Heuristic Solutions to Technical Issues Associated with Clustered Volatility Prediction using Support Vector Machines

Karen Hovsepian
Department of Computer Science
New Mexico Tech
Socorro, NM 87801
E-mail: karen@nmt.edu

Dr. Peter Anselmo
Department of Management
New Mexico Tech
Socorro, NM 87801
E-mail: anselmo@nmt.edu

Abstract—We outline technological issues and our findings for the problem of prediction of relative volatility bursts in dynamic time-series utilizing *support vector classifiers* (SVC). The core approach used for prediction has been applied successfully to detection of relative volatility clusters. In applying it to prediction, the main issue is the selection of the SVC training/testing set. We describe three selection schemes and experimentally compare their performances in order to propose a method for training the SVC for the prediction problem. In addition to performing cross-validation experiments, we propose an improved variation to sliding window experiments utilizing the output from SVC's decision function. Together with these experiments, we show that accurate and robust prediction of volatile bursts can be achieved with our approach.

I. INTRODUCTION

In [1] we proposed a multi-layer framework for detection of relative clustered volatility, based on supervised learning with support vector classifiers (SVC) [2-3], designed to be automated, deterministic and efficient. We showed experimentally that it had a high rate of detection of relative clustered volatility (RCV) and could be rapidly and easily deployed in on-line and near-real-time applications.

The approach also easily lends itself to be applied to the much harder problem of prediction of RCV, since pattern recognition is at the heart of it. The only required change is in the type of patterns that the SVC algorithm must discern. In this paper we present the technical questions tied to the setup of the prediction framework and our proposed heuristic solutions to these questions.

Specifically, for the prediction problem we are faced with the challenge of SVC training/testing set selection. We need to decide on the choice of categories in the universe of time-series segments. This choice is crucial since the more robust and non-biased the training data are, the more effective the final system is. For detection, the two categories are **relatively volatile** (RV) and **relatively non-volatile** (RN). This is a relatively obvious categorization, since RV is what we want to detect and RN, in this formulation, is the opposite of RV. For prediction,

it is much less apparent what categories imposed on the time-series segments are most opposite to each other in the sense of least pattern overlap. The experimentally derived answers to this question and smaller questions tied to it are the crux of this paper.

Next, we briefly describe the overall approach. After that, we will introduce the three training set selection schemes and their motivations. Finally, we'll summarize the experiments and the results, which lead to the choice of the optimal scheme, given the alternative training/testing regimes we propose.

II. APPROACH

The approach described in [1] is briefly summarized in this section. The focal component of the approach, whether it's applied to detection or prediction, is the classification of time-series segments using SVC. For detection, the assumption is that all time-series segments fall into one of two categories: relatively volatile (RV) and relatively non-volatile (RN). This assumption then dictates the creation of the SVC training set.

Using the previously mentioned GARCH model, coupled with χ^2 significance tests, we carefully choose the time-series examples of each category (RV & RN). The first part of this step is a simple fitting of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [4-5] model to the raw time-series data. For the most part the default GARCH parameters[1] can be used. Upon retrieving the conditional standard deviations, we use χ^2 tests to select the data points, for which the conditional standard deviations are significantly greater than the average conditional standard deviation. An unbroken sequence of such data points, assuming it is longer than some user-defined constant, are assigned into training set's RV class. Segments of data points with conditional standard deviations not significantly greater than the average are assigned into RN class. This necessary step represents the

---

[1] $p = 1, q = 1,$

nontrivial setup and preparation of the system and, thus, needs to be taken only once. Once trained, no GARCH fitting is required and the application of our system cuts down to an automated feeding of a new segment in question into the system.

Because the lengths of the segments in the training set are not all equal, application of SVC with such a training set is impossible without some reformulation of the SVC Kernel or transformation/standardization of the data. This is due to the encoding of the time-series segments as vectors. Each time-series data point becomes a vector component/dimension. Thus, all segments need to be of same length, so they may be encoded into vectors of same dimensionality. This, however, is not the case with the current training set. To resolve this issue we must apply a crucial standardization step, which aims to make all segments in the training set of same length. To achieve this standardization, we compute the Power Spectrum Density Estimates (PSDE), via the "periodogram" [6], of the segments. By mapping the time-domain segments into the frequency domain, this scheme standardizes the lengths of the signals and, as a secondary benefit, creates a common global context for the data.

The final step of the approach is the actual SVC training with the training set of standardized examples of RV and RN, as selected with post-GARCH χ^2 tests. This step, like the GARCH fitting step, requires meticulous search for the best SVC parameters, ones which seed the highest accuracy of classification. The testing is done on a data set which, like the training set, contains examples of each class. Once the best parameters are chosen and the most accurate decision function is trained, the setup and preparation of the system are completed, and from then on the model is ready to identify past volatility and to detect ensuing volatile clusters in their early onset.

III. TRAINING SET SELECTION SCHEMES

In its basic form, SVC is a binary classification technique. The training/testing data are composed of two classes, conventionally labeled as +1 and -1. Before training can begin, we need to supply a dataset with examples from these two classes. We extract these examples from the raw time-series. Examples of class +1 are segments occurring before volatility clusters, pre-RV. This is an intuitive choice, as the patterns we seek for prediction are most likely found in the segments occurring before the RV segments. One interesting point however is whether we should choose segments occurring immediately before the RV regions or some horizon before. On the other hand, it is more difficult to settle on the choice of examples for class -1. The reason is that it is not certain whether we should choose segments with patterns that we assess are not present in class +1, or simply segments not chosen in class +1. In the first case, the follow-up question is where such patterns are located.

To help resolve the above question, we explore three schemes, each with a different definition of class -1 and perform several experiments, aimed at highlighting the optimal scheme.

Table I contains the descriptions of the classes in the first scheme. The choice for class -1 is motivated by the idea that segments that are taken from time-series with no GARCH effects possess patterns most in contrast to the patterns in class +1. Indeed, these segments are neither pre-RV nor RV, and are guaranteed not to possess pre-RV patterns, as there are no RV bursts in their surroundings. To determine if a raw time-series had any relative clustered volatility, we used Engle's Test for GARCH effects [5]. Without GARCH effects, it would be impossible for volatility to persist; this persistence is the central feature of volatility clusters, meaning that it is a prerequisite for the formation of relative volatility clusters.

TABLE I

TRAINING SET SELECTION SCHEME #1

Class	Name	Description
+1	Pre-RV	Segments occurring immediately before an RV segment, as selected for the detection task – a segment (w/ lengths above user-defined constant) composed of data points with the conditional S.Ds. above the average conditional S.D. It may contain RV segments, which are not longer than the user-defined constant.
-1	Non-Volatile	Segments of random length.

Our criticism for this scheme is that some patterns may be present in class +1 and absent in class -1 because class +1 and class -1 observations are extracted from different raw time-series. These confounding patterns could in effect introduce a bias in the SVC training.

To address the potential bias criticism of the first scheme, we consider the second scheme, where we chose the class -1 segments from the same time-series as the pre-RV segments. The second scheme is outlined in Table II.

TABLE II

TRAINING SET SELECTION SCHEME #2

Class	Name	Description
+1	Pre-RV	As scheme #1.
-1	Pre-Pre-RV	Segments, occurring immediately before the Pre-RV segments chosen for class +1.

Finally, in the third scheme, outlined in Table III, the premise is to truly divide the whole space of time-series segments into class +1, pre-RV bursts, and all else. The next section presents the experimental conditions and results for each scheme in the hopes of putting forward the best choice for the training/testing set selection scheme.

TABLE III

TRAINING SET SELECTION SCHEME #3

Class	Name	Description
+1	Pre-RV	As scheme #1.
-1	All Else	Segments of random length, between user-defined constraints, chosen at random from the rest of the time-series data, once class +1 has been selected.

IV. EXPERIMENTAL CONDITIONS AND RESULTS

The raw time-series data are borrowed from the financial domain, representing the inter-day foreign exchange (FX) rates for a number of currencies and commodities. These data are available from an on-line database [7], which contains FX rates for 81 currencies and commodities. In financial time-series, volatility clusters have been found to affect long-term financial trends, socio-economic development, and living standards [8-13].

To facilitate the running of the experiments, we have developed a comprehensive volatility analysis tool, called *VolatilityAnalyst* ® [14], which allows the user to easily perform the necessary tasks and experiments via the software GUI. The SVC module in VolatilityAnalyst is an interface to LIBSVM [15], a well-known implementation of SVC and other *support vector machines* algorithms.

We report the results of the cross-validation (CV) experiments which were performed to assess the accuracy of the trained SVC model on a testing set, derived identically to the training set.

A. Cross-Validation Experiments

In the CV experiments, the accuracy score is calculated based on the results of *n* sub-experiments, where each sub-experiment consists of a different training and testing set. A larger *n* is preferred for smaller datasets. In all three schemes the datasets were around 1000 observations in size. For such intermediate-sized datasets, setting *n* to 5 is an acceptable option. Through these experiments we hope to choose the optimal training scheme before we move on to other experiments. Table IV summarizes the best CV results.

TABLE IV

CV RESULTS FOR THE THREE TRAINING SET SELECTION SCHEMES

Scheme	False Pos	False Neg	Training Time	SVC Parameters
#1	1%	3%	5s	C: 44400 Kernel: RBF$\|\gamma=40000$
#2	17%	13%	10s	C: 35000 Kernel: RBF$\|\gamma=10500$
#3	24%	18%	30s	C: 10400 Kernel: RBF$\|\gamma=650000$

The process of searching for optimal SVC parameters was simplified through the LIBSVM parameter selection tool [15]. For the specified range of the parameters, the tool generates the contour plot of cross-validation accuracy. Using the plot, one can zero in on the parameters that yield the best results. Figure 1 is the contour plot for scheme #1 showing the contours of the top CV accuracies.
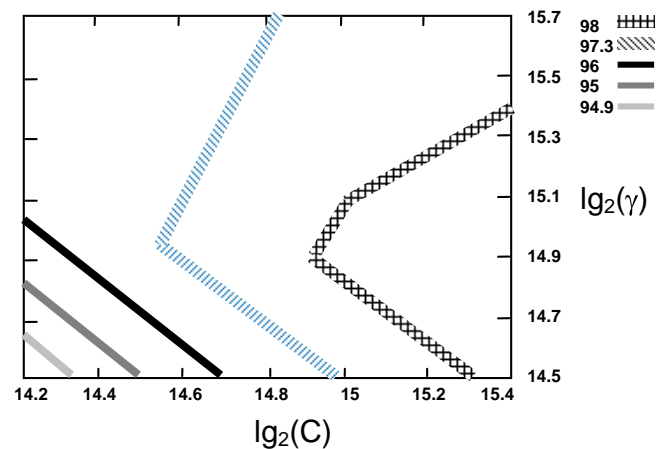

Fig. 1. CV contour plots for the range of SVC parameters that yield the highest CV accuracy of 98%.

The table summarizes the finding that scheme #1 defines the optimal choice of categories. The reason for the low performance of schemes #2 and #3 may lie in the possibility that in both schemes there is a lot of overlap of patterns between class -1 and class +1. In scheme #2, this is evident if one realizes that many pre-pre-RV examples may in fact be examples of pre-RV, since the boundaries between the two classes are not clearly defined. Similarly, in the third scheme, incorporating all the segments not included in class +1 into class -1 may poorly separate the relevant patterns.

It is important to note that the criticism of potential bias in scheme #1 may still be valid, even with such high reported CV accuracy. After all, the testing is performed on data that may include the same confounding patterns present in the training data.

To address this criticism and also to test the consistency and the practical applicability of the framework, we performed sliding windows experiments, simulating the real-time application of the system. These experiments are helpful in showing the ability of our system to predict well known RV bursts. They can also help understand the cases when misclassifications occur.

B. Sliding Windows Experiments

In the sliding windows experiments, we select a specific RV example, start some time before it and commence to classify the windows – time-series segments – using the pre-trained SVC model, as they slide towards the RV burst. We record if and how soon a window is classified as class +1, or pre-RV. In addition, we run the experiment on a section of the time-series that do not have any RV bursts, to

test if the results are consistent and to ensure that random segments are not classified as pre-RV.

Rather than keeping the size of each window throughout the experiment constant, we can test several windows of varying data-point length and make the final judgment based on the window for which the SVC decision had the highest confidence. To measure this confidence we use the SVC decision function:

$$\hat{y} = sign\left(\sum_{i}^{l_{SV}} \lambda_i y_i K(x_i, x) + b \right) \quad (1)$$

where x is the new test observation, x_i and y_i are the ith training observation and the corresponding class, $K()$ is the Kernel function, λ_i is the ith Lagrangian of the training problem and l_{SV} is the number of support vectors. If the sign is positive, the class is +1 and if it's negative, the class is -1. If we remove the sign function, however, the absolute value would measure the strength/confidence of the classification, which is what we use for ranking the windows and choosing the optimal one. Figure 2 illustrates the variable-sized sliding windows experiments.
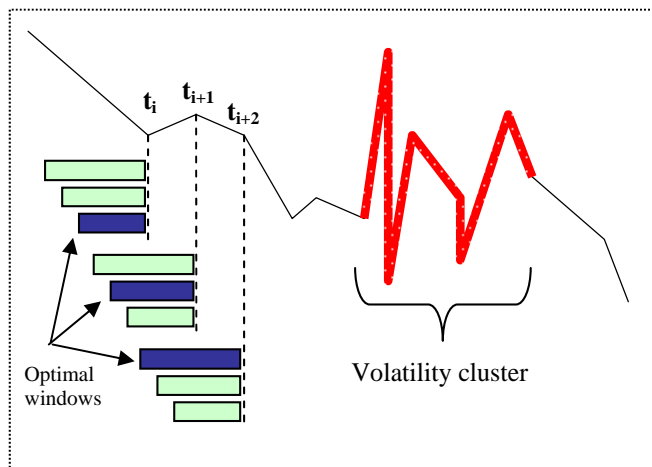
Fig. 2. Variable-sized sliding windows experiment.

The figure illustrates how at each tick, i.e. tick t_i, we test several windows and choose the optimal one. Whatever the class of this window, we accordingly make the final judgment of whether a prediction has taken place or not. This is repeated for each new tick. The benefit of this setup is that it recognizes that in some cases a smaller window may be preferred to make sure that irrelevant noisy patterns are not included in the classification decision, while in other cases a larger window may be needed so that it contains sufficient pre-RV patterns for an unambiguous prediction. These two aspects make the system more flexible when recognizing the patterns for pre-RV class.

In figure 3 we plot the prediction results for the three schemes. The horizontal axis measures how many ticks before the RV segments the system first classified the optimal window as pre-RV.
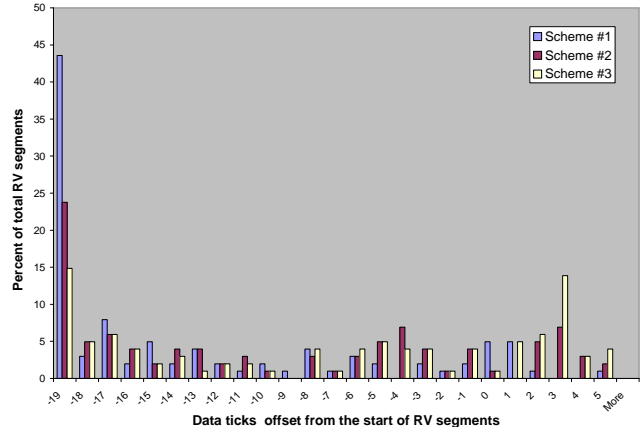
Fig. 3. Variable-sized sliding windows experiment results for the three schemes.

As we can see, the results of Table IV are in tune with the rolling windows experiments. Scheme #1 comes out as the most robust way for training the SVC decision function for the prediction, as 88% of the RV segments were predicted, with 43% predicted 19 ticks before the cluster onset. In the second scheme only 82% were predicted. Finally with the third scheme, only 67% were predicted. In the majority of the cases, most of the windows after the first pre-RV classification were also classified as pre-RV. Note in figure 3 that windows 19 ticks prior to the RV segments seem to possess most of the pre-RV patterns. We don't check to see what happens before 19 ticks because the gaps between many RV segments are not much larger than that, and we don't want to confuse the classification by identifying the classes of other RV segments.

Any bias that may be present in scheme #1 due to its definition of class -1 is insignificant as compared to its successful separation of the relevant pre-RV patterns from all other ones.

We also tested segments which did not precede RV segments just to see the consistency of the performance. These results and the results of the variable-sized sliding window experiments are summarized in table V.

TABLE V

SUMMARY OF ROLLING WINDOW EXPERIMENTS FOR THE THREE SCHEMES

Scheme	Sliding Windows
#1	Consistently classifies the windows occurring on average 19 data points before a RV segment as Class pre-RV.
#2	Results inconsistent and non-robust. Many classifications of pre-RV windows, which do not occur before an RV region.
#3	Results are worse than scheme #2. Seemingly random classification of segments as pre-RV.

V. CONCLUSIONS

We have presented what we view as the most intuitive choices for the SVC training/testing datasets and carried out performance comparisons between them. According to several experiments, we concluded that scheme #1 gives the most robust definition for the SVC categories. In the future, we plan to introduce several classes for a more refined prediction of various levels of clustered volatility. Thus, we will once again have to solve the question of optimal dataset selection. We will also continue testing and experimenting with relatively clustered volatility prediction in combination with our established detection scheme.

Besides the comparison, the paper shows that accurate and consistent prediction of RV segments is possible with our approach. Especially promising in this regard is the use of variable-sized windows for custom-fitting the windows to only the relevant patterns. There is however an execution time cost to doing so, as more windows need to be tested with each new data point. In a near real-time scenario, this could pose some problems. However, parallelizing the testing of all the windows, which is already in our future projects pipeline, could greatly reduce the execution time.

REFERENCES

[1] K. Hovsepian, P. Anselmo, and S. Mazumdar, "Support Vector Classifier Approach for Detection of Clustered Volatility in Dynamic Time-Series," New Mexico Tech, Tech. Rep., 2005. [Online]. Available: http://www.nmt.edu/~karoaper/AI_RVC.pdf

[2] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[3] B. Boser, I. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Fifth Annual Workshop on Computational Learning Theory*. Pittsburgh: ACM, 1992, pp. 144–152.

[4] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

[5] R. F. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, pp. 987–1007, 1982.

[6] S. Kay, *Modern Spectral Estimation.* Englewood Cliffs, NJ: Prentice Hall, 1988.

[7] Policy Analysis Computing and Information Facility In Commerce (PACIFIC) at University of British Columbia, http://pacific.commerce.ubc.ca/xr/.

[8] J. Aizenman and N. Marion, "Volatility and investment: Interpreting evidence from developing countries," *Economica*, vol. 66, no. 262, 1999.

[9] J. Aizenman and A. Powell, "Volatility and financial intermediation," *Journal of International Money and Finance*, June, 2002.

[10] F. R. B. of Kansas City, *Financial Market Volatility and the Economy*. Books for Business, December, 2001.

[11] G. Ramey and V. Ramey, "Cross-country evidence on the link between volatility and growth," *The American Economic Review*, vol. 85, no. 5, pp. 1138–1151, December 1995.

[12] F. Black, *Business cycles and equilibrium*. Cambridge, MA: Blackwell, 1987.

[13] M. Leonard, "Uncertainty and optimal consumption decision," *Econometrica*, vol. 39, no. 1, pp. 179–185, January 1971.

[14] VolatilityAnalyst ® software package, http://www.nmt.edu/˜karen/volanalyst/.

[15] LIBSVM – A Library for Support Vector Machines, Chih-Chung Chang and Chih-Jen Lin, http://www.csie.ntu.edu.tw/~cjlin/libsvm/7]

T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.