Stock Option Pricing Using Bayes Filters

Lin Liao

liaolin@cs.washington.edu

Abstract

When using Black-Scholes formula to price options, the key is the estimation of the stochastic return variance. In this paper we discuss an approach based on Bayes filters which combines the GARCH model and the implied volatilities. Empirical experiments demonstrate the better pricing accuracy of this approach. Furthermore, we show that we can re-estimate the parameters of the dynamics system using Expectation-Maximization algorithm.

1 Introduction

Option is a financial contract that gives the holder the right to buy (*call option*) or sell (*put option*) an asset for a certain price (*strike price*) on (*European style*) or before (*American style*) a certain date (*maturity date*). Option trading allows the investors to bet on future events and to reduce the financial risks. However, what the contract is worth is anything but trivial. In the early 1970s, Fischer Black, Myron Scholes, and Robert Merton made a major breakthrough in the pricing of stock options. This involved the development of what has been known as the Black-Scholes (BS) model. By making a few assumptions, BS model provides the following formula for an European call option [6]:

$$c = S\Phi(d_+) - Ke^{-rT}\Phi(d_-) \qquad (1)$$

where

$$d_+ = \frac{ln(\frac{S}{K}) + (r + \frac{v}{2})T}{\sqrt{vT}}$$

$$d_- = d_+ - \sqrt{vT}$$

where c is the call option price, S is the current market price of the stock, K is the strike price, r is the risk-free interest rate, , v is the stock return variance [1], T is the time to maturity, and $\Phi()$ is the cumulative distribution function for the standard normal distribution. The only parameter in the BS formula that cannot be directly observed is the variance v [2]. In the

---

[1]Usually this formula is expressed using the standard deviation of the stock return, a.k.a the *volatility*. Estimating the variance and the volatility are essentially the same thing and thus we will use the two terms interchangingly in the paper.

[2]Although a few literatures also think the risk-free interest rate is not observable [9], most literatures treat it as

original derivation of BS formula, v is treated simply as a constant. But this simplification is far from perfect. In practice, the variance v, like the stocks price, is a stochastic variable. Unlike the stock's price, this variable cannot be directly observed. Therefore, the key to use BS formula is to estimate the volatility, or equivalently, the variance v.

There are basically two ways of determining the variance v of stock returns, one of which is to estimate it from time series of the stock prices, and the other to calibrate the model prices to market option prices. The v, and hence the model option prices, obtained in these two ways may not coincide, and this is often the case. It is not surprising that calibrated model prices are closer to observed market prices, but they say little about the connection to the actual dynamics of the underlying asset. It is, after all, the value of the underlying asset at expiration that determines the option pay-off.

In this paper, we discuss the approach of pricing options within the framework of Bayes filters. This approach combines the above two ways and utilize both historical stock prices and the market option prices. As we will show, our approach is not only able to predict the option prices closer to the market values than the other two ways, but also able to re-estimate the dynamics of the underlying asset by taking into account the information from the market option prices.

In the next section, we first introduce the two basic ways of estimating the variance v and then present our approach which is based on Bayes filters. In the third section, we further discuss how to re-estimate the asset dynamics through parameter learning using EM algorithm. Then the empirical results are shown in the fourth section. Finally we will discuss related work and conclude the paper.

2 Bayes Filters for Option Pricing

If we assume the Black-Scholes (BS) formula is true, the problem of option pricing becomes the problem of estimating the variance of stock returns, which is a stochastic variable. As we have mentioned in the introduction, most existing approaches can be classified into two categories.

---

observable as what we do in this paper.

2.1 Variance Estimation from Historical Stock Prices

In the first category, the variance v is estimated from the time series of the stock return, which can be computed easily from the historical stock prices. More specifically, suppose we have the stock prices from day 0 to N as $S_0, S_1, ..., S_N$, then the stock return for day n is defined as

$$u_n = \frac{S_n - S_{n-1}}{S_{n-1}} \quad 1 \le n \le N$$

Our objective is to estimate v_n, which is the variance for day n, using the returns up to day n. It is commonly agreed that more weight should be given to the recent data. One of such models is the GARCH(1,1) model [3] proposed by Bollerslev in 1986. In Garch(1,1), v_n is calculated from a long-run average variance rate V_L, the most recent return u_n and the previous variance estimate v_{n-1}. The equation for GARCH(1,1) is [6]

$$v_n = \gamma V_L + \alpha u_n^2 + \beta v_{n-1} \quad (2)$$

where γ is the weight assigned to V_L, α is the weight assigned to u_n^2 and β is the weight assigned to v_{n-1}. Because the weights must sum up to one, we have

$$\gamma, \alpha, \beta > 0$$
$$\gamma + \alpha + \beta = 1$$

Setting $\omega = \gamma V_L$, the model can also be written

$$v_n = \omega + \alpha u_n^2 + \beta v_{n-1} \quad (3)$$
$$\omega, \alpha, \beta > 0$$
$$\alpha + \beta < 1$$

The key to use GARCH(1,1) model is the estimation of the parameters ω, α and β. One way to do that is the maximum-likelihood method, which involves choosing the values for the parameters that maximize the likelihood of the return series [6].

2.2 Variance Estimation from Market Option Prices

In practice, traders usually work with what are known as *implied volatilities*. These are the volatilities (or variances) implied by option prices observed in the market. If we think the market option price is the "true" price, from the BS formula, we have

$$c_n = BS(v_n)$$

where c_n and v_n are the market option price and variance for day n, respectively. Then we can calculate v_n as

$$v_n = BS^{-1}(c_n)$$

---

[3]GARCH stands for Generalized AutoRegressive Conditional Heteroscedasticity and $(1, 1)$ indicates that v_n is based on the most recent observation of the stock return and the most recent estimate of the variance rate.

Although the closed form of $BS^{-1}()$ does not exist, we can easily use some numerical methods, e.g. Newton-Raphson method, to compute the implied volatility [1].

There are a few places that implied volatility may introduce errors. First, traders usually calculate the implied volatility for day n and use it to predict the option price for day $n + 1$. Because the volatilities for two successive days are not identical, the prediction cannot be prefect. Second, the market is not ideal and the market price may deviate from the "true" value, especially, some "outliers" prices may cause the implied volatilities far from the real values.

2.3 Variance Estimation Using Bayes Filters

GARCH model estimates the variance from the underlying stock returns, while implied volatilities use only the market option prices. They use very different information and neither of them is prefect. Therefore, it may work better if we can combines the two approaches and use all the valuable information. One way of doing that is the Bayes filter.

Bayes filters probabilistically estimate the state of a dynamic system from the control information and noisy observations. In the context of variance estimation, the state is the variance v, the control signals are the stock returns u and the observations are the market option prices c. With the Markov assumption, the belief of the variance v can be estimated efficiently using the following recursive equation

$$p(v_n|u_{1:n}, c_{1:n}) \propto p(c_n|v_n) \int p(v_n|v_{n-1}, u_n) \bullet$$
$$p(v_{n-1}|u_{1:n-1}, c_{1:n-1})dv_{n-1}$$

where $p(v_{n-1}|u_{1:n-1}, c_{1:n-1})$ is the estimate for day $n - 1$ using all the information up to that day, $p(v_n|v_{n-1}, u_n)$ is the system dynamics and $p(c_n|v_n)$ is the observation model. We specify the dynamics using the GARCH(1,1) model plus a noise term w_n

$$v_n = \omega + \alpha u_n^2 + \beta v_{n-1} + w_n \quad (4)$$

and the observation model using the BS formula plus a noise term z_n

$$c_n = BS(v_n) + z_n \quad (5)$$

The graphical model for such a dynamic system is shown in Figure. 1.

2.4 Extended Kalman Filters

Bayes filter is a very powerful framework that has a number variants for different kinds of dynamics and observation models. A popular one is called Kalman filter which assumes linear system dynamics and observation models as well as Gaussian noises. As to the variance estimation in our case, we cannot directly apply Kalman filter because the observation model, i.e. BS formula, is obviously non-linear. Forturenately, we
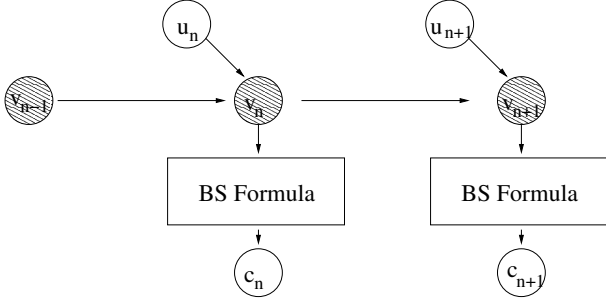
Figure 1: *Graphical model for variance estimation where shaded variables are hidden.*

can solve this problem by using an extension of the Kalman filter, called extended Kalman filters (EKF), which relaxes the assumption of linear system by performing local linearization. Like Kalman filters, EKF also provide closed-form solutions to the estimation problem [10]. Thus, if we describe the variance estimation problem using Eq. (4) , (5) and assume the noises, w_n and z_n, follow Gaussian distributions, we can solve the problem efficiently using EKF.

In EKF, the estimate of the state is depicted as a Gaussian distribution and at each time it performs a *prediction* step and a *correction* step.

Suppose

$$w_n \sim \mathcal{N}(0, R)$$
$$z_n \sim \mathcal{N}(0, Q)$$

where R and Q are the variance of the noises , and the estimation of variance v on day $n-1$ is

$$v_{n-1} \sim \mathcal{N}(\mu_{n-1}, \Sigma_{n-1})$$

where μ_{n-1} is the mean and Σ_{n-1} is the variance (i.e. the *variance* of the *variance* estimation).

In the prediction step, the estimate before seeing any new observations is updated only according to the dynamics. In this case, the new estimate $\bar{v}_n$ is simply a linear combination of two Gaussians (see Eq. 4). Therefore, $\bar{v}_n$ is also a Gaussian and the mean $\bar{\mu}_n$ and variance $\bar{\Sigma}_n$ are updated as

$$\bar{\mu}_n = \omega + \alpha u_n^2 + \beta \bar{\mu}_{n-1} \quad (6)$$
$$\bar{\Sigma}_n = \beta^2 \Sigma_{n-1} + R \quad (7)$$

In the correction step, when the new observation c_n comes, the estimate $\bar{v}_n$ is updated and we get the variance estimate for day n. More specifically, the update rules are as follows:

$$H_n \equiv \left. \frac{\partial c_n}{\partial v_n} \right|_{v_n = \bar{\mu}_n}$$
$$= 0.5 S \bar{\mu}_n^{-1/2} T^{1/2} \phi(d_+) \quad (8)$$
$$K_n = \frac{\bar{\Sigma}_n H_n}{\bar{\Sigma}_n H_n^2 + Q} \quad (9)$$
$$\mu_n = \bar{\mu}_n + K_n c_n \quad (10)$$
$$\Sigma_n = (1 - K_n H_n)\bar{\Sigma}_n \quad (11)$$

Eq. (8) does local linearization by computing the partial derivative of Eq. (1) and Eq. (9) calculates the Kalman gain K_n.

To summarize, EKF begins with a prior distribution of v_0, repeats the prediction and the correction steps as new information comes and thus the up-to-date estimate v_n takes into account all the information up to day n.

3 Parameter Learning Using EM

So far, we have shown that Bayes filters are able to utilize more information than GARCH model or implied volatilities. Another advantage of Bayes filters is that they are able to adjust the dynamics of the underlying asset through parameter learning. As explained in Sec. 2.1, the standard way of estimating the GARCH(1,1) parameters does not take into account the market option price, thus it is possible that the model option prices deviate significantly from the market values. Within the framework of Bayes filters, we are capable of re-estimating the parameters in Eq. (4), i.e. ω, α, β, as well as the variance of the dynamics noise, R [4]. There are more than one way of doing that. In this project, an estimation method based on Expectation-Maximization (EM) algorithm is used.

EM algorithm is widely used to deal with the learning problem with missing features (e.g. the return variance v in our case). It is an iterative approach that has an E-step and an M-step at each iteration. In a nutshell, each E-step estimates the expectation of the log likelihood of all the features given all the observations and each M-step updates the parameters to maximize the expectation calculated in the E-step.

3.1 E-Step:

Let $X = v_{1:N}$, $Y = \{c_{1:N}, u_{1:N}\}$, $\Theta = \{\omega, \alpha, \beta, R\}$. It is not hard to derive the expectation of the log likelihood as [5]

$E[\log p(X, Y|\Theta)|Y]$
$$= \sum_{n=1}^{N} (-\log R - \frac{1}{R} E((v_n - (\omega + \alpha u_n^2 + \beta v_{n-1}))^2|Y) \quad (12)$$

The key in (12) is the expectation term which can be further expressed as

$E((v_n - (\omega + \alpha u_n^2 + \beta v_{n-1}))^2|Y)$
$$= E(v_n^2|Y) + \beta^2 E(v_{n-1}^2|Y) - 2\beta E(v_{n-1}v_n|Y)$$
$$- 2(\omega + \alpha u_{n-1}^2)(E(v_n|Y) - E(v_n|Y)) +$$
$$(\omega + \alpha u_{n-1}^2)^2 \quad (13)$$

---

[4] Although in principle we can also estimate the parameter observation noise variance Q, it is much more complex because the observation model is non-linear [7]

[5] Since we are not going to update the observation model parameters, the likelihood here only considers the dynamics uncertainty.

To evaluate (13), we need to compute $E(v_n|Y)$, $E(v_n^2|Y)$ and $E(v_n v_{n-1}|Y)$ for $1 \leq n \leq N$ which requires the operation of standard *Extended Kalman Smoothing*.

3.2 M-Step:

In the M-step, we want to update the parameters to maximize the expectation in Eq. (12) which is a function of the parameters. In [3], a closed-form solution is given for such a optimization problem. But that result is not applicable to our situation because we have some constraints on the parameters. In particular, the M-step can be regarded as a non-linear optimization problem as

$$\text{maximize} \quad E[\log p(X, Y|\Theta)|Y]$$
$$\text{subject to}$$
$$\omega, \alpha, \beta, R > 0$$
$$\alpha + \beta < 1$$

This problem can be solved using some general optimization techniques such as Gradient method. The standard EM algorithm requires finding the global optimums in the M-step in order to converge. However, such kinds of techniques are only able to find local optimal solutions. Fortunately, this is not a big trouble because the Generalized EM algorithm [5] claims the algorithm has the same convergence properties if at each M-step we can guarantee to increase the likelihood. Thus, the local optimum in our case is enough.

4 Experiments

4.1 Date Set

The data used in the empirical analysis are the S&P 500 index option [8] from year 1991 to 2002. The database is huge. For each day, it has option quotes for both call and put options. Then for each of them, there exist a number of different expiration dates. And for each given expiration date, there are a few possible strike prices. Then for the given expiration date and strike price, it includes the price quotes for "ask","bid","high","low" and "closing".

In our experiment, we simply pick one option for each date. The procedure is the following. First, we only use the call options. Second, we choose the expiration dates in Mar, June, Sep and Dec, whichever is the closest to the trading date. For example, if the trading date is in Jan, we will pick the expiration date in Mar; if the trading date is in May, we will pick the expiration date in Jun; and so on. This is more or less arbitrary and mainly for simplifying the data processing. Third, after fixing the expiration date, among the options with different strike prices, we then pick the option which has the largest trading volume. This is important because for many options, the volume is zero or very low. Hence the market prices may deviate

a lot from the "true" prices. Finally, after picking the option, we simply select the closing price quotes.

Another data source necessary for the experiments is the risk-free interest rate (see Eq. (1)). We use the London InterBank Offer Rate (LIBOR) 3-month rate for U.S. dollars [2].

One limitation of the data set is that we do not have the dividends data. Thus we assume no dividends for the index. So far we are not sure yet how much this will affect our results.

We separate the data into two parts: the training set and the testing set. The training set includes the data from year 1991 to 1998 and is used for GARCH model parameter estimation and Bayes filter parameter learning. The testing set includes the data from year 1999 to 2002 and is used for comparisons of different approaches.

4.2 Results

We want to compare the pricing performance of the three approaches discussed in this paper: GARCH(1,1) model, implied volatilities and the Bayes filter. The measure we use here is the relative price error which is defined as

$$\frac{|BS(\hat{v}) - c|}{c}$$

where $\hat{v}$ is the estimate of the return variance and c is the market option price.

The cumulative distributions of the relative price error for the three approaches are shown in Fig. 2.
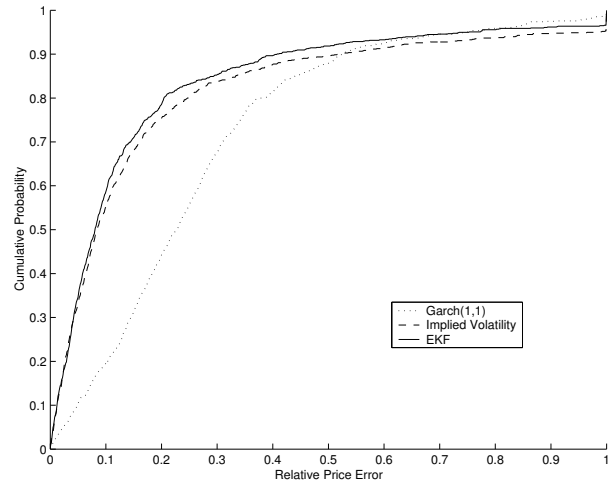
Figure 2: *Cumulative distribution function of relative price error for the three approaches.*

It is obvious to see that the curves for implied volatilities and Bayes filters are significantly more compact than GARCH(1,1) model and further more the Bayes filters perform better than implied volatilities.
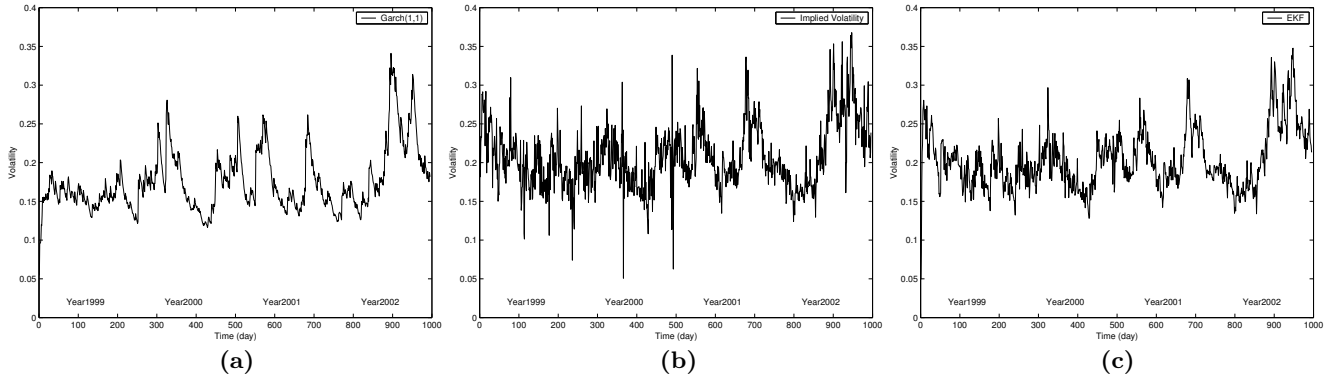
Figure 3: *The volatilities estimated by the three approaches. (a) GRACH(1,1) model; (b) Implied Volatility; (c) Extended Kalman Filters*

The average relative price errors are also listed in the following table.

Method	Avg. Relative Price Error
GARCH(1,1)	0.2829
Implied Volatility	0.2331
EKF	0.2125

The better performance of Bayes filters can be explained by comparing the volatility estimates as shown in Fig. 3.

There are a few observations in the graph. First, the GARCH(1,1) model tends to under-estimate the market volatilities because it does not consider any information from the market prices.

Second, implied volatilities have a lot of jumps and may introduce big pricing errors when outliers exist.

Third, Bayes filters estimates combines the advantages of the other two. The curve is closer to the market values than the GARCH(1,1) model and smoother than the implied volatility.

The second thing we want to evaluate is the performance of the learned model. On the training data, the learned model does improves the estimation. It reduces the average of price error from 0.201 to 0.181. But when applying the learned model on the test data, the average of the price error is 0.217, which is slightly worse than the model without learning. One possible explanation is that those parameters have some small drifts that cannot be captured by the offline learning.

5   Related Work

There have been a few research that puts the problem of stock return variance estimation within the framework of Bayes filters.

[9] used unscented Kalman filters, another extension of Kalman filters, to estimate the volatilities for option pricing. But they did not use any specific dynamics and did not discuss the problem of parameter learning.

[4] also applied Bayes filtering to estimate stochastic volatilities. Although they also used BS formula as the observation model, they used Heston model for the dynamics instead of GARCH model in our case.

What's more their learning algorithm is very different from ours. They estimated the dynamics parameters by augmenting the state space to include the parameters. The advantage of doing it is that they can have online estimation of the parameters while our EM algorithm only supports offline learning. The disadvantage is that the state space increases exponentially with the number of state variables such that they use more complex algorithm, e.g. Markov Chain Monte Carlo (MCMC), to perform inference.

6   Conclusions and Future Work

In this paper, we have shown it is possible to estimate volatilities from both the historical data and the market prices using Bayes filters. The inference in such a dynamic system can be done efficiently with the Gaussian noise assumption and the extended Kalman filter algorithm. Empirical analysis shows better pricing accuracy can be achieved. Furthermore, using EM algorithm, we are able to re-estimate the parameters of the dynamics which takes into account the market option prices.

Our approach assumes BS model is the *true* model as describing the relation between volatilities and option prices. In fact it may not be really *true*. Recent studies have shown that more data-driven models, such as Neural Networks, can be used to model the relation and outperform the BS model. The flexibility of Bayes filter allows us to use the Neural Networks as the observation model without changing our parts of the system. In that case, we may not be able to use extended Kalman filters to give closed-form solution, but we can still do inference using methods such as particle filters or MCMC.

7   Acknowledgements

References

[1] N.A. Chriss. *Black-Scholes and Beyond: Option Pricing Models.* IRWIN Professional Publishing, 1997.

[2] LIBOR data. `http://www.economagic.com/libor.htm`.

[3] Rohlicek J.R. Digalakis, V. and Ostendorf. M. Ml estimation of a stochastic linear systems with the em algorithm and its application to speech recognition. *IEEE Transactions on Speech and Audio Processing*, 1992.

[4] Martin. G.M. Forbes, C.S. and J. Wright. Bayesian estimation of a stochastic volatility model using option and spot prices. *Monash Econometrics and Business Statistics Working Papers.*

[5] t. Hebert and R. Leahy. A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors. *IEEE Trans. Med. Imag*, 8, 1989.

[6] John C. Hull. *Options, Futures, and Other Derivatives.* Prentice Hall, 2002.

[7] S.T. Roweis and Z. Ghahramani. An em algorithm for identification of nonlinear dynamical systems. *Kalman Filtering and Neural Networks*, 2000.

[8] Introduction to S&P 500. `http://www.cboe.com/micro/spx/`.

[9] E.A. Wan and R Van der Merwe. Unscented kalman filter. In *Kalman Filtering and Neural Networks*, chapter 7. Wiley Publishing, 2001.

[10] G. Welch and G. Bishop. Technical report, Department of Comp. Sc. and Engg., Univ. of North Carolina at Chapel Hill, 2002.