Time series analysis on return of spot gold price

Team member:
Tian Xie (#1371992)
Zizhen Li(#1368493)

Contents
Exploratory Analysis ........................................................................................................... 2
Data description ................................................................................................................... 2
Data preparation .................................................................................................................. 2
Basics Stats ........................................................................................................................ 2
Unit Root Test, ACF, PACF and ARCH effect test ................................................................ 3
Model fitting ........................................................................................................................ 6
ARMA model ....................................................................................................................... 6
ARMA(0,1)~GARCH(1,1) model ....................................................................................... 8
APARCH(3,0) model .......................................................................................................... 10
Residual analysis and model diagnostics .......................................................................... 10
Non tech analysis .............................................................................................................. 13
Forecast analysis ............................................................................................................... 14
Analysis of the results and discussion ............................................................................ 14
Appendix ............................................................................................................................. 15
ARMA(0,1)~EGARCH(1,1) model .................................................................................... 15
ARMA(0,1)~IGARCH(1,1) model ..................................................................................... 16
ARMA(0,1)~TGARCH(1,1) model .................................................................................... 17
ARMA(0,0)~ GARCH(1,1) model .................................................................................... 18
ARMA(0,0)~ EGARCH(1,1) model ................................................................................ 19
ARMA(0,0)~ EGARCH(3,0) model ................................................................................ 20
R CODE ............................................................................................................................. 21
Exploratory Analysis

Data description
Source: [http://research.stlouisfed.org/fred2/graph/?id=GOLDAMGBD228NLBM](http://research.stlouisfed.org/fred2/graph/?id=GOLDAMGBD228NLBM)

Properties:
- Range: price at 10:30 AM (London Time) each day from 04-01-1968 to 10-24-2013
- Unit: US Dollar
- Levels: Price for gold up to two decimal points
- Collected by: Federal Reserve of United States

Data preparation
Calculate the log return for gold price and remove NA values from the time series.

Basics Stats
```
> basicStats(nlrt)
   x
nobs 11533.000000
NAS  0.000000
Minimum -0.160286
Maximum  0.125345
1. Quartile -0.004882
3. Quartile  0.005438
Mean  0.000307
Median  0.000000
Sum   3.544766
SE Mean 0.000121
LCL Mean 0.000070
UCL Mean 0.000544
Variance 0.000168
Stdev   0.012978
Skewness 0.065722
Kurtosis 13.191215
```

Analysis: Values for the mean and standard deviation suggest that at a 1% confident level we reject the null hypothesis that the mean for the underlying time series is zero. Kurtosis of 13.12 shows the time series is leptokurtic distribution.
> normalTest(nlrt, method=c("jb"))

Title:
Jarque - Bera Normality Test

Test Results:
STATISTIC:
    X-squared: 83662.0729
P VALUE:
    Asymptotic p Value: < 2.2e-16

Analysis: Values at two tails are far off the normal line which is coincided with the large kurtosis value in the basic stats output. The p-value for the Jarque-Bera normal test as expected is way less than 0.05 suggesting that the time series is not normally distributed.

Unit Root Test, ACF, PACF and ARCH effect test

> adfTest(nlrt)

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
    Lag Order: 1
STATISTIC:
    Dickey-Fuller: -78.0665
P VALUE:
    0.01

> adfTest(nlrt, 3)

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
    Lag Order: 3
STATISTIC:
    Dickey-Fuller: -53.6614
P VALUE:
    0.01

> adfTest(nlrt, 5)

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
    Lag Order: 5
STATISTIC:
    Dickey-Fuller: -44.7119
P VALUE:
    0.01

Analysis: Null hypothesis that the underlying time series has unit root is rejected at lag 1, 3 and 5. No difference process is needed to apply to the data.
Analysis: Even though ACF and PACF value suggest an ARMA(1,1) model, result of PACF comes up with an ARMA(0,1) model. Both models will be tested in the following sections. Also, it appears to be an ARCH effect in the log return time series.
Model fitting

ARMA(1,1), ARMA(0,1), ARMA(0,1)~GARCH(1,1)/ IGARCH(1,1)/ EGARCH(1,1)/ GJR-GARCH(1,1)/ apARCH(1,1)/ TGARCH(1,1)/ csGARCH(1,1), ARMA(0,0)~ GARCH(1,1)/ IGARCH(1,1)/ EGARCH(1,1)/ GJR-GARCH(1,1)/ APARCH(1,1)/ TGARCH(1,1)/ csGARCH(1,1), GARCH(3,0), apGARCH(3,0) and TGARCH(3,0) have been applied to fit the data with t-distribution for residuals. Model APGARCH(3,0) appears to be the best model to fit the data.

ARMA model

> m1 <- auto.arima(coredata(nlrt), ic=c("bic"), trace = TRUE)

ARIMA(2,0,2) with non-zero mean : -67446.16
ARIMA(0,0,0) with non-zero mean : -67463.39
ARIMA(1,0,0) with non-zero mean : -67472.4
ARIMA(0,0,1) with non-zero mean : -67472.98
ARIMA(1,0,1) with non-zero mean : -67463.2
ARIMA(0,0,2) with non-zero mean : -67464.05
ARIMA(1,0,2) with non-zero mean : -67454.47
ARIMA(0,0,1) with zero mean : -67475.29
ARIMA(1,0,1) with zero mean : -67465.49
ARIMA(0,0,0) with zero mean : -67466.28
ARIMA(0,0,2) with zero mean : -67466.29
ARIMA(1,0,2) with zero mean : -67456.66

Best model: ARIMA(0,0,1) with zero mean

Outcome of the auto.arima process is coincided with that of eacf process but conflict with ARMA(1,1) model suggested by the individual ACF and PACF test.
Analysis: For all these test results, they all look the same to the counterparts. What noticed is that residuals for these two models do not follow a normal distribution and have a clear ARCH effect. Thus, a family of GARCH models is deployed to catch the ARCH effect in the residuals.
**ARMA(0,1)~GARCH(1,1) model**

### Optimal Parameters

|     | Estimate | Std. Error | t value | Pr(>|t|) |
|-----|----------|------------|---------|----------|
| mu  | 0.000050 | 0.000053   | 0.94279 | 0.34579  |
| ma1 | -0.071107| 0.009250   | -7.68759| 0.00000  |
| omga| 0.000001 | 0.000000   | 4.10686 | 0.00004  |
| alpha|0.103893  | 0.006150   | 16.89205| 0.00000  |
| beta| 0.895107 | 0.006199   | 144.38995| 0.00000  |
| shape|4.540827  | 0.159823   | 28.41157| 0.00000  |

Model expression

\[ r_t = 0.00005 + a_t - 0.071107a_{t-1}, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t_{4,54} \]

\[ \sigma_t^2 = 0.000001 + 0.103893a_{t-1}^2 + 0.895107\sigma_{t-1}^2 \]

**Analysis:** With three outliers on the right tail, residuals generally can be considered follow student distribution. The model also suggest the mean is equal to zero as what is suggested in auto.arima process.
Analysis: The ARMA(0,1)~GARCH(1,1) shows a strong correlation among residuals, even though arch effect in the residuals no longer exists. The other models in the GARCH model family with ARMA(0,1) model expressing the mean has the same problem as this ARMA(0,1)~GARCH(1,1) model. Thus, ARMA(0,1) model is dropped to build a pure GARCH model to fit the data.
The mean in the APARCH is suppressed because null hypothesis of the mean being zero cannot be rejected. The fitted APARCH(3,0) model is

\[ r_t = 0 + \alpha_t , \quad \alpha_t = \sigma_t \epsilon_t , \quad \epsilon_t \sim t^*_{3.19} \]

\[ \sigma_t^{0.95} = 0.0057 + 0.3522(\alpha_{t-1} + 0.0166\alpha_{t-1})^{0.95} + 0.3125(\alpha_{t-2} - 0.0506\alpha_{t-2})^{0.95} + 0.3265(\alpha_{t-3} + 0.0055\alpha_{t-3})^{0.95} \]

| Robust Standard Errors:                                | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------------------------------------------|----------|------------|---------|----------|
| omega                                                   | 0.005773 | 0.002909   | 1.98426 | 0.047226 |
| alpha1                                                  | 0.352294 | 0.025295   | 13.9274 | 0.000000 |
| alpha2                                                  | 0.312494 | 0.024189   | 12.9191 | 0.000000 |
| alpha3                                                  | 0.326523 | 0.022594   | 14.4514 | 0.000000 |
| gamma1                                                  | 0.016640 | 0.039517   | 0.4211  | 0.673692 |
| gamma2                                                  | -0.050623| 0.043918   | -1.1527 | 0.249049 |
| gamma3                                                  | 0.005488 | 0.025174   | 0.2179  | 0.827438 |
| delta                                                   | 0.949329 | 0.111929   | 8.4815  | 0.000000 |
| shape                                                   | 3.185173 | 0.127239   | 25.0330 | 0.000000 |

Residual analysis and model diagnostics

![ACF of Standardized Residuals](image)

Figure 3 (from APGARCH(3,0))
Analysis: What noticed in Figure 3 and Figure 4 is that the model does not have significant ACF value until lag 14. It indicates adequate reliability for the model in a short term.

Analysis: The residuals density of the model has a positive excess kurtosis meaning a fatter tail than normal distributed density. However, it has a bell shape close to normal one. The following QQ-Plot analysis will provide a further insight.
Analysis: The QQ-Plot justify the use of t-distribution for residuals which appears to fit to the normal line in the plot.
Non-tech analysis

The purpose of this project is designed to characterize and model observed time series data of gold prices. Same as stock prices, the gold prices have been very volatile, and the volatility varies overtime. In this project, we managed to test our data with several different ‘Autoregressive Conditional Heteroskedasticity’ models to measure the volatility cluster, trend, fluctuation, and analyze the impact of shocks to see if we can forecast the gold price volatility for future periods. Therefore, we can deploy these models to provide volatility measures in portfolio selection, risk management and pricing estimations.

Generally, financial time series often exhibits of low volatility followed by high volatility. This type of process is referred as volatility clustering. In order to capture the unequal variance in the squared error term of the expected values, we tried to use these models to fit our data: ARMA(1,1), ARMA(0,1), ARMA(0,1)~GARCH(1,1)/ IGARCH(1,1)/ EGARCH(1,1)/ GJR-GARCH(1,1)/ apARCH(1,1)/ TGARCH(1,1)/ csGARCH(1,1), ARMA(0,0)~ GARCH(1,1)/ IGARCH(1,1)/ EGARCH(1,1)/ GJR-GARCH(1,1)/ APARCH(1,1)/ TGARCH(1,1)/ csGARCH(1,1), GARCH(3,0), apGARCH(3,0) and TGARCH(3,0). The model APGARCH(3,0) appeared to be the best model among them. The APGARCH stands for asymmetric power generalized autogressive conditional heteroskedaticity. The power term is dedicated to capture volatility clustering by magnifying the outliers, which are the extreme values under extraordinary circumstances. The leverage parameter shows the amplitude of unparalleled response of the conditional variance towards negative versus positive shocks, for instance, the weakening of US dollars is a positive shock on gold prices. In addition, the model even captures the asymmetric effect of equal magnitude of positive and negative shocks produce an unequal response of gold price.

Here is the result of our model

\[ r_t = 0 + \alpha_t, \quad \alpha_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim t^{*3.19} \]

\[ \sigma_t^{0.95} = 0.0057 + 0.3522(\alpha_{t-1} + 0.0166\alpha_{t-1})^{0.95} + 0.3125(\alpha_{t-2} - 0.0506\alpha_{t-2})^{0.95} + 0.3265(\alpha_{t-3} + 0.0055\alpha_{t-3})^{0.95} \]

Based on the APGARCH model assumption, our model elucidates that under most of circumstances, the past positive shocks have deeper impact on current conditional volatility than past negative shocks. And they also seem to be more persistent than negative shocks because of the positive and significant alpha parameters. The result perfectly explains the weak dollar syndrome, when economic condition is unstable, investors tend to invest in gold since it is least correlated with equity markets. Unlike the equity markets, the positive shocks create a larger response than negative shocks of equal magnitude. The price of gold normally rises as a result of increased hedging positions after market crisis. Therefore, positive changes in the
price of gold are associated with negative financial news. The volatility is transmitted from the other markets to the gold market is asymmetrical.

**Forecast analysis**

As the graph shown above, we compare the rolling forecast versus the actual forecast within two sigma range. Notably, the actual forecast of gold return fluctuates around the x axis within the two sigma bands with non constant variance. However, the rolling forecast shows a straight linear curve perfectly overlays with the x axis since the p-value of mean in our previous models ARMA and ARIRIMA are nonsignificant. The mean is suppressed in our final APARCH (3,0), indicating gold has zero return overtime.

**Analysis of the results and discussion**

In conclusion, we narrowed down our model selection to ARMA based on preliminary ACF and PACF analysis, and the results suggest the the existence of ARCH effect in the log return time series. Then we tested all the GARCH model models try to capture the ARCH effect in the residuals. However, the residual analysis did not give us adequate results to fit the assumptions in any models we tested. Additionally, even though the ARCH effect is removed in ARMA(0,1)~GARCH(1,1) , but the mean is still shows nonsignificant p-value commensurate to other models. Therefore, we dropped the mean in the APARCH model since zero value of mean cannot be rejected. The APARCH model is the best fitted model based on ACF and residual analysis, as we stated in the report.
Appendix

ARMA(0,1)−EGARCH(1,1) model

ACF of Standardized Residuals

ACF of Squared Standardized Residuals
ARMA(0,1)~IGARCH(1,1) model
ARMA(0,1)~TGARCH(1,1) model
ARMA(0,0)~ GARCH(1,1) model
ARMA(0,0)–EGARCH(1,1) model

ACF of Standardized Residuals

ACF of Squared Standardized Residuals
ARMA(0,0)~ EGARCH(3,0) model
R CODE
library(zoo)
library(forecast)
library(fBasics)
library(fUnitRoots)
library(rugarch)
library(fGarch)
library(TSA)

#There is a conflict between package TSA and package rugarch

#Package TSA has to be detached first before use plot() to produce ACF figure for model of rugarch class.
setwd("E:/courses/csc425/hwork/project")
myd <- read.csv("fredgraph.csv")
names(myd) <- c("date", "price")
ts <- zoo(myd$price, as.Date(myd$date))
lrt <- log(ts / lag(ts, -1, na.pad = TRUE))
nlrt <- lrt[!is.na(lrt)]
plot(nlrt)
adfTest(nlrt)
adfTest(nlrt, 3)
adfTest(nlrt, 5)
Box.test(coredata(nlrt), lag = 1, type = "Ljung")
Box.test(coredata(nlrt), lag = 6, type = "Ljung")
Box.test(coredata(nlrt), lag = 12, type = "Ljung")
acf(coredata(nlrt), 20)
acf(coredata(nlrt)^2, 20, main = "ACF for squared log return")
acf(abs(coredata(nlrt)), 20, main = "ACF for absolute value of log return")
pacf(coredata(nlrt),10)
eacf(coredata(nlrt))

#ARMA(0,1) model
m1 <- auto.arima(coredata(nlrt), ic=c("bic"), trace = TRUE)
tsdia(m1, gof.lag=20)
qqnorm(m1$residual, main ="QQ Plot for residuals of ARMA(0,1) model")
qqline(m1$residuals, col=2)
acf(m1$residuals, main ="ACF value for the residuals of ARMA(0,1)")
acf(m1$residuals^2, main ="ACF value for the squared residuals of ARMA(0,1)")
acf(abs(m1$residuals), main ="ACF value for absolute value of residuals of ARMA(0,1)")

#ARMA(1,1) model
m2 <- arima(coredata(nlrt), order=c(1,0,1))
tsdia(m2, gof.lag=20)
qqnorm(m2$residual, main ="QQ Plot for residuals of ARMA(1,1) model")
qqline(m2$residuals, col=2)
acf(m2$residuals, main ="ACF value for the residuals of ARMA(1,1)")
acf(m2$residuals^2, main ="ACF value for the squared residuals of ARMA(1,1)")
acf(abs(m2$residuals), main ="ACF value for absolute value of residuals of ARMA(1,1)")

#ARMA(0,1)~GARCH(1,1) model
sgch.spec = ugarchspec(variance.model=list(model="sGARCH", garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,1)),distribution.model="std")
msg <- ugarchfit(sgch.spec, coredata(nlrt))
plot(msg)
igch.spec = ugarchspec(variance.model=list(model="iGARCH", garchOrder=c(1,1)), mean.model=list(armaOrder=c(0,1)), distribution.model="std")

mig <- ugarchfit(igch.spec, coredata(nlrt))

plot(mig)

egch.spec = ugarchspec(variance.model=list(model="eGARCH", garchOrder=c(1,1)), mean.model=list(armaOrder=c(0,1)), distribution.model="std")

meg <- ugarchfit(egch.spec, coredata(nlrt))

plot(meg)

tgch.spec = ugarchspec(variance.model=list(model="fGARCH", submodel="TGARCH", garchOrder=c(1,1)), mean.model=list(armaOrder=c(0,1)), distribution.model="std")

mtg <- ugarchfit(tgch.spec, coredata(nlrt))

plot(mtg)

#ARMA(0,0)~GARCH(1,1) / GARCH(3,0) model

sgch2.spec = ugarchspec(variance.model=list(model="sGARCH", garchOrder=c(1,1)), mean.model=list(armaOrder=c(0,0), include.mean=F), distribution.model="std")

msg2 <- ugarchfit(sgch2.spec, coredata(nlrt))

msg2.fcst <- ugarchforecast(msg2, n.ahead=20)

plot(msg2)

sgch2.spec = ugarchspec(variance.model=list(model="sGARCH", garchOrder=c(3,0)), mean.model=list(armaOrder=c(0,0), include.mean=F), distribution.model="std")

msg2 <- ugarchfit(sgch2.spec, coredata(nlrt))

plot(msg2)

igch2.spec = ugarchspec(variance.model=list(model="iGARCH", garchOrder=c(1,1)), mean.model=list(armaOrder=c(0,0)), distribution.model="std")

mig2 <- ugarchfit(igch2.spec, coredata(nlrt))

plot(mig2)
egch2.spec = ugarchspec(variance.model=list(model="eGARCH", garchOrder=c(1,1)),
           mean.model=list(armaOrder=c(0,0)),distribution.model="std")

meg2 <- ugarchfit(egch2.spec, coredata(nlrt))

plot(meg2)

ggch2.spec <- ugarchspec(variance.model=list(model="gjrGARCH", garchOrder=c(3,0)),
           mean.model=list(armaOrder=c(0,0), include.mean=F),distribution.model="std")

mgg2 <- ugarchfit(ggch2.spec, coredata(nlrt))

plot(mgg2)

apch2.spec <-ugarchspec(variance.model=list(model="apARCH", garchOrder=c(3,0)),
           mean.model=list(armaOrder=c(0,0), include.mean=F),distribution.model="std")

map2 <- ugarchfit(apch2.spec, coredata(nlrt))

plot(map2)

tgch2.spec <- ugarchspec(variance.model=list(model="fGARCH", submodel="TGARCH",
           garchOrder=c(3,0)), mean.model=list(armaOrder=c(0,0)), distribution.model="std")

mtg2 <- ugarchfit(tgch2.spec, coredata(nlrt))

plot(mtg2)

csgch2.spec <- ugarchspec(variance.model=list(model="csGARCH", garchOrder=c(1,1)),
           mean.model=list(armaOrder=c(0,0)), distribution.model="std")

mcsg2 <- ugarchfit(csgch2.spec, coredata(nlrt))

plot(mcsg2)

masgch2.spec <- ugarchspec(variance.model=list(model="sGARCH", garchOrder=c(2,1)),
           mean.model=list(armaOrder=c(0,0), include.mean=F), distribution.model="norm")

masg2 <- ugarchfit(masgch2.spec, nlrt)

plot(masg2)

#Forecast

map3 <- ugarchfit(apch2.spec, nlrt, out.sample=100)

map3.fcst <- ugarchforecast(map3, n.ahead=100, n.roll=100)
plot(map3)

plot(map3.fcst)