March 2017 | 28

AR-GARCH, VaR and Portfolio Selection

CHEN Wanli Desiree

GOH Zhi Wen Nicholas

Lu HE

Yumin ZHOU

Executive Summary

This paper is intended to use the AR (1)-GARCH (1,1) model to estimate the return series of a portfolio of 4 stocks, i.e. IBM, Apple (“AAPL”), CVS and Walmart (“WMT”) for the period from 3 January 2006 to 31 December 2015. The models provide recommendations on (i) how to estimate the portfolio VaR and (ii) how to optimize the portfolio weights by maximizing expected daily returns and minimizing daily 99% VaR.

Key Findings

Models and Estimated Parameters

We used AR (1)-GARCH (1,1) on the **Training Dataset (1st half period from 3 January 2006 to 31 December 2010)** to estimate the parameters. Multi-variate models were derived from the return series of the 4 stocks, while uni-variate models were derived from a single return series which represented the daily portfolio return on an equally-weighted basis.

For robustness, both Gaussian Distribution and Student-t Distribution are used and compared in Table 1 as follows:

Table 1: AR (1)-GARCH (1,1) Estimated Parameters on 1st Half Dataset

	Multi-variate										Uni-variate	
	IBM		AAPL		CVS		WMT		Portfolio		Portfolio	
Distribution	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
μ	0.000744	0.000847	0.002821	0.002170	0.000802	0.000745	0.000410	0.000214	NA	NA	0.001159	0.001265
ar1	-0.054966	-0.050905	0.009512	-0.004253	-0.040941	-0.054975	-0.027649	-0.025071	NA	NA	-0.040581	-0.032132
ω	0.000005	0.000003	0.000010	0.000004	0.000022	0.000015	0.000002	0.000001	NA	NA	0.000004	0.000003
α	0.103517	0.089535	0.081256	0.071068	0.118508	0.081748	0.051453	0.054694	NA	NA	0.100433	0.09413
Β	0.875081	0.897162	0.903892	0.926294	0.835608	0.875610	0.940025	0.940249	NA	NA	0.878166	0.891415
Shape	NA	6.402012	NA	6.305381	NA	4.952232	NA	5.222558	NA	NA	NA	7.313540
Log-Likelihood									14036.67	14307.84	3779.025	3795.498
Akaike Information Criteria									22.254	22.676	5.9953	6.0199

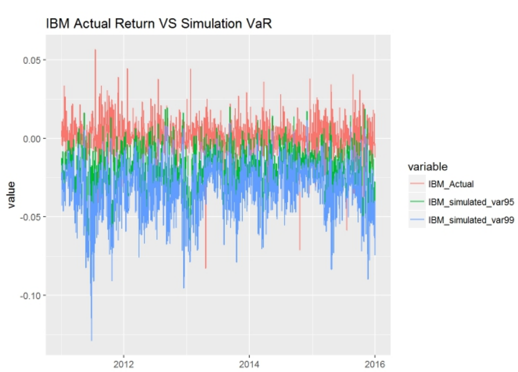
Note: (1): Gaussian Distribution; (2): Student-t Distribution, NA means Not applicable.

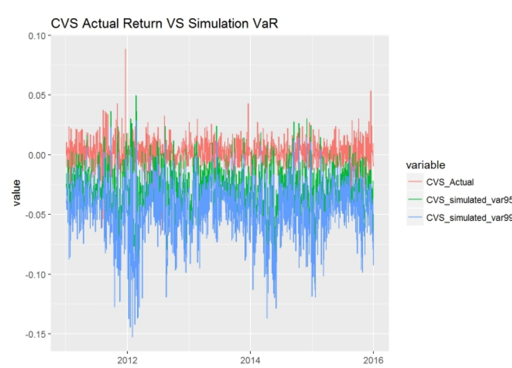
Based on the **maximum log-likelihood**, **Student-t Distribution** gave a slightly better fit as compared to Gaussian Distribution under both multi-variate and uni-variate scenarios.

Calculations of Out-of-Sample VaR

For multi-variate models, we used the estimated parameters from above to simulate **1258*4** data points, where 1258 represents the number of days for later out-of-sample comparisons with the **Test Dataset (2nd half period from 1 January 2011 to 31 December 2015)**; and obtained the daily **mean** and **covariance matrix.** Similarly, for uni-variate models, we simulated **1258** data points for portfolio returns and obtained the estimated daily mean and volatility.

Sample outputs of the simulated VaRs under Gaussian Distribution models are shown in the graph below:





At the respective confidence level (95% or 99%), a **violation** occurs when the daily return in the Test Dataset **goes below** the simulated VaR. The ***percentage of violation*** was calculated and indicates whether the simulated VaR provides an accurate estimation of the probability of occurrence of extreme losses.

We have compared the simulated VaRs with the respective actual daily returns in the Test Dataset in Table 2 as follows:

Table 2 Out-of-Sample VaR – Percentage of Violation

	95% VaR				99% VaR			
	Multi-variate		Uni-variate		Multi-variate		Uni-variate	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
IBM	0.1160572	0.1049285	NA	NA	0.05166932	0.0508744	NA	NA
AAPL	0.1073132	0.0580286	NA	NA	0.02941176	0.0206677	NA	NA
CVS	0.0715421	0.0802862	NA	NA	0.02464229	0.0310016	NA	NA
WMT	0.1025437	0.0786963	NA	NA	0.03656598	0.0381558	NA	NA
Portfolio	0.0341812	0.0357711	0.0937997	0.1057234	0.00953895	0.00953895	0.03974563	0.04213037

Note: (1): Gaussian Distribution; (2): Student-t Distribution, NA means Not applicable.

Please refer to the section Methodology for the details on how VaR was simulated and how percentage of violation was obtained.

Results show that

- At both 95% and 99% VaR, the percentage of violation under **multivariate models** was much closer to 5% and 1%, respectively, as compared to uni-variate models.
- Gaussian Distribution and Student-t Distribution gave consistent estimation.

To further quantify the results between Gaussian Distribution and Student-t distribution, we performed the **Binomial Test** with results in Table 3 as follows:

Table 3 Out-of-Sample VaR – Binomial Test

	95% VaR				99% VaR			
n	n=1258							
Expected	63 (1258 * 5%)				13 (1258 * 1%)			
	Multi-variate		Uni-variate		Multi-variate		Uni-variate	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
IBM	146(2.2e-16)	132(4.4e-15)	NA	NA	65(2.2e-16)	64(2.2e-16)	NA	NA
AAPL	135(2.9e-16)	73(0.1953)	NA	NA	37(1.4e-08)	26(0.0008724)	NA	NA
CVS	90(0.000928)	101(5.2e-06)	NA	NA	31(1.1e-05)	39(1.4e-09)	NA	NA
WMT	129(3.92e-14)	99(1.3e-05)	NA	NA	46(2.1e-13)	48(1.4e-14)	NA	NA
Portfolio	43(0.007949)	45(0.01963)	118(1.4e-10)	133(1.5e-15)	12(1)	12(1)	50(8.3e-16)	53(2.2e-16)

Note: (1): Gaussian Distribution; (2): Student-t Distribution, NA means Not applicable.

Based on the **p-value**, we further confirmed that

- Multi-variate models, compared to uni-variate models, provide a better estimation of VaR at both 95% and 99%.
- For **95% VaR**, the multi-variate model under **Student-t Distribution** gave a better estimation than Gaussian Distribution.
- For **99% VaR**, the multi-variate models under Gaussian Distribution and Student-t Distribution gave **equally good** estimations.

Portfolio Optimization and Recommendation

Without optimization, each stock contributed to 25% of the initial portfolio value. The portfolio was optimized using quadratic programming to maximize the Sharpe ratio on the efficient frontier. The weights of each stock in the optimized portfolio, and the parameters of the optimized portfolio and non-optimized portfolio are shown in Table 4 as follows:

Table 4 Portfolio Optimization

	Optimized Weight		Mean (µ)		Volatility (σ)		99% VaR		Ratio = µ / |99%Var|	
Distribution	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
IBM	0.237217	0.245604	NA		NA		NA		NA	
AAPL	0.244709	0.226642	NA		NA		NA		NA	
CVS	0.227326	0.255575	NA		NA		NA		NA	
WMT	0.290748	0.272179	NA		NA		NA		NA	
Portfolio:										
Optimized			0.00027863	0.00029922	0.01080123	0.01053723	-0.022800	-0.022948	0.0122	0.0130
Benchmark	Equally-weighted at 0.25		0.0002934105		0.01094895		-0.0229716		0.0128	

Note: (1): Gaussian Distribution; (2): Student-t Distribution, NA means Not applicable.

The optimized portfolio built on **Gaussian Distribution** resulted in both **lower expected returns** and **lower volatility** than the non-optimized portfolio. Meanwhile, the optimized portfolio built on **Student-t Distribution** resulted in **higher expected returns** and **lower volatility** than the non-optimized portfolio.

By further looking at the **normalized ratios of expected return divided by the absolute of 99% VaR**, **Student-t Distribution** gave a better optimized result as compared to Gaussian distribution. As such, we have **beaten the benchmark model**.

Robustness testing

For robustness testing, we swop the **Training** and **Test** datasets to derive estimated parameters on the other half period. Results are compared in Table 5 and Table 6. Sample 2 means that we fit the AR (1)-GARCH (1,1) on the 2^{nd} half period, obtained the parameters and used the parameters to simulate data points for comparison with 1^{st} half actual data points.

Table 5 Robustness - Out-of-Sample VaR – Percentage of Violation

	95% VaR								99% VaR							
	Multi-variate				Uni-variate				Multi-variate				Uni-variate			
	(1)		(2)		(1)		(2)		(1)		(2)		(1)		(2)	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
IBM	0.1161	0.1271	0.1049	0.1231	NA	NA	NA	NA	0.0517	0.0580	0.0509	0.0683	NA	NA	NA	NA
AAPL	0.1073	0.1740	0.0580	0.1485	NA	NA	NA	NA	0.0294	0.0921	0.0207	0.0802	NA	NA	NA	NA
CVS	0.0715	0.1787	0.0803	0.1676	NA	NA	NA	NA	0.0246	0.0953	0.0310	0.1025	NA	NA	NA	NA
WMT	0.1025	0.1342	0.0787	0.1493	NA	NA	NA	NA	0.0366	0.0635	0.0382	0.0794	NA	NA	NA	NA
Portfolio	0.03418	0.0786	0.0358	0.0945	0.0938	0.1676	0.1057	0.1565	0.0095	0.0365	0.0095	0.0381	0.0398	0.0945	0.0421	0.0969

Results show that

- At both 95% and 99% VaR, the percentage of violation under models fitted on Sample 1 was closer to 5% and 1%, respectively, as compared to models fitted on Sample 2 models.
- Comparing Sample 2 models, multivariate models are closer to 5% and 1% respectively, as compared to uni-variate models. This also holds when we compared Sample 1 models.
- Gaussian Distribution and Student-t Distribution gave consistent estimations for both Sample 1 and Sample 2.

Table 6 Robustness - Out-of-Sample VaR – Binomial Test

	95% VaR								99% VaR							
	Multi-variate				Uni-variate				Multi-variate				Uni-variate			
	(1)		(2)		(1)		(2)		(1)		(2)		(1)		(2)	
	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)	Sample 1 (n = 1258)	Sample 2 (n = 1259)
IBM	146 (2.2e-16)	160 (2.2e-16)	132 (4.4e-15)	155 (2.2e-16)	NA				66 (0.7466)	73 (2.2e-16)	64 (2.2e-16)	86 (2.2e-16)	NA			
AAPL	135 (2.9e-16)	219 (2.2e-16)	73 (0.1953)	187 (2.2e-16)	NA				37 (1.4e-08)	116 (2.2e-16)	26 (0.0008724)	101 (2.2e-16)	NA			
CVS	90 (0.000928)	225 (2.2e-16)	101 (5.2e-06)	211 (2.2e-16)	NA				31 (1.1e-05)	120 (2.2e-16)	39 (1.4e-09)	129 (2.2e-16)	NA			
WMT	129 (3.92e-14)	169 (2.2e-16)	99 (1.3e-05)	188 (2.2e-16)	NA				46 (2.1e-13)	80 (2.2e-16)	48 (1.4e-14)	100 (2.2e-16)	NA			
Portfolio	43 (0.007949)	99 (1.341e-05)	45 (0.01963)	119 (9.056e-11)	118 (1.4e-10)	211 (2.2e-16)	133 (1.5e-15)	197 (2.2e-16)	12 (1)	46 (2.161e-13)	12 (1)	48 (1.416e-14)	50 (8.3e-16)	119 (2.2e-16)	53 (2.2e-16)	122 (2.2e-16)

Based on the **p-value** from Binomial Tests, we further confirmed that

- For sample 2, multi-variate models, as compared to uni-variate models, provide a better estimation of VaR at both 95% and 99%. This also hold for models fitted on Sample 1.

- For Sample 2, the models had low p-values. Hence, we reject the hypothesis that the models are good. Comparing the models for Sample 2, the multi-variate models under Gaussian Distribution gave a better estimation than Student-t Distribution for both 95% and 99% VaR.

Methodology

Based on the simulated data points, Out-of-Sample VaRs are calculated as follows:

$$VaR\_simulated_{i,s,confidence\_level} = u\_simulated_{i,s} + z_{confidence\_level} * \sqrt{Covariance\_simulated_{i,s}}$$

where s ∈ {IBM, AAPL, CVS, WMT}, $i \in \{1{:}\ 1258\}$, $z_{95\%} =- 1.645$ and $z_{99\%} =- 2.326$

By comparing the simulated VaRs with the respective actual daily returns in the Test Dataset, we derived the ***Percentage of Violation*** from the following:

$$Percentage\_of\_Violation_{i,s,confidence\_level} = \frac{Violation_{i,s,\ confidence\_level}}{Respective\ set\ of\ data\ points}$$

$$Violation_{i,s,\ confidence\_level} = \begin{cases} 1 & \text{If } VaR\_simulated_{i,s} > u\_actual_{i,s} \\ 0 & \text{Otherwise} \end{cases}$$

Portfolio was optimized by finding the highest sharpe ratio on the portfolio efficient frontier. We used quadratic programming to help us arrive at the solution. Please refer to Appendix III Quadratic Formulation for more information.

Visualization of data

For comparison, we plotted the individual stock return series, the out-of-sample actual returns vs simulated VaRs. Please refer to Appendix I for details.

Positive Definite Tests

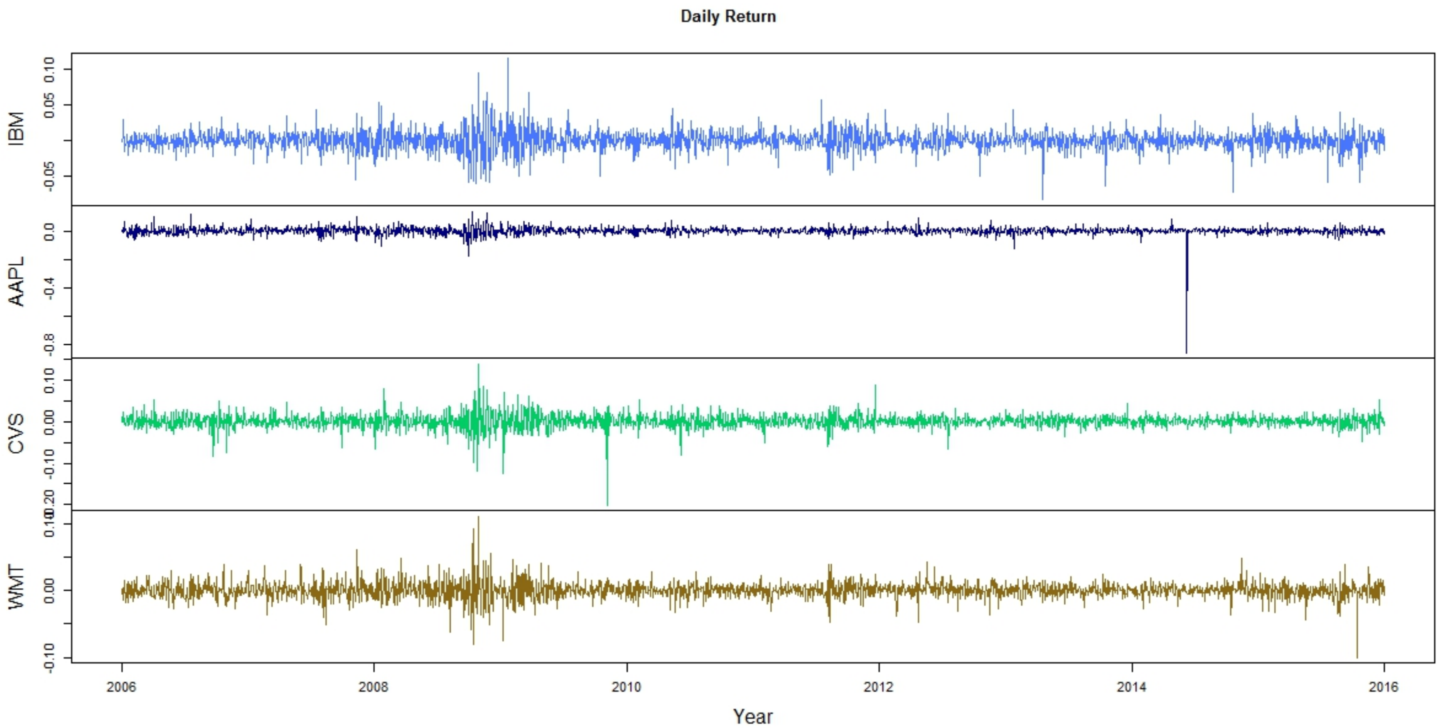
We have run stationary tests to check for stationarity and ensure that simulated covariance matrixes are positive definite. Please refer to Appendix II for the sample codes.

Assumptions and Model Risks

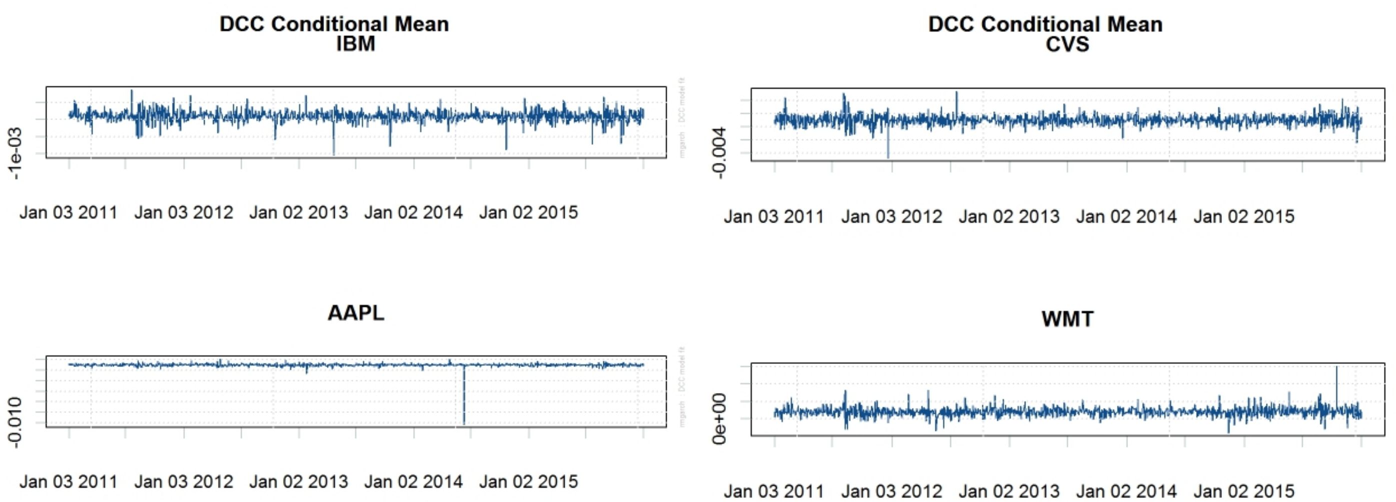
- The time series data is assumed to be stationary. Stationarity tests were performed with satisfactory results. However, there is a small chance it might be a false positive.
- A VaR at high confidence level (i.e. 99.9%) will be highly dependent on the outlier of the fitted distributions, as such, this is highly dependent on the number of samples generated from the multivariate distribution. If we generate too little data points, we might not be able to capture the extreme values, leading to poor estimation.
- We have assumed that the returns follow a normal or student t distribution. In the real world, this distribution relationship may not hold.
- We have assumed that AR-GARCH model parameters remain consistent. This may not be reflective of the real world.

Appendix I Data and Results Visualization

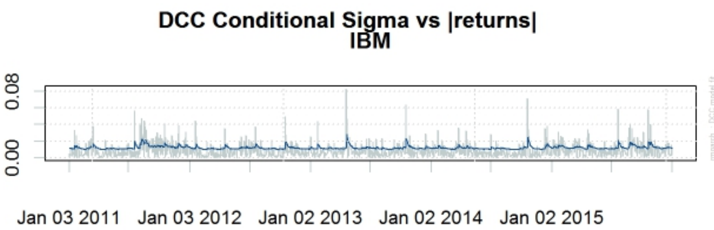
Daily return of the 4 stocks from 2006 to 2015

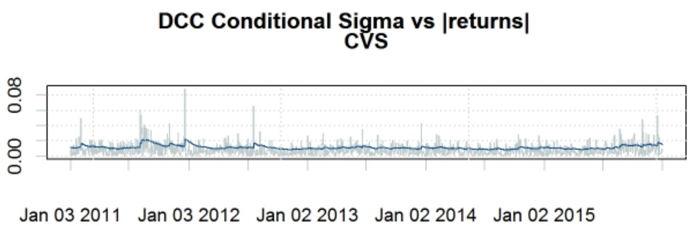


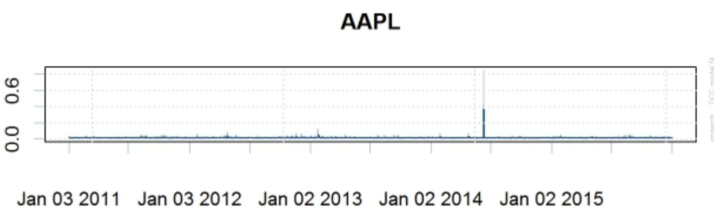
DCC Conditional Mean (Using Sample 1)

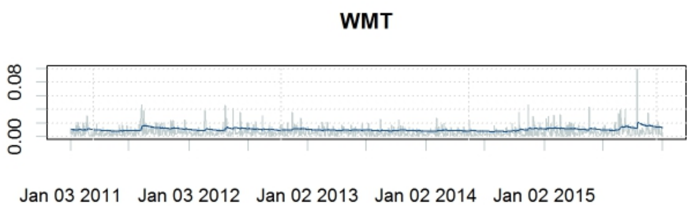


DCC Conditional Sigma vs Absolute Returns (Using Sample 1)



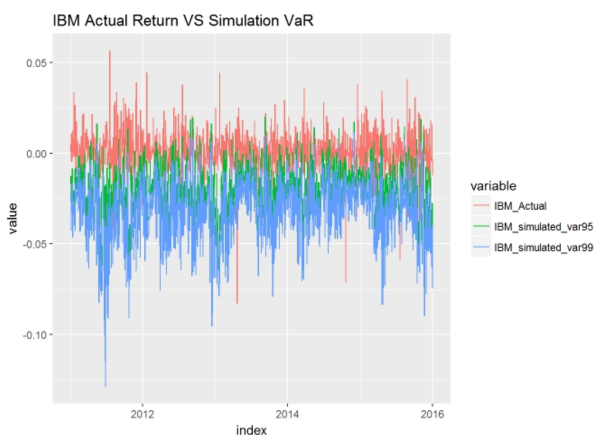


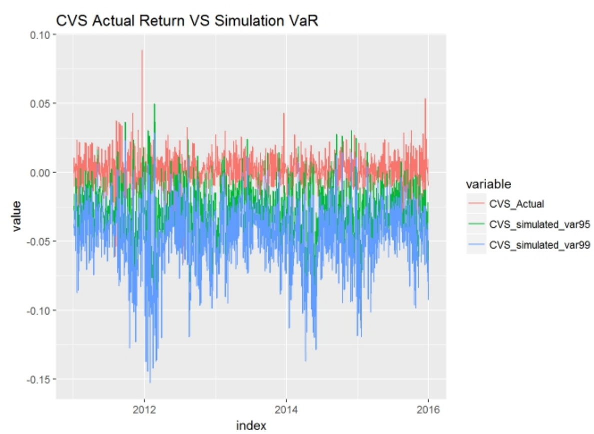


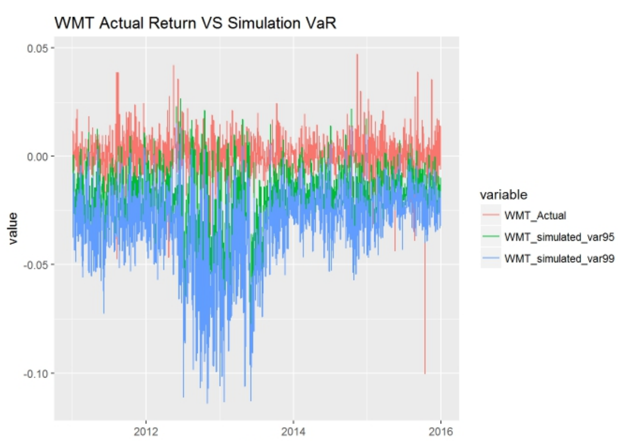


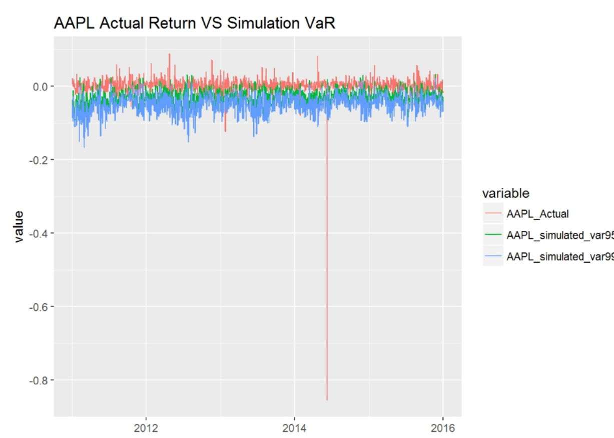
Out-of-Sample Actual Returns vs Simulated VaRs (using Sample 1)

Multi-variate models on Gaussian Distribution

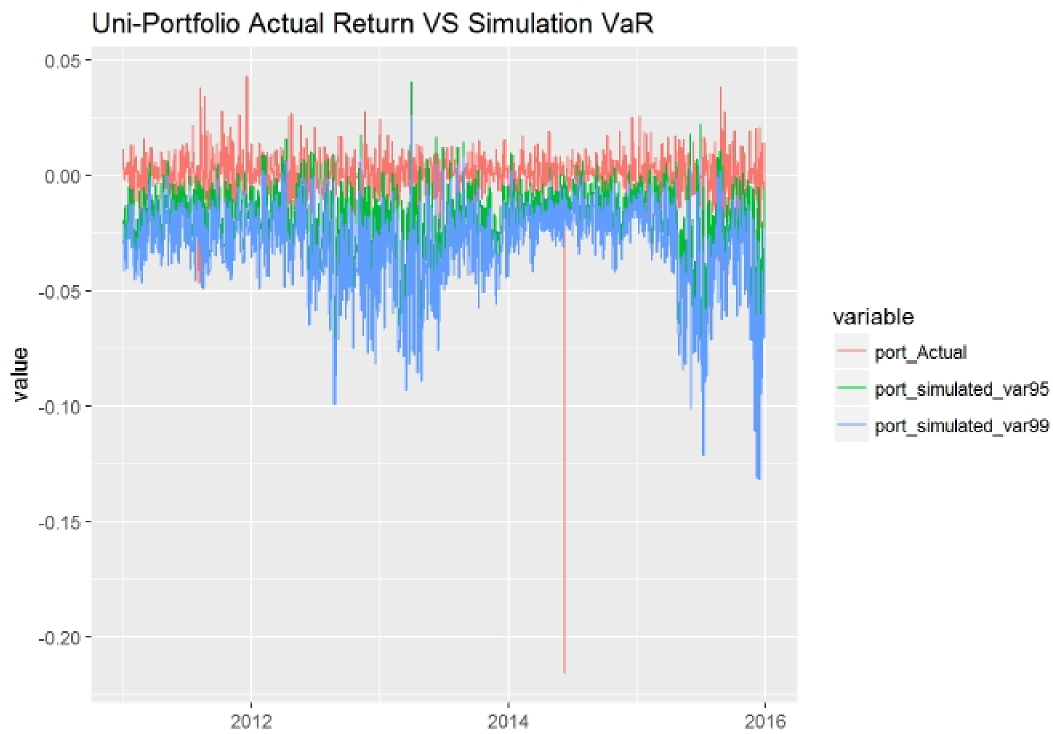






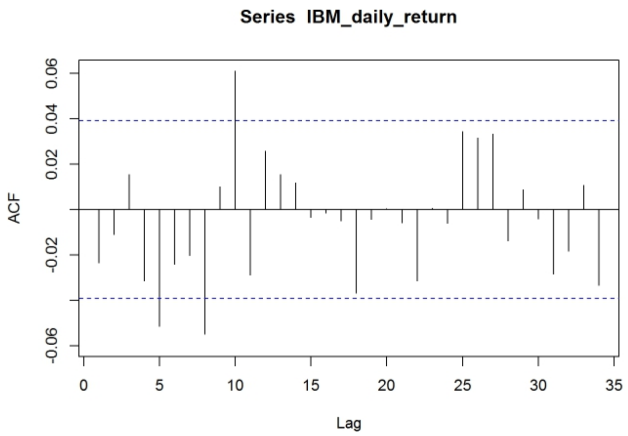


Uni-variate models on Gaussian Distribution

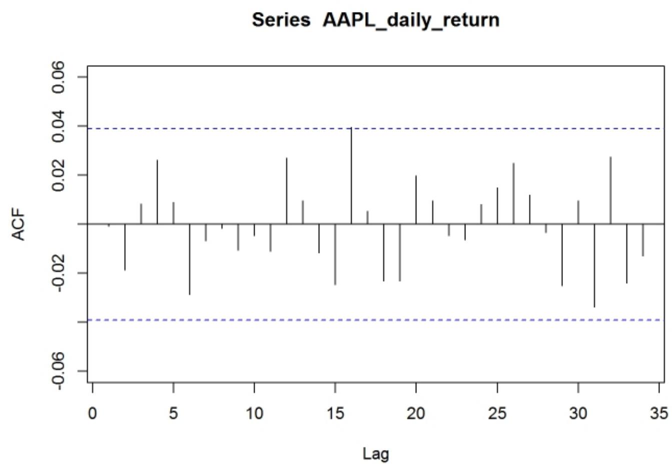


Appendix II – Stationarity Tests and Positive Definite Tests

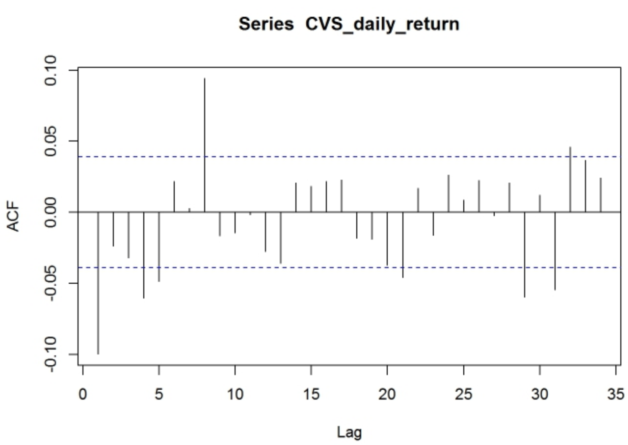
```
##
##  Augmented Dickey-Fuller Test
##
## data:  IBM_daily_return
## Dickey-Fuller = -13.661, Lag order = 13, p-value = 0.01
## alternative hypothesis: stationary
```



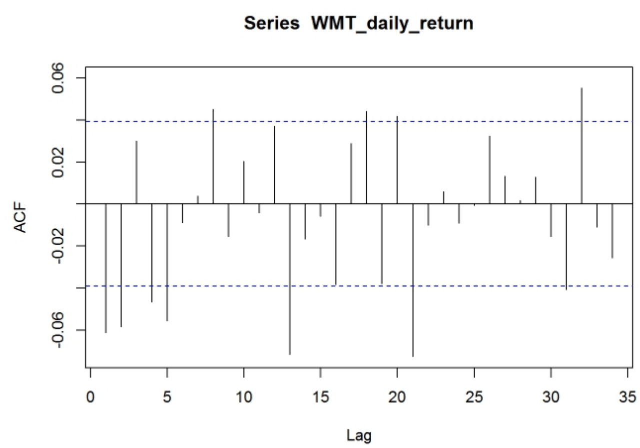

```
##
##  Augmented Dickey-Fuller Test
##
## data:  AAPL_daily_return
## Dickey-Fuller = -13.41, Lag order = 13, p-value = 0.01
## alternative hypothesis: stationary
```




```
##
##  Augmented Dickey-Fuller Test
##
## data:  CVS_daily_return
## Dickey-Fuller = -14.06, Lag order = 13, p-value = 0.01
## alternative hypothesis: stationary
```




```
##
##  Augmented Dickey-Fuller Test
##
## data:  WMT_daily_return
## Dickey-Fuller = -14.282, Lag order = 13, p-value = 0.01
## alternative hypothesis: stationary
```




```
#test for correlation across different period
library(Matrix)
posi_matrix <- data.frame()
for (i in 1:1258)   {
A<- simulated_data_covariance[,,i]
posi_matrix[i,1]<-isPositiveDefinite(A)
}
length(which(posi_matrix[,1]=="FALSE"))
```

```
## [1] 0
```

```
#all matrix is positive definite matrix
```

Appendix III - Quadratic Formulation

The Quadratic Model

Suppose that there are n different assets. The rate of return of asset i is a random variable with expected value m_i. The problem is to find what fraction x_i to invest in each asset i in order to minimize risk, subject to a specified minimum expected rate of return.

Let C denote the covariance matrix of rates of asset returns.

The classical mean-variance model consists of minimizing portfolio risk, as measured by

$$\frac{1}{2}x^T C x$$

subject to a set of constraints.

The expected return should be no less than a minimal rate of portfolio return r that the investor desires,

$$\sum_{i=1}^{n} m_i \, x_i \geq r,$$

the sum of the investment fractions x_i's should add up to a total of one,

$$\sum_{i=1}^{n} x_i = 1,$$

and, being fractions (or percentages), they should be numbers between zero and one,

$$0 \leq x_i \leq 1, \quad i = 1 \ldots n.$$

Since the objective to minimize portfolio risk is quadratic, and the constraints are linear, the resulting optimization problem is a quadratic program, or QP.