in the region $(0,\infty)\times(0,\tau]$ as well.

Another issue to be considered here is the fact that the dependence of the option premium of the volatility values is not uniform in the region $(0,\infty)\times(0,\tau]$. The option value depends little on the volatility values with small T and K far from S_{init} (see Figure 1). It is convenient to view this as follows: there exists a region centered on S_{init} within which the volatility values are significant in pricing and hedging (region D in the picture bellow). We can at most expect to approximate well the local volatility function in this region from the market option data.

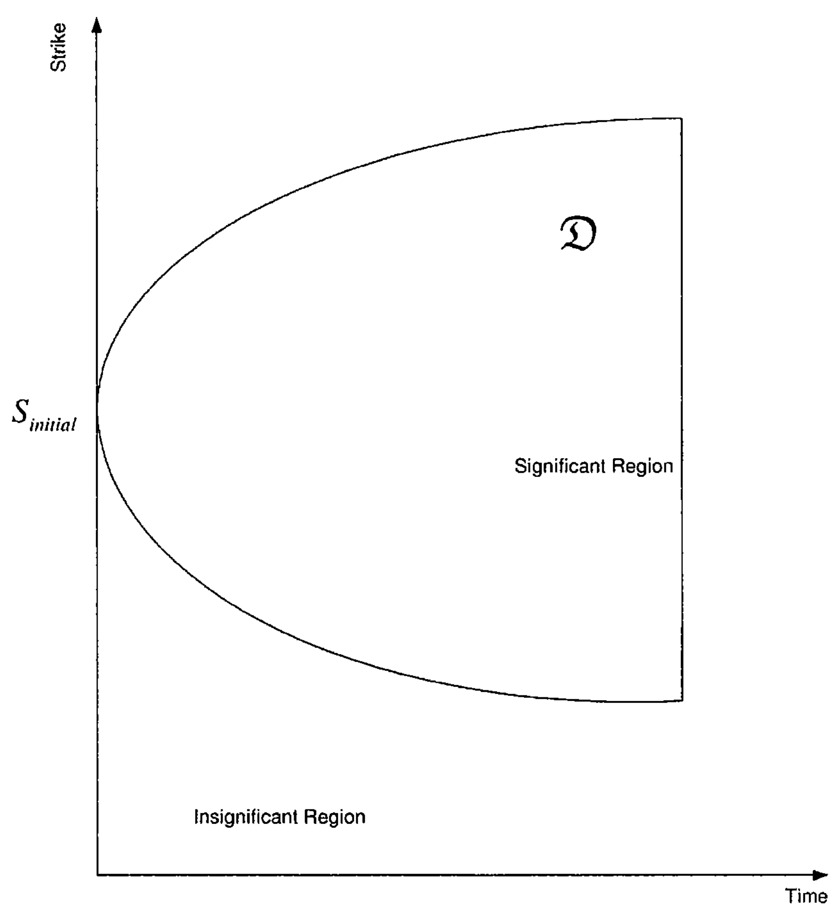


FIGURE 1. The local volatility in the region D is significant for pricing and hedging

3. The computational procedure.

We want to approximate the local volatility surface $\sigma^*(s,t)$ with a cubic spline by solving (4) for the vector $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_m)$. The values of the vector $\bar{\sigma}$ are determined by fitting the observable option values as closely as possible. Although, in theory, the volatility for any option can be approximated using this procedure, using anything else than European Options requires approximation of the volatility by means of PDE discretization techniques or binomial tree approach. Since we don't want to complicate things too much we will concern ourselves only with the European case.

The freedom in the problem consists in the choice of λ in equation (4). To get to this end one might use one of the three criterions described in Section 2.1. It is notable to observe that all three methods give approximately the same estimate for λ, though the speed with which each converges is different with the bootstrap being the slowest. Once λ has been chosen, the local volatility is approximated by the corresponding bicubic spline.

For the European case we will observe $\bar{v}_j$'s from the market and assume that they are the true price perturbed by normal errors ε_j. Then we will find the vector $\bar{\sigma}$ inverting these values using a modified bisection method, and use the found values to fit a bicubic spline.

4. Computational Examples

We now describe some computational experience with the proposed method for reconstructing the local volatility function $\sigma^*(s,t)$ from limited observation data. The method is implemented in R (a open source, clone of S-plus).

4.1. Reconstructing Local Volatility, a synthetic example.

We will consider an artificially constructed example. We consider European calls with different strike prices and maturity dates (it is the same example used by Lagnado & Osher 1997, and by Coleman & all 1999).

In this example the underlying is assumed to follow the diffusion process (1), where the volatility function $\sigma^*(s,t)$ is a function of only the stock, $\sigma^*(s,t) = \alpha/s$, with α =15. Let the initial underlying price be $S_{init} = 100$, the risk-free interest rate r=0.05, and the dividend rate q=0.02.

We consider as market option data, 22 European call options on S. Eleven options have half-year maturity with strike prices ranging from 90 to 110 in increments of 2 units and another eleven options have one year maturity with the same strikes. Thus the option and maturity vectors are:

$$K = [\,90, 92, \ldots, 110, 90, 92, \ldots, 110] \in R^{22}$$

$$T = [\,0.5, 0.5, \ldots, 0.5, 1, 1, \ldots, 1] \in R^{22}$$

For the European Call Option with continuous paid dividends an analytic formula exists, and we set the market European call price $\bar{v}_j$ equal to this value plus a random error. That is:

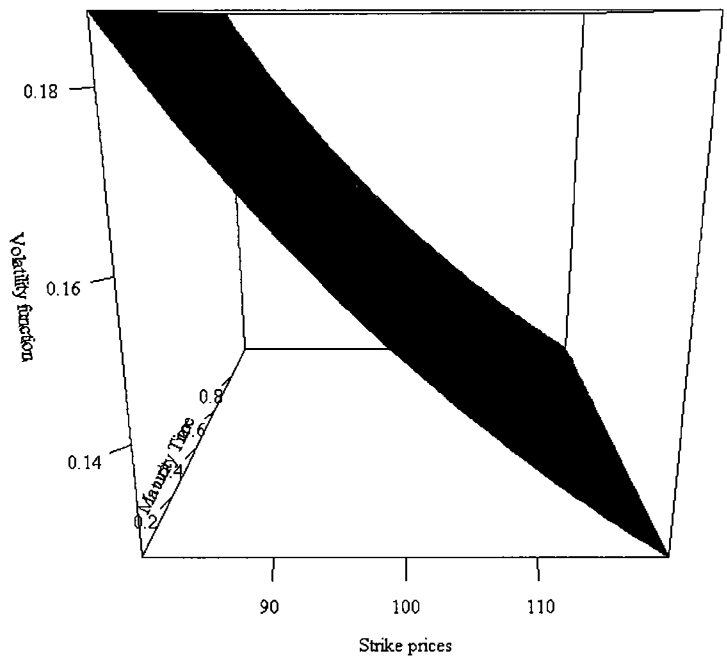
$$\bar{v}_j = \text{Exact price of option}_j + \varepsilon_j$$

where ε_j's are normally distributed with variance 1.

Note that this is not quite the same example as from the articles cited above where the market option price is set *exactly equal* to the value given by the formula and all the discussion and results look like taken from a fairy tale.

Figure 2 presents the true volatility function together with the estimated volatility. The next plot (Figure 3) is an interesting one, it plots the estimated accuracy of the reconstructed volatility function. I did not use the true data that is unknown in the general case, but used bootstrap again to estimate it. We can see, as we expected, that the best accuracy is attained in the region D described in Section 2.3. The fit has the same level of accuracy for a fixed level of maturity simply because the original volatility is not dependent on it.

Plot of the true local volatility function



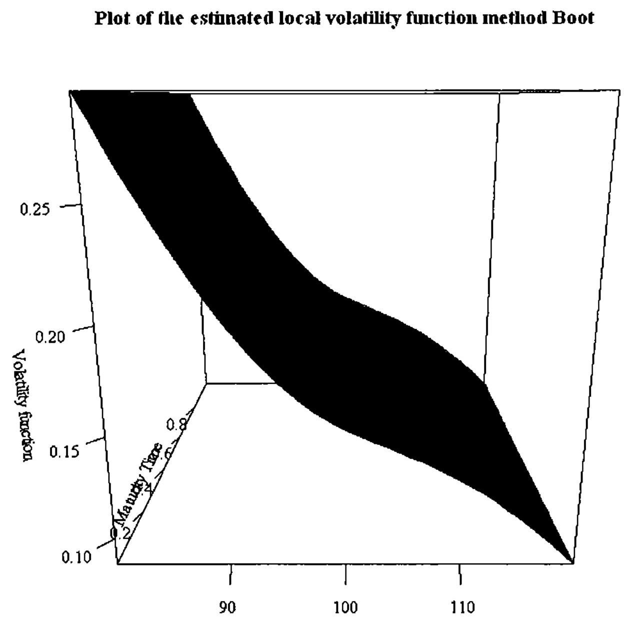


FIGURE 2: The true and the reconstructed local volatility

Next we looked to sections in the volatility surface for fixed levels of Maturity: at T=0, T=0.518, and T=1. The fit looks alike for all three Maturity dates proving that the fitted volatility function recognized no dependency on time.

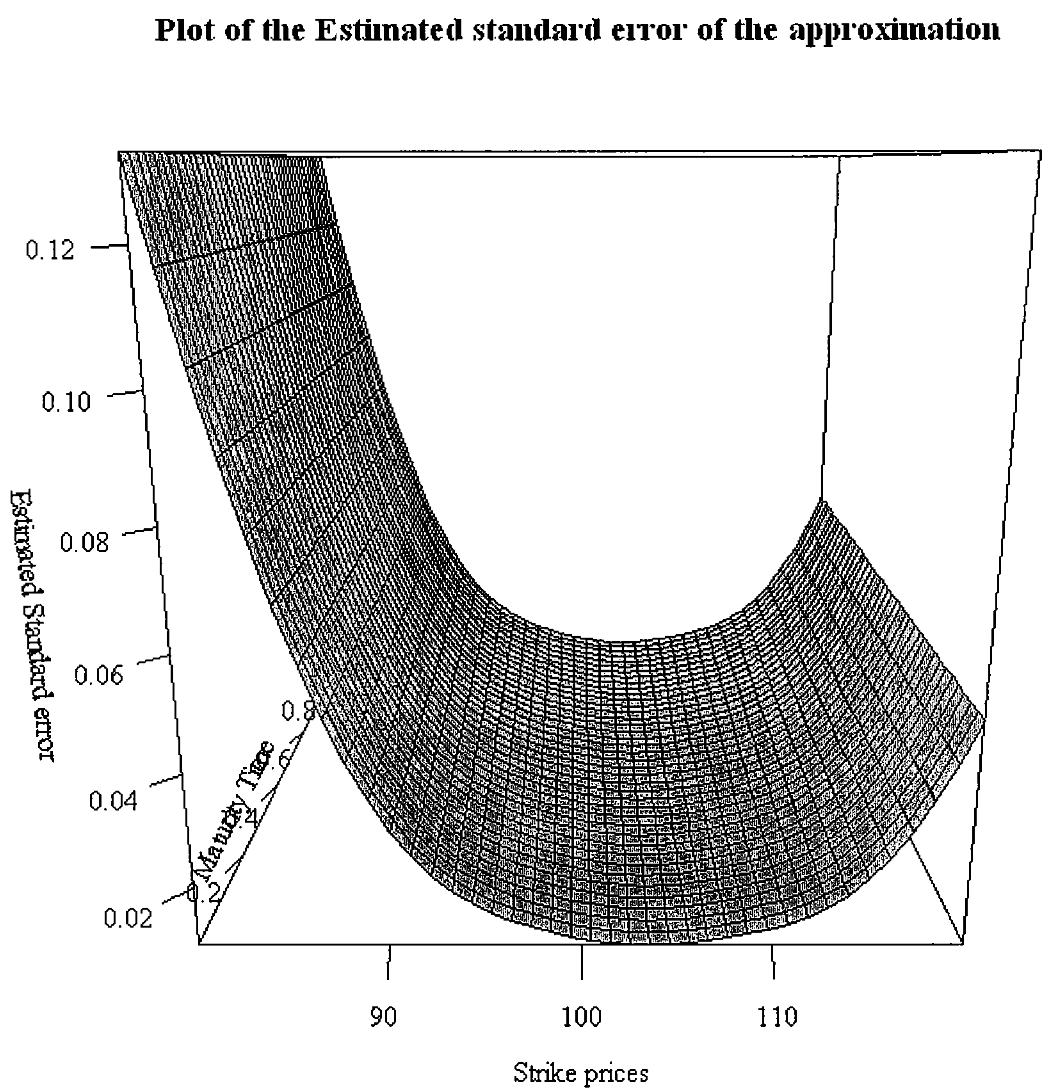


FIGURE 3: The estimated standard error

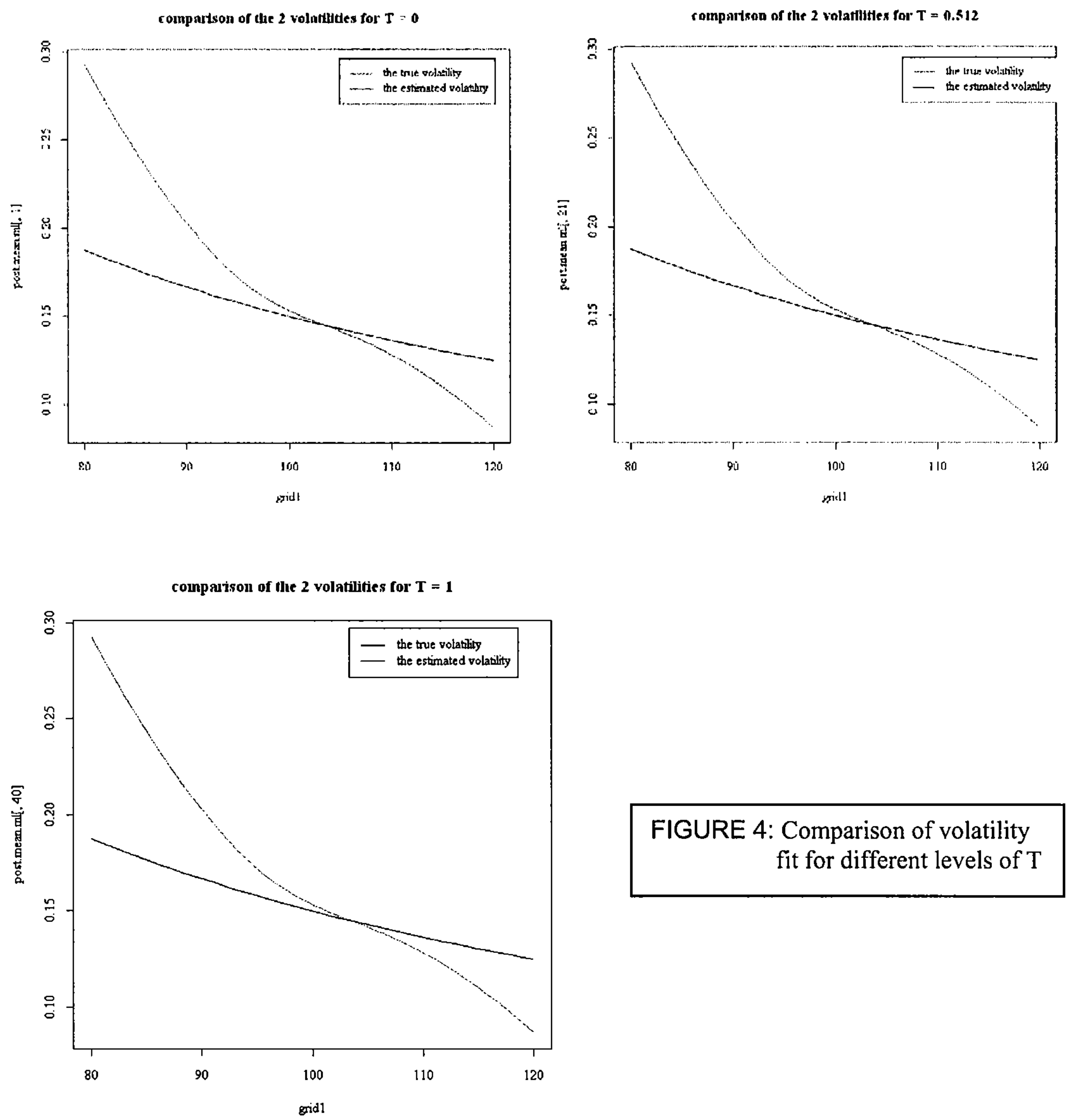


FIGURE 4: Comparison of volatility fit for different levels of T

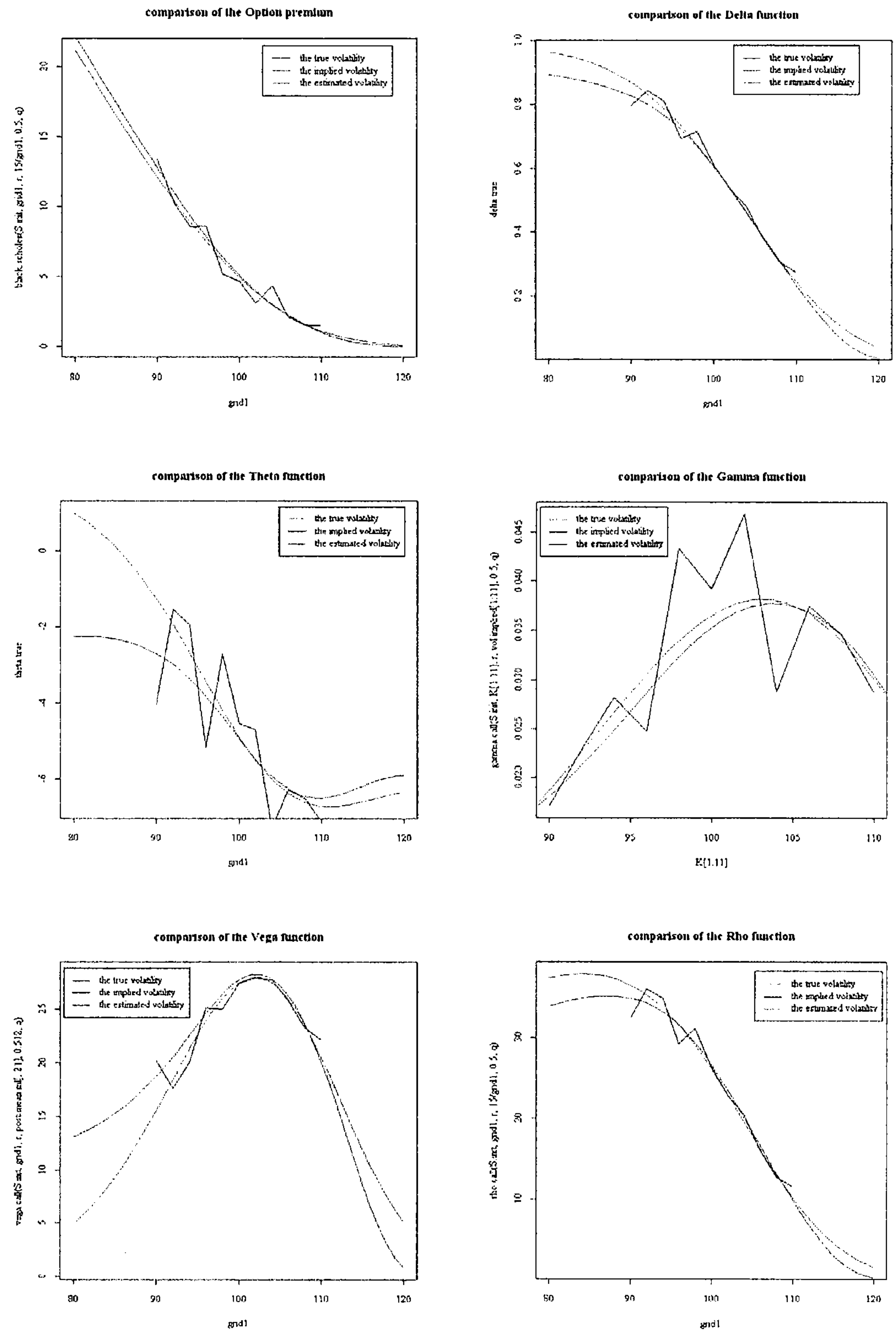


FIGURE 5: Comparison of hedging factors using the true, reconstructed and implied volatilities

Next we looked to the hedge factors and the option price computed for an option with Maturity 0.5 years (because there we can compute the implied volatility). We used true, reconstructed and implied volatility. The plots are given in Figure 5.

The findings are remarkable. Not only we can use the reconstructed volatility to predict the price way outside the range of the implied volatility but the predicted values using the reconstructed volatility are way better than using the implied volatility (the current market practice).

4.2. An S&P 500 Example

Here we used a more realistic example of reconstructing the local volatility using real data example specifically S&P 500 index European call options from October 1995. It is the same data given in Anderson & Ratcliffe (1998) and Coleman & all (1999).

The initial index, interest rate and dividend rate are set as in the cited articles:

$$S_{init} = \$590, \quad r = 0.06, \qquad q = 0.0262.$$

Using all the data with maturities ranging from .175 years (2 months) to 5 years we fit the local volatility function and the result is given in Figure 6. Also the estimated standard error is plotted in Figure 7. We note from these plots two interesting facts. Although the original data (implied volatility) varies a little with the Maturity, the difference is very small, and for the fitted volatility function practically disappears. We can see this phenomenon from the plots in Figure 8.

Also from the estimated standard error we see once again that the best fit is obtained in a region of type D as stipulated in Section 2.3. Noteworthy is the large error of the fit for large values of strike price (deep out of the money). We kind of expected this since there is a large uncertainty regarding those types of options.

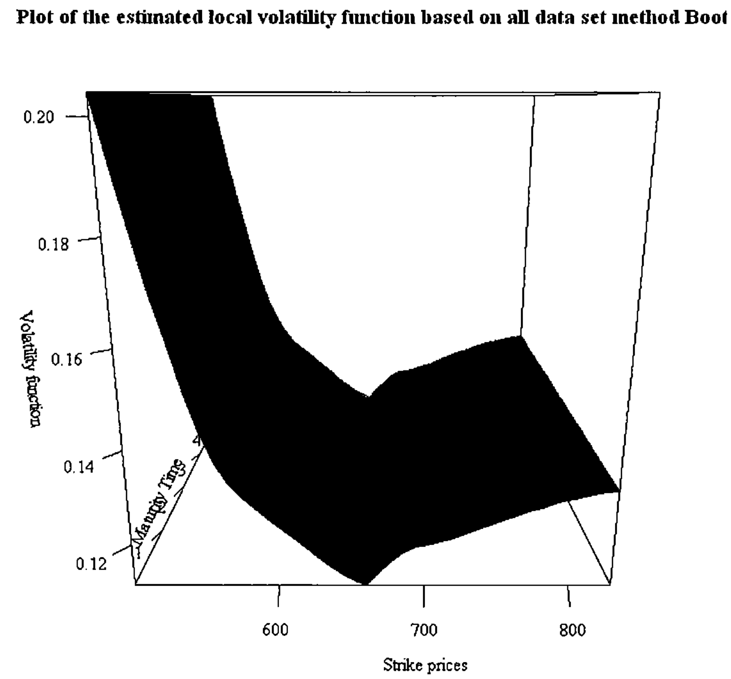


FIGURE 6: Plot of the estimated volatility surface

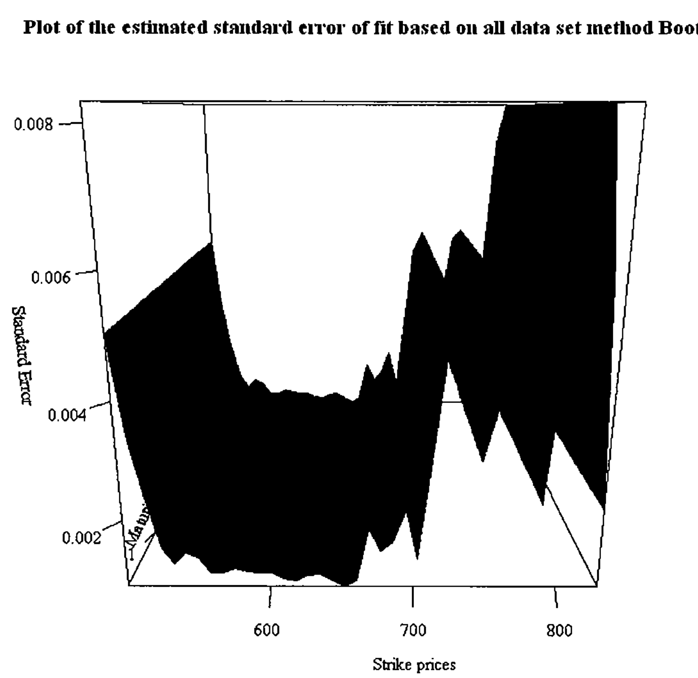


FIGURE 7: The estimated Standard error of the fit

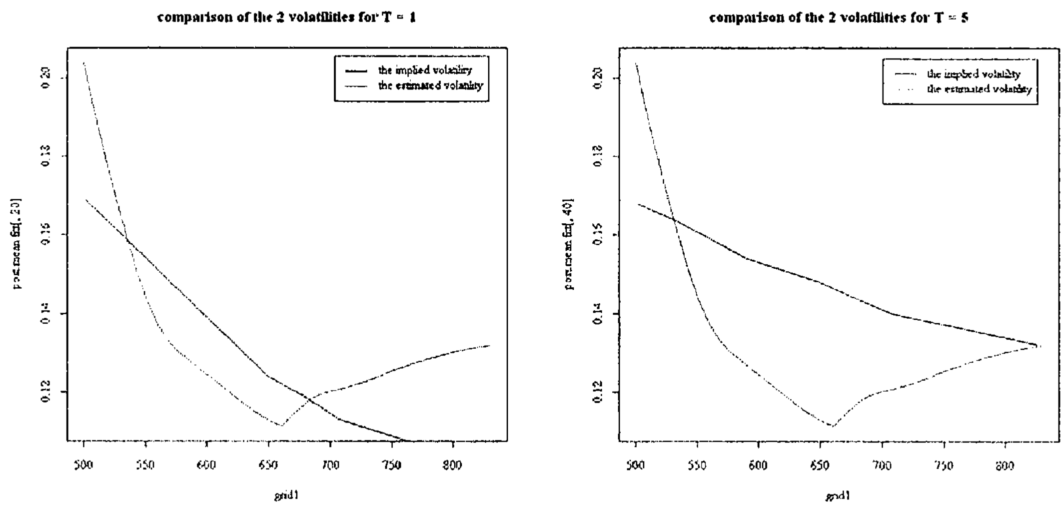


FIGURE 8: Comparison of the fitted curve for fixed values of Maturity

5. Concluding remarks

Assuming the 1 factor diffusion model we reconstructed the local volatility surface using 2 dimensional splines. We showed using a synthetically constructed example that the method gives better results than just using the implied volatility (the current practice on the market). The largest error that the implied volatility induces is observed in the option price and the hedging factors, where it counts the most from the practical reasons. Furthermore, our results look better than the ones previously published, considering the large deviation of prices from their assumed true values.

All the computer work necessary was done on a home, midrange value computer using R (freeware clone to Splus) and an optional package GSS (General Smoothing Splines) courtesy of the author Professor Chong Gu (Dept. Statistics, Purdue University). This fact emphasizes the practical value of this paper.

REFERENCES

1. L. Andersen and R. Brotherton-Ratcliffe (1998). The equity option volatility smile: An implicit finite-difference approach. *Journal of Computational Finance*, **1**(2) pp. 5-32
2. F. Black and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy,* **81**, pp. 637-659.
3. F. Derman and I. Kani (1994) Stochastic Implied Trees: Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility, *International Journal of Theoretical and Applied Finance*, pp. 17-22
4. B.Dupire (1994) Pricing with a Smile, *Risk*, **17**, pp. 18-20
5. J. Hull (1997). *Options, Futures, and Other Derivatives*, (3rd ed.), Prentice Hall
6. C. Gu and C. Qiu (1993). Smoothing spline density estimation: Theory. *Annals of Statistics*, **21**, pp. 217-234
7. R. Lagnado and S. Osher (1997). Reconciling differences. *Risk*, **10**, pp. 79-83.
8. D. Lamberton and B. Lapeye (1996). *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall
9. R. Merton (1973). The theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4**, pp. 141-183.
10. M. Rubinstein (1994), Implied Binomial trees, *Journal of Finance*, **49**, pp. 771-818
11. G. Wahba (1975). Smoothing noisy data by spline functions. *Numerical Mathematics,* **24**, pp. 383-393

12. G. Wahba (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. *Journal of Royal Statistical Society Series B*, **45**, pp. 133-150