Journal of Southeast University (English Edition) Vol. 26 No. 3 pp. 489−493 Sept. 2010 ISSN 1003−7985

Recovering implied risk-neutral probability density function using SVR

Hu Xiaoping[1] Cui Hairong[1,2] Zhu Lihua[1] Wang Xinyan[1]

([1] School of Economics and Management, Southeast University, Nanjing 211189, China)
([2] College of Engineering, Nanjing Agricultural University, Nanjing 210032, China)

Abstract: Using support vector regression (SVR), a novel non-parametric method for recovering implied risk-neutral probability density function (IRNPDF) is investigated by solving linear operator equations. First, the SVR principle for function approximation is introduced, and an SVR method for solving linear operator equations with knowing some values of the right-hand function and without knowing its form is depicted. Then, the principle for solving the IRNPDF based on SVR and the method for constructing cross-kernel functions are proposed. Finally, an empirical example is given to verify the validity of the method. The results show that the proposed method can overcome the shortcomings of the traditional parametric methods which have strict restrictions on the option exercise price; meanwhile, it requires less data than other non-parametric methods, and it is a promising method for the recover of IRNPDF.

Key words: support vector regression; option prices; implied risk-neutral probability; linear operator equation; non-parametric method

In recent years, many domestic and foreign research results have shown that the market risk preference (i.e. market sentiment), the equity prices and market returns have a very strong correlation. The market risk preference is a key factor in promoting asset price change, and market information, thus, is implicit in financial asset prices. How to measure the risk preference degree in financial asset prices becomes the subject which researchers have been concerned with.

These approaches to derive risk-neutral probability from observed option prices can be broadly classified as parametric and non-parametric techniques and are reviewed by Jackwerth[1]. Parametric methods choose a distribution family (or a mixture of distributions) and then try to identify the parameters for those distributions that are consistent with the observed prices. In order to achieve this method, continuous exercise prices from zero to infinity are required. However, the exercise price of option trading is discrete and in the limited range. Therefore, at present, most efforts are made on inferring and estimating the entire distribution through the exercise price interpolation within and outside the scope[2-4]. In addition, one can also specify the random process of the parameters, and then recover the parameters from the observed market option prices, while the risk-neutral probability can be inferred by the random process. For example, the classic Black-Scholes option pricing model assumes that stock prices follow the geometric Brownian motion. For example, Duffie et al.[5-6] put forward the stochastic volatility model with a jump-diffusion process.

Another method involves non-parametric techniques. In the absence of the relevant asset price random process, option pricing formula and prior restrictions on the price distribution, non-parametric techniques seem more flexible. For example, Aït-Sahalia and Lo[7] provided a non-parametric option pricing formula of kernel regression, and thus obtained risk-neutral probability. Hutchinson et al.[8-10] used a neutral network to non-parametric option pricing. Haven et al.[11] used the wavelet analysis method to derive risk-neutral probability functions from option prices. In addition, Galati et al.[12] used the Hermite polynomial method proposed by Milne and Madan[13] to recover the US dollar/Japanese yen forward exchange rate risk-neutral probability distribution density function. Corrado et al.[14-15] used the Edgeworth series expansion method to recover the risk-neutral probability density function of the futures. This method can derive the implied risk-neutral probability distribution density function of the lognormal distribution as a reference distribution.

According to the existing literature[1, 16], the common idea to recover implied risk-neutral probability is to use option cross-sectional data to explore state price through the discovery function of the option to the price. Parametric methods depend on the process of option prices, and have strict assumptions. Although non-parametric methods do not have too many restrictions on the data, generally large amounts of data are required in order to recover the risk-neutral probability and results are often poor under the conditions of small samples. In this paper, a new non-parametric method is put forward to extract the risk-neutral probability using the support vector regression technique, and the feasibility of the method is proved empirically. It overcomes the shortcomings of the traditional parametric methods which have strict restrictions on the option exercise price, and meanwhile, it requires less data than other non-parametric methods.

1 Method

1.1 Support vector regression

Support vector regression (SVR) is a statistical or machine learning theory based on classification[17]. To illustrate SVR, a typical regression problem is formulated. Regression is to obtain the relationship between input and output according to input-output data set (x_i, y_i) $(i=1, 2, \dots, l)$,

Received 2010-03-05.

Biography: Hu Xiaoping (1971—), male, doctor, associate professor, hxpnj@163.com.

Foundation item: The National Natural Science Foundation of China (No. 70671025).

Citation: Hu Xiaoping, Cui Hairong, Zhu Lihua, et al. Recovering implied risk-neutral probability density function using SVR[J]. Journal of Southeast University (English Edition), 2010, 26(3): 489−493.

where x_i is a vector of the model input, q_i is the actual value and represents the corresponding scalar output, and l is the total number of data patterns. The objective of the regression analysis is to determine a function $f(x)$, so as to accurately predict the desired (target) outputs q.

In support vector regression, first the inputs are nonlinearly mapped into a high dimensional feature space F wherein they are correlated linearly with the outputs. SVR considers the following linear estimation function:

$$f(x)= w\cdot\Phi(x)+b \tag{1}$$

where w is the weight vector, b is a constant, $\Phi(x)$ denotes a mapping function in the feature space, and $w\cdot\Phi(x)$ describes the dot production in the feature space F. In SVR, the problem of nonlinear regression in the lower dimensional input space x is transformed into a linear regression problem in a high dimensional feature space F.

A number of loss functions such as the Laplacian, Huber's Gaussian and ε-insensitive can be used in the SVR formulation. Among these, the robust ε-insensitive loss function L_ε given below is the most commonly adopted.

$$L_\varepsilon(f(x),q)=\begin{cases}|f(x)-q|-\varepsilon & \text{If } |f(x)-q|\geqslant\varepsilon\\ 0 & \text{Otherwise}\end{cases} \tag{2}$$

where ε is a precision parameter representing the radius of the tube located around the regression function $f(x)$.

The weight vector w and constant b in Eq. (1) can be estimated by minimizing the following regularized risk function,

$$R(C)=C\frac{1}{n}\sum_{i=1}^{n}L_\varepsilon(f(x_i),q_i)+\frac{1}{2}\|w\|^2 \tag{3}$$

where $L_\varepsilon(f(x_i),q_i)$ is the ε-insensitive loss function in Eq. (3); $\frac{1}{2}\|w\|^2$ is the regularization term which controls the trade-off between the complexity and the approximation accuracy of the regression model to ensure that the model possesses an improved generalized performance; C is the regularization constant used to specify the trade-off between the empirical risk and the regularization term. Both C and ε are user-determined parameters.

Two positive slack variables ξ_i and ξ_i^*, $i=1,2,\cdots,l$ can be used to measure the deviation $(q_i-f(x_i))$ from the boundaries of the ε-insensitive zone. Namely, they represent the distance from actual values to the corresponding boundary values of the ε-insensitive zone. By using slack variables, Eq. (3) is transformed into the following constrained form:

$$\min R_{reg}(f)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*) \tag{4}$$

$$\text{s.t.}\quad \begin{aligned} & q_i-w\cdot\Phi(x_i)-b\leqslant\varepsilon+\xi_i\\ & w\cdot\Phi(x_i)-b-q_i\leqslant\varepsilon+\xi_i^*\\ & \xi_i,\ \xi_i^*\geqslant 0\qquad i=1,2,\cdots,l\end{aligned}$$

By using Lagrangian multipliers and Karush-Kuhn-Tucker conditions on Eq. (4), it thus yields the following dual Lagrangian form. Maximize the following function

$$W=-\varepsilon\sum_{i=1}^{l}(\alpha_i^*+\alpha_i)+\sum_{i=1}^{l}q_i(\alpha_i^*-\alpha_i)-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(x_i,x_j) \tag{5}$$

$$\text{s.t.}\quad \begin{aligned} & \sum_{i=1}^{l}\alpha_i^*=\sum_{i=1}^{l}\alpha_i\\ & 0\leqslant\alpha_i^*\leqslant C\qquad i=1,2,\cdots,l\\ & 0\leqslant\alpha_i\leqslant C\qquad i=1,2,\cdots,l\end{aligned}$$

Hence, the general form of the SVR-based regression function can be written as

$$f(x,\alpha,\alpha^*)=\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)K(x_i,x)+b \tag{6}$$

The support vector regression's generalization ability can be controlled (even in high-dimensional space) by controlling the two parameters C and ε.

1.2 Solving linear operator equation

Known linear operator equation

$$Af(t)=F(x) \tag{7}$$

where operator A is one-to-one mapping from the Hilbert space E_1 to the Hilbert space E_2. Eq. (7) will be solved in the following situations: Suppose that the function $F(x)$ on the right of Eq. (7) is unknown, and a number of observed points with the error are known, i.e.,

$$(x_1,F_1),\cdots,(x_l,F_l) \tag{8}$$

Now minimize the functional

$$R_\gamma(f,F)=\frac{1}{l}\sum_{i=1}^{l}L(Af(t)|_{x_i}-F_i)+\gamma(Pf\cdot Pf) \tag{9}$$

in order to estimate the solution of Eq. (7) from the data (8), where $L(Af-F)$ is a loss function, and P is a non-generating operator. Let $\phi_1(t),\cdots,\phi_n(t),\cdots$ and $\lambda_1,\cdots,\lambda_n,\cdots$ is the eigen function and the eigen value of the self-adjoint operator P^*P, i.e., $P^*P\phi_i=\lambda_i\phi_i$.

Expand the solution of Eq. (7) as

$$f(t)=\sum_{k=1}^{\infty}\frac{\omega_k}{\sqrt{\lambda_k}}\phi_k(t) \tag{10}$$

Substitute Eq. (10) into the functional $R_\gamma(f,F)$, i.e.,

$$R_\gamma(f,F)=\frac{1}{l}\sum_{i=1}^{l}L\left(A\left\{\sum_{k=1}^{\infty}\frac{\omega_k}{\sqrt{\lambda_k}}\phi_k(t)\right\}\Bigg|_{x_i}-F_i\right)+\gamma\sum_{k=1}^{\infty}\omega_k^2 \tag{11}$$

and denote

$$\varphi_k(t)=\frac{\phi_k(t)}{\sqrt{\lambda_k}}$$

The above problem can be regarded as the function set

$$f(t,w)=\sum_{r=1}^{\infty} w_r\varphi_r(t)=w\cdot\Phi(t) \tag{12}$$

Minimize the functional

$$R_\gamma(f,F)=\frac{1}{l}\sum_{i=1}^{l} L(|A(w\cdot\Phi(t))|_{x=x_i}-F_i|)+\gamma(w\cdot w) \tag{13}$$

where $w=(w_1,\dots,w_N,\dots)$; $\Phi(t)=(\varphi_1(t),\dots,\varphi_N(t),\dots)$. The operator A maps the function set (12) into the function

$$F(x,w)=Af(t,w)=\sum_{r=1}^{\infty} w_r A\varphi_r(t)=\sum_{r=1}^{\infty} w_r\psi_r(x)=(w,\Psi(x)) \tag{14}$$

which is linear in another feature space $\Psi(x)=(\psi_1(x),\dots,\psi_N(x),\dots)$, where $\psi_r(x)=A\varphi_r(t)$.

To find the solution of Eq. (7) (i.e. to find the coefficient vectors) in the function set $f(t,w)$, minimize the functional in the function set $F(x,w)$ (i.e. the image space),

$$D(F)=C\sum_{i=1}^{l}(|F(x_i,w)-F_i|_\varepsilon)^k+(w\cdot w)\quad k=1,2$$

then define the solution of Eq. (12) (i.e. the preimage space) by using the parameters w. To achieve this method, we use the cross-kernel function in conjunction with the kernel function, thus the kernel function is

$$K(x_i,x_j)=\sum_{r=1}^{\infty}\psi_r(x_i)\psi_r(x_j) \tag{15}$$

and the cross-kernel function is

$$\mathcal{K}(x_i,t)=\sum_{r=1}^{\infty}\psi_r(x_i)\varphi_r(t) \tag{16}$$

Support vectors x_i, $i=1,2,\dots,N$ and the corresponding coefficients $\alpha_i^*-\alpha_i$ can be obtained by using the kernel function (15). The approximation vector of the support vector regression is[17]

$$w=\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)\Psi(x_i)$$

Substituting w into Eq. (12), we obtain

$$f(t,\alpha,\alpha^*)=\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)(\Psi(x_i),\Phi(t))=\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)\mathcal{K}(x_i,t) \tag{17}$$

2 Solving Implied Risk-Neutral Density Function

Extracting the implied risk-neutral probability for call option is to restore the implied risk-neutral probability density function $f(s)$ from $\int_k^{\infty}(s-k)f(s)\,ds=C(k)$ (For convenience, here the discount factor is assumed to be 1), where s is the underlying asset price, k is the exercise price, and $C(k)$ is the option value.

Solving steps are as follows.

Step 1 Define the corresponding regression problem in the image space.

Step 2 Select the kernel function $K(x,y)$. The kernel function of Mikhlin[18]

$$K(x,y)=\sum_{i=0}^{\infty}q^iH_i(x)H_i(y)=\frac{1}{\sqrt{\pi(1-q^2)}}\exp\left\{\frac{2xyq}{1+q}-\frac{(x-y)^2q^2}{1-q^2}\right\} \tag{18}$$

is used in this paper. This kernel function not only has a good smoothness, but also its local approximation performance guarantees approximation accuracy to some extent.

Step 3 Solve the corresponding cross-kernel function $\mathcal{K}(x_i,t)$. By $\psi_r(x)=A\varphi_r(t)$, then $\int_x^{\infty}(t-x)\varphi_r(t)\,dt=\psi_r(x)$, i.e.,

$$\int_x^{\infty}t\varphi_r(t)\,dt-\int_x^{\infty}x\varphi_r(t)\,dt=\psi_r(x) \tag{19}$$

Calculate the derivative of Eq. (19) with respect to x,

$$-x\varphi_r(x)+\int_{\infty}^{x}\varphi_r(t)\,dt+x\varphi_r(x)=\psi_r'(x) \tag{20}$$

i.e., $\int_{\infty}^{x}\varphi_r(t)\,dt=\psi_r'(x)$.

Furthermore, calculate the derivative of Eq. (20) with respect to x,

$$\varphi_r(x)=\psi_r''(x)$$

and then the cross-kernel function is

$$\mathcal{K}(x_i,t)=\sum_{r=1}^{\infty}\psi_r(x_i)\varphi_r(t)=\sum_{r=1}^{\infty}\psi_r(x_i)\psi_r''(t)=\frac{\partial^2}{\partial t^2}K(x_i,t)$$

According to the kernel function (18), we can obtain the cross-kernel function,

$$\mathcal{K}(x_i,t)=\frac{\partial^2}{\partial t^2}K(x_i,t)=\left(-\frac{2q^2}{\sqrt{\pi(1-q^2)}(1-q^2)}+\frac{\left(\frac{2x_iq}{1+q}+\frac{2(x_i-t)q^2}{1-q^2}\right)^2}{\sqrt{\pi(1-q^2)}}\right)\cdot\exp\left\{\frac{2x_itq}{1+q}-\frac{(x_i-t)^2q^2}{1-q^2}\right\} \tag{21}$$

Step 4 Use the support vector regression method and kernel function $K(x,y)$ to solve the regression problem (find support vector x_i, $i=1,2,\dots,N$ and the corresponding coefficients $\beta_i=\alpha_i^*-\alpha_i$, $i=1,2,\dots,N$).

Step 5 Determine the solution using these support vectors and the corresponding coefficients

$$f(t)=\sum_{i=1}^{N}\beta_i\mathcal{K}(x_i,t) \tag{22}$$

3 Empirical Analysis

This article uses 23 options trading data based on different exercise prices on September 12, 2007, provided by the

Chicago Board Options Exchange. Options contracts matured on January 9, 2008, and their underlying asset is the Manitowoc stock.

Support vector regression needs to select two parameters, the penalty parameter C and the error ε of insensitive loss function. This paper selects $C = 50$, $\varepsilon = 0.1$, and lets $q = 8.5 \times 10^{-4}$. We use Matlab (R2008a) and obtain 21 support vectors (91.3%) and coefficients $\beta_i = \alpha_i^* - \alpha_i$, $i = 1, 2, \cdots, 21$. The values of the coefficients are as follows: 50.0000, 50.0000, 50.0000, 50.0000, 50.0000, −15.8684, −50.0000, −50.0000, −50.0000, −50.0000, −50.0000, −50.0000, 0.0000, 50.0000, 50.0000, 50.0000, 35.9814, −0.0000, −50.0000, −50.0000, −2.2380, 50.0000, −15.3312. Fig. 1 shows the relationship of option price C(k) and exercise price k based on support vector regression.

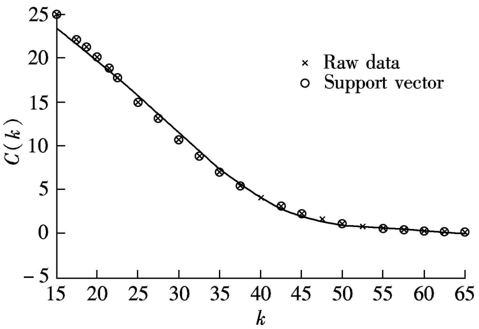

Fig. 1 Support vector regression results of option price C(k)

By using the above coefficients $\beta_i = \alpha_i^* - \alpha_i$, $i = 1, 2, \cdots, 23$ and the cross-kernel function (20) according to $f(t) = \sum_{i=1}^{N} \beta_i K(x_i, t)$, the corresponding implied risk-neutral probability density function is obtained. Fig. 2 shows the implied risk-neutral probability density function f(t). It can be seen from Fig. 2 that the implied risk-neutral probability density function is multi-peak. This result implies that there are inconsistencies between the above implied risk-neutral probability density function and the implied risk-neutral probability based on BS formula, where the implied risk-neutral probability measure is a log-normal distribution. This result also means that market participants' anticipations on the future price of the underlying asset cannot be described by a single-peak-like normal distribution, and should be a multi-peak probability distribution.

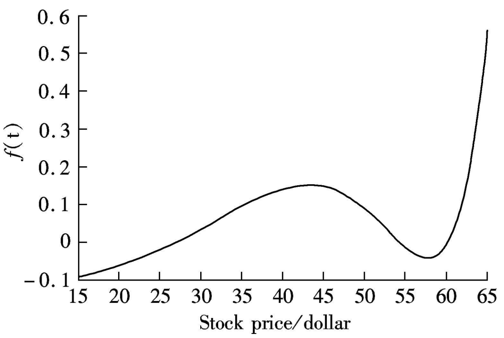

Fig. 2 Implied risk-neutral probability density function f(t)

4 Conclusion

This paper develops and tests a new way of recovering the risk-neutral probability density function (PDF) of an underlying asset from its corresponding option prices. Our approach is a nonparametric method based on support vector regression. The core inversion problem is to solve a linear operator equation. The proposed method can overcome the shortcomings of the traditional parametric methods which have strict restrictions on the option exercise prices. Furthermore, unlike other non-parametric methods, the method requires few amounts of data. The last empirical research proves the feasibility of the proposed method and shows that the probability density function curve is multi-peak. It is of great significance to restore the risk-neutral probability density function.

References

[1] Jackwerth J C. Option implied risk-neutral distributions and implied binomial trees: a literature review [J]. Journal of Derivatives, 1999, **7**(2): 66–82.

[2] Shimko D. Bounds of probability [J]. Risk, 1993, **6**(4): 33–37.

[3] Bates D S. The crash of 87: Was it expected? The evidence from options markets [J]. Journal of Finance, 1991, **46**(1): 1009–1044.

[4] Bliss R, Panigirtzoglou N. Testing the stability of implied probability density functions [J]. Journal of Banking and Finance, 2002, **26**(1): 381–422.

[5] Duffie D, Pan J, Singleton K. Transform analysis and asset pricing for affine jump-diffusions [J]. Econometrica, 2000, **68**(1): 1343–1376.

[6] Melick W R, Thomas C P. Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf crisis [J]. Journal of Financial and Quantitative Analysis, 1997, **32**(1): 91–115.

[7] Ait-Sahalia Y, Lo A W. Nonparametric estimation of state-price densities implicit in financial asset prices [J]. Journal of Finance, 1998, **53**(1): 499–547.

[8] Hutchinson J M, Lo A W, Poggio T. A nonparametric approach to pricing and hedging derivative securities via learning networks [J]. Journal of Finance, 1994, **49**(1): 851–889.

[9] Garcia R, Gencay R. Pricing and hedging derivative securities with neural networks and a homogeneity hint [J]. Journal of Econometrics, 2000, **94**(1/2): 93–115.

[10] Gibson R, Gencay R. Model risk for European-style stock index options [J]. IEEE Transactions on Neural Networks, 2007, **22**(18): 193–202.

[11] Haven E, Liu X, Ma C, et al. Revealing the implied risk-neutral MGF from options: the wavelet method [J]. Journal of Economic Dynamics and Control, 2009, **33**(3): 692–709.

[12] Galati G, Melick W, Micu M. Foreign exchange market intervention and expectations: the yen/dollar exchange rate [J]. Journal of International Money and Finance, 2005, **24**(6): 982–1011.

[13] Milne F, Madan D. Contingent claims valued and hedged by pricing and investing in a basis [J]. Mathematical Finance, 1994, **4**(3): 223–245.

[14] Corrado C J, Su T. Implied volatility skews and stock index skewness and kurtosis implied by S&P 500 index option prices [J]. The Journal of Derivatives, 1997, **4**(4): 8–19.

[15] Flamouris D, Giamouridis D. Estimating implied PDFs from American options on futures: a new semiparametric approach [J]. Journal of Futures Markets, 2002, 22(1): 1-30.
[16] Bahra B. Implied risk-neutral probability density functions from option prices: theory and application[EB/OL]. (1997-07)[2010-02-26]. http://ssrn.com/abstract=77429.
[17] Vapnik V N. Statistical learning theory[M]. New York: Wiley, 1998.
[18] Mikhlin S G. Mathematical physics and technology[M]. London: Pergamon Press, 1964.

基于SVR的隐含风险中性概率密度函数提取

胡小平[1] 崔海蓉[1,2] 朱丽华[1] 王新燕[1]

([1]东南大学经济管理学院,南京 211189)
([2]南京农业大学工学院,南京 210032)

摘要:利用支持向量回归机(SVR),通过求解线性算子方程,提出了一种全新的非参数类恢复隐含风险中性概率密度函数的方法.首先,介绍了支持向量回归机应用于函数逼近的基本原理,当仅知算子方程右边函数的一些函数值而不知其函数形式时,描述了基于支持向量回归机的线性算子方程求解方法.然后,给出了基于支持向量回归机的隐含风险中性概率密度函数求解原理及交叉核函数的构建方法.最后,通过实证研究,验证了该方法的有效性.研究结果表明,所提方法克服了传统参数类方法对期权执行价格有严格限制的缺陷,同时对数据量的要求也比其他非参数类方法少,是一种很有前景的还原隐含风险中性概率方法与手段.

关键词:支持向量回归机;期权价格;隐含风险中性概率;线性算子方程;非参数方法
中图分类号:F830.9