黄金价格时间序列分析

辛柏赢 2020111753

一、 案例背景

黄金是一种历史悠久的贵金属，自古代即被人类开采用作货币、装饰等多个方面。在现代金融市场中，黄金作为一种特殊的商品，也起到了举足轻重的作用。黄金本身具有的稀缺性和稳定的物理性质等，使其逐渐成为公认的避险资产[1]。其价格也随着国际政治形势、经济情况等外部情况发生着不同情况的波动。还有部分投机交易者利用黄金价格的波动特性进行投机交易，以赚取价差、获得利润。总的而言，黄金在金融市场中发挥着重要的作用，既可以作为一种价值稳定的资产抵抗风险，但同时也可以作为投机性交易和投资组合多元化的重要部分。

具体而言，本报告中以 AU9999 黄金作为主要研究对象。AU9999 为一种质量极高的黄金，其含金量可达 99.99%。也由于其较高的纯度，因此相比其它类型的黄金价格也会相对更高。

此外，本报告中除了关注该金融资产的资产价格外，对于其价格的波动率也是研究的重点。这是因为波动率良好地刻画了金融市场中某项资产风险的大小，是对期权等多种金融衍生品定价的重要因素。此外其对于在险价值（VaR）等的定量研究也有重要的作用。在一些衍生品市场中，例如 CBOE（Chicago Board of Option Exchange），波动率已经成为一种可以交易的期货产品[2]。

二、 数据介绍与描述性统计

（一） 数据获取与说明

本报告数据由 CSMAR 数据库（China Stock Market & Accounting Research Database, http://cndata1.csmar.com/）整理提供。数据内容为 2002 年 10 月到 2023 年 5 月上海黄金及贵重金属交易市场中 AU9999 黄金月收盘价（单位元/克）。数据总长度为 248，无缺失值或异常值。

在本报告中，预留了最后 3 期的数据作为外样本，其余作为内样本对模型进行拟合。

（二） 描述性数据分析

由下图 1 的时序图中可见，整体而言 AU9999 黄金（下简称黄金）价格自 2002 年至约 2012 年呈现波动上升趋势，最高达 361 元。后价格有所回落，在 260 元区间左右上下波动。在 2019 年末黄金价格开始由于新冠疫情等因素冲击第二次高升，最高点在

2020 年 7 月份达 428.97 元。尽管随后有所下降，但目前依然保持在了历史高位水平。

图 1 黄金价格时序图

除价格之外，在上文的讨论中，我们还将重点关注该金融产的收益波动率。首先给出资产（对数）收益的定义：若记资产t时刻的价格为P_t，则相应的对数收益率为：

$$r_t = \log P_t - \log P_{t-1} \quad (1)$$

进一步，我们认为波动率即为收益率的条件标准差。

在图 2～图 4 中，分别给出了黄金价格对数收益率的时序图、直方图及 Q-Q 图，如下所示。由图可见，整体而言经过对数、差分变换，序列基本呈现水平平稳，其波动的方差也基本在一定范围内进行。但较为特殊的是，该对数收益序列有一定的分布集群特征，即序列的波动倾向于“扎堆”出现，在历史波动较大时更倾向于出现波动率较大的变化，而在波动率较小的时期则更倾向于小范围的波动。除此之外观察其分布状态，可以发现整体而言其分布基本形状依然是近似钟型的，均值水平在 0 左右。但需要关注极端收益（或极端损失）部分，由 Q-Q 图可见该对数收益率依然存在着一定的厚尾特征。这也符合常见金融资产的分布特性。

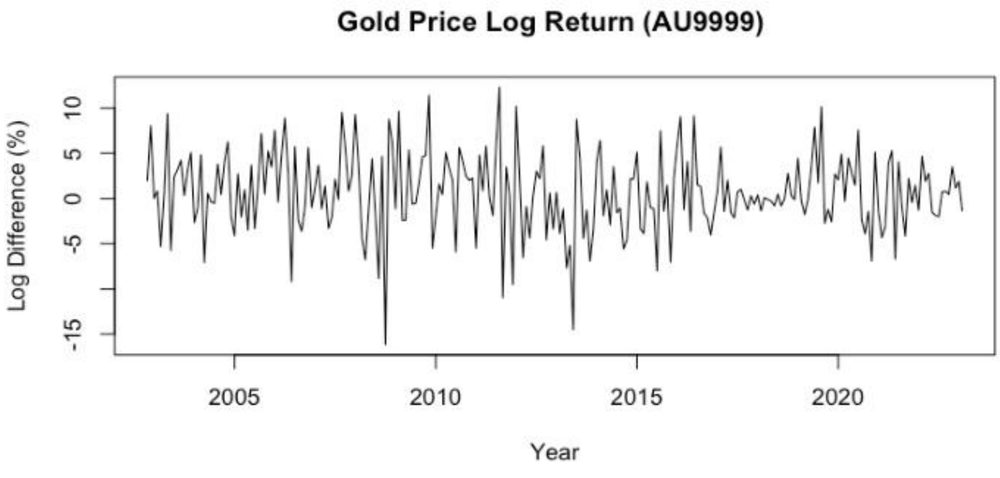

图 2 黄金价格对数收益率时序图

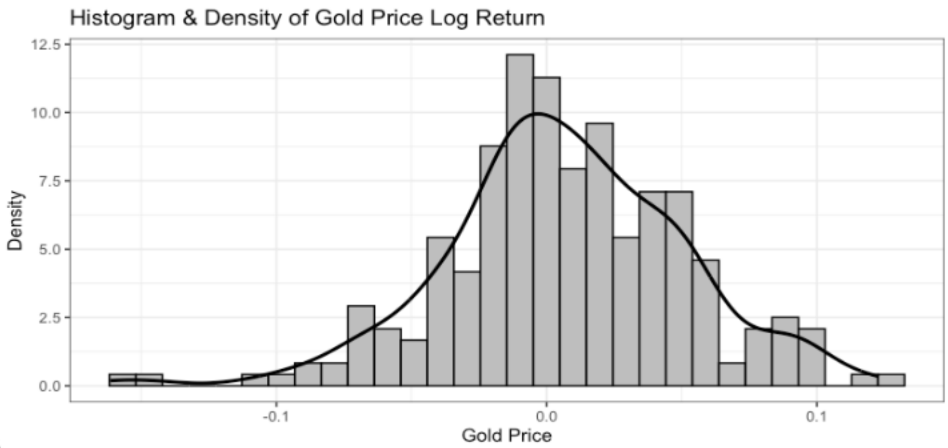

图 3 黄金价格对数收益率直方图

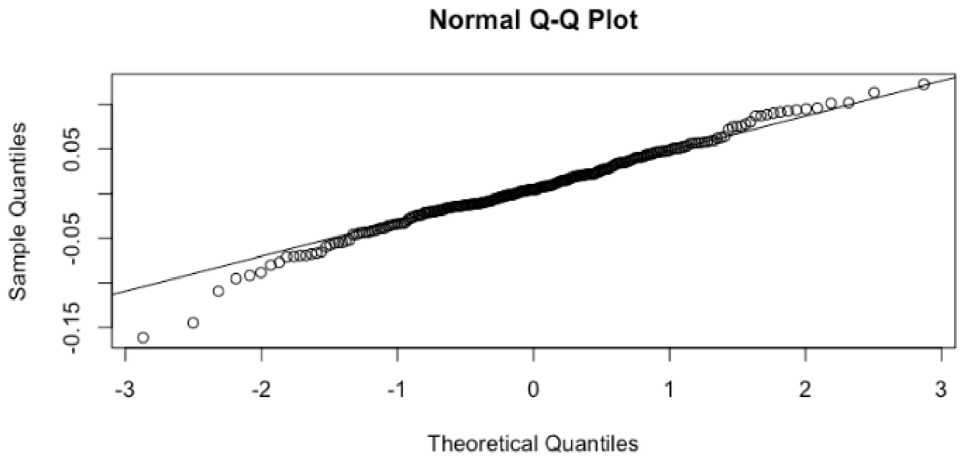

图 4 黄金价格对数收益率 Q-Q 图

三、 随机时间序列建模及预测

（一） 模型识别与模型定阶

由上述的分析可知，该黄金价格的收益率尽管基本而言是平稳的、前后不相关的，但其集中波动的特性暗示其彼此之间并不是完全独立的。为验证该特点我们进行模型识别如下。

首先作对数收益率序列的 ACF 图、PACF 图如图 5 所示。由图可见，除了在较高阶滞后的位置零星出现了略微凸出的 ACF、PACF 值外，其他各值都落在了零水平线之内，基本可以认为该对数收益率序列是不存在相关性的白噪声序列。进一步验证该直观感受，对序列连续进行直到 24 阶滞后的 Ljung-Box 检验，其检验的 p 值均大于 0.15，可以认为有较为充分的统计学理由支持上述白噪声观点。

考虑到其上述的波动性特征，这里对序列的 ARCH 效应进行检验。这里分别对序列取绝对值及平方，再计算其对应的 ACF，对应结果如图 6 所示。由此可以看出，有

多个相关值均超出了相应界限。进一步对收益率序列进行直到 24 阶的 McLeod-Li 检验，其对应阶数与 p 值如表 1 所示。由表可见，可以认为在 10%的显著性水平下可以拒绝原假设，认为序列整体之间存在显著的相关性；在 5%的显著性水平下该结论也基本成立。综合 ACF 及时序图等多方面判断，基本可以认为黄金价格的对数收益率是存在着明显的 ARCH 效应的。

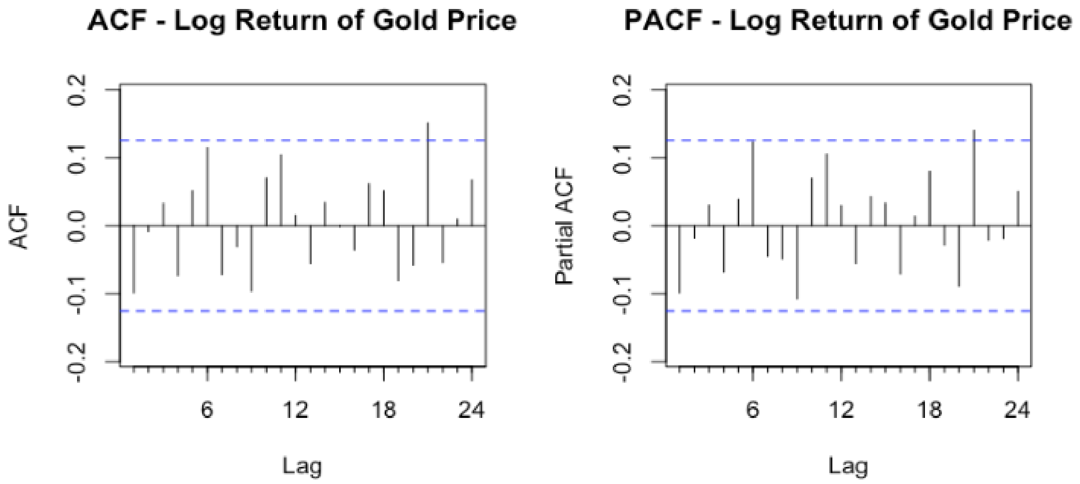

图 5 对数收益率的 ACF 与 PACF

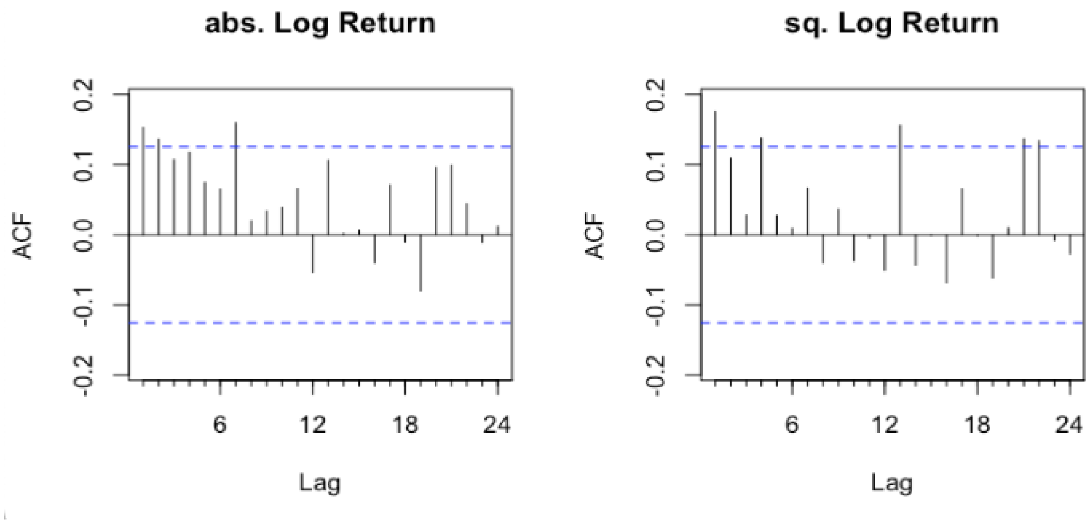

图 6 经绝对值/平方处理后的序列 ACF

表 1 Mcleod-Li 检验结果

滞后阶数	p-value
6	0.01524 (**)
12	0.09813 (*)
18	0.06499 (*)
24	0.02762 (**)

注：**对应 5%显著水平，*对应 10%显著水平

根据上述判断，该序列的对数差分整体呈现白噪声形势，对其进行 ARIMA 建模的意义不大。综合上述的 ARCH 效应，下面将尝试对模型的波动率进行建模，构建诸如 ARCH、GARCH 等相应模型。特别地，在沿用上述记号的基础上，记$\mathcal{F}_t$为截止到t时刻的信息集，其中包含截止到该时刻到全部收益率信息。另记$\{\varepsilon_t\}$为$\mathcal{F}_t$可测的不相关白噪声序列。

首先尝试拟合 ARCH(q)模型如下：

$$\begin{cases} r_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 \end{cases} \quad (2)$$

由 ARCH(q)的性质可知，若序列$\{r_t\}$符合一个 ARCH(q)模型，则对应的$\{r_t^2\}$可满足 AR(q)模型。故由此为依据进行定阶。

对该对数收益率求解 PACF 如图 7 所示。由图式可知，在第 1, 13, 21 阶滞后处 PACF 取值较为突出。另一方面，由于 21 阶滞后阶数过高，不可避免将引入过多参数，故暂时只考虑尝试拟合 ARCH(1), ARCH(13)。此外参考信息准则再次尝试定阶，给出的参考建议为拟合 ARCH(2)模型。综上所述，接下来将针对这三项阶数尝试拟合模型。

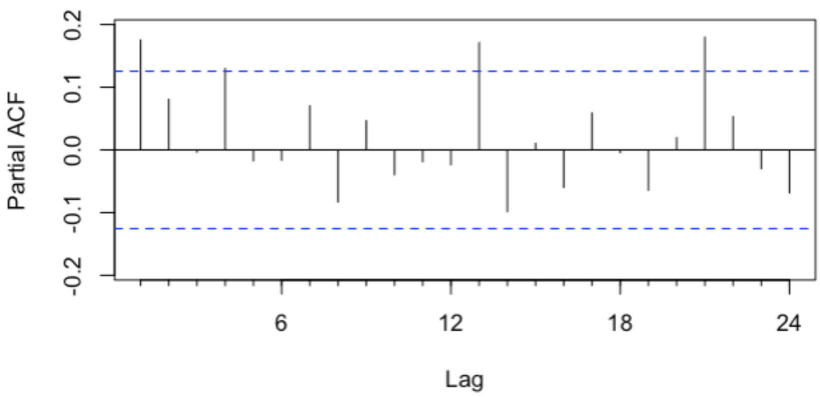

图 7 对数收益平方 PACF

此外，考虑如下所示的 GARCH(p,q)模型：

$$\begin{cases} r_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j r_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \end{cases} \quad (3)$$

可以证明，对于一个 GARCH(p,q)模型可以等价地写作一个 ARCH(∞)模型。故在通常情况下拟合一个 GARCH(1,1)即可满足模型的相应需求。

（二） 模型估计

在模型拟合估计中，进一步假定ε_t为独立同分布与标准正态分布，综合采用极大似然（MLE）以及伪极大似然（Quasi-MLE）准则进行拟合估计。

在 ARCH(1)的模型估计中，α系数对应的 p 值为 0.610，在 5%的水平下是不显著的。在 ARCH(13)的模型估计中 5%水平显著的系数主要集中于前四阶中，且较高阶数项的系数基本趋近于 0。尝试拟合 ARCH(2)，经计算发现其各阶系数显著性表现良好，具体结果展示如下表 2 所示。综合上述内容，暂定拟合该 ARCH(2)模型。

此外 GARCH(1,1)模型也基本较好的进行了拟合，具体结果如表 3 所示。两模型对应信息准则 ARCH(2)表现整体略优于 GARCH(1,1)，但差距较小。

表 2 ARCH(2)模型拟合结果

参数	估计值	p-value
μ	0.006524	0.007728 (**)
ω	0.001336	0.000000 (**)
α_1	0.115932	0.053249 (*)
α_2	0.214787	0.020953 (**)

注：**对应 5%显著水平，*对应 10%显著水平；对应信息准则：AIC = -3.4022, BIC = -3.3444

表 3 GARCH(1,1) 模型拟合结果

参数	估计值	p-value
μ	0.005962	0.009521 (**)
ω	0.000111	0.1553885
α_1	0.178534	0.0056345 (**)
β_1	0.776987	0.000000 (**)

注：**对应 5%显著水平，*对应 10%显著水平；对应信息准则：AIC = -3.4192, BIC = -3.3614

（三） 模型检验与评估

图 8，图 9 分别展示了 ARCH(2)及 GARCH(1,1)两模型建模的各个方面拟合结果及残差分析。为叙述方便，从左至右按行进行排序，依次记作子图 1～12。

对于 ARCH(2)模型，子图 1 展示了序列数据及上下估计的两倍条件标准差区间。可见绝大多数的序列数据均落在了该区间之内，也说明了模型较充分地对波动率进行了拟合。子图 8 展示标准化残差序列的分布直方图及标准正态分布对应的密度曲线，子图 9 展示的对应的 Q-Q 图。综合两图可以看到，整体而言基本符合在模型拟合时做出的正态假设，但需要指出从 Q-Q 图中依然体现出了一定的厚尾倾向。子图 10、11 分别为标准化残差序列与标准化残差序列平方的 ACF 图，由图中基本可以看出标准化残差序列基本不存在任何的序列相关性，但平方标准化残差序列在一些较低的阶数处依然存在着不是特别显著的相关性。为进一步确认，对标准化残差序列及标准化残差平方序列分别进行加权 Ljung-Box 检验，检验结果的 p 值均大于 0.05，基本都处于

0.20 水平以上，因此可以认为残差序列基本是不存在序列或平方序列相关性。

对于 GARCH(1,1)模型，大致的模型拟合效果内容与 ARCH(2)模型相似，并且可以看到其残差序列的相关性更低，在此角度而言 GARCH(1,1)的拟合更为充分。此外，该模型的残差序列亦通过了类似的 Ljung-Box 检验。

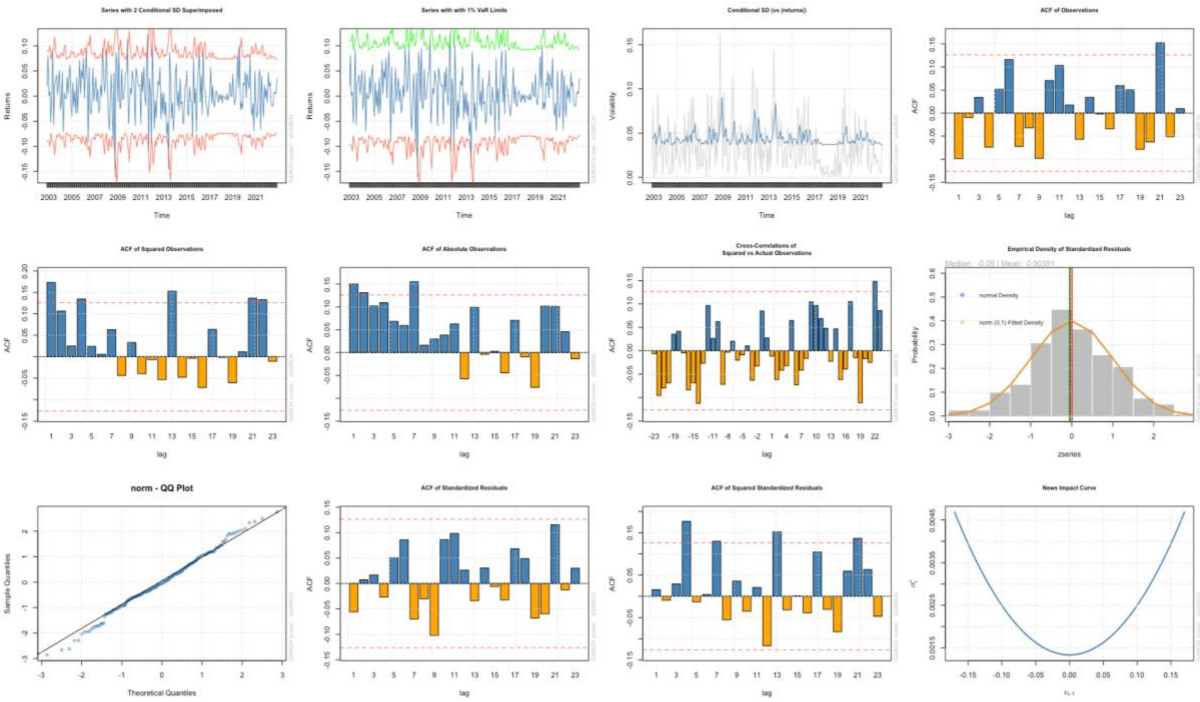

图 8 ARCH(2) 模型拟合效果及残差检验图

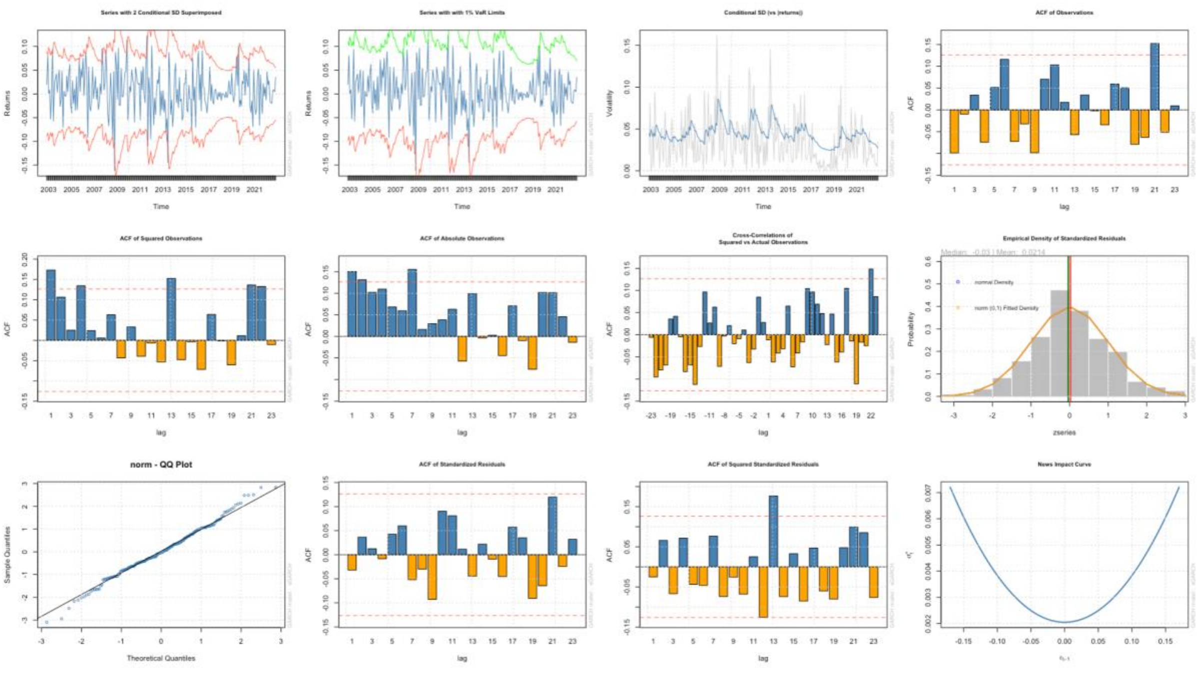

图 9 GARCH(1,1) 模型拟合效果及残差检验图

（四） 预测

与均值建模不同，在波动率模型中，我们这里更关注模型的条件方差。因为正如前述背景所言，在金融市场中，波动率反映了资产的风险大小，对于资产的定价等有重要意义。根据上述所拟合的 ARCH(2)及 GARCH(1,1)模型，可得到黄金价格对数收益率的波动率预测结果汇总如下表所示。此外，由于波动率作为潜在变量无法直接通过观测得到，这里简单地使用直到t期的序列无条件方差作为第$t+1$期的波动率的条件方差的粗略参考，亦整合在下表中。

表 4 预测结果

	模型波动率预测		无条件标准差
	ARCH(2)	GARCH(1,1)	
T+1	0.03784	0.02905	0.04430
T+2	0.04096	0.03029	0.04421
T+3	0.04287	0.03142	0.04438

若以该无条件标准差作为参考，并且考虑模型本身拟合的信息准则等，这里可以粗略地认为 ARCH(2)模型在本例中的预测效果略优于 GARCH(1,1)模型。

四、 总结与讨论

本报告通过对 AU9999 黄金价格的分析，重点讨论了其对数收益率的时间趋势情况。通过检验发现序列具有较明显的波动聚集情况，而其本身的模型相关性不强。因此这里选择对序列的波动率进行建模。事实上该建模方向在金融资产中也是自然的，波动率作为资产风险的代表，在金融市场中亦有着重要的作用。最终我们拟合了 ARCH(2)模型与 GARCH(1,1)两波动率模型，并对外样本的波动率进行了预测。

需要指出的是，由于对于均值部分没有进行特殊的建模处理，因此对于序列的未来取值无法做出较好的预测，因此也相应无法通过 MSE、MAE 等准则进一步对拟合效果进行讨论。在后续的工作中可以尝试对模型的随机趋势进一步建模，如构建 ARMA-GARCH 或更复杂的随机模型，以得到更全面的预测结果。

五、 参考文献

[1] BAUR D G, LUCEY B M. Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold [J]. Financial Review, 2010, 45(2): 217-29.
[2] HULL J C. Risk Management and Financial Institutions [M]. 6th Edition ed., 2023.

六、 程序代码

```
library(readxl)
library(TSA)
library(ggplot2)
library(forecast)
library(tseries)
library(FinTS)
library(rugarch)
library(fGarch)

rm(list = ls())

########################################
#          Comments Rules:           #
#                                   #
# # -------- #. SECTION NAME --------  #
#                                   #
# ## ------ #.a Unit Name ------      #
#                                   #
# #: Function Description             #
#                                   #
# #" Natural Language Explanation"     #
#                                   #
########################################

# -------- 0. Load Data ----------------

setwd("/Users/xinby/Desktop/Sufe/TSA/ts_proj2/proj2")
dat <- as.matrix(read_excel("tdat.xlsx",col_names = F,sheet = "gold"))
vec <- c(dat)
vec
gold <- ts(vec, start=c(2002,10), freq=12)
gold.frame <- data.frame(gold)
ld.gold.frame <- data.frame(diff(log(gold)))

dat.t <- as.matrix(read_excel("tdat.xlsx",col_names = F,sheet = "test"))
vec.t <- c(dat)
```

```r
vec.t
gold.t <- ts(vec, start=c(2002,10), freq=12)

# -------- 1. Descriptive Stat --------

## ------ 1.a For original data -------

plot(gold,xlab = 'Year', ylab = "Price", main="Gold Price (AU9999)") #:TS plot

qqnorm(gold);qqline(gold) #:Q-Q plot

ggplot(data=gold.frame,aes(x=gold.frame$gold,y=..density..))+
  geom_histogram(bins=30,color="black",fill="gray")+
  geom_density(size=1)+
  theme_bw()+
  xlab('Gold Price')+
  ylab("Density")+
  scale_y_continuous(labels = scales::percent_format())+
  ggtitle("Histogram & Density of Gold Price") #:Histogram & Density plot

## ------ 1.b For Diff Data ------
plot(diff(gold) , xlab = 'Year', ylab = "Price Difference", #:TS plot
     main="Gold Price Difference (AU9999)" )

## ------ 1.c For Log Diff Data ------
plot(diff(log(gold)*100),xlab = 'Year', ylab = "Log Difference (%)",  #:TS plot
     main="Gold Price Log Return (AU9999) ")
qqnorm(diff(log(gold)));qqline(diff(log(gold))) #:Q-Q plot

ggplot(data=ld.gold.frame,aes(x=ld.gold.frame$diff.log.gold..,y=..density..))+
  geom_histogram(bins=30,color="black",fill="gray")+
  geom_density(size=1)+
  theme_bw()+
  xlab('Gold Price')+
  ylab("Density")+
  ggtitle("Histogram & Density of Gold Price Log Return") #:Histogram & Density plot

# -------- 2. Model Recoginition ----------------

## ------ 2.a Stationarity Test ------

par(mfrow = c(1, 2))
forecast::Acf(diff(log(gold)),main="ACF - Log Return of Gold Price") #:ACF
  #"From ACF, apart from lag=21, others all stay inbetween -> WN "#
```

```
  #"cannot fit an MA"#
forecast::Pacf(diff(log(gold)),main="PACF - Log Return of Gold Price") #:PACF
  #"From PACF, basically all stay inbetween"#
  #"cannot fit an AR"#
  #"ARMA model not appliable"#
par(mfrow=c(1,1))

Box.test(diff(log(gold)),lag=6,type = "Ljung")
Box.test(diff(log(gold)),lag=12,type = "Ljung")
Box.test(diff(log(gold)),lag=18,type = "Ljung")
Box.test(diff(log(gold)),lag=24,type = "Ljung") #:Ljung-Box
  #"From Ljung-Box with different lags, affirm there's no relationships"#

t.test(as.vector(diff(log(gold)))) #:t.test
  #"t test, H0: mean=0 -> rejected, mean (approx.)= 0.006"

## ------ 2.b ARCH Effect Test ------

par(mfrow = c(1, 2))
forecast::Acf(abs(diff(log(gold))),main="abs. Log Return")
forecast::Acf((diff(log(gold)))^2,main ="sq. Log Return") #:ACF
par(mfrow=c(1,1))

Box.test(diff(log(gold))^2,lag=6,type = "Ljung")
Box.test(diff(log(gold))^2,lag=12,type = "Ljung")
Box.test(diff(log(gold))^2,lag=18,type = "Ljung")
Box.test(diff(log(gold))^2,lag=24,type = "Ljung") #:Mcleod Test

## ------ 2.c Model ARCH ------

forecast::Pacf((diff(log(gold)))^2,main="") #:PACF
  ##"consider arch(1),arch(13)"
auto.arima(diff(log(gold))^2,seasonal = F,stationary = T,max.q=0) #aic
  ##"consider arch(2)
garchFit(~ 1 + garch(1,0), data=c(diff(log(gold))), trace=FALSE)
garchFit(~ 1 + garch(13,0), data=c(diff(log(gold))), trace=FALSE)
garchFit(~ 1 + garch(2,0), data=c(diff(log(gold))), trace=FALSE)

fit.arch2=ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(2, 0)),
                   mean.model = list(armaOrder = c(0, 0),include.mean = T),
                   distribution.model = "norm")

gold.arch2<-ugarchfit(spec = fit.arch2, data = diff(log(gold.t)), out.sample = 3, solver =
'hybrid', fit.control = list(stationarity = 1))
```

```
gold.arch2
plot(gold.arch2,which='all')

### ---- 2.c.b Predict ----

forecast.norm.arch2=ugarchforecast(gold.arch2, data = gold.t, n.ahead = 3, n.roll = 0,
out.sample = 3)
forecast.norm.arch2

## ------ 3 GARCH -------
garchFit(~1+garch(1,1),data=c(diff(log(gold))),trace=F)
fit.garch11=ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
                   mean.model = list(armaOrder = c(0, 0),include.mean = T),
                   distribution.model = "norm")
gold.garch11<-ugarchfit(spec = fit.garch11, data = diff(log(gold.t)), out.sample = 3,
solver = 'hybrid', fit.control = list(stationarity = 1))
gold.garch11
plot(gold.garch11,which='all')
forecast.norm.garch11=ugarchforecast(gold.garch11, data = gold.t, n.ahead = 3, n.roll = 0,
out.sample = 3)
forecast.norm.garch11

sqrt(var(diff(log(gold.t[0:-3]))))
```