Analysis of Selected Stocks' Risk and Future Performance Using Copula and Time Series Methods

GR5261: Statistical Methods in Finance, Group 8

I. Introduction

In our project, we try to answer two major concerns when investors are given selected stocks to construct a portfolio: First, how risky are they? Second, how will they probably perform in the future? In the first part of this report, two major risk measures, Value at Risk (VaR) and Conditional VaR (CVaR), are assessed using copula models. In the second part, the return of stocks are predicted using time series models, including ARIMA and ARIMA-GARCH(1,1) model. The whole process of model selection, diagnostics and relevant tests are performed.

VaR measures the maximum potential loss on a group of securities over some time period T, given a specified probability. CVaR is the weighted average of the "extreme" losses in the tail of the distribution of possible returns, beyond the value at risk (VaR) cutoff point. Major ways to calculate them include historical method and Monte-Carlo simulation method. Historical method treats all historical observations equally and re-organize actual historical returns, putting them in order from worst to best, and then take the quantile according to the given probability α. Monte Carlo VaR is fitting the model to risk factors and calculate VaR by simulation. It uses a computer program to generate a series of random numbers to predict scenarios (or market conditions). The value of the portfolio being assessed is then calculated for each set of market conditions generated. Both two methods have their drawbacks. For further explanation and comparison of them, you may refer to #2 in appendix.

Sklar's theorem [4] states that a collection of marginal distributions can be coupled together via a copula to form a multivariate distribution. A copula is a multivariate CDF whose univariate marginal distributions are all Uniform(0,1). For example, the student's t copula, which we decide to use for risk assessment purpose in the following steps, has such form: $C_{v,P}^{t}(\boldsymbol{u}) := \boldsymbol{t}_{v,P}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_d))$, where $\boldsymbol{P}$ is a correlation matrix, $\boldsymbol{t}_{v,P}$ is the joint CDF of $\boldsymbol{X} \sim \boldsymbol{t}_d(v, \boldsymbol{0}, \boldsymbol{P})$ and t_v is the standard univariate CDF of a t-distribution with v degrees of freedom. Since copula can describe multivariate dependence, it's natural to think that it may be applicable in portfolio risk measurement [3].

Autoregressive integrated moving average (ARIMA (p, d, q)) model, is a generalization of an autoregressive moving average model with differential term d. When d=0, it's equal to the ARMA model. The model are as follows: $Y_t = (1-B)^d$, where B is the differential term. $Y_t - \alpha_1 Y_{t-1} - \cdots - \alpha_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$. One important assumption of ARIMA model is the homoscedasticity. If ARIMA model is assumed for the error variance, the model is a GARCH model. The model is as follows: $\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$

When we combine ARIMA model (for Yt) and GARCH model (for variance) together, it becomes an ARIMA-GARCH model. It can capture the trend of value and volatility of a time series.

Our dataset is the stock prices of VZ, GOOGL, MSFT from 2010-01-01 to 2018-12-31 and FB from 2012-05-21 to 2018-12-31 (ever since it was first listed), which are available on Yahoo Finance. The risk assessment part uses partial of it (2017-01-01 to 2018-12-31) and the time series part uses the whole dataset for training. Our computer programming is done in R.

II. Methodology

Part One: Risk Assessment

To make the data from different stocks comparable and stable, we transform the price p into a log-return series r: $r_{t,i} = \ln \frac{p_{t,i}}{p_{t-1,i}}$, where $i \in [1,d]$, d is the number of assets, $t \in [1,T]$ is a time point, T is the maximum time.

Next, we investigate the univariate marginal distribution of each log-return series, and conduct proper transformation of the log-return and generate uniform marginal distributions, in order to fit the data to a copula in the next step. A common way [1,2] is to generate pseudo observations: $u_{t,i} = \frac{rank(r_{t,i})}{T+1}, i \in [1,d], t \in [1,T]$. By doing this, each $u_{t,i}$ has value between 0 and 1, and each u_i approximately follows Uniform(0,1) distribution.

Now that we have uniform marginal distributions, we can fit data to a copula. We try several copulas and decide on which to use according to the goodness-of-fit test results based on Cramer-von Mises statistic S_n.

After a copula is fitted, we use the following algorithm [2] to calculate the risk measurements. To fully understand it, it can be thought of as a combination of historical and simulation method: It has the process of generating random samples from specified condition (a given copula), while in the final step it takes the quantile of hypothetical profit & loss, which seems more like historical method.

Algorithm 1: Computation of Risk measures by a Copula

Input: Log-returns $\{r_{i,t}\}$, weights w_i of optimal portfolio, $i \in [1,d]$, where d is the number of dimensions characterized by the number of stocks in our portfolio, level α for VaR_α and $CVaR_\alpha$ calculation.

1: Generate a sample of pseudo-observations $\{\hat{u}_{j,s}\} \in [0,1]^d, j \in [1,d], s \in [1,S]$ is the number of random samples.

2: Transform simulated pseudo-observations to univariate quantiles: Let $k = 1, 2, \dots, K$, where $K = S \cdot d$,

3: **for** $i \in [1,d]$ { **for** $j \in [1,d]$ { **for** $s \in [1,S]$ **do** Set $\hat{s}_{i,k} \leftarrow \mathbb{Q}_{\hat{u}_{j,s}}(r_i)$ **End for } End for } End for**

*Explanation: this line means to set $\hat{s}_{i,k}$ to be the value of the $\hat{u}_{j,s}{}^{\text{th}}$ quantile of the log-return series r_i.

4: Compute the portfolio Profit & Loss series: **for** $k \in [1,K]$ **do** $P\&L_k = \sum_{i=1}^{d} \hat{s}_{i,k} \cdot w_i$ **End for**

5: Calculate VaR_α and $CVaR_\alpha$ of the hypothetical Profit & Loss series (i.e. taking the quantile of the Profit & Loss series)

Output: VaR_α and $CVaR_\alpha$ of simulated $CVaR_\alpha$ of the Profit & Loss series.

Part Two: Time Series Prediction

The procedure of conducting time series prediction [5] is shown in figure 1. It's a standard procedure of how to use time series models to predict log-return of stocks. The outcomes are shown through the functions and plots. Besides, the weakness of the model will also be drawn.

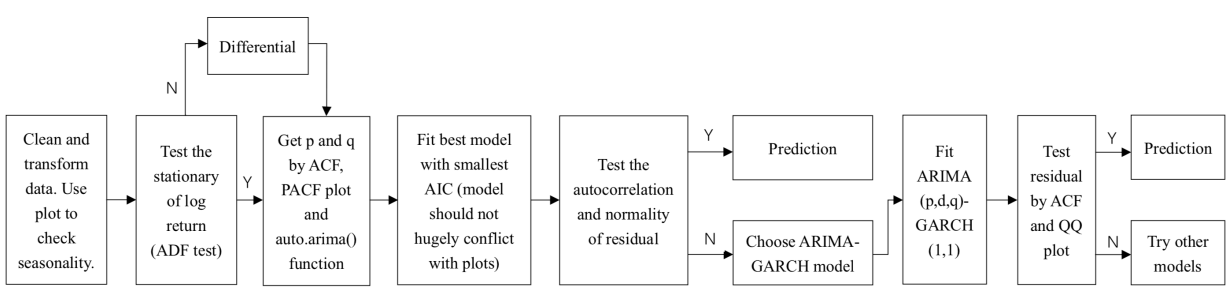

Fig. 1 Time series prediction procedure

III. Implementation and Result

Part One: Risk Assessment

Step 1. Investigation of marginal distribution

In our method we are not interested in the specific analytical form of marginal distributions. (Indeed we have some investigation on the non-normality of marginal distribution, and you can find it in #4 of attachment.) What we have to know is that they are not uniform, and thus we generate pseudo-observations for fitting copula using maximum pseudo likelihood method.

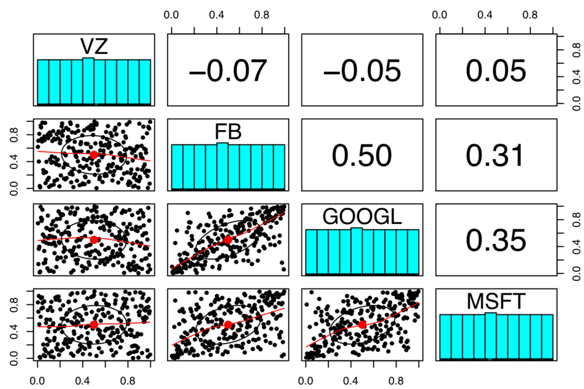

Fig. 2 Pseudo-observations overview

Step 2. Fitting copula models to data

We can verify from figure 2 that pseudo observations we generate from historical observations are very close to uniform distribution, which can be used as the input for copula models. Different copulas are attempted, and the goodness-of-fit test result shows that t copula is the best fit. Therefore, we only consider to use t copula in further steps.

Copula	GoF Test Statistic	p-value
Gaussian	0.03378	0.1733
Student's t	0.03107	0.2802
Clayton/Gumbel /Joe/Frank	0.14297/0.21535/ 0.37422/0.18157	All <0.05

Table 1 GoF test result of different copulas

Step 3. Conducting risk assessment using copula

We first try to calculate the VaR and CVaR on a single day (Jan 2, 2018). Table shows the result, together with historical risk for comparison, and standard deviation obtained from a bootstrap process.

Level, %	Copula VaR/CVaR, %	Historical VaR/CVaR, %	SD of Copula VaR/CVaR, %
99	-2.5381937/-3.253638	-1.884848/-2.210461	0.16471956/0.17023056
95	-1.3418583/-2.127460	-0.9005257/-1.600650	0.06932138/0.11300544
90	-0.8879561/-1.600566	-0.6030257/-1.172797	0.05218309/0.08091348

Table 2 Single day (Jan 2, 2018) risk assessment result

Then we conduct risk assessment over a time period (Jan 2, 2018 to Dec 31,2018). In this process, a new copula is fitted at every time point from Jan 1, 2018 to Dec 31, 2018, using data of 1 year looking back from that time point. The results are shown in figure 3, together with historical and Monte-Carlo value for comparison. Judging from table 2 and figure 3, the result given by copula is close to historical value, and they share similar movement patterns. Moreover, the copula risk is always a little larger (in terms of absolute value) than historical value, which means copula has superior risk prediction ability than traditional empirical (historical) method and Monte Carlo method, because it means we will lose no more than we predict by copula model.

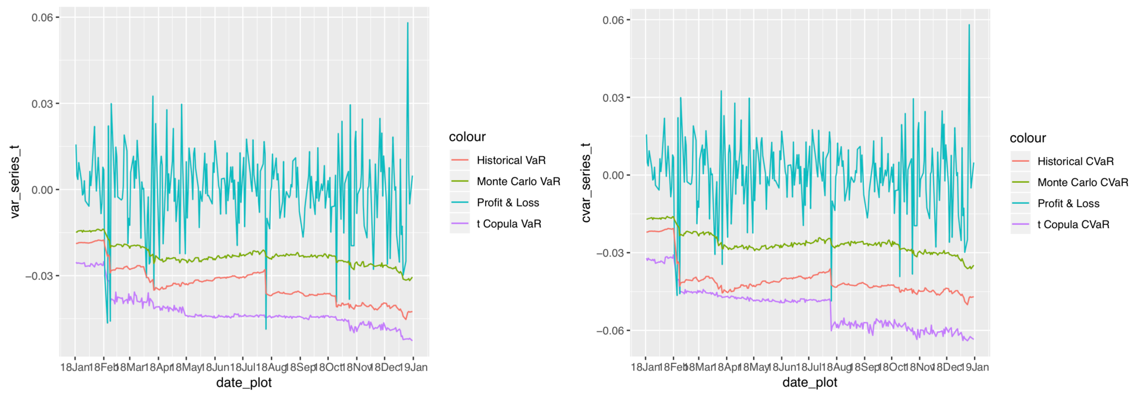

Fig. 3 Risk assessment over 1 year (2018)

Part Two: Time Series Prediction

We need to perform the procedure for four times. If some plots are needed only one example will be shown. Some plots of more importance are presented in #3 of attachment, and others are in #5 of attachment.

Step 1. The stock prices are transformed to log-return without NAs. All the plots show no seasonality effect in the series.

Step 2. Performs the Augmented Dickey-Fuller test for the null hypothesis of a unit root of a univariate time series x. Since all the P-value are less than 0.05 for lag 1 to lag 8, the null hypothesis is rejected and the series are all stationary.

Step 3. Plot ACF and PACF for time series to get p,q and compare with the result of auto.arima(). If the auto fit result is severely conflict with plot, it should not be considered. Take MSFT as an example.

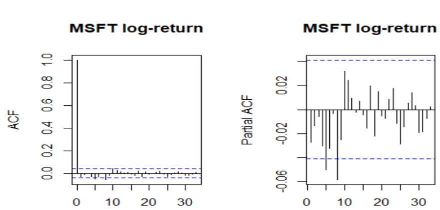

Fig. 4 ACF and PACF of MSFT

From the plot, p=8 and q=8. From the auto fit function, p=1 and q=1. Even though the AIC of second result is better than the first one, it's conflict with the plot. So, the model of MSFT is ARIMA(8,0,8). Using the same process, we get the following models. FB: ARIMA(2,0,2) ; VZ: ARIMA(2,0,0); GOOGL: ARIMA(0,0,8).

Step 4. After fitting the model, the residual should be independent and normal distributed. The ACF plot and Box-Pierce test are used to check autocorrelation of residual. Histogram and Q-Q plot are used to check the normality.

The null hypothesis of Box-Pierce test is independence in a given time series. The p-value for all test are greater than 0.05, which show the residual has no autocorrelations. The ACF plots for residual also show the same result. However, all the histogram and Q-Q plot show that the residual are not normal. These indicate that we should choose an ARIMA - GARCH model.

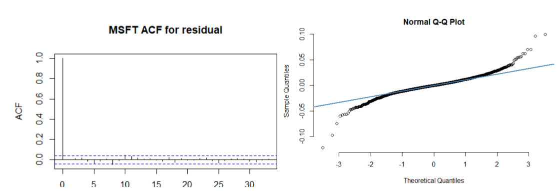

Fig. 5 ACF and QQ plot for residuals of ARIMA for MSFT

Step 5. Since all the residual of ARIMA model don't past the test for normality, the next step is to fit an ARIMA- GARCH model. The p and q in ARIMA are from previous model and the GARCH is arbitrary chosen to be GARCH (1,1) based on the reference. Since the log-return have fat tail, we use t-distribution in this model, rather than the frequently used normal distribution.

Also, we take MSFT as an example. From the ACF plot of standard residuals and QQ plot of residuals, we know that the residuals don't have autocorrelation and satisfy normal distribution.

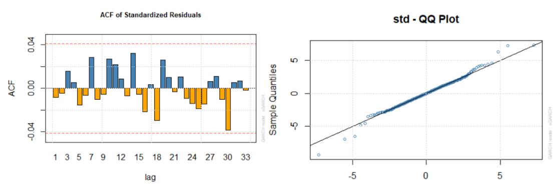

Fig. 6 ACF, QQ plot for residuals of ARIMA-GARCH for MSFT

Step 6. Use the fitted model to do the prediction

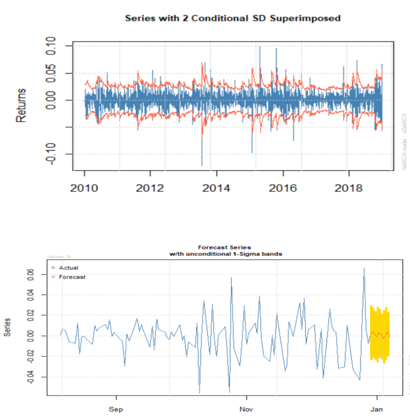

Fig. 7 fitted and predicted plot of ARIMA-GARCH for MSFT

The first plot is based on the historical data, it shows our model explains most information of original time series. The second plot shows that the volatility of the time series is to large that our model can just capture part of them. This can be verified by the plot of fitted value vs. true value.

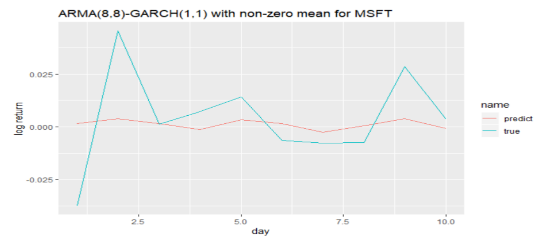

Fig. 8 true and predicted value of ARIMA-GARCH for MSFT

Comparison of ARIMA and ARIMA-GARCH

The model can capture some the trend of the time series, but the true volatility is too large for this model. Besides, corresponding to the prediction of ARIMA model, it does better in capturing the trend. This is because the GARCH(1,1) term in the model fitted the different volatility in the model, rather than regard them as same.

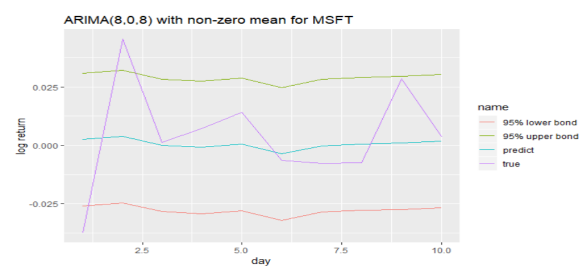

Fig. 9 true and predicted value of ARIMA for MSFT

IV. Conclusion

In the first part of our project, we verify that copula model is a possible approach to measure portfolio risk because of its ability of describing multivariate dependence. Compared with the result from historical and Monte Carlo method, copula method gives close but a little bit more conservative risk estimation, which indicates that copula has superior risk prediction ability than traditional method, making it especially a great risk control tool for risk-averse investors.

In our second part, we give prediction of stock return using ARIMA and ARIMA-GARCH model, and we succeed in capturing some trends. Nevertheless, the models are not quite precise, and for VZ the prediction is bad. For future improvement, rather than arbitrary use GARCH(1,1), we may need to fit a better model for the volatility in the time series. Also, when fitting the ARIMA-GARCH model, the standard T distribution is used to model the distribution of time series. Because, even though we fit and test the specific distribution, the "rugarch" package in R is not flexible for that, so more work could be done to achieve a more accurate prediction result.