Combing Extended Kalman Filters and Support Vector Machines for Online Option Price Forecasting

Shian-Chang Huang

Department of Business Administration

National Changhua University of Education, Taiwan

Abstract

This study combines extended Kalman filters (EKFs) and support vector machines (SVMs) to implement a fast online predictor for option prices. The EKF is used to infer latent variables and makes a prediction based on the Black-Scholes formula, while the SVM is employed to capture the nonlinear residuals between the actual option prices and the EKF predictions. Taking option data traded in Taiwan Futures Exchange, this study examined the forecasting accuracy of the proposed model, and found that the hybrid model is superior to traditional feedforward neural network models, which can significantly reduce the root-mean-squared forecasting errors.

Keywords: Online forecast, Extended Kalman filter, Support vector machine, Feedforward neural network

1. Introduction

The valuation of financial assets and the pricing of financial derivatives are one of the most important areas in financial studies in recent years. Evidence of this importance can be found in the large number of publications. Moreover, option markets are the most popular shares of financial institutions. To explore the markets well and improve investment yields, modelling and online forecasting of option prices is very important for practitioners.

Due to the high risk in trading options, an accurate online forecast of option prices is important for an investor to control and hedge his risk. Traditionally, the literature of online forecasting of option prices focus on using Kalman filters or extended Kalman filters to infer those latent variables in the Black-Scholes model and then make predictions. However, option prices are nonlinear functions of the underlying stock price process. Traditional Kalman filters require linearity and Gaussian conditions of the measurement and state processes. Thus, it's not suitable for online option price forecasting. Extended Kalman filters use a first order linear approximation to simplify nonlinear functions encountered in the measurement equation. But there are still some nonlinear residuals which can't be captured by the EKF, thus one needs a hybrid model to reduce the prediction errors.

In recent years, neural networks (NN) has found useful applications in financial time-series analysis and forecasting. However, there are several disadvantages for traditional NN models: (I) dependency on a large number of parameters, e.g,

network size, learning parameters and initial weights chosen, (II) possibility of being trapped into local minima resulting in a very slow convergence, and (III) over-fitting on training data resulting in a poor generalization ability.

Recently, the support vector machine (SVM) method (Vapnik [8], Cristianini, N. and J. Shawe-Taylor [3], Schoelkopf, Burges, and Smola [7]), another form of neural networks, has been gaining popularity and has been regarded as the state-of-the-art technique for regression and classification applications. It is believed that the formulation of SVM embodies the structural risk minimization principle, thus combining excellent generalization properties with a sparse model representation. SVM is based on some beautifully simple ideas and provide a clear intuition of what learning from examples is all about. More importantly they are showing high performances in practical applications. Support vector regression (SVR), as described by Vapnik [8], exploit the idea of mapping input data into a high dimensional reproducing kernel Hilbert space (RKHS) where a linear regression is performed. The advantages of SVR are: the presence of a global minimum solution resulting from the minimization of a convex programming problem; relatively fast training speed; and sparseness in solution representation.

Different forecasting models can complement each other in capturing different patterns appearing in one data set, and both theatrical and empirical studies have concluded that a combination of forecast outperforms individual forecasting models ([4, 1, 5]). Typical examples of combining forecasting are Terui and Dijk [6] and Zhang [9]. Terui and Dijk [6] presented a linear and nonlinear time series model for forecasting the US monthly employment rate and production indices. Their results demonstrated that the combined forecasts outperformed the individual forecasts. Zhang [9] combined the ARIMA and feedforward neural networks models in forecasting typical time series data, and showed the superiority of the combined forecasting model.

This study presents a hybrid model that combines EKFs and SVMs to forecast option prices. In the hybrid approach, the EKF serves as a state estimator and makes predictions based on the Black-Scholes formula. The residuals between the actual prices and the Black-Scholes model are fed into the SVM in the hybrid model, and the SVM is conducted to further reduce the prediction errors. Empirical results of this study demonstrated that the hybrid model outperforms another hybrid model based on EKFs and feedforward neural

networks (EKF+FFNN). The empirical results revealed that the EKF can't capture all of the patterns in the option prices, and the performance of EKF+FFNN is also not good. Only the new hybrid model (EKF+SVM) can significantly reduce the forecasting errors.

The remainder of the paper is organized as follows. Section 2 describes the option price modelling and forecasting methods, including the BS model and the SVM. Section 3 describes the hybrid prediction model, and displays the empirical results with discuss on some empirical findings. Conclusions are given in Section 4.

2. Option Price Modelling and Forecasting

Black and Scholes [2] established the price of an european call option through a well known formula, which is the solution to a second order partial differential equation. This closed analytical solution conferred elegance to the proposed formulation and multiplied in extensive and complementary studies. Black and Scholes [2] assumed that the underlying stock price follows a geometric Brownian motion with constant volatility,

$$\frac{dS}{S} = \mu dt + \sigma dW_t, \tag{1}$$

where μ is the expected return and W_t is the Brownian motion. In their paper, Black and Scholes derived the following partial differential equation which governs prices of a call option or a put option,

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf, \tag{2}$$

where r is the riskless interest rate. The solutions for the call and put option prices are thus

$$C = SN(d_1) - Ke^{-r\tau}N(d_2), \tag{3}$$
$$P = -SN(-d_1) + Ke^{-r\tau}N(-d_2), \tag{4}$$

where C is the call option price, P the put option price, K the strike price, and $\tau = T - t$ the maturity. Parameters d_1 and d_2 are as follows:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}, \tag{5}$$
$$d_2 = d_1 - \sigma\sqrt{\tau}. \tag{6}$$

The above equations can be represented as a state-space form:

$$\begin{aligned} C_t, P_t &= F(S_t, K, T - t, r(t), \sigma(t)) + \epsilon_t, \\ r(t) &= r(t-1) + \varsigma_t, \\ \sigma(t) &= \sigma(t-1) + \eta_t, \end{aligned}$$

where r, σ can be viewed as the hidden states. C, P the output observations are a nonlinear function of $S_t, K, T - t, r(t), \sigma(t)$. τ, K are treated as the parameters or the input signals. Consequently, an EKF is employed to approximate F linearly. The EKF infers the latent states $r(t), \sigma(t)$ and simultaneously makes predictions.

2.1. Support Vector Machines

The support vector machines (SVMs) were proposed by Vapnik [8]. Based on the structured risk minimization (SRM) principle, SVMs seek to minimize an upper bound of the generalization error instead of the empirical error as in other neural networks. Additionally, the SVMs models generate the regress function by applying a set of high dimensional linear functions. The SVM regression function is formulated as follows:

$$y = w\phi(x) + b, \tag{7}$$

where $\phi(x)$ is called the feature, which is nonlinear mapped from the input space x to the future space. The coefficients w and b are estimated by minimizing

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}||w||^2, \tag{8}$$

where

$$L_\varepsilon(d, y) = \begin{cases} |d - y| - \varepsilon & |d - y| \geq \varepsilon, \\ 0 & \text{others}, \end{cases}, \tag{9}$$

where both C and ε are prescribed parameters. The first term $L_\varepsilon(d, y)$ is called the ε-intensive loss function. The d_i is the actual option price in the ith period. This function indicates that errors below ε are not penalized. The term $\frac{C}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i)$ is the empirical error. The second term, $\frac{1}{2}||w||^2$, measures the smoothness of the function. C evaluates the trade-off between the empirical risk and the smoothness of the model. Introducing the positive slack variables ξ and ξ^*, which represent the distance from the actual values to the corresponding boundary values of ε-tube. Equation (8) is transformed to the following constrained formation:

$$\min_{w,b,\xi,\xi^*} R(w, \xi, \xi^*) = \frac{1}{2}w^T w + C\left(\sum_{i=1}^{N}(\xi_i + \xi_i^*).\right) \tag{10}$$

Subject to

$$w\phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^*, \tag{11}$$
$$d_i - w\phi(x_i) - b_i \leq \varepsilon + \xi_i, \tag{12}$$
$$\xi_i, \xi_i^* \geq 0. \tag{13}$$

After taking the Lagrangian and conditions for optimality, one obtain the following dual problem:

$$\begin{aligned} \max_{\alpha,\alpha^*} R_D(\alpha_i, \alpha_i^*) &= \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i}^{N}(\alpha_i - \alpha_i^*) \\ &- \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j), \end{aligned}$$

with the constraints

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0,$$
$$0 \leq \alpha_i \leq C,$$
$$0 \leq \alpha_i^* \leq C,$$
$$i = 1, ..., N.$$

α_i and α_i^* are called Lagrangian multipliers. The dual representation of the model becomes

$$y = f(x, \alpha, \alpha^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) + b, \qquad (14)$$

where $K(x, x_i)$ is called the kernel function. α_i, α_i^* are the solution to the dual problem and b follows from the complementarity Karush-Kuhn-Tucker (KKT) conditions.

The value of the kernel is equal to the inner product of two vectors x_i and x_j in the feature space, such that $K(x_i, x_j) = \phi(x_i)\phi(x_j)$. Any function that satisfying Mercer's condition (Vapnik [8]) can be used as the Kernel function. The Gaussian kernel function

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \qquad (15)$$

is specified in this study, because Gaussian kernels tend to give good performance under general smoothness assumptions.

3. Experimental Results and Analysis

The data used in this research are the option prices on the Taiwan composite stock index traded in Taiwan Futures Exchange (TWIFEX). The transaction data of call and put option prices from 16 September 2004 to 14 June 2005 with expiration on 15 June 2005 were studied.

The true dynamics of stock prices is not easily modelled. Therefore, a hybrid strategy combining both Black-Scholes model and the support vector regression is a good alternative for forecasting option prices. Compared with general classes of non-parametric regression methods such as global parametric methods (i.e. linear regression), adaptive computation methods (i.e. projection pursuit regression), and neural networks, support vector regression were among the most accurate and efficient regression techniques. In the hybrid model, both the Black-Scholes model and the support vector machine have different capabilities to capture data characteristics in different domains. Thus, the overall forecasting performance can be significantly improved.

The hybrid prediction model of call option price C_t can be represented as follows:

$$\widetilde{C}_t = \widetilde{BS}_t + \widetilde{f}_t, \qquad (16)$$

where $\widetilde{BS}_t$ is the Black-Scholes model prediction, $\widetilde{f}_t$ the nonlinear SVM prediction, and $\widetilde{C}_t$ the overall prediction. Let δ_t represent the residual between the actual option price and the BS model at time t, namely,

$$\delta_t = C_t - \widetilde{BS}_t. \qquad (17)$$

TABLE I

FORECASTING PERFORMANCE OF EVERY MODEL ON THE THREE CALL OPTIONS

	$K = 6200$	$K = 6000$	$K = 5800$
Trivial Predictions	0.00230	0.00416	0.00552
Pure EKF Predictions	0.00229	0.00338	0.00376
EKF+FFNN Predictions	0.00159	0.00252	0.00217
EKF+SVM Predictions	0.00121	0.00191	0.00159

The lagged residual, stock price, strike price, and maturity are fed into the SVM, which is employed to capture the residual dynamics and further reduces the prediction errors, that is,

$$\delta_t = \widetilde{f}(S_t, K, T - t, \delta_{t-1}) + \varepsilon_t, \qquad (18)$$

where ε_t is the final residual.

In the experiment, the EKF was trained in a sequential manner, while the SVM was trained in a batch manner. The last eighty samples of every data set were used as the test set, and one hundred data points before the prediction day were used as the training set for the SVM. The input features of the SVM were stated as in equation (18), The lagged residual δ_{t-1}, stock price S_t, strike price K, and maturity $T - t$.

In this study, only one-step-ahead forecasting is considered. One-step-ahead forecasting can prevent problems associated with cumulative errors from the previous period for out-of-sample forecasting. Traditional performance indices such as MSE (mean squared error), RMSE (root mean squared error), MAE (mean absolute error), and MAPE (mean absolute percent error), can be used as measures of forecasting accuracy. In this study, the RMSE were used to measure the model performance. RMSE is defined as

$$RMSE = \left(\frac{1}{N}\sum_{t=1}^{N}(d_t - \widehat{d}_t)^2\right)^{1/2}, \qquad (19)$$

where N is the number of forecasting periods, d_i is the normalized actual option price at period t, namely, $d_i = \frac{C_i}{K}$ or $\frac{P_i}{K}$, where C_i is the actual call price and P_i the actual put price. $\widehat{d}_t$ is the forecasting price at period t.

Table 1 displayed the forecasting performance of the trivial model (which takes the current option price as the next day prediction), the pure EKF model, and two hybrid models for call option prices with different Ks. Table 2 provided the performance of these models on predicting put option prices. Figures 1 and 2 plotted selected results on the EKF and SVM predictions. Figure 1 plotted the normalized actual prices, and the predicted prices of the EKF on the option of K=6000. Figure 2 displayed the SVM predictions on the residuals of Figure 1.

The results in Table 1 and 2 indicated that the performance of the EKF were better than the trivial model. For the two hybrid models, the EKF model served as a state estimator and made predictions based on the BS model. Then, the residuals δ_{t-1} between the actual prices and the BS model predictions

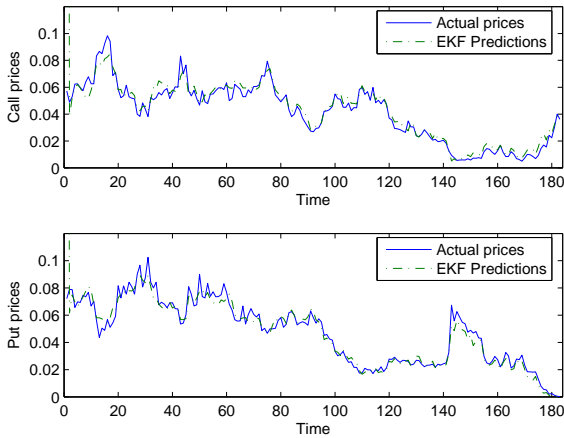|	$K = 6200$	$K = 6000$	$K = 5800$
Trivial Predictions	0.00667	0.00543	0.00379
Pure EKF Predictions	0.00437	0.00346	0.00260
EKF+FFNN Predictions	0.00317	0.00234	0.00169
EKF+SVM Predictions	0.00304	0.00160	0.00095

Fig. 1. **EKF predictions on option prices of** $K = 6000$

Fig. 2. **SVM predictions on the BS model residuals of** $K = 6000$

were fed into the feedforward neural network (FFNN) or the SVM in these hybrid models. FFNN or SVM were conducted to model the residual dynamics and thus further reduce the prediction errors. This study trained the FFNN and SVM by the data including $S_t/K, T - t, \delta_{t-1}$. From Tables 1 and 2, one can compare the forecasting performance of different models. These results indicated that these hybrid model outperformed the pure EKF model in terms of all three options, and the EKF+SVM model outperformed the EKF+FFNN model. These results revealed that neither the EKF nor the FFNN model can capture all of the nonlinear patterns in the option data, but the SVM model with enough flexibility can model these patterns and significantly reduce the overall forecasting errors.

From Figure 1, one can observe that the BS model residuals are still quite nonlinear. Comparing the forecasts of FFNNs and SVMs in these nonlinear residuals, SVMs outperformed FFNNs and were the best choice to capture these nonlinear residual dynamics.

4. Conclusions

This study proposes to hybrid an EKF and a SVM as an option price predictor. The EKF serves as a state estimator and a predictor based on the Black-Schole formula, and the nonlinear price characteristics which can't be captured by the BS model is captured by the SVM. Compared with another hybrid model based on EKF+FFNN, the proposed hybrid model can capture the maximum amount of dynamics in the option prices.
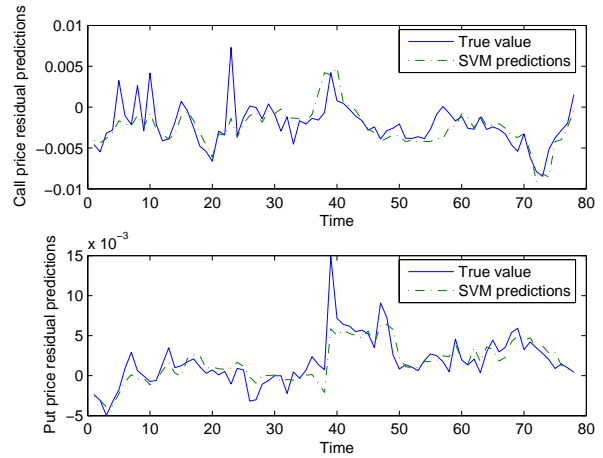
Due to the high risk in trading options, an accurate online forecast of option prices is important for an investor to control and hedge his risk. Options are nonlinear financial instruments. Their price dynamics are quite nonlinear. Linear prediction tools such as the Kalman filter is not a good choice for online forecasting. This study found that the new hybrid model was superior to traditional hybrid models (such as EKF+FFNN). Compared with EKF+FFNN, the new hybrid model significantly reduced the root-mean-squares errors.

The powerful framework established in this study can be applied to high-frequency financial data, and other online forecasting problems. A challenging future task is the Bayesian selection of optimal weighting between the EKF and the SVM by data, and to train the SVM by a Bayesian method to adjust its internal parameters.

REFERENCES

[1] Clemen, R., Combining forecasts: a review and annotated bibliography with discussion. *International Journal of Forecasting*, 1989, 5, 559-608.
[2] Black, F. and Scholes, M. S., The pricing of options and corporate liabilities, *Journal of Political Economy*, 1973, 81, 637-654.
[3] Cristianini, N. and Shawe-Taylor, J., An Introduction to Support Vector Machines, Cambridge University Press, 2000.
[4] Lawerence, M. J., Edmundson, R. H., and O'Connor, M. J., The accuracy of combining judgemental and stastical forecasts, *Management Science*, 1986, 32, 1521-1532.
[5] Makridakis, S., Why combining works?, *International Journal of Forecasting*, 1989, 5, 601-603.
[6] Terui, N. and Dijk, H. K., Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 2002, 18, 421-438.
[7] Schoelkopf, B., Burges, C. J. C., and Smola, A. J., Advances in kernel methods - support vector learning, MIT Press, 1999, Cambridge, MA.
[8] Vapnik, V. N., The Nature of Statistical Learning Theory, 1995, New York, Springer-Verlag.
[9] Zhang, G. P., Times series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, 50, 159-175.