

GARCH prediction using spline wavelet support vector machine

Ling-Bing Tang · Huan-Ye Sheng · Ling-Xiao Tang

Received: 29 January 2008 / Accepted: 29 January 2009 / Published online: 20 February 2009
© Springer-Verlag London Limited 2009

Abstract Volatility forecasting is vital important in finance to reduce risk and take better decisions. This paper proposes a spline wavelet support vector machine (SWSVM) to forecast the volatility of financial time series based on generalized autoregressive conditional heteroscedasticity model. An admissible spline wavelet kernel is constructed by incorporating the wavelet technique and spline theory into support vector machine (SVM). Since spline wavelet function can yield features that describe the stock time series both at various locations and at varying time granularities, the SWSVM gains the cluster feature of volatility well. Compared with Gaussian kernel in the standard SVM, the applicability and validity of spline wavelet kernel in SWSVM are confirmed through computer simulations and experiments on real-world stock data.

Keywords Volatility forecasting · GARCH · Spline wavelet support vector machine

1 Introduction

Time series prediction is important and challenging task in finance. Compared with traditional time series analysis with emphasis on modeling the conditional first moment, Engle [1] and Bollerslev [2] developed generalized autoregressive conditional heteroscedasticity (GARCH) model to take the dependency of the conditional second moments into modeling consideration. Since this accommodates the increasingly important demand to explain and to model risk and uncertainty in financial time series, GARCH model has been main tool for volatility forecasting [3–8]. To enhance the forecasting performance of GARCH farther, Perez-Cruz proposed GARCH-SVM model and proved that forecasting volatility using support vector machine (SVM) is not only feasible but also effective [9].

SVM have originally been used for classification purposes but their principles can be extended easily to the task of regression and time series prediction [10–12]. The prediction performance of SVM is greatly dependent upon kernel [13]. There are many kinds of existent support vector kernels such as the Gaussian and polynomial kernels used to map the data in input space to a high-dimensional feature space in which the problem becomes linearly separable. Since wavelet function can describe stock time series both at various locations and at varying time granularities [14–17], it should describe the cluster feature of volatility well. One of the basic methods for constructing wavelets involves the use of spline functions, which are probably the simplest functions with small supports [18, 19]. Therefore, it is valuable for us to research the problem of whether a desirable performance could be achieved if we combine SVM with spline wavelet theory. In this paper, we construct a novel spline wavelet kernel for SVM to forecast volatility.

L.-B. Tang (✉) · H.-Y. Sheng
Department of Computer Science and Engineering,
Shanghai Jiao Tong University, 800 Dongchuan Rd.,
Minhang District, 200240 Shanghai, China
e-mail: lingbingtang@gmail.com

L.-B. Tang
Department of Computer and Electronic Engineering,
Hunan Business College, Yuelu Rd.,
Yuelu District, 410205 Changsha, China

L.-X. Tang
School of Economics, Changsha University of Science
and Technology, 45 Chiling Rd.,
Tianxin District, 410076 Changsha, China

The objective of this paper is to evaluate the performance of spline wavelet kernel in spline wavelet support vector machine (SWSVM) for volatility prediction according to GARCH model by comparing it with the Gaussian kernel in SVM. This paper is organized as follows: Sect. 2 provides a brief introduction to the multi-resolution theory of wavelets. Section 3 describes how to construct spline wavelet kernel and prove that it is admissible support vector kernel. Section 4 discusses about the experimental results on simulating and real data sets, followed by conclusions in the last section.

2 Multi-resolution theory

Multi-resolution analysis is the decomposition of the Hilbert space $L_2(\mathbb{R})$ to the nested sequence of closed subspaces $\{V_j\}_{j \in \mathbb{Z}}$, which satisfy the following relations [15]:

- (1) $\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots$
- (2) $\text{close} \left\{ \bigcup_{j \in \mathbb{Z}} V_j \right\} = L_2(\mathbb{R}), \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\}$
- (3) $\forall f \in L_2(\mathbb{R})$ and $\forall j \in \mathbb{Z}, \quad f(x) \in V_j \Leftrightarrow f(2x) \in V_{j-1}$
- (4) $\forall f \in L_2(\mathbb{R})$ and $\forall k \in \mathbb{Z}, \quad f(x) \in V_0 \Leftrightarrow f(x - k) \in V_0$
- (5) $\exists \varphi \in V_0$ so that $\{\varphi(x - k)\}_{k \in \mathbb{Z}}$ is the Riesz basis in V_0 .

Dilations and translations of scaling function $\varphi(x)$, the $\{\varphi_{j,k}(x) = \varphi(2^{-j}x - k)\}_{k \in \mathbb{Z}}$ is the Riesz basis in V_j . Let us denote by W_j the complement of the subspace V_j in the space V_{j-1} , there is $V_{j-1} = V_j \oplus W_j, j \in \mathbb{Z}$. Similarly, the Riesz basis of W_j is $\{\psi_{j,k}(x) = \psi(2^{-j}x - k)\}_{k \in \mathbb{Z}}$ by dilations and translations of wavelet function $\psi(x)$.

3 Spline wavelet kernel and SWSVM

Theorem 1: Let $x \in \mathbb{R}$, the scaling function $\varphi^{(N)}(x)$ is B-spline of the order $(N - 1)$ with the compact support interval $[0, N]$ and the integer division $\{x_k = k, k \in \mathbb{Z}\}$, the recursion formula is valid [20]

$$\varphi_k^{(N)}(x) = \frac{x - k}{N - 1} \varphi_k^{(N-1)}(x) + \frac{k + N - x}{N - 1} \varphi_{k+1}^{(N-1)}(x), \quad N = 2, 3, \dots, \quad (1)$$

$$\varphi_k^{(1)}(x) = A_{[k,k+1)}(x) = \begin{cases} 1, & k \leq x < k + 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $\varphi_k^{(N)}(x) = \varphi^{(N)}(x - k)$.

Theorem 2: Compactly supported spline wavelet $\psi^{(N)}(x)$ can be expressed by the wavelet equation $\psi^{(N)}(x) = \sum_k d(k) \varphi^{(N)}(2x - k)$, where scaling function is B-spline $\varphi^{(N)}(x)$ given in (1), and coefficients $d(k)$ are equal to [19]

$$d(k) = \frac{(-1)^k}{2^{N-1}} \sum_{l=0}^N \binom{N}{l} \varphi^{(2N)}(k - l + 1), \quad (2)$$

$$k = 0, \dots, 3 - 2$$

Theorem 3: Let $\psi^{(N)}(x)$ be a mother wavelet, let j and k denote the dilation and translation, respectively. All calculations are done on integer divisions, with the finest resolution level division $\{x_k = k, k = 0, \dots, 2^m\}$. If $s, t \in \mathbb{R}^M$, then dot-product wavelet kernels are:

$$K(s, t) = \prod_{i=1}^M \sum_{j=1}^J \sum_{k=1-N}^{L_j-N} \psi_{j,k}^{(N)}(s_i) \psi_{j,k}^{(N)'}(t_i), \quad L_j = 2^{m-j} \quad (3)$$

Proof: Let $x^1, \dots, x^J \in \mathbb{R}^M$ and $a_1, \dots, a_l \in \mathbb{R}$,

$$\begin{aligned} & \sum_{p,q}^l a_p a_q K(x^p, x^q) \\ &= \sum_{p,q}^l a_p a_q \prod_{i=1}^M \sum_{j=1}^J \sum_{k=1-N}^{L_j-N} \psi_{j,k}^{(N)}(x_i^p) \psi_{j,k}^{(N)'}(x_i^q) \\ &= \left(\sum_p^l a_p \prod_{i=1}^M \sum_{j=1}^J \sum_{k=1-N}^{L_j-N} \psi_{j,k}^{(N)}(x_i^p) \right) \\ & \quad \times \left(\sum_q^l a_q \prod_{i=1}^M \sum_{j=1}^J \sum_{k=1-N}^{L_j-N} \psi_{j,k}^{(N)'}(x_i^q) \right) \\ &= \left(\sum_p^l a_p \prod_{i=1}^M \sum_{j=1}^J \sum_{k=1-N}^{L_j-N} \psi_{j,k}^{(N)}(x_i^p) \right)^2 \\ & \geq 0 \end{aligned}$$

Hence, dot-product kernels satisfy Mercer’s condition. Therefore, it is admissible support vector kernel.

It’s well known to all that the Gaussian kernel need to compute a l by l kernel matrix, $O(l^2)$. The major difference on complexity of spline wavelet kernel is on the constructing stage to gain Multi-resolution frame. Hence, if there is the number h of frame elements describing the frame-based kernel, the computational complexity can be calculated by $O(h^2 \times l^2)$.

4 Experiment analysis

Two simulated data sets are examined in the first series of experiment. Each data set consists of 1,560 samples generated by a GARCH (1,1) model:

$$y_t = \mu + \sigma_t \varepsilon_t \quad (4)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (5)$$

where y_t is daily return and ε_t is innovation, an uncorrelated process with zero mean and unit variance. For the sake of simplicity, the mean of a financial return series μ is often

neglected. In addition, the parameters ω , α and β must satisfy $\omega > 0$, $\alpha, \beta \geq 0$ to ensure that the conditional variance σ_t^2 is positive. The experimental setup is as follows: $\mu = 0$, $\omega = 0.1$, $\alpha = 0.4$ and $\beta = 0.5$ and a disturbance term ε_t distributed first as Gaussian and then as a Student's t with four degrees of freedom (kurtosis = 4). This second distribution tries to model the excess of kurtosis that appears in real financial series. Every time series consists of 1,560 samples. They are referred to as Data-1 and Data-2.

The data examined in the experiment are composed of the following daily indices: DAXINDEX, FRCAC40, FTSE100, JAPDOWA and SPCOMP. These stock market indices P_t are then transformed into daily returns y_t by 100 times their log differences:

$$y_t = 100 \ln(p_t/p_{t-1}) \tag{6}$$

All the index data encompass the period from 1 January 1992 to 31 December 1997. There are 1,560 observations for each time series of daily return. Each whole data set is divided into several overlapping training and testing sets according to the walk-forward testing routine [21]. Each training and test set is moved forward through the time series by 130 observations, in which there are a total of 520 observations in the training set, 520 observations in the test set. The optimal values of C , ε and γ in SVM with Gaussian kernel are chosen based on fivefold cross validation. The same method is also used in the SWSVM to choose C , ε and the dilation j separately. All parameters are shown in Table 1. The results are collated and the best results are recorded as follow, which are gained from the first sets (January 1992–January 1996).

The prediction performance is evaluated using the following statistical metrics: normalized mean squared error (NMSE), normalized mean absolute error (NMAE) and the hit rate (HR). These metrics are calculated as follows:

$$\text{NMSE} = \sqrt{\frac{\sum_{t=1}^N (\hat{\sigma}_t^2 - y_t^2)^2}{\sum_{t=1}^N (y_{t-1}^2 - y_t^2) \sum_{t=1}^N (y_{t-1}^2 - y_t^2)}} \tag{7}$$

$$\text{NMAE} = \frac{\sum_{t=1}^N |\hat{\sigma}_t^2 - y_t^2|}{\sum_{t=1}^N |y_{t-1}^2 - y_t^2|} \tag{8}$$

$$\text{HR} = \frac{1}{N} \sum_{t=1}^N q_t, q_t = \begin{cases} 1 & : (\hat{\sigma}_t^2 - y_{t-1}^2)(y_t^2 - y_{t-1}^2) \geq 0 \\ 0 & : \text{else} \end{cases} \tag{9}$$

where N represents the total number of data points in the test set. $\hat{\sigma}_t^2$ denotes the predicted conditional variance. \hat{y} represents the predicted return. y denotes the actual return. The NMSE relates the mean square error of the predicted volatility $\hat{\sigma}_t^2$ by SVM to the mean square error of the naive model $\hat{\sigma}_t^2 = y_{t-1}^2$. The NMAE is more robust against outliers in comparison with NMSE. They are the measures of the deviation between the actual and predicted values. The smaller the values of them, the closer are the predicted time series values to the actual values. On the contrary, the larger the value of HR as a measure of how often the model gives the correct direction of change of volatility, the better is the performance of prediction.

The smaller cross-validation error is the stronger generalization ability of model is. After we gained parameters by cross-validation method to avoid over-fitting, we firstly inspect the prediction result of training set, which is supplement for prediction result of test set.

The results on the training set are listed in Table 2. It can be observed that in all the daily indices, the smaller values of NMSE are in the spline wavelet kernel. The smaller values of NMAE are in spline wavelet kernel with the exclusion of FTSE100 and JAPDOWA, too. As for HR, only in Data-1 and DAXINDEX the larger values are in Gaussian kernel. A paired-test [22] is performed to determine if there is significant difference between the two kernels based on the NMSE of the training set. The calculated t -value indicates that spline wavelet kernel outperforms Gaussian kernel with 5% significance level for a one-tailed test.

The results on the test set in Table 3 provide a better basis for a comparison of the two kernels where over-fitting issues may be neglected. As expected, the results of the test set are worse than those of the training set in terms of

Table 1 Parameters of model

Stock indices	Gaussian kernel			Spline wavelet kernel		
	C	Epsilon	Gamma	C	Epsilon	j
Data-1	35.1344	0.000011	0.3038	17.5702	0.000081	2
Data-2	13.2920	0.000045	0.7825	23.3347	0.000059	3
DAXINDEX	1.0451	0.012609	0.1589	1.7266	0.003512	3
FRCAC40	11.8356	0.003906	0.3526	1.4822	0.005975	2
FTSE100	6.4264	0.006807	0.6361	2.3901	0.010869	2
JAPDOWA	1.3561	0.012609	0.7324	10.6866	0.015511	2
SPCOMP	1.6044	0.009708	0.3717	3.3726	0.007431	1
Scope	$[2^{-20}, 2^8]$	$[2^{-24}, 2^{-1}]$	$[2^{-20}, 1]$	$[2^{-20}, 2^8]$	$[2^{-24}, 2^{-1}]$	$[1, 4]$

Table 2 Results on the training set

Stock indices	Gaussian kernel			Spline wavelet kernel		
	NMSE	NMAE	HR	NMSE	NMAE	HR
Data-1	0.8844	0.9157	0.6683	0.8731	0.8606	0.6264
Data-2	0.8721	0.9682	0.5134	0.8216	0.9362	0.5832
DAXINDEX	0.7649	0.8486	0.7073	0.7559	0.8390	0.7055
FRCAC40	0.7960	0.9574	0.6525	0.6955	0.7566	0.7228
FTSE100	0.6980	0.6738	0.7883	0.6587	0.7080	0.7893
JAPDOWA	0.8506	0.6556	0.6663	0.8189	0.8205	0.7217
SPCOMP	0.7683	0.9643	0.6457	0.7378	0.8546	0.6971
<i>t</i> -Value	3.3431 > $t_{0.05,6} = 1.9432$					

Table 3 Results on the test set

Stock indices	Gaussian kernel			Spline wavelet kernel		
	NMSE	NMAE	HR	NMSE	NMAE	HR
Data-1	0.6033	0.6417	0.6444	0.6582	0.5952	0.7156
Data-2	0.9360	0.9136	0.4939	0.9035	0.8823	0.5231
DAXINDEX	0.7484	0.7960	0.7502	0.7170	0.8117	0.7689
FRCAC40	0.7724	0.8745	0.6845	0.7058	0.8152	0.7065
FTSE100	0.8296	0.5361	0.7844	0.6893	0.7750	0.7908
JAPDOWA	0.8906	0.7148	0.6246	0.7932	0.6755	0.6948
SPCOMP	0.7780	0.9941	0.6688	0.7229	0.8321	0.6795
<i>t</i> -Value	2.2835 > $t_{0.05,6} = 1.9432$					

NMSE, NMAE and HR. But the similar conclusion can still be achieved. The table illustrates that apart from Data-1, the smaller values of NMSE are founded in spline wavelet kernel. And the smaller values of NMAE are founded in spline wavelet kernel except for DAXINDEX and FTSE100. The larger values of HR all occurred in spline wavelet kernel. And a paired-test on the NMSE of the test set also shows that spline wavelet kernel outperforms Gaussian kernel with 5% significance level for a one-tailed test.

The squared observations y_t^2 and the predicted values of $\hat{\sigma}_t^2$ from both two kernels for the test sets are illustrated in Figs. 1 and 2, where only the FRCAC40 is drawn, since it has well representative values of NMSE and NMAE in all daily indices. In this investigation, all of the parameters are seen in Table 1. It is clear that both Gaussian and spline wavelet kernel are enough to grasp the features reflected by the naive model. The predictions made by both kernels are very similar according to Figs. 1 and 2, although, as we can see in Table 3, performance of spline wavelet kernel is better than that of Gaussian kernel.

All experiments are run on 1.6 MHz Intel processors with 2 GB main memory under window XP professional. Training time in CPU-seconds of SVM is 3.5688 while that of WSVM is $3.1182e + 003$. WSVM is more slowly than

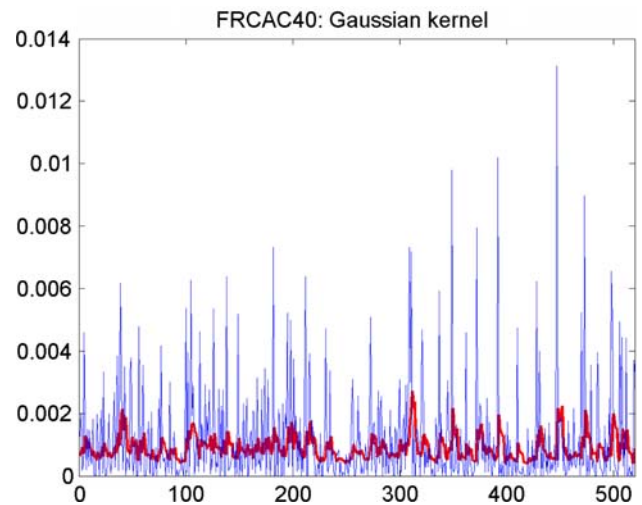


Fig. 1 Squared observations and forecasted volatility by Gaussian kernel

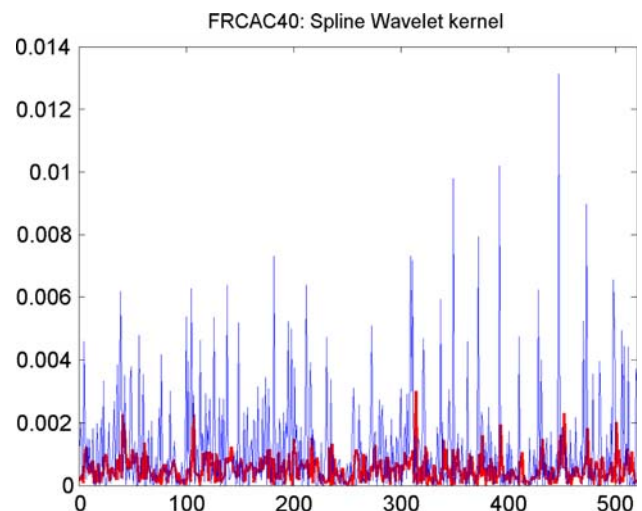


Fig. 2 Squared observations and forecasted volatility by spline wavelet kernel

SVM because better prediction performance gained by constructing wavelet kernel based on multi-scale theory is at the cost of the time consuming.

5 Conclusion and discussion

An effective spline wavelet kernel which we combine spline theory and wavelet method with SVM to construct for volatility forecasting is presented in this paper. The existence of spline wavelet kernel is proven firstly. And then the forecasting performance is evaluated by using two simulated data sets and five real daily indices. As demonstrated in the experiment, spline wavelet kernel forecasts significantly better than Gaussian kernel in all aspects. The

superior performance of spline wavelet kernel to the Gaussian kernel mostly lies in that spline wavelet is a set of bases that can approximate arbitrary functions. Future work will involve a theoretic analysis of the multi-scale frame on spline. More sophisticated spline wavelet kernel which can closely follow the volatility cluster will be explored for further improving the performance of SWSVM in volatility forecast.

Acknowledgments We gratefully acknowledge the many helpful discussions and suggestions we have had with Prof. Dr. Kuno Egle in Department of Econometrics, Statistics and Mathematical Finance, School of Economics and Business Engineering, University of Karlsruhe. We also wish to thank the generous support provided by Baden-Württemberg state government in Germany.

References

- Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* 50:987–1008. doi:[10.2307/1912773](https://doi.org/10.2307/1912773)
- Bollerslev T (1986) A generalized autoregressive conditional heteroskedasticity. *J Econom* 31:307–327. doi:[10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Day TE, Lewis CM (1998) The behavior of the volatility implicit in the prices of stock index options. *J Financ Econ* 22:103–122. doi:[10.1016/0304-405X\(88\)90024-4](https://doi.org/10.1016/0304-405X(88)90024-4)
- Franses PH, van Dijk D (1996) Forecasting stock market volatility using (nonlinear) GARCH models. *J Forecast* 15:229–235. doi:[10.1002/\(SICI\)1099-131X\(199604\)15:3<229::AID-FOR620>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1099-131X(199604)15:3<229::AID-FOR620>3.0.CO;2-3)
- Harvey CR, Whaley RE (1991) S&P100 index option volatility. *J Finance* 46:1551–1561. doi:[10.2307/2328872](https://doi.org/10.2307/2328872)
- Hull J, White A (1987) The pricing of options on assets with stochastic volatilities. *J Finance* 42:281–300. doi:[10.2307/2328253](https://doi.org/10.2307/2328253)
- Li WK, Mak TK (1994) On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. *J Time Ser Anal* 15:627–636. doi:[10.1111/j.1467-9892.1994.tb00217.x](https://doi.org/10.1111/j.1467-9892.1994.tb00217.x)
- Poterba JM, Summers LH (1986) The persistence of volatility and stock market fluctuations. *Am Econ Rev* 76:1142–1151
- Perez-Cruz F, Afonso-Rodriguez JA, Giner J (2003) Estimating GARCH models using support vector machines. *Quant Finance* 3(3):163–172
- Smola AJ, Schölkopf BA (1998) Tutorial on support vector regression. NeuroCOLT technical report NC-TR-98-030. Royal Holloway College, London
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
- Vapnik VN (1998) Statistical learning theory. Wiley, New York
- Schölkopf B, Burges CJC, Smola AJ (1999) Advances in kernel methods. The MIT Press, London
- Daubechies I (1990) The wavelet transform: time-frequency localization and signal analysis. *IEEE Trans Inform Theory* 36(5):961–1005. doi:[10.1109/18.57199](https://doi.org/10.1109/18.57199)
- Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
- Mallat S (1998) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11:674–693. doi:[10.1109/34.192463](https://doi.org/10.1109/34.192463)
- Mallat S (1998) A wavelet tour of signal processing. Academic Press, Boston
- Chui CK (1992) An introduction to wavelets. Academic Press, London
- Chui CK (1997) Wavelets: a mathematical tool for signal analysis. SIAM, Philadelphia
- Zavjalov YS, Kvasov BI, Miroshnicenko VL (1980) Spline function methods. Nauka, Moscow
- Kaastra I, Boyd M (1996) Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10(3):215–236. doi:[10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9)
- Montgomery DC, Runger GC (1999) Applied statistics and probability for engineers. Wiley & Sons, New York