CMS (2006) 3:147–160
DOI 10.1007/s10287-005-0005-5

ORIGINAL PAPER

Support vector machine as an efficient framework for stock market volatility forecasting

Valeriy V. Gavrishchaka · Supriya Banerjee

Published online: 16 February 2006

Abstract Advantages and limitations of the existing models for practical forecasting of stock market volatility have been identified. Support vector machine (SVM) have been proposed as a complimentary volatility model that is capable to extract information from multi-scale and high-dimensional market data. Presented results for SP500 index suggest that SVM can efficiently work with high-dimensional inputs to account for volatility long-memory and multiscale effects and is often superior to the main-stream volatility models. SVM-based framework for volatility forecasting is expected to be important in the development of the novel strategies for volatility trading, advanced risk management systems, and other applications dealing with multi-scale and high-dimensional market data.

1. Introduction

Availability of high-resolution and multi-source data increases in many fields of practical interest including financial industry. However, it is well-known that the majority of advanced statistical and machine learning algorithms, including neural networks (NN), can encounter a set of problems called "dimensionality curse" when applied to high-dimensional data (Bishop 1995). Nonstationarity of the time series can also impose significant limitations on data available for training that often leads to poor generalization ability of the model. The latter feature is especially relevant for financial applications.

A promising algorithm that can tolerate high-dimensional and incomplete data is support vector machine (SVM) (Vapnik 1995, 1998). SVMs have recently been receiving significant interest due to excellent results in various applications (Cristianini and Shawe-Taylor 2000). SVM combines the training efficiency and simplicity of linear algorithms with the accuracy of the best nonlinear techniques as well as systematic approach for optimal generalization. In many practical applications SVMs can tolerate high-dimensional and/or incomplete data and often demonstrate performances superior to the best available techniques including classical NNs (Cristianini and Shawe-Taylor 2000). Recent successful applications of

V.V. Gavrishchaka (✉) · S. Banerjee
Science Applications International Corporation, McLean, VA 22102, USA
E-mail: valeriyg@alexandra.net
E-mail: supriya.banerjee@saic.com

SVM-based adaptive systems include space–weather forecasting (Gavrishchaka and Ganguli 2001), image/object classification (Pontil and Verri 1998), face detection and recognition (Osuna et al. 1997), text categorization (Joachims 1998), process identification in high-energy physics (Vannerem et al. 1999), cancer diagnostic and prognosis (Mangasarian et al. 1995), gene classification (Brown et al. 1999), as well as many other scientific, engineering, medical, and biological applications.

Financial time-series forecasting is another challenging area where advantages of the SVM-based systems could be very important. Although some financial applications of the SVM have been reported (Edelman 2001; Fan et al. 1999; Van Gestel et al. 2001), the full range of potential SVM applications in finance remains largely unexplored. Recently we have proposed SVM-based model for the volatility forecasting from the multi-scale and high-dimensional market data (Gavrishchaka and Ganguli 2003). Application of our model to foreign exchange data have demonstrated that SVM can efficiently extract information from the high-dimensional inputs of lagged returns and in many regimes can outperform the main-stream volatility models.

Similar SVM-based volatility model can also be applied to stock market. Stock volatility is a very important quantity for derivative pricing, value-at-risk calculations in portfolio risk management, and as one of the components used for decision making in trading systems. Although many stylized facts of foreign exchange and stock market data are similar, stock market has important distinct signatures. This includes such universal features as pronounced leverage effect (Bouchaud et al. 2001) and importance of trading volume data that is widely accepted by technical analysts. Individual stock dynamics may also have some specific and less understood signatures as well as significant psychological component in its response to the market news. Due to its ability to extract information from the high-dimensional and incomplete data, SVM is well-suited to incorporate many of the above effects in the unified framework for stock market volatility forecasting.

In this paper we expand our previous work on SVM-based volatility model (Gavrishchaka and Ganguli 2003) to demonstrate its performance on stock market data. Daily closing prices of the SP500 index have been chosen for this purpose. It has been shown that SVM-based volatility model can be comparable and often superior to the main-stream models such as generalized ARCH (GARCH) and its generalizations. Other possible configurations of the SVM-based model for the further improvement of the volatility forecasting are also discussed.

2. Stylized facts of the stock market data

In this section we define the main measures used to characterize financial time series and describe their universal properties revealed in numerous empirical studies. The time series that will be used in this paper is shown in Figure 1a. These are daily closing prices of the SP500 index from 11 October 1999 to 11 November 2003. Nonstationarity of the moving average of the time series is clear from this figure. The more practical quantity is the return given by

$$r_i = \ln(S_i/S_{i-1}), \tag{1}$$

where i is an index of a homogeneous time sequence (e.g., the end of each trading day) and S_i is the closing index or stock price at time t_i. The daily return time series corresponding to Figure 1a is shown in Figure 1b. The moving average of the return time series is almost stationary and close to zero. Another important quantity of the financial time series is volatility. Optimal definition of the realized volatility depends on the particular application and

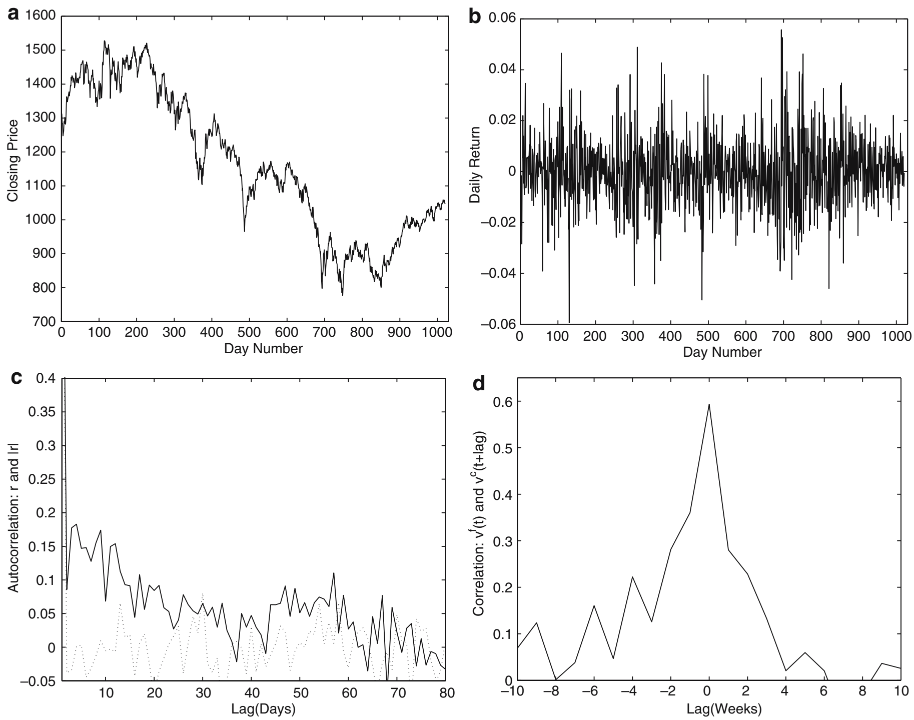

Fig. 1 **a** Closing price. **b** Daily returns. **c** Autocorrelation function of daily returns (*dotted line*) and absolute returns (*solid line*). **d** lead–lagged correlations of the fine and coarse-grain volatilities as a function of positive/negative lag in weeks, obtained from the SP500 index data (from 11 October 1999 to 11 November 2003)

properties of the time series of interest. In many cases realized volatility at time t_i is defined as a standard deviation of returns in some interval $[t_{i-n}, t_i]$. For purposes of this paper we consider realized volatility to be $v_i = |r_i|$ or $v_i = r_i^2$ that is a reasonable choice in many other applications as well (Dacorogna et al. 2001).

Extensive empirical studies of the stock market data revealed several universal or stylized facts. Returns have been found to have only very short-range correlation with typical characteristic time as small as just a few minutes (Bouchaud and Potters 1999; Dacorogna et al. 2001; Mantegna and Stanley 2000). This absence of linear correlation is illustrated in Figure 1c where dotted line represents the autocorrelation function of daily returns computed from SP500 closing price. Here data from Figure 1b have been used. On the other hand volatilities (e.g., represented by absolute values of returns) are clustered and have long-range memory (up to several months) (Bouchaud and Potters 1999; Dacorogna et al. 2001; Mantegna and Stanley 2000). The volatility autocorrelation function exhibits hyperbolic (power-law) behavior. This is illustrated in figure 1c for the same SP500 data (solid line). The other important fact is that probability density function of returns is fat-tailed and leptokurtic at small time scales (from several minutes to several days) and approaches Gaussian at larger scales (Bouchaud and Potters 1999; Dacorogna et al. 2001; Mantegna and Stanley 2000). Volatilities have also been found to be negatively correlated with corresponding returns. This leverage effect characteristic for stock market data has been clarified in recent detailed empirical studies (Bouchaud et al. 2001).

Recently introduced heterogeneous market hypothesis (Dacorogna et al. 2001) suggests that traders (market participants) with different time horizons are interested in the volatility on different time grids. A coarse time grid reflects the view of a long-term trader and a fine time grid that of a short-term trader. The "coarse" (v^c) and "fine" (v^f) volatilities can be defined as

$$v^c(t_i) = |\sum_{j=1}^{n} r(\Delta t^*, t_{i-1} + j\Delta t^*)|, \tag{2}$$

and

$$v^f(t_i) = \sum_{j=1}^{n} |r(\Delta t^*, t_{i-1} + j\Delta t^*)|, \tag{3}$$

where $\Delta t^* = \Delta t/n$, $\Delta t = t_i - t_{i-1}$, the first return argument is the time scale over which return is computed, and the second argument is the time of this return measurement. For example, if we consider weekly volatility measures (on business time scale), then v^c is given by $|\sum_{i=1}^{5} r_i|$ and v^f by $\sum_{i=1}^{5} |r_i|$, where r_i is a daily return at the i-th day.

An important effect found for both foreign exchange (Dacorogna et al. 2001) and stock (Arneodo et al. 1998) markets is asymmetric lead–lag correlation of volatilities. Lagged correlation is a linear correlation of the two time series one of which is shifted (lagged) in time. Lagged correlation reveals causal relations and information flow structures in the market. To illustrate effect of asymmetric volatility correlations we consider fine volatility defined by averaged absolute returns over five working days and coarse volatility defined as absolute return over a full (working) week (i.e., $n = 5$ in equations (2), (3)). Lead–lagged correlations of these volatilities obtained from SP500 closing price (from Figure 1a) are shown in Figure 1d. We see a clear asymmetry: the coarse volatility predicts fine volatility better than the other way around, i.e., information flows from large to small scales. This is consistent with heterogeneous market hypothesis since short-term traders can react to clusters of coarse volatility, while the level of fine volatility does not affect strategies of long-term traders.

In the next section we review some of the existing volatility models. Although there is no universal volatility model that incorporates or explains all of the stylized market facts, different models focus on different set of features that finally determines their accuracy and applicability scope.

3. Limitations of the existing volatility models

There are two general classes of volatility models in a widespread use: deterministic and stochastic models. Deterministic models consider volatility (conditional variance) to be a deterministic function of the past returns (and/or other observables) that are described by some stochastic process (e.g., Wiener process) (Bollerslev 1986; Engle 1982; Engle and Patton 2001; Tsay 2002). Stochastic volatility models describe volatility by its own stochastic process (Masoliver and Perello 2001; Tsay 2002). Stochastic volatility models are more flexible than deterministic models. However, it is significantly more difficult to analyze and calibrate these models from the available market data. Therefore stochastic volatility models can be impractical for some applications and are not yet widely accepted by practitioners. In this paper we will consider only deterministic models.

A common example of the deterministic volatility models is autoregressive conditional heteroskedastic (ARCH)-type models (Bollerslev 1986; Engle 1982). These models assume a particular stochastic process for the returns and a simple functional form for the volatility. Volatility in these models is an unobservable (latent) variable. The most widely used model of this family is GARCH process (Bollerslev 1986). GARCH(p, q) process defines volatility as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \tag{4}$$

where return process defined as

$$r_t = \sigma_t \epsilon_t. \tag{5}$$

Here ϵ_t is an identically and independently distributed (i.i.d.) with zero mean and variance 1. The most common choice for the return stochastic model (ϵ_t) is a Gaussian (Wiener process). However, to take into account realistic fat-tailed return distributions, GARCH model is also used with Student-t distribution of returns. Parameters α_i and β_i from equation are estimated from historical data by maximizing the likelihood function which depends on the assumed return distribution.

GARCH and other ARCH-type processes is the most common choice of the volatility model by practitioners. GARCH process can reproduce a number of known stylized volatility facts including clustering and mean reverting. Explicit specification of the stochastic process and simplified (linear) functional form for the volatility allows to do simple analysis of the model properties and its asymptotic behavior. However assumptions of the ARCH-type models also impose significant limitations. For example, GARCH(p, q) model does not cover leverage and general nonlinear effects. Model parameter calculation from the market data is practical only for low-order models (small p and q), i.e., in general, it is difficult to capture direct long memory effects. As discussed later volatility multiscale effects are also not covered. Finally, the model gives unobservable quantity that leads to difficulty in quantifying the prediction accuracy and comparison with other models.

To resolve some of GARCH limitations, a number of extensions have been proposed and used by practitioners. For example, since GARCH model depends only on the absolute values of returns (r_i^2), it does not cover leverage effect. Different forms of leverage effect have been introduced in EGARCH (Nelson 1991), TGARCH (Zakoian 1994), and PGARCH (Ding et al. 1993) models that are given below.

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|r_{t-i}| + \gamma_i r_{t-i}}{\sigma_{t-i}} + \sum_{i=1}^{q} \beta_i \ln(\sigma_{t-i}^2), \tag{6}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{p} \gamma_i S_{t-i} r_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \tag{7}$$

$$\sigma_t^d = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|r_{t-i}| + \gamma_i r_{t-i})^d + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^d, \tag{8}$$

Here γ_i is a leverage parameter. For TGARCH model $S_{t-i} = 1$ for $r_{t-i} < 0$ and $S_{t-i} = 0$ otherwise. For PGARCH d is a positive number that can also be estimated together with α_i, β_i, and γ_i coefficients. However, majority of the mentioned limitations cannot be resolved in a self-consistent fashion.

A number of nonlinear extensions of the ARCH-type framework have been proposed. For example, Donaldson and Kamstra (1997) proposed NN-based volatility model. They found that a proper modeling of nonlinearities captures volatility effects that are overlooked by traditional models like GARCH and its extensions.

C. Schittenkopf et al. (1998, Submitted) added a detailed analysis of the distributional assumptions underlying NN-based volatility models. They found that models with nonGaussian distributions (mixture of gaussians or Student-t) are superior to those with Gaussian distributions. This is due to heteroskedastic nature of the financial time series and fat-tail nature of return distribution. NonGaussian (mixture of gaussians) models have been formulated as mixture density NNs (Bishop 1995) where appropriate generalization of a simple Gaussian loss function (mean squared error) is made. In some regimes, mixture density NNs have been shown to perform significantly better than GARCH-type models. Potential limitations of the NN-based models can be related to high-dimensional inputs ("dimensionality curse", Bishop 1995) in such applications as small-scale volatility forecasting.

One of the significant limitations of the existing ARCH-type and similar deterministic volatility models is their inability to capture the heterogeneity of traders acting at different time horizons. For example, if the empirical data can be described as generated by one GARCH process at one particular data frequency, the dynamics of the data sampled at any other frequency is theoretically determined by temporal aggregation (or disaggregation) of the original process. These derived processes at different frequencies can be compared to empirically estimated processes at the same frequencies. Significant deviation between theoretical and empirical results reject hypothesis of only one GARCH process responsible for data generation (Dacorogna et al. 2001; Engle and Patton 2001). In other words model parameters obtained for the data of different frequencies are significantly different. It means that there is more than one relevant frequency in the volatility generation. This is manifestation of the presence of many independent volatility components in the data, i.e., the signature of market heterogeneity.

As discussed earlier there is asymmetry in the interaction between volatilities measured at different frequencies (see Figure 1d). A coarsely defined volatility predicts a fine volatility better than the other way around. This effect is not present in a simple GARCH model. More complex types of ARCH models have to be developed to account for the heterogeneity that is especially pronounced in high-frequency data. One of such approaches is the Heterogeneous Autoregressive Conditional Heteroskedasticity (HARCH) model (Dacorogna et al. 2001).

The HARCH process has a variance equation based on multiscale returns, i.e., returns computed over time intervals of different sizes

$$\sigma_t^2 = c_0 + \sum_{j=1}^{n} c_j \left(\sum_{i=1}^{j} r_{t-i} \right)^2, \tag{9}$$

where return process is still given by (5) and c_j are parameters of the model. The terms of (9) reflect the component structure of the market in a natural way. HARCH model is rather different from the typical ARCH model. For example HARCH(2) model can be rewritten in two forms:

$$\sigma_t^2 = c_0 + c_1 r_{t-1}^2 + c_2 (r_{t-1} + r_{t-2})^2, \tag{10}$$

or

$$\sigma_t^2 = c_0 + (c_1 + c_2) r_{t-1}^2 + c_2 r_{t-2}^2 + 2 c_2 r_{t-1} r_{t-2}. \tag{11}$$

The last form (11) can be identified as ARCH(2) model plus an important mixed term $r_{t-1} r_{t-2}$, i.e., signs of returns matter.

HARCH can reproduce empirical behavior of lagged correlations as well as the long memory of volatility. This is a qualitative difference between GARCH model and its variations.

For example, fractionally integrated GARCH (FIGARCH) process (Baillie et al. 1996) has been designed to model the long memory but cannot reproduce the lead–lag correlations since it is still based on returns measured over one time scale.

Although HARCH model is able to capture multiscale nature of volatility, application of the HARCH model in its original form can be computationally prohibitive especially for high-frequency volatility. This is due to many free parameters (corresponding to different market components) that need to be estimated from the market data. For example modeling of the intraday volatility can easily result in hundreds of free parameters since small-scale volatility depends on many larger scale volatilities. To make HARCH model practical, additional restriction on the number of independent market components has to be applied. This is done by clustering adjacent components and assuming the coefficients c_j to be equal across the same cluster. No more than seven clusters (components) are usually considered (Dacorogna et al. 2001).

In the next section we review our SVM-based volatility model (Gavrishchaka and Ganguli 2003) that can relax a number of restrictive assumptions of the GARCH/HARCH models including limitation on the number of independent market components.

4. SVM-based volatility model

Recently we have proposed SVM-based volatility model and applied it to foreign exchange market data (Gavrishchaka and Ganguli 2003). Here the same model is used for stock market volatility forecasting. We start with a short description of our original model.

SVMs developed by Vapnik Vapnik (1995, 1998) have recently been receiving significant interest due to excellent results in various applications (Cristianini and Shawe-Taylor 2000). SVM is a combination of a kernel-based approach and a structural risk minimization (SRM) principle (Vapnik 1995, 1998). First step is a nonlinear mapping from the input to a higher-dimensional feature space. Kernel-based approach allows to represent the discriminant function in high-dimensional feature space without explicit dependence on the feature space dimensionality. Kernel-based machine decouples the number of free parameters (related to the machine capacity) from the size of the input space which can be very large or even infinite. SRM provides solid theoretical grounds for optimization of the SVM generalization ability that is often superior to other approaches used in machine learning algorithms.

In general, training of the SVM for classification and support vector regression (SVR) reduces to the minimization problem with constraints that is a typical quadratic programming problem (Chang et al. 2000). Application of the SVR also involves finding adequate loss function. Loss function should not only be able to correctly approximate noise distribution of the modeled data but also have a suitable form for optimization algorithm used in a particular SVR implementation. The most common choice is the original ϵ-insensitive loss function (ϵ-ILF) (Cristianini and Shawe-Taylor 2000; Vapnik 1995, 1998) which is similar to loss functions used in the field of robust statistics. It has been shown (Pontil et al. 1998) that the use of ϵ-ILF is justified under assumption that the noise is a superposition of Gaussian processes. This noise model is quite suitable for heteroskedastic market data we are interested in, and ϵ-ILF will be used in our volatility model.

The optimization problem for the ϵ-SVR is given by

$$\min_{\alpha,\alpha^*}\left[\frac{1}{2}(\alpha-\alpha^*)^T Q(\alpha-\alpha^*)+\epsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)\right],$$

$$\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)=0,\quad 0\le\alpha_i,\alpha_i^*\le C, i=1,\ldots,l \tag{12}$$

Here $C>0$ is a regularization parameter (soft margins), $(y_i,\mathbf{x_i})$ is a training set, l is a number of training samples, $Q_{ij}\equiv K(\mathbf{x}_i,\mathbf{x}_j)$ is a positive semidefinite matrix, and K is a kernel function representing inner product of the feature vectors. ϵ-ILF is given by $L^\epsilon(x,y,f)=|y-f(x)|_\epsilon=\max(0,|y-f(x)|-\epsilon)$, where $f=\sum_{i=1}^{l}(-\alpha_i+\alpha_i^*)K(x_i,x)+b, x\in R^n$, $y\in R^1$, and $b\in R^1$ is a constant. Approximation function f is equivalent to the hyperplane in the feature space implicitly defined by the kernel K that solves the optimization problem (12).

Although SVM training is a typical quadratic programming problem, due to the specifics of SVM applications such as large data sets and high density of the Q-matrix, standard algorithms can become impractical. Recent developments mainly include algorithms that employ various decomposition techniques (Cristianini and Shawe-Taylor 2000), where at any time a fixed size subset of α_i is updated, while others are kept constant. Various heuristics are used for choosing a working set at each step. Here we use an algorithm described by Chang et al. (2000) and implemented as LIBSVM library (www.csie.ntu.edu.tw/ cjlin/libsvm).

Applicability of the SVM (SVR) model to our problem is based on the assumption that a measure of stock or index volatility σ can be described as a nonlinear function F of a time series of returns r:

$$\sigma_i^2=F[r_{i-1},r_{i-2},...,r_{i-p}], \tag{13}$$

where index $i-j$ correspond to time $(t_i-j\,dt)$, dt is a time lag interval and $T=p\,dt$ is a total length of the memory for previous inputs. In the following, SVM(p) notation will be used for the above model. Since F could be any nonlinear function this framework automatically covers multiscale dependencies in a more general form than HARCH model.

Practical usage of the described SVM models requires specification of the volatility σ in (13). In this paper, we adopt the most common choice as $\sigma_i^2=r_i^2$. However, in general, other practical volatility measures can be used in the described framework. For example, SVM can be trained on σ_i time series that is calculated using intraday return data from day i (Andersen et al. 2000).

We need also to ensure that trained SVM model will always output nonnegative numbers for σ^2. This is achieved by choosing mapping function as

$$\sigma_i^2=\exp(F[r_{i-1},r_{i-2},...,r_{i-p}]). \tag{14}$$

In operation terms it means that SVM is trained on $\ln(r^2)$ instead of r^2, and exp mapping is applied to the SVM output.

Since the main advantage of the SVM is its ability to handle high-dimensional data, SVM-based volatility model can cover long memory and multiscale effects without restrictive assumptions required by other models. For example, unlike HARCH model (Dacorogna et al. 2001), SVM will not require strict limitations on the number of independent market components. Our volatility model (13) also includes an important leverage effect (Bouchaud et al. 2001) in the most general form. As discussed later, SVM model has an advantage over other frameworks to study effects of trading volume and other factors on the accuracy of the volatility forecasting.

5. Results

In this section we will illustrate the ability of the SVM-based volatility model to efficiently extract information from high-dimensional market data and compare its performance to the common benchmark models. As benchmark we use the main-stream GARCH-type models (4)–(8) as well as "naive" model that uses current volatility value as a next step prediction. Since only (1,1) input configuration of GARCH-type models is widely used in practice (due to stability and technical simplicity), we also restrict our benchmark GARCH models to this input dimensionality.

In this example we use SP500 data shown in Figure 1. Steps of our analysis include the choice of a 750 day time window and shifting this window with a step of 10 business days to illustrate model sensitivity to the training and test data. SP500 data from 11/10/1999 to 12/10/2003 have been used. First 500 days of data in each window are used for training and validation. Remaining 250 days of data are used as test sets.

Cross-validation procedure is used to optimize SVM parameters such as regularization parameter C, ϵ-parameter of the loss function, coefficients of the kernel function, and the type of the kernel function itself. Optimization is performed with respect to a linear correlation coefficient between model outputs and corresponding real data. Final conclusions of the model performance are based on the results obtained on the test set.

As mentioned in the previous section, SVM model in the form (13) incorporates both long memory and multiscale effects. To demonstrate that SVM can efficiently extract information from lagged return vectors of high-dimension we trained SVMs with small and large number of lagged returns as inputs. We found that model with $\sim$ 10 inputs demonstrates stable performance in most regimes. Therefore in the following only SVM(10) model will be used. We also found that among common kernel types radial basis function is an optimal choice for our application. Kernel based on radial basis function is given by: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2)$, where γ is a constant.

Our main purpose was to demonstrate stable performance of the SVM that is comparable or superior to the benchmark models. The search of the optimal parameters (C, ϵ, and γ) that warranty stable generalization ability of the SVM is often difficult and application-dependent task (Chapelle and Vapnik 2001). In our problem, variance of the SVM out-of-sample performance with optimal parameter values obtained from n-fold ($3 \leq n \leq 10$) cross-validation procedure, was above that of the GARCH-type benchmark models. We found that reduction of the parameter range available for optimization leads to significant stabilization of the SVM performance. This procedure makes SVM model more conservative. However, even the most conservative model (with constant set of parameters for all regimes) demonstrates performance that is consistently superior to the naive and GARCH(1,1) models and comparable to the best model from the pool of GARCH-type models.

In Figure 2, we compare performance of SVM(10) with a fixed set of parameters ($C = 20$, $\epsilon = 0.15$, and $\gamma = 0.015$) to that of the naive and GARCH(1,1) models (solid, dashed, and dash-dotted lines, respectively). Correlation of the model prediction (σ) with a real data ($|r|$) for different SP500 data windows shifted from a base window by a variable number of business days: (a) chronological order, (b) sorted distribution. SP500 data are used as described above. It should be mentioned that GARCH outputs latent (unobservable) variable. Therefore measuring GARCH model performance with respect to realized r^2 (or $|r|$) time series is an approximation frequently used in practice (Tsay 2002).

It is clear from Figure 2, that SVM(10) significantly outperforms naive model at all times and consistently outperforms GARCH(1,1) model as well. It is also clear that performance

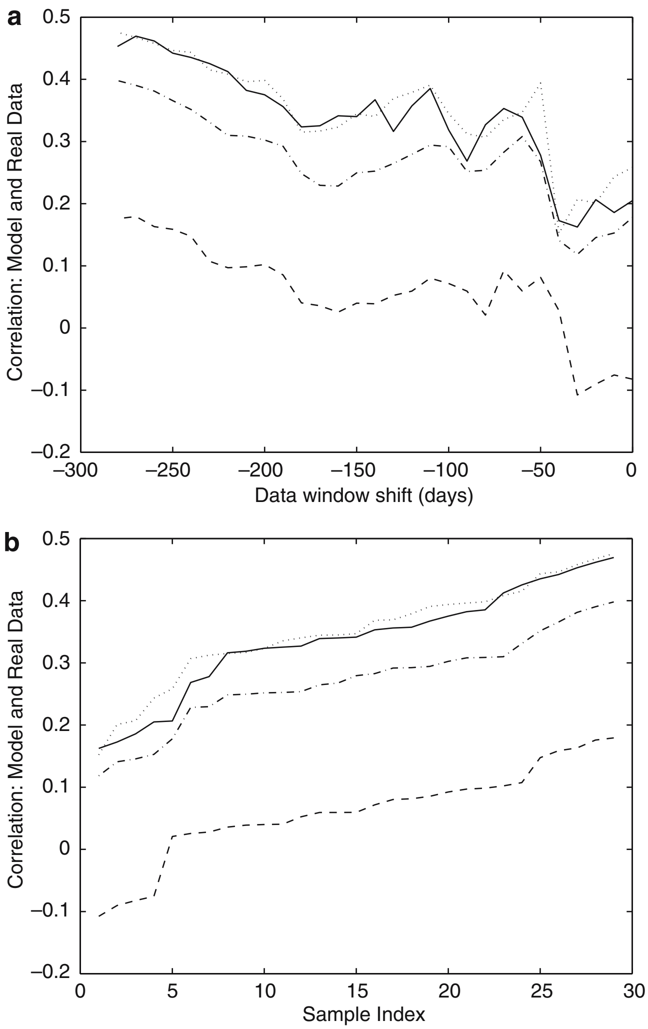

Fig. 2 Linear correlation of real and model volatilities for different SP500 data windows shifted from a base window by a variable number of business days. **a** Chronological order. **b** Sorted distribution. Results for SVM(10), naive, GARCH(1,1), and the best GARCH-type models are shown by *solid*, *dashed*, *dot-dashed*, and *dotted lines*, respectively. SVM(10) with constant parameters ($C = 20$, $\epsilon = 0.15$, and $\gamma = 0.015$) is used. Each window consists of 500 training and 250 testing data points. Windows are shifted with a step of 10 business days. Starting window (shift=0) ends at 12 October 2003

of both SVM and GARCH model are quite sensitive to the data set used. Superiority of the SVM model remains largely unchanged for other periods of the SP500 time series.

Out-of-sample performance of the best model from the pool of GARCH(1,1), EGARCH(1,1), TGARCH(1,1), and PGARCH(1,1) models is shown in Figure 2 by dotted line. The choice of the best model is based on its performance on training set as would be the case in practical applications. The best model changes as data window shifts. In most cases, the best model is either EGARCH or TGARCH. It is clear that performance of the SVM(10) model remains comparable to the best GARCH model across different regimes.

Table 1

Model	25%	50%	75%
SVM(10)	13.2	16.4	22.5
GARCH(1,1)	12.9	14.1	14.9
EGARCH(1,1)	11.7	13.6	16.7
TGARCH(1,1)	9.02	12.0	14.5
PGARCH(1,1)	10.5	12.6	14.4

Ljung–Box Statistics for r_i^2/σ_i^2 time series

Table 2

Model	25%	50%	75%
SVM(10)	0.14	0.22	0.40
GARCH(1,1)	0.13	0.23	0.28
EGARCH(1,1)	0.17	0.21	0.25
TGARCH(1,1)	0.13	0.17	0.22
PGARCH(1,1)	0.14	0.18	0.22

Ljung–Box Statistics for r_i^2/σ_i^2 normalized to the static for the original r_i^2 time series

A typical diagnostics used to examine quality of the volatility model is autocorrelation structure of the time series r_i^2/σ_i^2 (Tsay 2002). As illustrated in Figure 1c, r_i^2 exhibits significant autocorrelation. According to (5), good volatility model is supposed to produce r_i^2/σ_i^2 time series with minimal autocorrelation. A standard test for the time series autocorrelation is m-lag Ljung–Box statistic that decreases for smaller autocorrelation (Tsay 2002). For completeness, we provide Ljung–Box statistics for the volatility models considered here. However, this test alone should not be used as a final measure of the model quality since it is not robust to heavily tailed data and nonlinearity (Chen 2002). Moreover, when practical measures of volatility other than r_i^2 are used in SVM model, Ljung–Box statistic may have even more limited diagnostic power.

We have computed 15-lag Ljung–Box statistic for 250-day out-of-sample period of each overlapping data window used in Figure 2. Due to adaptive change of the model for each data window, it is not possible to use longer periods in a single calculation. Therefore, results for different windows vary significantly. Summary of the obtained Ljung–Box statistics for r_i^2/σ_i^2 time series (25, 50, and 75th percentiles) are shown in Table 1. More informative is statistic for r_i^2/σ_i^2 normalized to the statistic for the original r_i^2 time series. This is summarized in Table 2. It is evident from both tables that Ljung–Box statistics for different GARCH models and SVM are comparable and significantly reduced with respect to the original r_i^2 time series.

So far we compared GARCH models to the SVM with a fixed set of parameters. In Figure 3 performance of the SVMs with different procedures for partial parameter optimization is compared to the best GARCH model (dotted line). Results for SVM(10) with partially adjustable parameters ($C = [5, 10]$, $\epsilon = 0.2$, and $\gamma < 0.03$) based on 5-fold cross-validation and ($C = 15$, $\epsilon = 0.15$, and $\gamma < 0.03$) based on 10-fold cross-validation are shown by solid and dashed lines, respectively. It is clear that both SVM models remain to be comparable to the best GARCH models and even outperform it in certain regimes.

6. Discussion and Conclusion

In this paper, we addressed the problem of volatility forecasting from high-dimensional stock market data. SVM-based model was proposed as a possible complimentary approach

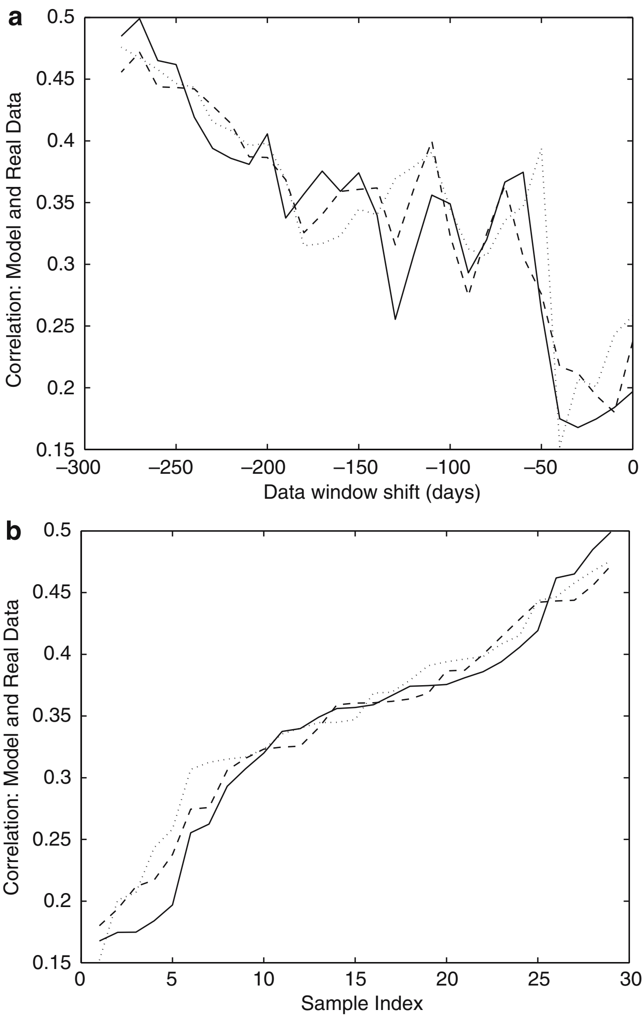

Fig. 3 Linear correlation of real and model volatilities for different SP500 data windows shifted from a base window by a variable number of business days. **a** Chronological order. **b** Sorted distribution. Results for SVM(10) with partially adjustable parameters ($C = [5, 10]$, $\epsilon = 0.2$, and $\gamma < 0.03$) based on 5-fold cross-validation and ($C = 15$, $\epsilon = 0.15$, and $\gamma < 0.03$) based on 10-fold cross-validation are shown by *solid* and *dashed lines*, respectively. Results for the best GARCH model are shown with *dotted line*. Each window consists of 500 training and 250 testing data points. Windows are shifted with a step of 10 business days. Starting window (shift=0) ends at 12 October 2003

to volatility forecasting. SVM combines the learning effectiveness of linear machines with the classification/regression power of the best nonlinear algorithms. Unlike typical nonlinear techniques such as NNs, the size of the SVM input space is decoupled from the number of free parameters and allows one to process high-dimensional data without encountering the "high-dimensionality curse". This makes SVM a possible model for processing real-time multiscale and high-frequency market data. SVM tolerance to incomplete data is another advantage of the SVM-based volatility model that can address the problem of the market data nonstationarity.

We reviewed the most important features of the stock market data and existing volatility models. Model limitations in describing volatility dynamics and ability to extract information from high-dimensional historical market data have been identified. Adequate description of such important volatility features as long term memory and asymmetric lead–lag correlation of volatilities (i.e., asymmetric information flow from large to small scales) leads to increasing dimensionality of the model and is one of the most challenging problems. Most of the existing approaches address this problem with rather restrictive assumptions to make the model computationally practical. These restrictions include limiting memory size, disregarding multiscale and nonlinear volatility effects, and limiting number of independent market components in some multiscale volatility models.

SVM's ability to handle high-dimensional and incomplete data allows to significantly relax those restrictions in the SVM-based volatility model discussed in this paper. Since this model imposes no significant restrictions on the length of the lagged vector of input parameters (memory size) and on the number of independent multiscale volatilities (market components), SVM model allows to study parameter regimes where other existing models become computationally unrealistic. Besides that SVM models can automatically include such effects as volatility dependence on the sign of the return (which is required to cover leverage effect) and general nonlinear effects that are not covered by the models currently used in practice.

Our results with SP500 index data indicate that SVM model can efficiently extract information from the inputs with large number of lagged returns. Our benchmark tests indicate that even conservative SVM(10) model can perform significantly better or comparable to both naive and GARCH(1,1) models. Performance of the SVM is also consistently comparable with the best model from the pool of GARCH, EGARCH, TGARCH, and PGARCH models.

SVMs considered here are quite conservative models with fixed input dimensionality and only partial optimization of other SVM parameters. More robust methods for a full parameter optimization (including input dimensionality) that warranty stable generalization could significantly improve results reported in this paper.

Since SVM is tolerant to high-dimensional inputs, our framework can be easily expanded by including other relevant input variables. These include trading volume, implied volatilities from derivative prices, and other external factors with additional information content. It would be interesting to compare performance of these heterogeneous volatility models based on SVM with those based on advanced NN algorithms (both feedforward and recurrent). Combination of different models using adaptive ensemble learning techniques (e.g., boosting) could also lead to more accurate volatility models. Further research in this area is warranted.

Acknowledgements This work is supported by Science Applications International Corporation.

References

Andersen TG, Bollerslev T, Diebold FX, Labys P (2000) Exchange rate returns standardized by realised volatility are (nearly) Gaussian. NBER Working Paper No: 7488

Arneodo A, Muzy JF, Sornette D (1998) Direct causal cascade in the stock market. Eur Phys J B 2:277

Baillie RT, Bollerslev T, Mikkelsen HO (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. J Econom 74:3

Bishop CM (1995) Neural networks for pattern recognition. Clarendon Press, Oxford

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31:307
Bouchaud JP, Potters M (1999) Theory of financial risk: from statistical physics to risk management. Cambridge University Press, Cambridge
Bouchaud JP, Matacz A, Potters M (2001) Leverage effect in financial markets: The retarded volatility model. Phys Rev Lett 87:228701
Brown M, Grundy W, Lin D, Cristianini N, Sugnet C, Ares M Jr, Haussler D (1999) Support vector machine classification of microarray gene expression data. University of California, Santa Cruz, technical report UCSC-CRL-99-09
Chang CC, Hsu CW, Lin CJ (2000) The analysis of decomposition methods for support vector machines. IEEE Trans Neural Netw 11:1003
Chapelle O, Vapnik V (2001) Choosing multiple parameters for support vector machines. Adv Neural Inf Process Syst 03(5)
Chen YT (2002) On the robustness of Ljung–Box and McLeod–Li Q tests: a simulation study. Econ Bull 3:1
Cristianini N, Shawe-Taylor J (2000) Introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
Dacorogna MM, Gencay R, Muller U, Olsen RB, Pictet OV (2001) An introduction to high-frequency finance. Academic Press, San Diego
Ding Z, Granger CW, Engle RF (1993) A long memory property of stock market returns and a new model. J Empir Finance 1:83
Donaldson RG, Kamstra M (1997) An artificial neural network-GARCH model for international stock return volatility. J Empir Finance 4:17
Edelman D (2001) Enforced-denial support vector machines for noisy data with applications to financial time series forecasting. In: Proceedings of the international conference on statistics, combinatorics and related areas and the 8th international conference of forum for interdisciplinary mathematics, school of mathematics and applied statistics, University of Wollongong, Wollongong, NSW 2522, Australia
Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50:987
Engle RF, Patton AJ (2001) What good is a volatility model? Quant Finance 1:237
Fan A, Hong D, Palanaswami M, Tan C (1999) A support vector machine approach to bankruptcy prediction: a case study. In: Proceedings of the 6th international conference on computational finance, New York
Gavrishchaka VV, Ganguli SB (2001) Support vector machine as an efficient tool for high-dimensional data processing: application to substorm forecasting. J Geophys Res 106:29911
Gavrishchaka VV, Ganguli SB (2003) Volatility forecasting from multiscale and high-dimensional market data. Neurocomputing 55:285
Joachims T (1998) Text categorization with support vector machines. In: Nedellec C, Rouveirol C (eds) Proceedings of the 10th European conference on machine learning (ECML). vol 1398, pp 137–142
Mangasarian OL, Street WN, Wolberg WH (1995) Breast cancer diagnosis and prognosis via linear programming. Oper Res 43(4):570
Mantegna RN, Stanley HE (2000) An introduction to econophysics: correlation and complexity in finance. Cambridge University Press, Cambridge
Masoliver J, Perello J (2001) A correlated stochastic volatility model measuring leverage and other stylized facts. cond-mat/0111334 v1
Nelson DB (1991) Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59:347
Osuna E, Freund R, Girosi F (1997) Training support vector machines: an application to face detection. In: Proceedings of computer vision and pattern recognition, pp 130
Pontil M, Verri A (1998) Object recognition with support vector machines. IEEE Trans PAMI 20:637
Pontil M, Mukherjee S, Girosi F (1998) On the noise model of support vector machine regression. CBCL Paper 168, AI Memo 1651, Massachusetts Institute of Technology, Cambridge
Schittenkopf C, Dorffner G, Dockner EJ (1998) Volatility prediction with mixture density networks. In: Niklasson L, Boden M, and Ziemke T, ICANN ’98 – proceedings of the 8th international conference on artificial neural networks, Springer, Berlin Heidelberg New York, pp 929
Tsay RS (2002) Analysis of financial time series. Wiley, New York
Van Gestel T, Suykens J, Baestaens D, Lambrechts A, Lanckriet G, Vandaele B, De Moor B, Vandewalle J (2001) Financial time series prediction using least squares support vector machines within the evidence framework. IEEE Tran Neural Netw 12:809 (Special Issue on Neural Networks in Financial Engineering)
Vannerem P, Mller KR, Schlkopf B, Smola A, Sldner-Rembold S (1999) Classifying LEP data with support vector algorithms. hep-ex/9905027
Vapnik V (1995) The nature of statistical learning theory. Springer Verlag, Berlin Heidelberg New York
Vapnik V (1998) Statistical learning theory. Wiley, New York
Zakoian JM (1994) Threshold heteroskedastic models. J Econ Dyn Control 18:931