

## KOSPI200 옵션의 내재변동성 추정

최지은<sup>1</sup>, 이장택<sup>2</sup>

<sup>12</sup>단국대학교 응용통계학과

접수 2014년 2월 9일, 수정 2014년 3월 23일, 게재확정 2014년 4월 1일

### 요약

옵션가격의 결정에 있어서 실제 변동성은 사후에 알 수 있는 정보이므로 대용값으로 내재변동성을 가장 많이 사용하는데 본 연구에서는 동일한 기초자산을 가진 옵션의 잔존만기와 행사가격을 이용하여 내재변동성을 추정하고자 한다. KOSPI200 옵션 데이터와 서포트벡터회귀, 나무모형 및 회귀모형을 통해 모형의 설명력을 평균제곱근오차 (RMSE)와 평균절대오차 (MAE)를 사용하여 살펴보았다. 그 결과 서포트벡터회귀와 MART의 성능이 최소제곱회귀보다 우수한 것으로 나타났으며, 서포트벡터회귀와 MART의 성능은 거의 비슷하였다.

주요용어: 내재변동성, 마트, 서포트벡터회귀, 코스피200, 평균절대오차, 평균제곱근오차.

### 1. 서론

#### 1.1. 연구배경

옵션가격의 결정에서 정확한 변동성 지표를 사용하려면 현시점에서 만기까지 기초자산의 변동을 나타내는 실제변동성을 구해야 하겠지만 만기가 되어야 알 수 있으므로 대신할 변동성이 필요한데, 실무에서 실제변동성의 대용값으로 가장 많이 사용되는 것은 내재변동성 (implied volatility)이다 (Kwon과 Lee, 2008).

내재변동성이란 Black과 Scholes (1973)가 제안한 옵션가격 결정모형을 역으로 활용하여 옵션의 시장가격으로부터 얻은 변동성으로, 옵션가격이 시장에서 결정될 때 시장참여자들의 기대가 반영된 변동성이다. Black과 Scholes (1973)가 제안한 옵션가격 결정모형에서는 동일한 기초자산을 가진 옵션은 항상 동일한 내재변동성을 갖는다고 가정하고 있지만, 실제 시장에서 거래되는 옵션은 행사가격별 내재변동성이 일정하지 않고 오히려 외가격 (out of the money) 옵션이나 내가격 (in the money) 옵션의 내재변동성이 등가격 (at the money) 옵션에 비해 높게 나타나는 변동성 미소 (smile) 현상이 널리 관측되고 있기 때문에 내재변동성을 기존의 옵션가격 결정모형 방법으로 정확히 계산할 수가 없다. 따라서 본 연구에서는 이러한 문제점을 최대한으로 극복하기 위해 새로운 방법으로 내재변동성을 추정하고 정확성을 검증해 보고자 한다.

<sup>1</sup> (448-701) 경기도 용인시 수지구 죽전동 126번지, 단국대학교 정보통계학과, 석사.

<sup>2</sup> 교신저자: (448-701) 경기도 용인시 수지구 죽전동 126번지, 단국대학교 응용통계학과, 교수.  
E-mail: jtlee@dankook.ac.kr

## 1.2. 연구 방법

Black과 Scholes (1973)는 유럽형 옵션의 가격  $C$ 를 결정하는 공식을 6개의 변수로 구성되는 함수인 기초자산의 가격 (the stock price)  $S$ , 콜옵션의 행사가격 (the strike price)  $K$ , 무위험이자율 (the risk free interest rate)  $r$ , 옵션의 잔존만기 (the time to the expiration of the option)  $T$ , 지급되는 배당률 (the dividend rate paid by the stock)  $q$ , 기초자산가격의 변동성 (the volatility of the stock)  $\sigma$ 로 제안하였다. 구체적인 함수는 식 (1.1)과 같으며,  $\Phi(\cdot)$ 은 표준정규분포의 분포함수를 의미한다.

$$C = Se^{-qT}\Phi(d_1) - Ke^{-rT}\Phi(d_2) \quad (1.1)$$

$$d_1 = \frac{\ln(S/K) + (r - q + \sigma^2/2)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}$$

일반적으로 기초자산가격의 변동성  $\sigma$ 의 추정치는 과거 데이터의 표준오차를 추정하여 구한다. 하지만 과거의 패턴이 앞으로도 지속된다는 보장이 없기 때문에 Black과 Scholes의 이론이 맞다고 가정하고 식 (1.1)을 이용하여 내재변동성 (implied volatility)이라고 지칭되는  $\sigma$ 를 추정하는 방법을 가장 많이 선호한다. 그러나 Black과 Scholes의 이론은 서로 다른  $K$ 와  $T$ 에 대해서 만기일까지  $r$ 과  $S$ 의 변동성은 변하지 않는다고 가정하지만 실제로  $\sigma$ 와  $S$ 의 산점도를 작성하여 보면  $K$ 가 다른 옵션들의 내재변동성  $\sigma$ 가 서로 다른 현상이 나타난다. 이와 같은 이유로  $\sigma$ 를  $K$ 와  $T$ 의 함수로 간주하여 변수들 간의 관계를 가장 잘 설명하는 함수를 찾는 것이 본 연구의 목적이다. 내재변동성의 추정에 관한 대표적인 해외연구들은 Rubenstein (1994)의 내재된 이항나무모형, Heston (1993)의 확률적 변동성모형, Duan (1995)의 GARCH 모형의 활용 등이 있다.

본 연구에서는 먼저 실제 시장에서 거래되는 옵션 KOSPI200지수 옵션의 일일 종가와 KOSPI200 지수, 3년 만기 국고채의 일일금리를 이용하여  $S, C, K, T, r, q$ 의 값을 식 (1.1)에 대입한 후, 역계산을 통하여 내재변동성  $\sigma$ 를 구한 후, 구해진  $\sigma$ 를 행사가격  $K$ 와 잔존만기  $T$ 의 함수로 간주하여 서로 다른 3가지 방법을 사용하여 추정하고 성능을 비교하였다. 내재변동성은 서포트벡터회귀 (support vector regression; SVR), 나무모형 (multiple additive regression trees; MART) 및 최소제곱회귀모형 (ordinary least squares; OLS)을 이용하여 추정하였으며, 모형의 성능을 비교하기 위해서 훈련자료를 이용하여 모형을 개발한 후, 평가자료를 모형에 적합하였다. 모형의 평가는 RMSE (root mean squared error)와 MAE (mean absolute error)와 같은 2가지 지표를 사용하였으며 분석에 사용한 프로그램은 MATLAB과 R을 사용하였다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 모형이론에 대한 소개로 SVR, MART와 OLS에 대해 약술하고, 3절에서는 3가지 방법을 이용하여 실제 KOSPI200 옵션 자료에 적용한 결과를 비교, 분석하였다. 마지막으로 4절에서는 결론을 제시하였다.

## 2. 모형 이론

본 연구에서 내재변동성의 추정을 위해 사용한 방법은 다음과 같다.

### 2.1. 서포트벡터회귀

Vapnik (1995, 1998)에 의하여 이론이 정립된 서포트벡터기계 (support vector machine; SVM)는 통계학, 생물정보학, 금융정보학 등의 여러 분야에서 분류와 함수근사를 위한 우수한 성능의 기법으로 많이 활용되고 있다. SVM은 연속적인 종속변수의 값을 예측하거나 설명하기 위한 회귀분석에도 널리 사용되는데, 이것을 서포트벡터회귀 (support vector regression; SVR)라고 한다. SVR에 대한 기본

내용은 Vapnik (1995, 1998) 등에 자세하게 설명되어 있으며, SVM의 응용에 대한 자세한 내용은 Kim 등 (2008), Jo 등 (2010), Hwang과 Shim (2010, 2011)에 설명되어 있다.

SVR은 독립변수들로 구성된 입력벡터  $\mathbf{x}_t \in R^p$ 와 종속변수  $y_t \in R, (t = 1, \dots, n)$ 로 이루어진 데이터에서 다음과 같은 형태로 주어진다.

$$f(\mathbf{x}) = \mathbf{w}'\phi(\mathbf{x}) + b, \quad \mathbf{w} \in R^p, \quad b \in R \quad (2.1)$$

식 (2.1)에서 함수  $\phi$ 는 현재의 입력 공간에서 더 높은 차원의 새로운 공간으로 대응시키는 비선형 특징사상함수를 의미하며,  $\mathbf{w}$ 와  $b$ 는 추정될 회귀계수들이다. Vapnik (1995, 1998)은  $K(\mathbf{x}_t, \mathbf{x}_s) = \phi(\mathbf{x}_t)'\phi(\mathbf{x}_s)$ 를 만족하는 커널함수  $K$ 와  $e$ -무감각 손실함수 ( $e$ -insensitive loss function)를 사용하여 다음과 같은 목적함수의 최소화를 고려하는 SVR을 제안하였다.

최소화 목적함수:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{t=1}^n (\xi_t + \xi_t^*) \quad (2.2)$$

제약조건:

$$y_t - \mathbf{w}'\phi(\mathbf{x}_t) - b \leq e + \xi_t, \quad \mathbf{w}'\phi(\mathbf{x}_t) + b - y_t \leq e + \xi_t^*, \quad \xi_t, \xi_t^* \geq 0, \quad i = 1, \dots, n. \quad (2.3)$$

식 (2.2)와 (2.3)을 이용하여 라그랑주 (Lagrange) 함수를 작성하면 다음과 같다.

$$L = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{t=1}^n (\xi_t + \xi_t^*) - \sum_{t=1}^n \alpha_t (e + \xi_t - y_t + \mathbf{w}'\phi(\mathbf{x}_t) + b) - \sum_{t=1}^n \alpha_t^* (e + \xi_t^* + y_t - \mathbf{w}'\phi(\mathbf{x}_t) - b) - \sum_{t=1}^n (\eta_t \xi_t + \eta_t^* \xi_t^*) \quad (2.4)$$

여기서 제약조건  $\alpha_t, \alpha_t^*, \eta_t, \eta_t^* \geq 0$ 는 라그랑주 배수이며, 라그랑주 함수  $L$ 을 주 변수 ( $\mathbf{w}, b, \xi_t, \xi_t^*$ )로 각각 편미분하면 다음과 같은 결과를 얻는다.

$$\mathbf{w} = \sum_{t=1}^n (\alpha_t - \alpha_t^*) \phi(\mathbf{x}_t), \quad \sum_{t=1}^n (\alpha_t - \alpha_t^*) = 0, \quad 0 \leq \alpha_t, \alpha_t^* \leq C. \quad (2.5)$$

식 (2.5)의 편미분 결과를 식 (2.4)에 대입하고, 커널함수를 이용하면 라그랑주 배수들은 다음 목적함수를 최대화함으로써 구해진다.

최대화 목적함수:

$$-\frac{1}{2} \sum_{t=1}^n \sum_{s=1}^n (\alpha_t - \alpha_t^*) (\alpha_s - \alpha_s^*) K(\mathbf{x}_t, \mathbf{x}_s) - e \sum_{t=1}^n (\alpha_t + \alpha_t^*) + \sum_{t=1}^n y_t (\alpha_t - \alpha_t^*) \quad (2.6)$$

제약조건:

$$\sum_{t=1}^n (\alpha_t - \alpha_t^*) = 0, \quad 0 \leq \alpha_t, \alpha_t^* \leq C, \quad i = 1, \dots, n. \quad (2.7)$$

따라서  $f(\mathbf{x})$ 의 최적값은 다음과 같이 표현된다.

$$f(\mathbf{x}) = \sum_{t=1}^n K(\mathbf{x}, \mathbf{x}_t) (\alpha_t - \alpha_t^*) + b \quad (2.8)$$

여기서 편의  $b$ 는 KKT 조건 (Karush-Kuhn-Tucker conditions; Kuhn과 Tucker, 1951)으로부터 다음과 같이 구해진다.

$$b = \frac{1}{n_s} \sum_{t \in I_s} [y_t - K(\mathbf{x}_t, X)(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - e \times \text{sign}(\alpha_t - \alpha_t^*)] \quad (2.9)$$

여기서  $X = \{\mathbf{x}_t\}_{t=1}^n$ ,  $\boldsymbol{\alpha}^* = (\alpha_1^*, \dots, \alpha_n^*)'$ 이고  $n_s$ 는  $I_s = \{t = 1, \dots, n \mid 0 < \alpha_t^* < C\}$ 의 크기이다.

## 2.2. MART

MART (multiple additive regression trees)는 예측을 위한 데이터 마이닝 기법 중의 하나이다. Friedman (1999, 2001, 2002)에 의해 개발된 이 알고리즘은 나무모형의 정확도를 높이는 부스팅 (boosting) 알고리즘을 최적화 알고리즘의 하나인 그래디언트 강하 (gradient descent) 알고리즘으로 해석하여 개발된 그래디언트 부스팅 (gradient boosting)을 이용하는 기저 학습기 (base learner)가 소수의 최종 노드를 가지는 이진나무모형인 앙상블 기법 (ensemble learning method)이다.

종속변수  $y$ 와 독립변수  $\mathbf{x} = \{x_1, \dots, x_p\}$ , 전체 자료 중에서 추출된 훈련자료  $\{y_i, \mathbf{x}_i\}_1^N$ 에 대하여 일반적으로 회귀함수  $\hat{f}(\mathbf{x})$ 를 추정하기 위해서는 모든  $(y, \mathbf{x})$ 값의 결합분포에 대한 어떤 손실함수  $L(y, f(\mathbf{x}))$ 의 기댓값을 최소화하는 함수  $f^*(\mathbf{x})$ 을 이용한다.

$$\hat{f}(\mathbf{x}) \cong f^*(\mathbf{x}) = \arg \min_{f(\mathbf{x})} E_{y, \mathbf{x}} L(y, f(\mathbf{x})) \quad (2.10)$$

기술적으로 부스팅은 추정치  $\hat{f}(\mathbf{x})$ 를 다음과 같은 형태로 간주하는데,

$$\hat{f}(\mathbf{x}) = \sum_{m=0}^M \beta_m h(\mathbf{x}; \mathbf{a}_m) \quad (2.11)$$

식 (2.11)에서 기저 학습기 함수  $h(\mathbf{x}; \mathbf{a})$ 는 모수가  $\mathbf{a} = \{a_1, a_2, \dots\}$ 인  $\mathbf{x}$ 의 단순함수로 주어진다. 확장 계수  $\{\beta_m\}_0^M$ 와 모수  $\{\mathbf{a}_m\}_0^M$ 는 모두 결합적으로 훈련 자료에 적합된다. 먼저 초기값  $f_0(\mathbf{x})$ 를 가정하고,  $m = 1, 2, \dots, M$ 에 대하여 다음 식들을 순차적으로 계산한다.

$$(\beta_m, \mathbf{a}_m) = \arg \min_{\beta, \mathbf{a}} \sum_{i=1}^N L(y_i, f_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; \mathbf{a})) \quad (2.12)$$

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; \mathbf{a}_m) \quad (2.13)$$

여기서  $L(y, f(\mathbf{x}))$ 는 식 (2.10)에서 정의한 손실함수이다.

그래디언트 부스팅은 근사적으로 식 (2.12)를 임의의 손실 함수  $L(y, f(\mathbf{x}))$ 에 대하여 2단계에 걸쳐 푸는데, 첫째로 함수  $h(\mathbf{x}; \mathbf{a})$ 를 식 (2.14)와 같이 정의되는 유사 잔차 (pseudo residual)에 대하여

$$\tilde{y}_{im} = - \left[ \frac{\partial L(y_i, f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)} \right]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})} \quad (2.14)$$

식 (2.15)를 최소제곱법을 이용하여 적합시킨다.

$$\mathbf{a}_m = \arg \min_{\mathbf{a}, \rho} \sum_{i=1}^N [\tilde{y}_{im} - \rho h(\mathbf{x}_i; \mathbf{a})]^2 \quad (2.15)$$

그 다음 주어진  $h(\mathbf{x}; \mathbf{a}_m)$ 에 대하여 계수  $\beta_m$ 의 최적값은 다음 식에 의해 결정된다.

$$\beta_m = \arg \min_{\beta} \sum_{i=1}^N L(y_i, f_{m-1}(\mathbf{x}_i) + \beta h(\mathbf{x}_i; \mathbf{a}_m)) \quad (2.16)$$

따라서 그래디언트 부스팅은 잠재적 난해 함수의 최적화 문제인 식 (2.12)를 제곱오차 손실  $L(y, f) = (y - f)^2$  또는 최소절대편차 손실  $L(y, f) = |y - f|$ 과 같은 일반화된 손실함수  $L$ 을 기저로 식 (2.16)와 같은 단순 모수 최적화 문제로 해결할 수 있다. 목표 함수  $f^*(\mathbf{x})$ 는 정의된 손실 함수  $L$  아래에서  $y|\mathbf{x}$ 의 최적 예측자 (best predictor)로 정의된다. MART는 기저 학습자  $h(\mathbf{x}; \mathbf{a})$ 가  $J$ -최종노드 회귀 또는 분류 회귀일 때 그래디언트 부스팅으로 접근하는 경우로 기본함수  $f(\mathbf{x})$ 에 대한 초기 추정치로부터 순차적으로 나무 모형을 더해 나간다. MART의 성능은 나무모형이 훈련 자료의 부분집합에서 임의로 선택된 유사 잔차에 대해 적합되는 확률적 요소의 도입으로 향상된다. 확률적 요소의 도입은 모형이 과적합되는 현상을 줄여주며 이상치의 영향을 덜 받게 해준다.

### 2.3. 최소제곱회귀

식 (2.17)과 같이 표시되는 종속변수  $y$ 를 설명할 수 있는  $p$ 개의 독립변수  $(x_1, \dots, x_p)$ 를 포함하는 회귀모형을 다중선형회귀모형이라고 한다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (2.17)$$

여기서  $\beta_0, \beta_1, \dots, \beta_p$ 는 회귀계수이고,  $\epsilon$ 는 확률오차이며 평균이 0이고 서로 독립이며 분산이  $\sigma^2$ 을 따른다고 가정한다. 회귀계수는 일반적으로 최소제곱법을 이용하여 추정하는데, 회귀계수의 최소제곱추정량은 최량선형불편추정량 (best linear unbiased estimator; BLUE)이 되기 때문이다.

## 3. 실증 분석

### 3.1. 자료 구성

KOSPI200은 한국주가지수선물 및 옵션의 거래대상으로 개발된 주가지수로서, 증권시장에 상장된 전체 종목 중에서 200종목으로 구성되었다. 본 연구에서 이용한 자료는 KOSPI200 지수가 기초자산인 KOSPI200 콜옵션으로, 최종거래일에만 권리행사가 가능한 유럽형 옵션이며, 2007년 1월 2일부터 2009년 5월 29일까지 거래가 발생한 KOSPI200 옵션의 일별 종가이다. 그리고 옵션 거래량이 없는 자료와 극단적인 옵션가격으로 인한 내재변동성의 과대추정을 피하기 위해서 잔존만기가 6일 미만 또는 260일 초과한 옵션은 분석에서 제외하였으며, 내재변동성은 MATLAB의 blsimpv 함수를 이용하여 산출하였는데 blsimpv 함수가 지정된 반복수 안에서 해를 찾지 못한 경우도 분석에서 제외하였다. 따라서 처음 고려된 40,013개의 자료 중에서 최종적으로 37,139개의 자료를 이용하여 분석을 실시하였다. 또한 내재변동성 산출 시에는 무위험 이자율이 필요한데, 본 연구에서는 한국거래소 (KRX)에서 발표한 3년 만기 국고채의 일일 이자율 (단위: 연리)로 대체하였다.

### 3.2. 분석 과정

내재변동성을 예측한 모형의 타당성을 높이기 위해 자료를 훈련자료와 평가자료로 8:2의 비율로 랜덤하게 분할하여 훈련자료만을 이용하여 모형을 개발한 후 평가 자료를 모형에 적용하는 과정을 10번 반복하였다. 한편  $i$ 번째 내재변동성을  $\sigma_i$ , 내재변동성의 추정량을  $\hat{\sigma}_i$ , 총 내재변동성의 개수를  $n$ 이라고 두면, 모형을 비교하는 평가지표로는 다음 식 (3.1)에 정의되는 RMSE (root mean squaresd error)와

MAE (mean absolute error)을 사용하였으며, 추정량들의 RMSE와 MAE를 구해서 구한 값이 가장 작은 것이 제일 좋은 추정량이라 할 수 있다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\sigma_i - \hat{\sigma}_i)^2}{n}}, \quad MAE = \frac{\sum_{i=1}^n |\sigma_i - \hat{\sigma}_i|}{n} \quad (3.1)$$

고려된 세 가지 모형의 모수들은 R (ver 2.12.2)를 사용하여 추정하였는데, SVR은 e1071 패키지의 svm 함수, MART는 mart 패키지의 mart 함수, OLS는 lm 함수를 각각 사용하였다. SVR을 사용하는 경우에 커널함수의 선택은 매우 중요한데 일반적으로 가장 많이 사용하는 커널함수인  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|)$ 로 정의되는 가우시안 RBF (Gaussian radial basis function)을 사용하였으며, SVR의 성능에 영향을 미치는 모수  $\gamma$ ,  $e$ 와 상수  $C$ 를 다음과 같이 설정하고 가장 좋은 성능을 보이는 3가지 모수의 값을 찾았다. 모수 값의 후보는 3개 모수의 영향을 다룬 Tay와 Cao (2001)를 참조하였으며, 좀 더 다양한 모수들의 조합을 고려하면 보다 개선된 결과를 SVR이 제공할 것으로 간주된다.

$$\gamma = \{0.1, 0.5, 0.75, 1\}, \quad C = \{1, 10, 100, 1000\}, \quad e = \{0.01, 0.05, 0.1, 0.5\}$$

수행결과는 한번 분석할 때마다  $4^3$ 개의 모형이 생성되고 RMSE가 가장 작은 모형을 선정하였다. 대입하는 모수의 값들은 여러 번의 반복 시행을 통해 임의로 지정하였으며, 자료를 랜덤하게 훈련자료와 평가자료로 분할하고 분석하는 작업을 10번 반복하였다. 그 결과 10번 모두 같은 모수 값인  $\gamma = 1$ ,  $C = 1000$ ,  $e = 0.5$ 일 때 SVR의 RMSE가 가장 작았다.

MART 모형에서는 입력된 자료가 함수 내부에서 임의로 훈련자료와 평가자료로 분할되며 훈련자료는 모형을 구축하기 위한 용도로만 사용되고, 평가자료는 모형의 반복횟수에 따른 MAE를 계산하여 가장 좋은 모형을 찾는데 사용된다. MART에서는 반복횟수가 계속 증가하면 훈련자료에 과적합된 모형이 개발되어서 평가자료에 모형을 적용했을 때 예측력이 떨어지는 문제가 발생하기 때문에 모형 구축에 쓰이지 않은 평가자료의 MAE가 가장 작은 모형을 선택한다. 본 연구의 경우에는 반복수가 200회를 초과하면 평가자료의 예측력이 감소하는 것을 알 수 있었는데, 최적모형이 도출되는 반복횟수는 25회에서 191회 사이로 나타났다.

OLS에서는 다양한 가설을 가정하여 행사가격  $K$ 과 잔존만기  $T$ 의 교호작용을 고려한 교호작용  $KT$ 항과 2차항  $K^2$ ,  $T^2$ 을 추가로 다음과 같은 모형을 고려하였다.

$$\sigma = \beta_0 + \beta_1 K + \beta_2 K^2 + \beta_3 T + \beta_4 T^2 + \beta_5 KT + \epsilon \quad (3.2)$$

하지만 10번 모두 공선성의 문제가 발생하여서 공선성을 해결한 다음과 같은 최종모형을 구할 수 있었는데, 따라서 본 연구에서는 식 (3.3)을 사용하였다.

$$\sigma = \beta_0 + \beta_1 K + \beta_2 T + \epsilon \quad (3.3)$$

### 3.3. 분석 결과

Table 3.1은 10번에 걸친 평가자료를 이용하여 계산한 모형평가지표 값이며, Table 3.2는 비교의 편이를 위하여 10번의 결과에 대한 평균과 표준오차를 구한 결과인데, \*\*는 가장 좋은 경우, \*는 2번째로 좋은 경우를 의미하며, 괄호 안의 값은 표준오차를 의미한다.

Table 3.2로부터 RMSE는 SVR이 0.1194로 가장 작게 나타났으며 그 다음 순서로는 MART가 0.1199로 SVR과 근소한 차이를 보였으며 OLS는 0.1303으로 상대적으로 성능이 떨어지는 것을 알 수 있었다. 한편 MAE는 MART가 0.0787로 가장 작게 나타났으며, SVR이 0.0797로 근소한 차이로 밀리

**Table 3.1** Estimation of RMSEs and MAEs for three models

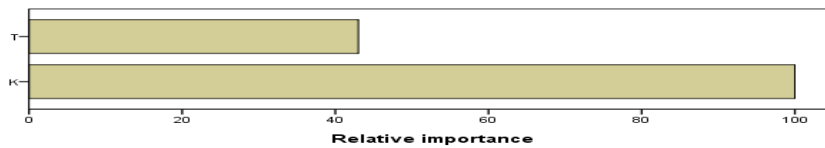
	RMSE			MAE		
	SVR	MART	OLS	SVR	MART	OLS
1	0.1214	0.1213	0.1320	0.0798	0.0787	0.0914
2	0.1179	0.1191	0.1297	0.0787	0.0778	0.0905
3	0.1186	0.1184	0.1289	0.0801	0.0790	0.0918
4	0.1201	0.1198	0.1319	0.0797	0.0786	0.0914
5	0.1237	0.1244	0.1340	0.0809	0.0798	0.0920
6	0.1166	0.1165	0.1273	0.0794	0.0782	0.0912
7	0.1184	0.1189	0.1295	0.0799	0.0791	0.0925
8	0.1221	0.1237	0.1331	0.0811	0.0804	0.0928
9	0.1160	0.1158	0.1280	0.0784	0.0772	0.0908
10	0.1189	0.1207	0.1286	0.0791	0.0784	0.0906

**Table 3.2** Average of RMSEs and MAEs for models (standard errors in parenthesis)

	SVR	MART	OLS
RMSE	0.1194** (0.0023)	0.1199* (0.0026)	0.1303 (0.0022)
MAE	0.0797* (0.0008)	0.0787** (0.0009)	0.0915 (0.0007)

며 역시 OLS가 0.0915로 가장 비효율적임을 알 수 있었다. 일반적으로 SVR은 변수들 사이의 해석이 용이하지 않으나 하지만 MART에서는 모형에서 사용된 독립 변수에 대한 정보를 그림으로 제공한다.

Figure 3.1은 MART에서 사용된 독립변수들 간의 상대적 중요도를 나타낸 것으로 내재변동성의 추정에 있어서 잔존만기  $T$ 는 행사가격  $K$ 에 비해 약 43%의 중요도를 갖는 것으로 나타났다.



**Figure 3.1** Relative importance in MART

Figure 3.2는 MART 모형을 이용하여 행사가격  $K$ 과 잔존만기  $T$ 가 주어졌을 때 추정된 내재변동성의 값을 나타내는데, 이 경우  $d$ 는 나머지 독립변수들의 평균 결합효과를 보정한 후의 여러 가지  $K$ 와  $T$ 값의 종속성을 설명하는 편종속 (partial dependences) 값이다. 행사가격이 증가함에 따라 추정된 내재변동성의 평균은 급격히 감소하다가 180일쯤부터 완만해진다. 반면에 잔존만기는 행사가격보다 전반적으로 완만하게 감소하는 것을 볼 수 있는데, 이를 통해서도 내재변동성의 추정에 있어 잔존만기보다 행사가격이 더 영향을 미치는 변수로 작용하는 것을 알 수 있다.

OLS에서는 앞서 개발한 회귀모형을 바탕으로 회귀계수를 통해 독립변수가 한 단위 변할 때 종속변수의 값이 어느 정도 변화하는지 알 수 있다. 이번 연구에서 개발한 회귀모형의 결정계수는 10개 모형의 평균이 0.3154로, 회귀모형이 자료의 약 32% 정도를 설명한다. Table 3.3은 10개의 OLS 모형에서의 추정된 회귀계수의 평균 및 표본표준오차를 나타낸다. 10개의 OLS 모형에서 모든 회귀계수는 유의수준 0.01에서 유의하고,  $K$ 와  $T$ 에 연관된 회귀계수는 모두 음수였으며, KOSPI200 지수가 기초자산인 KOSPI200 콜옵션은 Black과 Scholes의 내재변동성이 항상 동일하다는 이론과는 달리 행사가격과 잔존만기 둘 다 증가할수록 내재변동성은 감소하였다.

**Table 3.3** Estimated regression coefficient of OLS (standard errors in parenthesis)

intercept	$K$	$T$
0.81093 (0.00298)	-0.00212 (0.00001)	-0.00084 (0.00001)

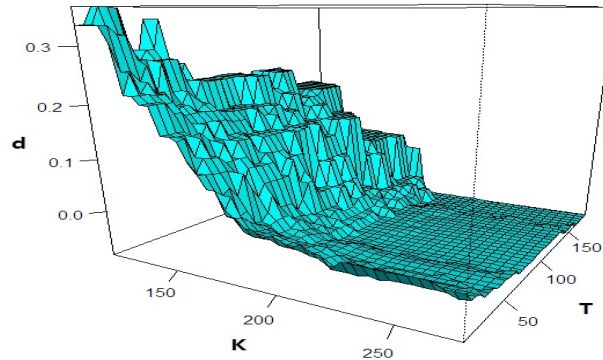


Figure 3.2 Estimation of implied volatility using MART

#### 4. 결론

본 논문에서는 동일한 기초자산을 가진 옵션의 내재변동성이 잔존만기와 행사가격에 따라 변화하는 변동성 미소 현상을 보완하기 위해 3가지 방법으로 어느 모형이 KOSPI200 지수 옵션의 내재변동성 추정에 있어 더 바람직한지 알아보았다. 그 결과 기계학습 방법인 SVR과 MART의 성능은 OLS보다 뛰어났으며, SVR과 MART의 성능 차이는 거의 비슷하였다. 하지만 SVR은 대용량 자료를 분석하는 경우에 있어서 다른 방법들보다 많은 시간을 소모하였으며 또한 모형 내부에 대한 정보를 공개하지 않는 단점을 가지고 있기 때문에 시간도 절약되고, 모형에 대한 설명도 가능한 MART를 사용하는 것을 권장하며, 또한 MART는 부스팅 기법을 사용하기 때문에 이상치의 영향을 다른 방법보다 덜 받아서 가공 안 된 원시자료에 대한 예측력은 다른 방법보다 뛰어날 것으로 간주된다. 한편 선형모형인 OLS의 성능도 RMSE나 MAE의 값이 다른 두 모형에 비해 근소한 차이가 나기 때문에 좀 더 간단하고 알기 쉬운 모형을 개발하고 싶다면 대안으로 고려될 수 있을 것이다. 아울러 내재변동성의 추정에 있어서 MART와 OLS를 통하여 잔존만기보다 행사가격이 더 영향을 미친다는 사실을 확인할 수 있었다.

Black과 Scholes의 내재변동성을 소수의 변수들을 이용하여 추정하는 것은 매우 의미 있는 일이라고 생각되며, 기계학습 방법인 SVR과 MART가 좋은 성능을 보이므로 새로 개발되고 있는 다양한 기계학습 방법을 내재변동성의 추정에 적용하면 옵션거래를 통한 투자이익의 극대화 등과 같은 보다 더 바람직한 결과를 얻을 수 있을 것으로 간주된다.

#### References

- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637-659.
- Duan, J. (1995). The GARCH option pricing model. *Mathematical Finance*, **5**, 13-32.
- Friedman, J. H. (1999). Stochastic gradient boosting, <http://www-stat.stanford.edu/~jhf/ftp/stobst.ps>.
- Friedman J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, **29**, 1189-1232.
- Friedman, J. H. (2002). *Tutorial: Getting started with MART in R*, <http://statweb.stanford.edu/~jhf/r-mart/tutorial/tutorial.pdf>.
- Heston, S. (1993). A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327-343.



- Hwang, C. and Shim, J. (2010). Semiparametric support vector machine for accelerated failure time model. *Journal of the Korean Data & Information Society*, **21**, 467-477.
- Hwang, C. and Shim, J. (2011). Cox proportional hazard model with L1 penalty. *Journal of the Korean Data & Information Society*, **22**, 613-618.
- Jo, D. H., Shim, J. and Seok, K. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of the Korean Data & Information Society*, **21**, 155-162.
- Kim, M., Park, H., Hwang, C. and Shim, J. (2008). Claims reserving via kernel machine. *Journal of the Korean Data & Information Society*, **19**, 1419-1427.
- Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, Berkeley, University of California Press, 481-491.
- Kwon, S. and Lee, J. (2008). The function of intraday implied volatility in the KOSPI200 options. *Asia-Pacific Journal of Financial Studies*, **37**, 913-948.
- Rubenstein, M. (1994). Implied binomial trees. *Journal of Finance*, **49**, 771-818.
- Tay, F. E. H. and Cao, L. J. (2001). Application of support vector machines in financial time series forecasting. *Omega*, **29**, 309-317.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.
- Vapnik, V. N. (1998). *Statistical learning theory*, Springer, New York.

## An estimation of implied volatility for KOSPI200 option

Jieun Choi<sup>1</sup> · Jang Taek Lee<sup>2</sup>

<sup>1,2</sup>Department of Statistics, Dankook University

Received 9 February 2014, revised 23 March 2014, accepted 1 April 2014

### Abstract

Using the assumption that the price of a stock follows a geometric Brownian motion with constant volatility, Black and Scholes (BS) derived a formula that gives the price of a European call option on the stock as a function of the stock price, the strike price, the time to maturity, the risk-free interest rate, the dividend rate paid by the stock, and the volatility of the stock's return. However, implied volatilities of BS method tend to depend on the stock prices and the time to maturity in practice. To address this shortcoming, we estimate the implied volatility function as a function of the strike price and the time to maturity for data consisting of the daily prices for KOSPI200 call options from January 2007 to May 2009 using support vector regression (SVR), the multiple additive regression trees (MART) algorithm, and ordinary least squares (OLS) regression. In conclusion, use of MART or SVR in the BS pricing model reduced both RMSE and MAE, compared to the OLS-based BS pricing model.

*Keywords:* Implied volatility, KOSPI200 option, MAE, MART, RMSE, SVR.

---

<sup>1</sup> Graduate student, Department of Statistics, Dankook University, Yongin 448-701, Korea.

<sup>2</sup> Corresponding author: Professor, Department of Statistics, Dankook University, Yongin 448-701, Korea. E-mail: [jtleed@dankook.ac.kr](mailto:jtleed@dankook.ac.kr)