

Stochastic Volatility Estimation of Stock Prices using the Ensemble Kalman Filter

Yudi Mahatma* and Ibnu Hadi

Program Studi Matematika, FMIPA, Universitas Negeri Jakarta

Email: *yudi_mahatma@uni.ac.id, ibnu_hadi@uni.ac.id

Abstract

Volatility plays important role in options trading. In their seminal paper published in 1973, Black and Scholes assume that the stock price volatility, which is the underlying security volatility of a call option, is constant. But thereafter, researchers found that the return volatility was not constant but conditional to the information set available at the computation time. In this research, we improve a methodology to estimate volatility and interest rate using Ensemble Kalman Filter (EnKF). The price of call and put option used in the observation and the forecasting step of the EnKF algorithm computed using the solution of Black-Scholes PDE. The state-space used in this method is the augmented state space, which consists of static variables: volatility and interest rate, and dynamic variables: call and put option price. The numerical experiment shows that the EnKF algorithm is able to estimate accurately the estimated volatility and interest rates with an RMSE value of 0.0506.

Keywords: stochastic volatility; call option; put option; Ensemble Kalman Filter.

Abstrak

Volatilitas adalah faktor penting dalam perdagangan suatu opsi. Dalam makalahnya yang dipublikasikan tahun 1973, Black dan Scholes mengasumsikan bahwa volatilitas harga saham, yang merupakan volatilitas sekuritas yang mendasari opsi beli, adalah konstan. Akan tetapi, para peneliti menemukan bahwa volatilitas pengembalian tidaklah konstan melainkan tergantung pada kumpulan informasi yang dapat digunakan pada saat perhitungan. Pada penelitian ini dikembangkan metodologi untuk mengestimasi volatilitas dan suku bunga menggunakan metode Ensemble Kalman Filter (EnKF). Harga opsi beli dan opsi jual yang digunakan pada observasi dan pada tahap prakiraan pada algoritma EnKF dihitung menggunakan solusi persamaan Black-Scholes. Ruang keadaan yang digunakan adalah ruang keadaan yang diperluas yang terdiri dari variabel statis yaitu volatilitas dan suku bunga, dan variabel dinamis yaitu harga opsi beli dan harga opsi jual. Eksperimen numerik menunjukkan bahwa algoritma ENKF dapat secara akurat mengestimasi volatilitas dan suku bunga dengan RMSE 0.0506.

Kata kunci: volatilitas stokastik; opsi beli; opsi jual; Ensemble Kalman Filter.

1. INTRODUCTION

Volatility is a measure of the uncertainty of future stock price movements so that volatility analysis is an important factor in the consideration of buying or selling an option. An option is a contract that gives the holder the right to buy or sell an asset at a predetermined period at an agreed price so that options are the most flexible investment tool that allows an investor to control a large number of shares at a cost, which is much lower [1]. There are two types of options, namely, call options and put options. A call option provides the right to buy several assets while a put option provides the right to sell several assets.

Several models have been developed to analyze the price movement of options, one of which is the Black-Scholes model [2] expressed by the equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0 \quad (1)$$

where $V = V(S, t)$ is the option value for the stock price $S \geq 0$ at $0 \leq t \leq T$, r is the risk-free interest rate, and σ is the volatility of the stock? In their paper, Fisher Black and Myron Scholes [2] assume that the volatility of the security underlying call options is constant. They then rely on unconditional volatility to formulate the equation. As is common practice today, they choose the standard deviation of stock returns as the empirical value of volatility. However, the researchers found that the volatility of returns is not constant but depends on the collection of information that can be used at the time of calculation. Several studies have been developed to estimate volatility. Racicot and Theoret [3] compared the volatility model based on the Kalman Filter with the stochastic volatility estimated by the GARCH model. The comparison results show that stochastic volatility is a better predictor than GARCH (1,1).

The GARCH method is also used by Ahmed and Suliman [4] to estimate the volatility of the return index on the Khartoum Stock Exchange (KSE). This index is modeled using the symmetric and taxymmetric GARCH models. The result is that the taxymmetric GARCH model shows better estimation results than the symmetric GARCH model. Elliott et al. [5] introduced a class of stochastic volatility models and developed a nonlinear filter with discrete-time to estimate hidden volatility based on observed returns. Burtnyak and Malyska [6] develop a method to forecast the option prices on assumption stochastic volatility. Kim et al. [7] forecast the volatility of nine leading cryptocurrencies using Bayesian Stochastic Volatility and several model GARCH, Nguyen et al. [8] estimate the stock return using a dynamic leverage stochastic volatility (DLSV) model, and Alghalith et al. [9] propose novel nonparametric estimators to estimate the stochastic volatility.

Stochastic volatility is one of the most developed classes at present, this is due to the irrelevance of the assumption of constant volatility in the real world. Therefore, this study will develop a methodology to estimate volatility and interest rates using the Ensemble Kalman Filter (EnKF) method. Awashie [10] uses the EnKF method to estimate volatility and interest rates based on the Black-Scholes model. However, in his thesis, Awashie did not fully explain the methodology and state-space used to estimate volatility and interest rates so that methodological differences may occur even though the methods used are the same. In this study, the estimation of volatility and interest rates is carried out simultaneously using EnKF with an augmented state-space based on data on call and put options.

2. METHODS

2.1. The Solution of Black-Scholes Equation

These are assumptions that must be made to use the Black-Scholes equation (1) i.e. the option can only be exercised at the expiration date, as it is a European option, constant composition returns are normally distributed, volatility is known and constant, there are efficient markets, the risk-free rate is known and constant, no taxes or transaction costs, there are no dividends during the life of the option. Let $C(S, t)$ is the call option price. The appropriate boundary conditions for $C(S, t)$ are $C(S, T) = \max\{S(T) - X, 0\}$ and $C(0, t) = 0$ for every $0 \leq t \leq T$. For the large value of S , the $C(S, t)$ value will be approached to S . For these boundary conditions, the $C(S, t)$ function that fulfill equation (1) is

$$C(S, t) = SN(d_1) - Xe^{-r(T-t)}N(d_2) \tag{2}$$

where $d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$, $d_2 = d_1 - \sigma\sqrt{T-t}$, and $N(d_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_i} e^{-\frac{z^2}{2}} dz$ is a standard normal distribution function. Function $C(S, t)$ satisfies the boundary conditions for the call option [1].

Let $P(S, t)$ is the put option. The boundary condition for $P(S, t)$ are $P(S, T) = \max\{X - S(T), 0\}$ and $C(0, t) = Xe^{-r(T-t)}$ for every $0 \leq t \leq T$. For the large value of S , the $P(S, t)$ value will be approached to 0. For this boundary condition, the $P(S, t)$ function that fulfill equation (1)

$$P(S, t) = Xe^{-r(T-t)}N(-d_2) - SN(-d_1). \tag{3}$$

Function $P(S, t)$ satisfies the boundary conditions for the put option. The solution of equation (1) for V is

$$V(S, t) = \frac{e^{(\sigma^2 - 2r)(T-t)}}{S}, \text{ for } V(S, t) = S \text{ and } V(S, t) = e^{rt}. \tag{4}$$

In this study, the state equation at time k has two variables: static and dynamic. In general, static variables consist of parameters to be estimated, namely volatility (v) and interest rates (r), while dynamic variables consist of observations used, namely the price of the call option (C) that follows equation (2) and the price of the put option (P) that follows equation (3). The value of the static variable does not change during the iteration of the EnKF algorithm, so it is expressed $v_k = v_{k-1}$ for volatility and $r_k = r_{k-1}$ for interest rates, but the value of dynamic variables is always updated as long as the algorithm is run. Therefore $C_k = f(v_{k-1}, r_{k-1})$ and $P_k = g(v_{k-1}, r_{k-1})$. Thus, a state-space can be formed as follows:

$$\text{State: } X_k = \begin{bmatrix} C_k \\ P_k \end{bmatrix} = \begin{bmatrix} f(v_{k-1}, r_{k-1}) \\ g(v_{k-1}, r_{k-1}) \end{bmatrix} + \begin{bmatrix} \eta_{k-1} \\ \varepsilon_{k-1} \end{bmatrix}, \quad \begin{bmatrix} \eta_{k-1} \\ \varepsilon_{k-1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}\right)$$

$$\text{Observation: } Y_k = \begin{bmatrix} C_k \\ P_k \end{bmatrix} = HX_k + \begin{bmatrix} \rho_k \\ \xi_k \end{bmatrix}, \text{ with } H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} \rho_k \\ \xi_k \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}\right)$$

If we insert a static variable (the variable to be estimated) into this state-space, an expanded state space will be formed and this state space will then be used to estimate volatility and interest rates simultaneously based on the call option price (C) and the put option price (P). Therefore, the augmented state-space representation is:

$$\text{State: } X_k = \begin{bmatrix} v_k \\ r_k \\ C_k \\ P_k \end{bmatrix} = \begin{bmatrix} v_{k-1} \\ r_{k-1} \\ f(v_{k-1}, r_{k-1}) \\ g(v_{k-1}, r_{k-1}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \eta_{k-1} \\ \varepsilon_{k-1} \end{bmatrix}, \quad \begin{bmatrix} \eta_{k-1} \\ \varepsilon_{k-1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}\right)$$

$$\text{Observation: } Y_k = \begin{bmatrix} C_k \\ P_k \end{bmatrix} = HX_k + \begin{bmatrix} \rho_k \\ \xi_k \end{bmatrix}, \quad \text{with } H = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} \rho_k \\ \xi_k \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}\right)$$

where η_k and ε_k are the model errors, while ρ_k and ξ_k are the observation errors, assuming that the model error and the observation error are independent of each other.

3. RESULTS

The observation is generated using parameters $v_{true} = 0.3$ and $r_{true} = 0.01$. The observation profile of the price of call options and put options used during the analysis stage is shown in Figure 1. The value S used to calculate the observed price for the call and put option is obtained using the equation

$$S(t_2) = S(t_1) e^{\left(\mu - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma \sqrt{t_2 - t_1} Z},$$

with $\mu = 0.05$ and $\sigma = 0.3$.

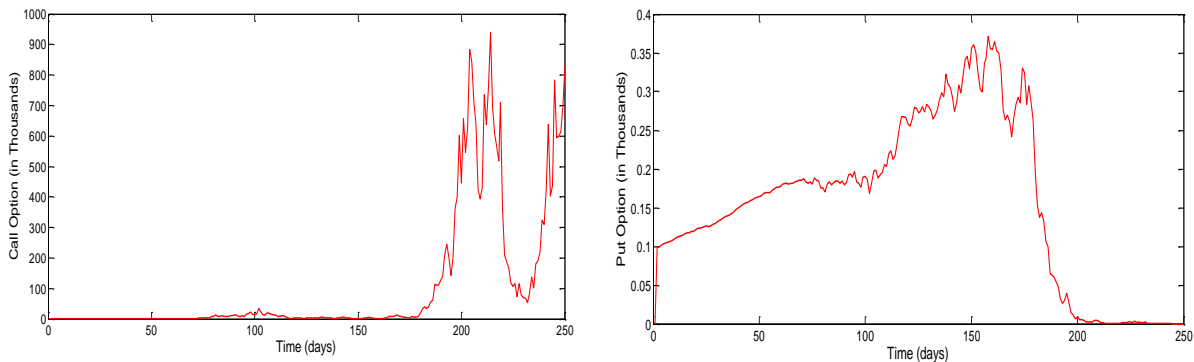


Figure 1. The observation of put options price (left) and observation of call options price (right).

The EnKF algorithm was carried out for $T = 250$ days with 250 observations and an ensemble size $n = 100$. When running the EnKF algorithm, the most common difficulty is guessing the distribution of the original ensemble. This guess can be greater or less than the actual parameter distribution. The initial ensemble distribution is generated from the normal distribution with the following parameters:

$$X_0^j = \begin{bmatrix} v_0^j \\ r_0^j \\ C_0^j \\ P_0^j \end{bmatrix} \sim N \left(\begin{bmatrix} 0.7 \\ 0.03 \\ 0.03 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.7 \end{bmatrix} \right)$$

In the EnKF scheme, the forecasting stage is to estimate the values that exist in the state vector which will then be updated or the dynamic variable values will be updated using observations at the analysis stage. The use of observations at the analysis stage is called data assimilation. The values in this state vector will continue to be corrected using sequential observations until all observations are assimilated. The improvement in the estimated volatility value is shown in Figure 2 (left). This figure shows that the volatility will be estimated sequentially based on the match between the update of the call option price and the put option with the available observations. The same applies to the estimated interest rates shown in Figure 3 (left). In both figures, it can be seen that initially, the estimated parameter values are following the distribution in the initial ensemble, then these values will be updated by assimilating the observations at the analysis stage, and at the end of the iteration, the values of volatility and interest rates will approach the real values, namely, v_{true} and r_{true} .

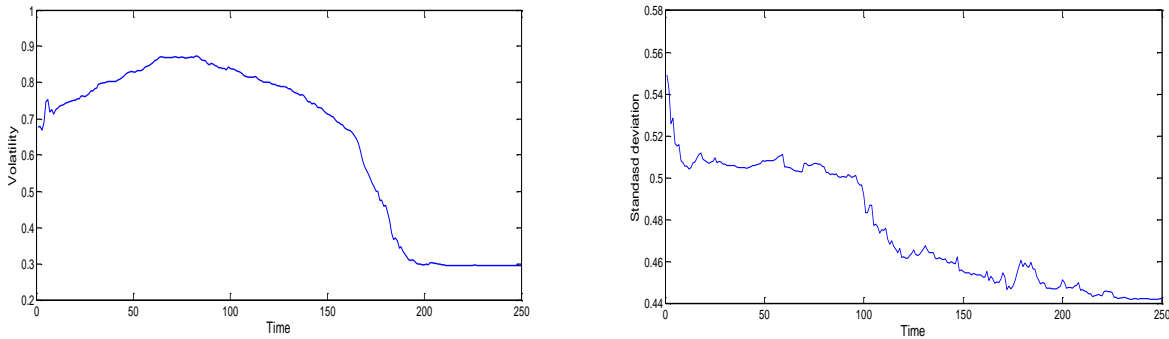


Figure 2. Estimation of volatility based on the observation of put and call options (left) and standard deviation of volatility estimates (right).

Figures 2 and 3 show a large variation in volatility and interest rates at the beginning of the iteration and then shrink after being corrected at the analysis stage. This improvement was caused by assimilating the two observations. From these two figures it can be seen that at the end of the iteration, the estimated volatility and interest rates are close to the real values, namely, v_{true} and r_{true} . This shows that the standard deviation will also have the same pattern, namely shrinking at the end of the iteration (figure 2 (right) and figure 3 (right)).

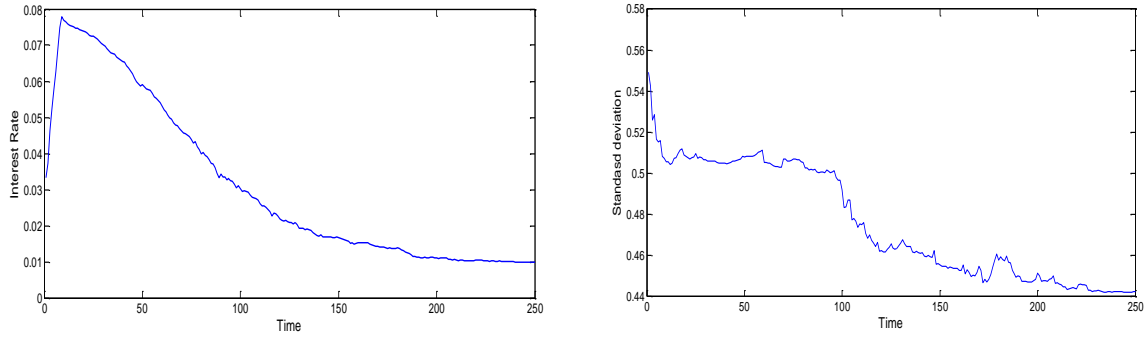


Figure 3. Estimation of interest rate based on the observation of put and call options (left), the standard deviation of interest rate estimates (right).

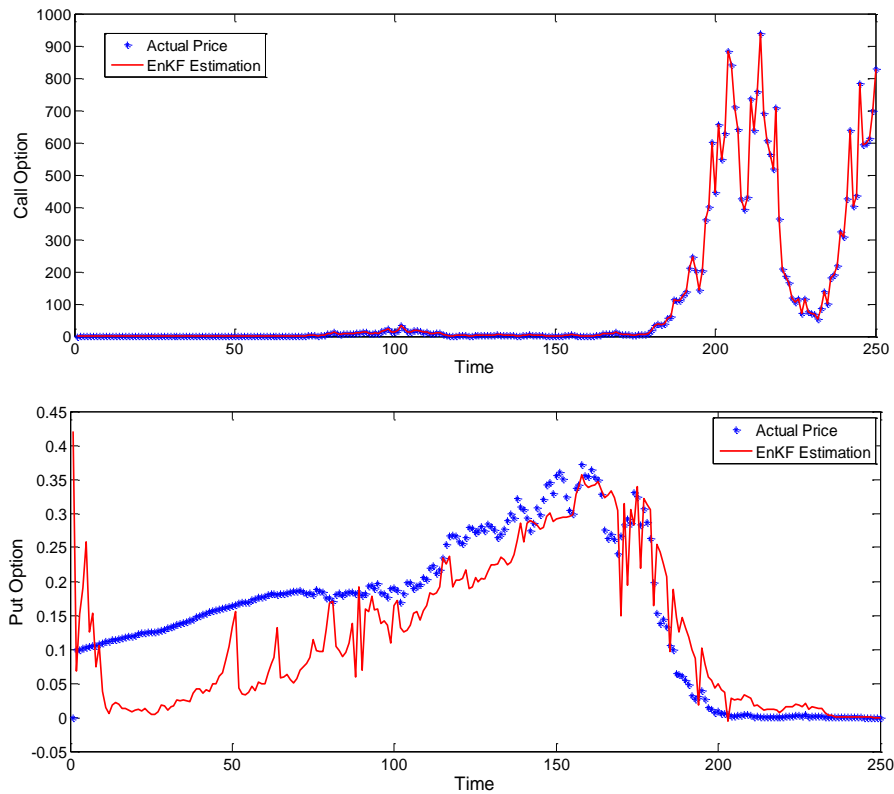


Figure 4. Update of call option value generated by EnKF vs observation (above), update of put option value generated by EnKF vs observation (below).

Figure 4 (above) shows the compatibility of the call option price update (the estimated call option value generated by EnKF) with the available call option observation. From this figure, it can be seen that the update of the call option price is very close to the observation, but not the update of the put option value. Figure 4 (below) shows the compatibility of the call option price update with the available call option observation. The update of the put option value at the beginning of the iteration is quite far from the observation and becomes closer to the observation at the end of the iteration. To assess the merits of this value closeness, one simple formula that we can use is RMSE (Root Mean Square

Error). The RMSE for the updated value of the observed call option is 0.0506 and the RMSE for the put option value is 41.4808. The RMSE value of the call option is higher than the RMSE value of the put option, this can be seen clearly from the two figures above.

4. CONCLUSION

In this research, a methodology for estimating the value of volatility and interest rates simultaneously using the Ensemble Kalman Filter (EnKF) has been built based on the observation of the price of the call option and the put option. The price of the call and put options used in the observation and forecast stage in the EnKF algorithm are calculated using the Black-Scholes equation solution. The state-space used is an extended state space consisting of static variables and dynamic variables. The static variable consists of the estimated parameters, namely the volatility (v) and the interest rate (r), and the dynamic variable consists of the observations used, namely the call option price (C) and the put option price (P). Based on these two variables, the state vector that can be formed is $X_k = (v_k \quad r_k \quad C_k \quad P_k)^T$.

Numerical experiments are carried out to test the ability of the methodology that has been developed to estimate volatility and interest rates. Numerical experiments show that the EnKF algorithm can estimate accurately the estimated volatility and interest rates approaching the real value at the end of the iteration and the call option update value at the analysis stage approaches the purchase option observation value with an RMSE value of 0.0506. The update of the put option value at the beginning of the iteration is a bit far from observation and becomes closer at the end of the iteration so that the resulting RMSE is higher than the RMSE update of the call option, namely 41.4808. However, this does not reduce the robustness of the methodology that has been developed as an alternative methodology for estimating volatility and interest rates.

REFERENCES

- [1] D. J. Higham, *An Introduction to Financial Option Valuation*, United Kingdom: Cambridge University Press, 2004.
- [2] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy*, vol. 81, no. 3, pp. 637-654, 1973.
- [3] F. E. Racicot and R. Theoret, "Forecasting Stochastic Volatility using the Kalman Filter: An Application to Canadian Interest Rates and Price-Earnings Ratio, AESTIMATIO," *The IEB International Journal of Finance*, vol. 1, pp. 28-45, 2010.
- [4] A. E. Ahmed and S. Z. Suliman, "Modelling Stock Market Volatility Using GARCH Models Evidence From Sudan," *International Journal of Business and Social Science*, vol. 2, no. 23, pp. 114-128, 2011.
- [5] R. J. Elliot, T. K. Siu and E. S. Fung, "Filtering a Nonlinear Stochastic Volatility Model," *Nonlinear Dyn*, vol. 67, pp. 1295-1313, 2012.
- [6] I. Burtnyak and A. Malyska, "CEV Model with Stochastic Volatility," *Journal of Vasyly Stefanyk Precarpathian National University*, vol. 6, no. 3-4, pp. 22-28, 2019.

- [7] J.-M. Kim, C. Jun and J. Lee, "Forecasting the Volatility of the Cryptocurrency Market by GARCH and Stochastic Volatility," *Mathematics*, vol. 9, no. 1614, <https://doi.org/10.3390/math9141614>, 2021.
- [8] K. Nguyen, T.-N. Nguyen and M.-N. Tran, "A dynamic leverage stochastic volatility model," *Applied Economics Letters*, vol. DOI: 10.1080/13504851.2021.1983127, <https://doi.org/10.1080/13504851.2021.1983127>, 2021.
- [9] M. Alghalith, C. Floros and K. Gkillas, "Estimating Stochastic Volatility under the Assumption of Stochastic Volatility of Volatility," *Risks*, vol. 8, no. 35, doi:10.3390/risks8020035 , 2020.
- [10] B. E. Awashie, Pricing Financial Options Using Ensemble Kalman Filter, Ghana: Master Thesis of Kwame Nkrumah University of Science and Technology, 2012.