 ISSN: 978-602-71169-7-9

Volatility Modelling Using Hybrid Autoregressive Conditional Heteroskedasticity (ARCH) - Support Vector Regression (SVR)

Hasbi Yasin[1], Tarno[2], and Abdul Hoyyi[3]

[1,2,3] Department of Statistics, Faculty of Science and Mathematics, Diponegoro University

Jl. Prof. Soedharto SH, Tembalang, Semarang 50275

hasbiyasin@live. undip.ac.id , tarno@undip.ac.id , and ahy_stat@undip.ac.id

Abstract: High fluctuations in stock returns is one problem that is considered by the investors. Therefore we need a model that is able to predict accurately the volatility of stock returns. One model that can be used is a model Autoregressive Conditional Heteroskedasticity (ARCH). This model can serve as a model input in the Support Vector Regression (SVR) model, known as Hybrid ARCH-SVR. This modeling is one of the alternatives in modeling the volatility of stock returns. This method is able to show a good performance in modeling the volatility of stock returns. The purpose of this study was to determine the stock return volatility models using a Hybrid ARCH-SVR model on stock price data of PT. Indofood Sukses Makmur Tbk. The result shows that the determination of the input variables based on the ARIMA (3,0,3)-ARCH (5), so that the SVR model consists of 5 lags as input vector. Using a this model was obtained that the Mean Absolute Percentage Error (MAPE) of 1,98% and R^2 =99,99%.

Keywords: ARCH; ARIMA; SVR; Volatility

Introduction

The Volatility is statically a standard deviation of returning stock that represents the share price returns [3],[12]. The higher the volatility, the higher the risk of profit or loss [5],[11]. The uncertainty value of the volatility in the financial markets leds to the need for a tool to foresee. Whilst the value at risk (VaR) is a concept that is used for measuring a risk in risk management. VaR can be simply defined as how much investors can lose their money during the investment period. In calculating VaR, the main problem to be solved is to determine a prediction of the volatility stock returns accurately which will be used as basis for calculating VaR.

According to Jorion [6], data stock returns have usually variances that are not constant at any point of time, called conditional heteroskedasticity. One of the financial time serie models that can accomodate heteroskedastisity is Autoagressive Conditional Heteroskedasticity (ARCH) which was introduced by Engle [4]. Whereas the more flexible model for modeling variance which is not constant is Generalized Autoregressive Conditional Heteroskedasticity (GARCH) proposed by Bollerslev [2]. GARCH structure consists of two equations, one is conditional mean equation which is ARCH standard model and the other is conditional variance equation that allows the variance changes anytime [13]. This model will be less optimal when used for prediction of stock return volatility. One of the forecasting method developed at this time is using Support Vector regressions (SVR). SVR is a non-linear approach that is based on machine learning. SVR is a modification of the Support Vector Machine (SVM) which is used for regression approach. The concept of SVR is maximizing hyperplane to collect data that can be support vector. One of the advantages is SVR able to overcome overfitting.

Therefore, this study will develop an alternative model that combines ARCH and SVR (Hybrid ARCH-SVR) for modeling the volatility shares of PT. Indofood Sukses Makmur Tbk, which later would be used to calculate Value at Risk (VaR).

Literature Review

Autoregressive Conditional Heteroskedasticity (ARCH)

Generally, ARCH models of order q is used to form the conditional variance models $\left(\sigma_t^2\right)$ at all time (t) based on the squared error at a time (t – 1) to (t – q). E.g. the average models are:

$$Z_t = \mu_t + e_t$$

According to Tsay[12] that μ_t is a expectation value Z_t conditional F_{t-1}, with $F_{t-1} = \left\{Z_{t-1}, Z_{t-2}, Z_{t-3}, \cdots, Z_2, Z_1\right\}$. So the models of ARMA(r,m) of Z_t are:

$$\begin{aligned}\mu_t &= E\left(Z_t \mid F_{t-1}\right)\\ &= \theta_0 + \sum_{i=1}^{r} \phi_i Z_{t-i} + \sum_{j=1}^{m} \theta_j e_{t-j}\end{aligned}$$

with:

X_t = return at a time -t

F_{t-1} = the entire set of information at a time -1 to- t-1

μ_t = expectation value X_t conditional F_{t-1}

e_t = residual ARMA at a time-t

Tsay[12] stated that ARCH model is a remnant e_t of the ARIMA model which is in the high order will be correlated, e_t could be describes as follows:

$$e_t = \varepsilon_t \sigma_t$$

$$e_t \mid F_{t-1} \sim iidN\left(0, \sigma_t^2\right)$$

$$\varepsilon_t \sim iidN\left(0,1\right)$$

Acquired conditional variance for e_t:

$$\begin{aligned}\operatorname{Var}\left(e_t \mid F_{t-1}\right) &= E\left(e_t^2 \mid F_{t-1}\right)\\ &= E\left(\varepsilon_t^2 \sigma_t^2 \mid F_{t-1}\right)\\ &= \sigma_t^2 E\left(\varepsilon_t^2 \mid F_{t-1}\right)\\ &= \sigma_t^2\end{aligned}$$

so that the conditional variance that defines the order q ARCH models, is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2$$

with q > 0, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for i = 1, 2, 3, …, q.

Support Vector Regression (SVR)

Support Vector Regression (SVR) is a development of SVM for regression case. The goal of SVR is to find out a function $f(x)$ as a *hyperplane* in the form of regression functions which correspond to all the input data by an error ε and made ε as thin as possible[10]. Suppose there is l data training, $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, l$ in which $\boldsymbol{x}_i$ an input vector $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n\} \subseteq \boldsymbol{\Re}^n$ and scalar output $y = \{y_i, \ldots y_l\} \subseteq \Re$ and l is the number of training data. With SVR, will be determined a function $f(x)$ which has the biggest variation ε from the actual target y_i, for all the training data. if ε equal to 0 then obtained a perfect regression equation [9].

The purpose of SVR ist o mapping input vector into the higher dimension [1]. For example a function below the regression line as the optimal hyperplane:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \varphi(\boldsymbol{x}) + b$$

with:

$\boldsymbol{w}$ = dimensional weight vector$l$

$\varphi(\boldsymbol{x})$ = function that maps x to the space with l dimension

b = bias

Kernel Function

Many techniques of data mining or *machine learning* developed with the assumption of linearity, so that the resulting algorithm is limited to linear cases. With K*ernel Trick*, the data x in the *input space* mapped to the *feature space* with higher dimension through φ [9].

- ✓ Linear:$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}.\boldsymbol{y}$
- ✓ *Polynomial:* $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}.\boldsymbol{y} + \boldsymbol{c})^{\boldsymbol{d}}$
- ✓ *Radial Basis Function* (RBF): $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = exp(-\gamma\|\boldsymbol{x} - \boldsymbol{y}\|^2)$, with $\gamma = \frac{1}{2\sigma^2}$
- ✓ *Tangent hyperbolic* (sigmoid): $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = tanh(\sigma(\boldsymbol{x}.\boldsymbol{y}) + c)$

$\boldsymbol{x}$ and $\boldsymbol{y}$ are two pairs of data from all parts of the training data. Parameter $\sigma, c, d > 0$, is constant. According toVapnik and Haykin, legitimate Kernel function provided by Mercer theory where these functions should be qualified continouos and positive definite [9].

ISSN: 978-602-71169-7-9

Selection Parameters

According to Leidiyana [7], *cross-validation* is a standard test that is performed to predict error rate. Training data are randomly divided into several parts with the same ratio then the error rate is calculated section by section, and then calculate the overall average error rate to get the overall error rate. The rate of error can be calculated with the following formula:

$$CV = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i} \right)^2$$

with:

$\hat{y}_{\neq i}$: fitting value y_i where the observation toi removed from the assessment process

y_i : actual value y on observation to i

in the *cross-validation*, known validation *leave-one-out* (LOO). In the LOO, data is divided into two subsets, one subset contains N-1 data for training and the rest oft he data for testing [9].

Hybrid ARCH-SVR

Hybrid ARCH-SVR is a combination model between SVR and ARCH, where ARCH models are used as an initial model for the determination of the input variables in the model SVR. Modeling a number of return data Y_t at the time $t_1, t_2, t_3, \cdots, t_n$ then used to estimate the value of the return at time t_{n+1}. One of the important things in ARCH-SVR model is determining the input variables. For example, to specify the input and the target of ARCH models (1). Suppose ARCH models (1) $\sigma_t^2 = \omega + \alpha_1 e_{t-1}^2$, then the used input is e_{t-1}^2 with the target σ_t^2. So that the model can be written $\sigma_t^2 = f\left(e_{t-1}^2\right)$.

Value at Risk (VaR)

Value at Risk (VaR) to return a single asset PT. Indofood Sukses Makmur, Tbk with a confidence level (1-α) and the holding period (hp), can be calculated using the formula:

$$VaR(1-\alpha, hp) = -Z_{1-\alpha} * S_0 * \sqrt{\sigma_t^2 * hp}$$

with:

S_0 = initial investment

σ_t^2 = The volatility of stock returns PT. Indofood Sukses Makmur, Tbk at the time *t*

Material & Methodology

Preparing daily stock return data PT Indofood Sukses Makmur Tbk.

Determining the independent variables based on the model of the best ARCH

Dividing the data into training data and testing the data to the percentage of a certain proportion.

Performing modeling stock returns using SVR method with kernel function, the values of kernel parameters and cost parameters and parameter optimization hyperplane epsilon for the training data.

Using the hyperplane with the best parameters obtained in the data testing.

Evaluating of regression models in testing using the coefficient of determination (R^2) and MAPE.

Results and Discussion

Result

In Modeling stock returns PT. Indofood Sukses Makmur, Tbk. Conducted by using GARCH models. Based on the results of data processing using MATLAB GUI program, it could be found that the identification initial model is ARMA (3,3) ARCH (3). But to obtain the best GARCH model, overfit process and underfit to parameter model used need to be done, and the results shown in Table 1.

Table 1. Determination of the best ARCH model for return stocks of PT. Indofood Sukses Makmur, Tbk.

NO	MODEL	AIC
1	ARMA(3,3) ARCH(3)	-3229.8079
2	ARMA(2,2) ARCH(3)	-3220.8947
3	ARMA(3,3) ARCH(4)	-3231.8740
4	ARMA(2,2) ARCH(4)	-3221.8036
5	**ARMA(3,3) ARCH(5)**	**-3263.0156**
6	ARMA(2,2) ARCH(5)	-3261.0707
7	ARMA(1,1) ARCH(1)	-3171.7286

The best model for modeling stock returns PT. Indofood Sukses Makmur, Tbk is a model ARMA (3,3) ARCH (5) which mathematically can be written as follows:

$$Z_t = 2.4017 \times 10^{-4} + 0.17698Z_{t-1} - 0.17352Z_{t-2} + 0,18189Z_{t-3} - 0,23633e_{t-1} + 0,09151e_{t-2} - 0,36459e_{t-3} + e_t$$

with $e_t \sim N\left(0, \sigma_t^2\right)$ and

$$\sigma_t^2 = 9,33 \times 10^{-5} + 0,14318e_{t-1}^2 + 0,20259e_{t-2}^2 + 0,3671e_{t-3}^2 + 0,010765e_{t-4}^2 + 0,17057e_{t-5}^2$$

Determination of Kernel function and parameters for hyperplane

The This study only used Kernel linear functions in linear at hyperplane SVR. The best parameters on the kernel function is determined by trying out some of the values in a specific range to build hyperplane. Optimized parameters is the value of C and the value of epsilon. The best parameters for the hyperplane determined by the smallest error value. From the selected parameters could be found that the best parameters for the hyperplane with linear kernel function is C = 10 and epsilon = 0.01. SVR modeling results with the parameter values obtained very high accuracy of the model, namely R^2=99.99% and MAPE = 1.98%. Visually, the results of prediction data can be seen in Figure 1. While the results of the predictive value of the stock return volatility can be seen in Figure 2. In those figures show that the data pattern has followed the same pattern so obtained SVR models used for prediction decent stock return volatility PT. Indofood Sukses Makmur Tbk.

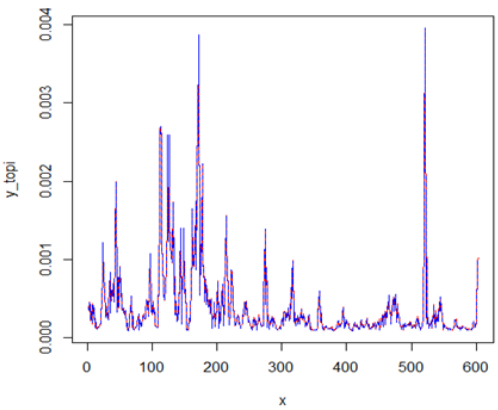

Figure 1. Plot of predicted and actual results

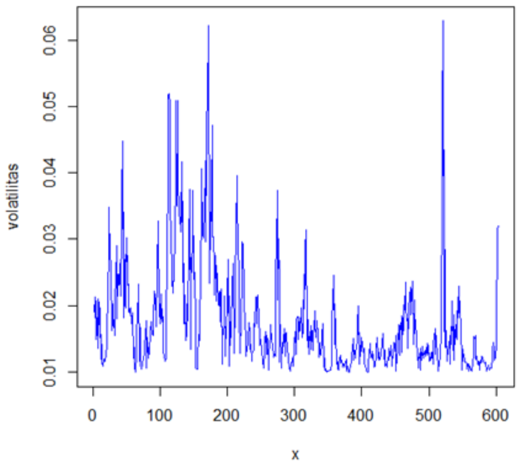

Figure 2. Plot of stock return volatility predicted results of PT. Indofood Sukses Makmur, Tbk

Calculation of VaR using the best model

Value at Risk (VaR) to return a single asset of PT. Indofood Sukses Makmur Tbk with a confidence level (1-α) and the holding period (hp) can be calculated using the formula:

$$VaR(1-\alpha, hp) = -Z_{1-\alpha} * S_0 * \sqrt{\sigma_{INDF}^2 * hp}$$

with:

S_0 = the value of the initial investment

σ_{INDF}^2 = The volatility of stock returns PT. Indofood Sukses Makmur, Tbk.

VaR return value shares of PT. Indofood Sukses Makmur, Tbk with a 95% confidence level and 1 day holding period is $VaR(95\%, 1) = -1,645 * S_0\sqrt{\sigma_{INDF}^2}$. Volatility estimation results to the data in the sample shown in Figure 2.

Conclusion

Estimation of the model inputs used to predict the volatility of stock returns PT. Indofood Sukses Makmur, Tbk is ARIMA (3,0,3) -ARCH (5). So that the SVR model consists of 5 lags as input vector. This method is capable of performing well in modeling the volatility of stock returns with MAPE of 1.98% and R^2 = 99.99%.

 ISSN: 978-602-71169-7-9

References

[1] Abe, S. 2005. *Support Vector Machine for Pattern Classification*. Springer - Verlag. London Limited.

[2] Bollerslev, T., Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, 1986, **31**: 307-327

[3] Engle, R.F., Autoregressive Conditional Heteroscedasticity with Estimates of Variance United Kingdom Inflation, *Econometrica*, 1982, Vol. 50, No.4: 987-1007.

[4] Engle, R.F. and S. Manganeli, *Value at Risk Models in Finance*, Working Paper Series No. 75 August 2001 European Central Bank. Germany.

[5] Holton, G., *Value at Risk, Theory and Practice*, Academic Press, Boston, 2003.

[6] Jorion, P., *Value at Risk: The New Benchmarking for Managing Financial Risk*. Mc Graw Hill, 2002.

[7] Leidiyana, H. 2013. Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Risiko Kredit Kepemilikan Kendaraan Bermotor*. Jurnal Penelitian Ilmu Komputer. Sistem Embedded & Logic*. Vol. 1. No. 1: 65-76.

[8] Morgan J.P., *Risk Metrics – Technical Document*, J. P. Morgan Global Research Foueth Edition, Reuters, 1996.

[9] Santosa, B.2007.*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.*Yogyakarta:Graha Ilmu.

[10] Scholkpof, B., and Smola A. 2003. *A Tutorial on Support Vector Regression*.

[11] Situngkir, H. dan Surya, Y., *Value at Risk yang Memperhatikan Sifat Statistika Distribusi Return*, Working Paper WDP 2006, Bandung Fe Institute.

[12] Tsay, *Analysis of Financial Time Series, second Edition*, Wiley Interscience, A John Wiley and Sons. Inc. Publication, USA, 2005.

[13] Wei, W.W.S., *Time Series Analysis: Univariate and Multivariate Methods*, Second Edition, Pearson Education Inc., Boston, 2006.