SINGAPORE MANAGEMENT UNIVERSITY

MSC. IN QUANTITATIVE FINANCE

OPTIMISED MODEL BASED ON GARCH, ARMA, SVM AND SVR

Authors
MEIYUE LI
XINRONG LI
THU TRANG LE
SUKANYA MUKHERJEE
DANMENG ZHANG

 $Supervisor \\ \text{Dr. Benjamin EE}$

Contents

1	Introduction	2
2	Datasets	2
	2.1 Stock Index ETF	2
	2.2 Source	2
3	Methodology	2
	3.1 ARMA-GARCH Model	2
	3.1.1 Definition	2
	3.1.2 Pros and Cons of Model	3
	3.2 Support Vector Machine	3
	3.2.1 Pros and Cons of Model	4
	3.3 Support Vector Regression	4
	3.3.1 Definition	4
	3.3.2 Pros and Cons of Model	4
	3.4 Risk Management	5
	3.5 Model Applications	5
	3.5.1 ARMA-GARCH	5
	3.5.2 SVM-GARCH	6
	3.5.3 SVR-GARCH	7
	3.6 Model Optimisation	8
	3.7 Further Comparison After Optimisation	9
4	Conclusion	11
5	Reference	11
6	Appendix	13

OPTIMISED MODEL BASED ON ARMA-GARCH, SVM AND SVR

July 14, 2018

1 Introduction

Historically, ARMA-GARCH model has been adopted by researchers to make predictions about electricity prices, or to forecast exchange rates. In the recent years, with the rise of machine learning, an increasing number of techniques related are being adopted to predict financial instruments prices more accurately. In this paper, ARMA-GARCH model is tested in the first place. Then two machine learning methods called Support Vector Classification (SVC) and Support Vector Regression (SVR) are respectively combined with GARCH to predict the return of index ETF products. As far as the test results can tell, the SVR-GARCH hybrid model proves to be the best one. The implementation is carried out using Python solely throughout the project.

2 Datasets

2.1 Stock Index ETF

19 Stock ETFs with the largest trading volume are tested in our project. Please note that all the prices are daily close price (subscription) with 5-day as the rolling window.

2.2 Source

The financial data is downloaded from Bloomberg Terminal, with the help of Spreadsheet Builder.

3 Methodology

3.1 ARMA-GARCH Model

The autoregressive—moving-average (ARMA) model introduced by Peter Whittle(1951) and the Generalized ARCH (GARCH) model introduced by Bollerslev(1986) are generally accepted for measuring volatility in financial models.

ARMA models have an unconditionally non-random and constant variance, which typically serves well in effectively representing homoscedastic data. The GARCH models feature variable variance that is non-random when conditioning in the past. ARMA and GARCH can be combined to model the dynamics of stock indices and their volatilities.

3.1.1 Definition

The general linear autoregressive moving average (ARMA (p, q)) model for conditional mean is expressed as

$$y_t = \mu + \sum_{i=1}^p \rho_i * y_{t-i} + \sum_{j=1}^q \theta_j * \epsilon_{t-j} + \epsilon_t$$
 (1)

where y_t is the time series needed to be modelled, μ is a constant, p is the number of autoregressive orders, q is the number of moving average orders, ρ_i is autoregressive coefficients, θ_j is moving average coefficients and ϵ_t is the error.

In the last formula, the term $\sum_{i=1}^{p} \rho_i * y_{t-i} + \sum_{j=1}^{q} \theta_j * \epsilon_{t-j}$ is the deterministic component that presents the forecast of the current state as the functions of past observations and errors. The term of error ϵ_t is the random component, which is commonly assumed to be zero mean and constant variance.

However, for some practical time series, the error terms ϵ_t do not satisfy the homoscedastic assumption of constant variance. The time-varying variance, which depends on the observations of the immediate past, is called conditional heteroscedastic variance. The conditional variance of ϵ_t , σ_t^2 , is defined as

$$\sigma_t^2 = Var_{t-1}(\epsilon_t) = E_{t-1}(\epsilon_t^2) \tag{2}$$

The generalized autoregressive conditional heteroscedasticity (GARCH (p, q)) prediction model for the conditional variance of ϵ_t is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i * \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j * \epsilon_{t-j}^2$$
 (3)

with constraints $\sum_{i=1}^{p} \beta_i * + \sum_{j=1}^{q} \alpha_j < 1, \ \alpha_0 > 0, \ \beta_i \ge 0 (i=1,2,\ldots,p), \ \alpha_j \ge 0 (j=1,2,\ldots,q).$

So, the GARCH-ARMA model can be summarized in the following equations

$$y_{t} = \mu + \sum_{i=1}^{p} \rho_{i} * y_{t-i} + \sum_{i=1}^{q} \theta_{j} * \epsilon_{t-j} + \epsilon_{t}, \epsilon_{t} \sim N(0, \sigma_{t}^{2})$$
(4)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \beta_i * \sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j * \epsilon_{t-j}^2$$
 (5)

$$\epsilon_t = \sigma_t Z_t, Z_t \sim N(0, 1) \tag{6}$$

3.1.2 Pros and Cons of Model

The simplicity and explicitness of ARMA-GARCH model structures have lots of advantages. It is shown that under the same confidence level, the ARMA-GARCH model can predict the more accurate confidence intervals compared to the traditional ARMA model. ARMA-GARCH model can capture volatility clustering effect of rates and hence is an appropriate model for forecasting it.

The major complication with the ARMA-GARCH is its estimation, the application of maximum likelihood estimation does require that the distributional assumptions be valid and that the estimation converges.

3.2 Support Vector Machine

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the

examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

3.2.1 Pros and Cons of Model

On the one hand, SVM does not assume that there is a probability density function over the return series but it adjusts the parameters relying on the empirical risk minimization inductive principle. By comparing with other models, such as GARCH, which performs well when the data is drawn from a Gaussian distribution, but if the probability density function is not Gaussian, then SVM will give a better result.

On the other hand, SVM is time-consuming and computational demanding mainly because SVM is a binary classifier. Also, compared with logistic regression and neural networks, SVM has high algorithmic complexity and extensive memory requirements when it comes to large-scale tasks.

3.3 Support Vector Regression

Support vector regression (SVR) is an efficient model for regression problems. It has many attractive properties, including sparsity, robustness and excellent generalization ability. The SVR formulation is modified in the sense of ridge regression and taking equality instead of inequality constraints in the problem formulation. As a result, a linear system instead of a quadratic programming problem is solved.

3.3.1 Definition

Given training set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ where $x_i \in R^d$ and $y_i \in R$, the goal of SVR is to find the prediction function $f(x) = w^T \phi(x)$ (the bias term is dropped by preprocessing the data with zero mean). The formulation of SVR can be expressed as

$$\min_{w,\xi(*)} \frac{1}{2} ||w||_2^2 + Ce^T(\xi + \xi^*) \tag{7}$$

s.t.

$$\langle \phi(x_i), w \rangle - y_i \le \epsilon + \xi_i$$
 (8)

$$y_i - \langle \phi(x_i), w \rangle \le \epsilon + \xi_i$$
 (9)

$$\xi_i, \xi_i^* \ge 0, i = 1, 2, \cdots, l.$$
 (10)

In most cases, formula (1) can be solved easily in the following dual formulation.

$$\min_{\hat{a}} \frac{C}{2} \hat{a}^T H \hat{a} + g^T \hat{a} \tag{11}$$

s.t.

$$0 \le \hat{a} \le e \tag{12}$$

Here,
$$H = \begin{bmatrix} K & -K \\ -K & K \end{bmatrix}$$
, $K \in \mathbb{R}^{l \times l}$ is the kernel function, $K_{ij} = K(x_i, x_j)$, $i, j \in \{1, 2, \dots, l\}$, and $g = \begin{bmatrix} \epsilon e - Y \\ \epsilon e + Y \end{bmatrix}$, $\hat{a} = \frac{1}{C} \begin{bmatrix} \alpha_* \\ \alpha \end{bmatrix}$, where α^* and α are Lagrange multipliers.

3.3.2 Pros and Cons of Model

The main advantage of SVR is the ability to generate nonlinear decision boundaries through linear classifiers while having a simple geometric interpretation. The cost function in SVR is formulated by equality instead of inequality constraints and an LS loss function. This reformulation significantly simplifies the problem in such a way that the solution is estimated by solving a simple linear system instead of the QP problem.

However, one of its potential drawbacks is that it is not robust to outliers due to the LS loss function. In real-world applications, outliers may arise due to various reasons, such as mechanical faults, fraudulent behavior, instrument error or human error. In order to decrease the influence of outliers, many methods have been proposed to improve the robustness of SVR for nonlinear regression applications.

3.4 Risk Management

Estimating the risk of loss to an algorithmic trading strategy, or portfolio of strategies, is of extreme importance for long-term capital growth. Many techniques for risk management have been developed for use in institutional settings. One technique in particular, known as Value at Risk or VaR, is put into use in the risk management of the strategy. The VaR equation is expressed as below

$$VaR_{\alpha}(x) = \inf (x \in \mathbb{R} : \mathbb{F}_{x}(x) > 1 - \alpha)$$

VaR provides an estimate, under a given degree of confidence, of the size of a loss from a portfolio over a given time period. A "portfolio" can refer to a single strategy, a group of strategies, a trader's book, a prop desk, a hedge fund or an entire investment bank. The "given time period" will be chosen to reflect one that would lead to a minimal market impact if a portfolio were to be liquidated.

The application of VaR in the project has addressed the worst loss at the 95% confidence level for all the exchange-traded funds discussed. Beginning with the daily returns over a year, the 95th percentile worst loss is recorded for each year. This provides the movement of the value at risk for each year for each ETF over its operation period.

It is observed that the VaR behaves almost similarly across the ETFs. This shows that ETFs usually do not vary much with their riskiness, so it is plausible that the choice across the ETFs be based majorly on returns.

3.5 Model Applications

3.5.1 ARMA-GARCH

The first try is given to the ARMA-GARCH model. Nevertheless, the exact value of the index is difficult to predict, so only the trend is expected to be found.

To begin with, the moving average is applied to manage the data. The rolling window is set to be 5 days, of which the reasons are as follows. First, if the data of a long period is applied, the value of the result will approach 0 very easily because the long-term expectation usually converges to 0. Thus, the data for a long period is not suitable for this research. Second, ror the rolling window of 5 days, 10 days and 20 days, the 5-day rolling window is the best. Then the parameters p and q are determined based on Bayes'Theorem. However, when the 5-day rolling window is adopted, there is a great probability that the time series will fluctuate and the process of moving average will be irreversible. So the strategy should be changed to avoid these two situations. The data of average return for 5 days will be used as the predicted value. In the meantime, due to the small quantity of 5-day data, the maximum number of parameters used in ARMA model and GARCH model are both 3.

Additionally, before putting GARCH model into practice, it can be found that the price decreases when the return fluctuation is large. Therefore, ARMA model is used to calculate the predicted value of the return. Besides, GARCH is applied in volatility prediction. Then, if the predicted value of volatility is larger than the current volatility, then the prediction of a downward trend could be made and the predicted value of the return next day equals ARMA model minus 2 times the result of GARCH model. If the predicted value of volatility is lower than the present volatility, a downward trend is more likely. Then the predicted value for next day equals ARMA model plus 8 times the result of GARCH model. It should be noted here that the number 2 is a value chosen

after comparing results given by different by other coefficients. The number 8 is a value chosen to make the model more sensitive to the bullish market. These two numbers can be dynamically changed according to real practice.

As for the trading strategy, if the predicted return for the next day is above 0, the investor can long the stock. On the contrary, if the value is below 0, the investor can short the stock. It should be noticed here that for simplicity, the transaction costs and other fees are not involved.

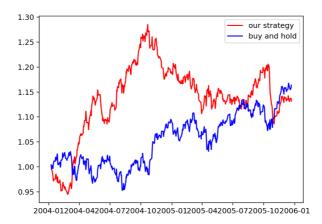


Figure 1: ARMA-GARCH hybrid model simple return (Data: SPY)

As is shown in the graph, the result is still far from perfect, because the strategy outperforms the market only during the bearish period, and when it comes to the bullish market, it fails to take effect.

The reasons should be put into the following points. First, even if GARCH does a great job in predicting volatility, the ARMA part cannot give a relatively accurate prediction of the return. Second, there is no dynamic adjustment about the buying and selling signals. Third, the ETF return is only closely related to the recent data, which means it is not so sensitive to the less recent data or, so to speak, the less recent data contains too much useless information. To sum up, a better model is expected using other techniques.

3.5.2 SVM-GARCH

Then the attention is shifted towards the Support Vector Machine method, and the corresponding answer is given in the following graph. It should be noted here only 6 factors have been used in the model: 5-day rolling window, volatility, open price, close price, high price and low price.



Figure 2: SVM-GARCH hybrid model simple return (Data: SPY)

A few reasons can account for the utter failure. First, if the six prices are used as factors of

SVM, the multicollinearity among them can also lead to inaccurate output. Second, the feature engineering work has never been paid a great attention to, let alone the parameter tuning part. In our case, only 6 factors are put into use, which put the result at a great risk of underfitting. The daily data of most of the macroeconomic variables or features are not available, which to a great extent limit the number of factors and thus hinders our feature-engineering work. So it is natural taht the performance of the SVM model is far from satisfactory.

3.5.3 SVR-GARCH

Since it is no longer realistic to apply SVM in our research, more thinking about another branch of support vector technique is taken into serious consideration. To be more specific, GARCH is combined with support vector regression method this time. The backtesting result is given as follows.

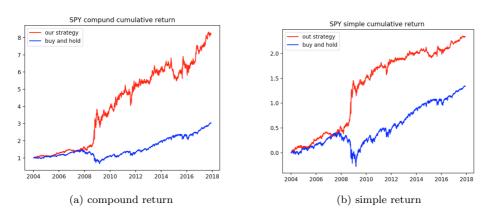


Figure 3: SVR-GARCH hybrid model before optimisation(Data: SPY)

At the first glance, it appears much better than its SVM counterpart, in both cumulative return and simple return. However, anyone who takes a closer look will find that the strategy only outperforms the benchmark after the year of 2008 and lose its advantage during the 2008 global financial crisis and the bullish time before 2008. In addition, the lines are frightfully volatile compared with the benchmark.

Then the data of DDM is applied in the model and give the results below, which further confirms our initial conclusion. Apparently, at this stage, all the work left to do is to smooth the lines and improve its performance during the bullish time prior to 2008.

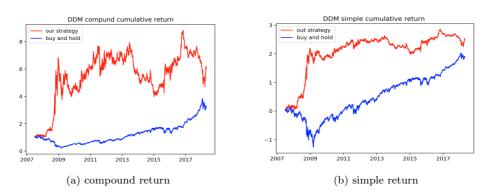


Figure 4: SVR-GARCH hybrid model before optimisation (Data: DDM)

3.6 Model Optimisation

To optimise the hybrid model, something called Bear-Bull-Coefficient (BBC) is introduced into the SVR-GARCH hybrid model. It is a coefficient which ranges from 0 to 4, with 0 represents the strongest market, 4 the weakest and 2 the initial value. The coefficient is dynamically changed according to the market performance. Further technical details about the idea can be found in the following part.

About BBC Adjustment:

if $P_{predict} > P_{t+1}$:

$$BBC = BBC + 2$$

else:

$$BBC = BBC - 2$$

About Trading Execution:

$$BBC = 2, \quad sup\{BBC\} = 4, \quad inf\{BBC\} = 0$$
 if
$$P_{predict} < P_t \text{ , and } Vol_{predict} \times f(BBC) > g(BBC, Vol_{last5days}) :$$

$$Sell: -1 \times h(BBC)$$

$$\text{where } f(BBC) = [1 - 0.1 \times (5 - BBC)],$$

$$g(BBC, Vol_{last5days}) = Vol_i, \quad i = BBC \in [0, 4], Vol_0 > Vol_1 > \dots > Vol_5$$

$$h(BBC) = BBC \times 0.2 \times (1 + \frac{BBC}{10}), \quad inf\{h(BBC)\} = 0.2, \quad sup\{h(BBC)\} = 1$$

else:

Buy:
$$+1$$
 (full position)

After optimisation, it can be seen that our strategy lines remain above the benchmark throughout the time and they appear to be less volatile or smoother than previously. So we decide to try the optimised strategy on more stock index ETFs.

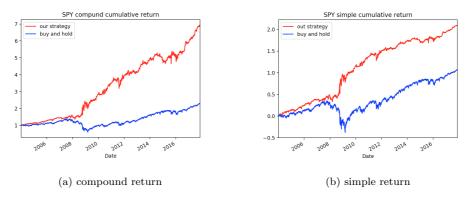


Figure 5: SVR-GARCH hybrid model after optimisation (Data: SPY)

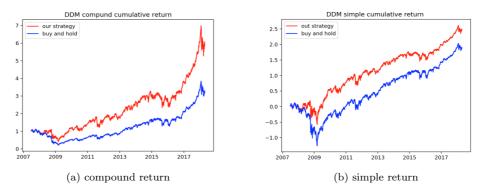


Figure 6: SVR-GARCH hybrid model after optimisation (Data: DDM)

3.7 Further Comparison After Optimisation

From the 19 index ETFs we have only selected three which we find the most typical to illustrate our backtesting results. They are generally categorized into three groups: Good, Draw and Bad. The complete sets of testing results can be found in the appendix.

(1) Good performance:

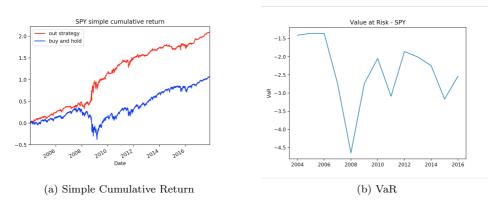
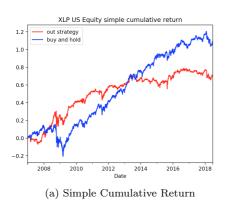


Figure 7: Good Performance (Data: SPY)

The good performance means that as our strategy manages to beat "buy-and-hold" most of the time, as shown by figure 7. Among all the good performance (10 ETFs) of appendix, however, it is apparent to find that it works really well if the underlying has demonstrated a trace of "trend", which means there is no large volatility there. It should be noticed that during the 2008 global financial Crisis, when most of the ETFs dropped down like manhole covers, our strategy generates considerable profits.

(2) Draw performance:



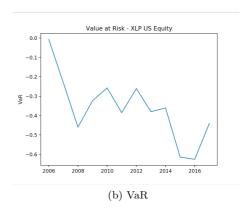


Figure 8: Draw Performance (Data: XLP)

As to Draw performance, it signifies a condition where it is not certain to say that our strategy can generate a higher return than buy-and-hold. Figure 8 shows that the return of our strategy is higher than "buy-and-hold" within the first half timeframe, but it suddenly reverses in 2013 and fails to outperform the benchmark for the rest of the time. It should be noted here that our strategy loses its effect when the index ETF shows a tremendous upward or downward trend.

(3) Poor performance:



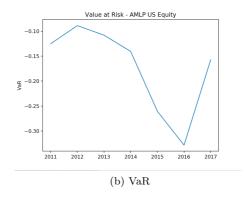


Figure 9: Draw Performance (Data: AMPL)

When it comes to Poor performance, our strategy never defeats the market. This is mainly because the benchmark shows an downward trend itself and the volatility is much larger than other cases. The two tables below present the five main indicators of backtesting results after optimisation. Please note that only eleven cases are selected here because they are more representative. They are divided into two groups. Table 1 consists of those produce positive annual return, Sharpe Ratio and information ratio, whilst Table 2 gives a relatively unsatisfactory result of the strategy.

In sum, they only further support the conclusion that the strategy is not so stable across the index ETFs. Even if it demonstrates a comparatively good performance in Table 1, the max drawdown is surprisingly high which means the strategy can be way too risky for the investors.

INDICATOR	SPY	USO	XLU	XLF	XLE	EFA
Annual Return	13.360%	1.773%	0.109%	14.284%	11.492%	12.008%
Max Drawdown	-44.889%	-89.310%	-25.700%	-71.760%	-45.380%	-36.910%
Sharpe Ratio	0.598	0.006	0.614	0.420	0.404	0.366
Information Ratio	0.556	0.298	0.206	0.246	0.160	0.213
Winning Odds	54.993%	47.630%	49.420%	50.150%	49.280%	48.920%

Table 1: Five main indicators of test results after optimisation for SPY, USO, XLU, XLF, XLE and EFA

INDICATOR	XLP	VXX	UVXY	XOP	AMLP
Annual Return	5.9429%	-45.9432%	-25.7815%	2.9400%	-3.1910%
Max Drawdown	-18.450%	-415.660%	-266.180%	-69.730%	-22.850%
Sharpe Ratio	0.319	-0.986	-0.287	0.043	0.988
Information Ratio	-0.274	0.213	0.753	-0.099	0.374
Winning Odds	47.710%	45.120%	45.860%	48.540%	50.850%

Table 2: Five main indicators of test results after optimisation for XLP, VXX, UVXY, XOP and AMLP

4 Conclusion

It is obvious our strategy works well in the "trend" market and fails to generate profit at volatile market. This makes sense and is identified with the disadvantages of machine learning model. When regime changes often, including volatile and placid, trending and mean reverting, the strategy performance declines. However, what should be admired in this strategy is that during the financial crash, our strategy avoids the crisis and continue to make high return in the good performance.

Through the project the whole group, for the very first time, learns how to apply machine learning techniques into practice and combine them with time series models. The P&L obtained from the strategy is surprisingly outstanding. However, it should be pointed out that the validity of results for real trading is debatable due to the following reasons:

First, transaction costs and other fees were not taken into consideration. If the positions taken are very high, then they can have a huge impact on the P&L.

The aspect of risk-analysis of the strategy can be further explored, but initial analysis of the risk-iness based on the Value at Risk of the strategy against the exchange traded funds, reveals that the strategy is consistently similarly risky, with little change across the etfs. This could indicate that the choice across ETFs could be based on expected return alone, without much weight to risk measured as VaR.

Due to the limited timeframe, the strategy was not backtested over other time periods to find out if it works out on a longer timeline. Our future work will cover this part and the strategy will be tested on other ETF products, individual stocks and commodities to see whether something more interesting could come into the view.

5 Reference

Cortes, Corinna, and Vladimir Vapnik, 1995. Support-vector networks. Machine Learning, 20(3), 273–297.

Chen, S and KIHO, J, 2007. Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns, 1-2

Steve R. Gunn, 1998. Support Vector Machines for Classification and Regression, 1.

Manish Kumar, 2002. Forcasting Stock Index Movement/ A Comparison of Support Vector Machines and Random Forest, 1-2, 5.

Altaf Hossain, Faisal Zaman, M. Nasser and M. Mufakhkharul. IslamComparison of GARCH, Neural Network and Support Vector Machine in Financial Time Series Prediction, 4.

Phichhang Ou and Hengshan Wang, 2010. Financial Volatility Forecasting by Least Square Support Vector Machine/ Based on GARCH, EGARCH and GJR Models: Evidence from ASEAN Stock Markets. International Journal of Economics and Finance, 2(1), 51-56.

Yi Yang, Yao Dong, Yanhua Chen and Caihong Li, 2014. Intelligent Optimized Combined Model Based on GARCH and SVM for Forecasting Electricity Price of New South Wales, Australia. Hindawi, 2014, 1-8.

Mergani Khairalla, Xu-Ning and Nashat T. AL-Jallad, 2017. Hybrid Forecasting Scheme for Financial Time-Series Data using Neural Network and Statistical Methods. International Journal of Advanced Computer Science and Applications, 8(9), 319-320, 326.

Fernando Perez-Cruz, Julio A Afonso-Rodriguez and Javier Giner, 2003. Estimating GARCH models using support vector machines. Quantitative Finance, 3(2003), 1-2, 9.

R. F. Engle. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica, 50, 987–1007.

T. Bollerslev, 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31(3), 307–327.

W.C. Wong, F. Yip, and L. Xu. 1998. Financial Prediction by Finite Mixture GARCH Model," Proceeding of Fifth International Conference on Neural Information Processing, 1351–1354.

Radha, S. and Thenmozhi, M. 2006. Forecasting Short Term Interest Rates Using Arma, Arma-Garch and Arma-Egarch Models. Indian Institute of Capital Markets 9th Capital Markets Conference Paper.

A.A.P. Santos, N.C.A. Costa Jr., and L.S. Coelho. 2007. Computational intelligence approaches and linear models in case studies of forecasting exchange rates, Expert Syst. Appl. 33(4), 816–823.

Jeongcheol Ha, Taewook Lee. 2011. NM-QELE for ARMA-GARCH models with non-Gaussian innovations, Statistics and Probability Letters, 81(6), 694-703.

Heping Liu, Jing Shi. 2013. Applying ARMA-GARCH approaches to forecasting short-term electricity prices, Energy Economics, 37,152-166.

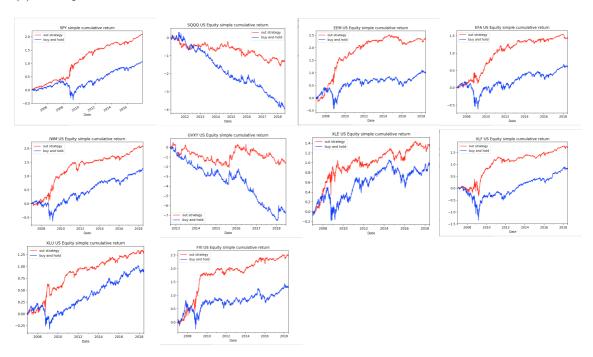
Xie, H., Li, D., Xiong, L. 2016. Exploring the regional variance using ARMA-GARCH models. Water Resources Management, 30(10), 3507-3518.

Neha, C. R., Tiwari, P. 2016. Forecasting volatility spillover of information technology sector stocks in india: An application of ARMA and GARCH model. International Journal of Business Analytics and Intelligence, 4(2), 36-44.

6 Appendix

(due to the page limit, the rest of the plots are presented here)

(1) Good performance:



(2) Bad performance:

