# 單變量黃金時間數列分析

林子立

2025-05-18

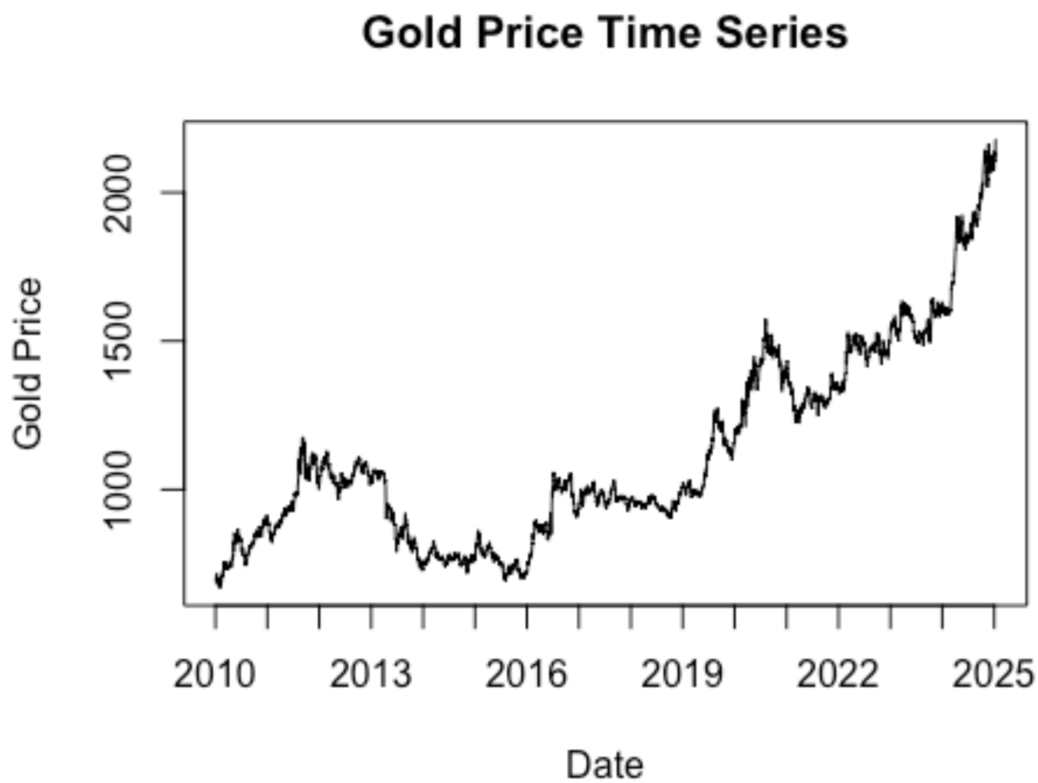
```
## ── Attaching packages ─────────────────────────────────────── tidyve
rse 1.3.2 ──
## ✔ ggplot2 3.5.1     ✔ dplyr   1.1.4
## ✔ tibble  3.2.1     ✔ stringr 1.5.0
## ✔ tidyr   1.3.0     ✔ forcats 0.5.2
## ✔ purrr   1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_co
nflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
##
## 載入套件：'zoo'
##
##
## 下列物件被遮斷自 'package:base':
##
##     as.Date, as.Date.numeric
##
##
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
##
## Rows: 14497 Columns: 2
## ── Column specification ────────────────────────────────────────────
────────────
## Delimiter: ","
## dbl  (1): LBMA Gold Prices - daily - euro - AM (LBMA/gold_D/gold_D_E
UR_AM)
## date (1): period
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.
## Rows: 10268 Columns: 2
## ── Column specification ────────────────────────────────────────────
────────────
## Delimiter: ","
## dbl  (1): DCOILWTICO
```

```
## date (1): observation_date
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.
##
## 載入套件：'aTSA'
##
##
## 下列物件被遮斷自 'package:tseries':
##
##     adf.test, kpss.test, pp.test
##
##
## 下列物件被遮斷自 'package:graphics':
##
##     identify
##
##
##
## 載入套件：'forecast'
##
##
## 下列物件被遮斷自 'package:aTSA':
##
##     forecast
```

#Step1: Exploratory Data Analysis

##Data Visualization

```
# 時間向量與價格
dates <- AU_OIL$Date
gold <- AU_OIL$GoldPrice_interp

# 繪圖（線圖，不畫 x 軸）
plot(dates, gold, type = "l",
     main = "Gold Price Time Series",
     ylab = "Gold Price",
     xlab = "Date",
     xaxt = "n")  # 不畫預設 x 軸

# 建立每年 1 月 1 日的日期向量
years <- seq(from = as.Date("2010-01-01"), to = as.Date("2025-01-01"),
by = "year")

# 標上年份作為 x 軸刻度
axis(side = 1, at = years, labels = format(years, "%Y"))
```



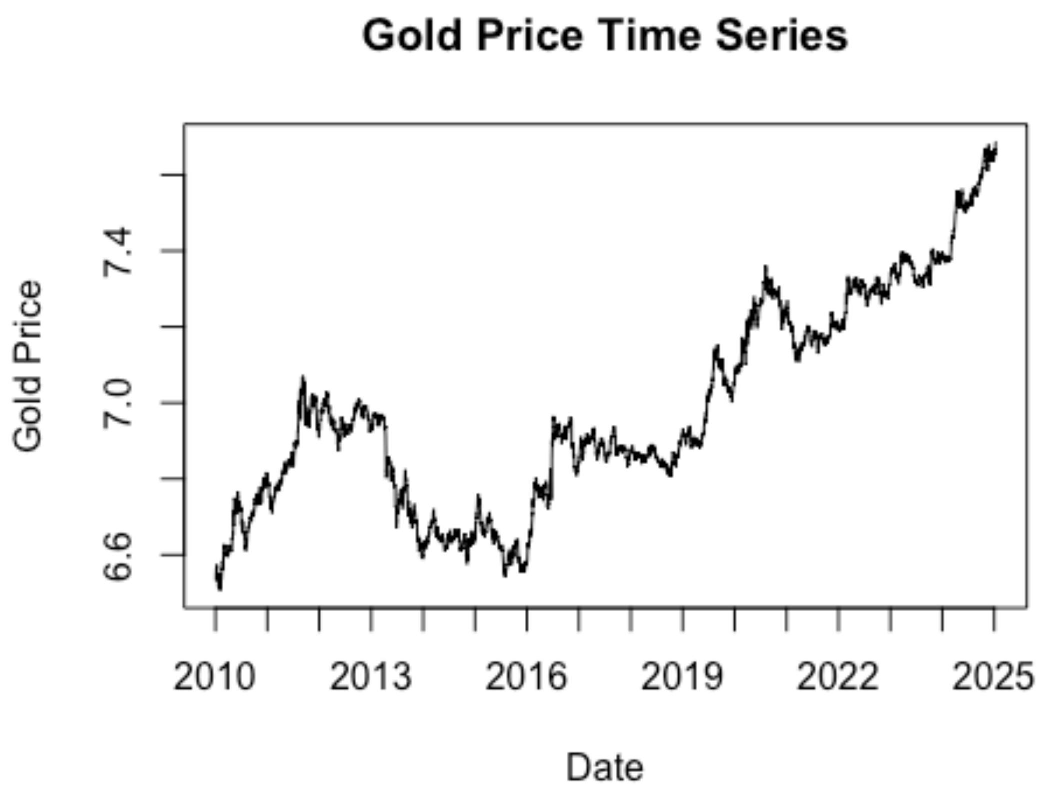

```
adf.test(gold)

## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##       lag  ADF p.value
##  [1,]   0 2.71    0.99
##  [2,]   1 2.52    0.99
##  [3,]   2 2.46    0.99
##  [4,]   3 2.50    0.99
##  [5,]   4 2.54    0.99
##  [6,]   5 2.46    0.99
##  [7,]   6 2.47    0.99
##  [8,]   7 2.46    0.99
##  [9,]   8 2.53    0.99
## [10,]   9 2.58    0.99
## Type 2: with drift no trend
##       lag  ADF p.value
##  [1,]   0 1.37    0.99
##  [2,]   1 1.18    0.99
##  [3,]   2 1.11    0.99
##  [4,]   3 1.16    0.99
```

```
## [5,]   4 1.24    0.99
## [6,]   5 1.17    0.99
## [7,]   6 1.20    0.99
## [8,]   7 1.16    0.99
## [9,]   8 1.25    0.99
## [10,]  9 1.31    0.99
## Type 3: with drift and trend
##       lag    ADF p.value
## [1,]   0 -0.133   0.990
## [2,]   1 -0.319   0.990
## [3,]   2 -0.383   0.987
## [4,]   3 -0.333   0.989
## [5,]   4 -0.288   0.990
## [6,]   5 -0.362   0.988
## [7,]   6 -0.345   0.989
## [8,]   7 -0.356   0.988
## [9,]   8 -0.287   0.990
## [10,]  9 -0.233   0.990
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

此圖呈現的是自 2010 年以來，LMBA 黃金每日拍賣價（AM）的價格。從圖中可以看到，價格以大週期來看可以以 2016 年為分界找到兩個趨勢。在 2016 年以前黃金價格經歷先升後降的趨勢；2016 年之後則是持續攀升。其中值得注意的現象是，黃金價格兩、三年就會存在一次跳動情形，造成短期價格劇烈變化的情況，如 2013 年終、2016 年底、2019 年底、2021 年初及 2024 年初。 除了部分時間存在劇烈跳動的價格變化以外，長期趨勢也存在單調遞增與增加速度加快的趨勢。從此觀點出發，合理懷疑此資料序列存在單根。近一步使用 Augmented Dickey-Fuller 檢定（以下簡稱 ADF 檢定），可以發現不管有無趨勢(Trend)或飄移（Drift）的加入，以及滯後期為 0-9 期之間，強烈的證據都指向單根存在於此序列資料中。為了符合對於穩定態（Stationary）資料的統計假設，以下使用常見的資料處理方法：取對數及差分進行探討。

```r
# 時間向量與價格
dates <- AU_OIL$Date
gold <- log(AU_OIL$GoldPrice_interp)

# 繪圖（線圖，不畫 x 軸）
plot(dates, gold, type = "l",
     main = "Gold Price Time Series",
     ylab = "Gold Price",
     xlab = "Date",
     xaxt = "n")  # 不畫預設 x 軸

# 建立每年 1 月 1 日的日期向量
years <- seq(from = as.Date("2010-01-01"), to = as.Date("2025-01-01"),
by = "year")
```

```r
# 標上年份作為 x 軸刻度
axis(side = 1, at = years, labels = format(years, "%Y"))
```



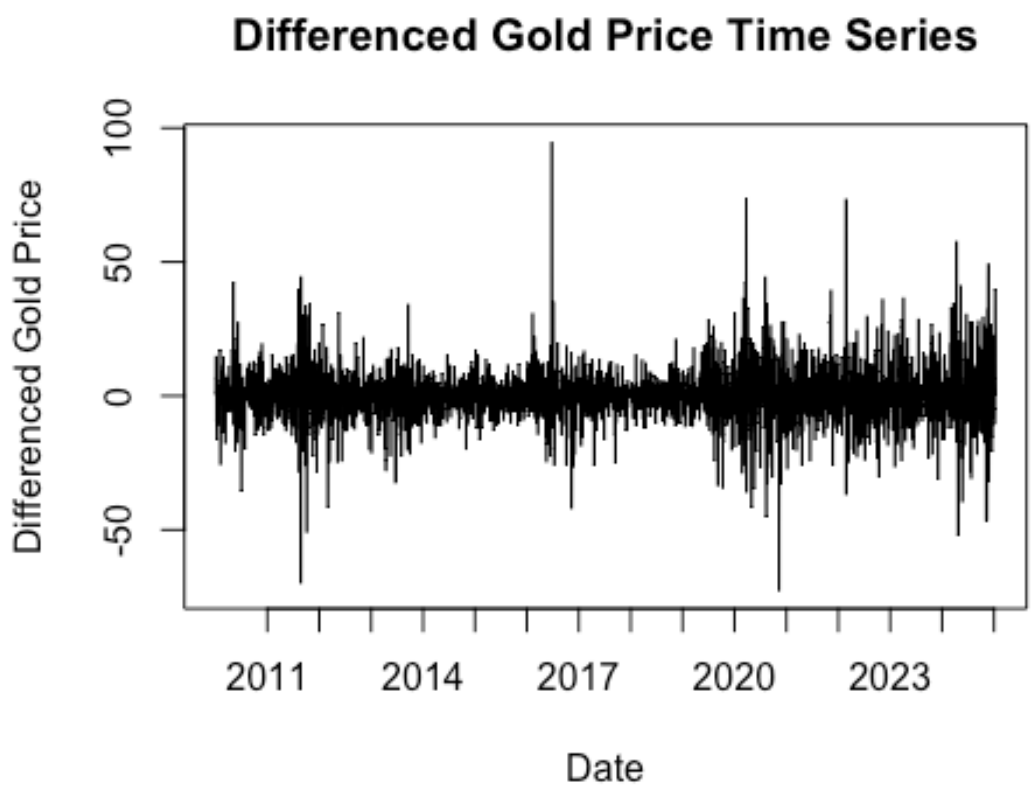

```
adf.test(gold)

## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##       lag  ADF p.value
##  [1,]   0 2.14   0.990
##  [2,]   1 2.01   0.990
##  [3,]   2 1.97   0.988
##  [4,]   3 1.99   0.989
##  [5,]   4 1.98   0.989
##  [6,]   5 1.92   0.986
##  [7,]   6 1.92   0.986
##  [8,]   7 1.92   0.987
##  [9,]   8 1.93   0.987
## [10,]   9 1.92   0.987
## Type 2: with drift no trend
##       lag      ADF p.value
```

```
##  [1,]   0  0.00958   0.956
##  [2,]   1 -0.11378   0.944
##  [3,]   2 -0.15154   0.939
##  [4,]   3 -0.12829   0.942
##  [5,]   4 -0.04057   0.952
##  [6,]   5 -0.09845   0.946
##  [7,]   6 -0.05907   0.951
##  [8,]   7 -0.11796   0.944
##  [9,]   8 -0.07078   0.950
## [10,]   9 -0.04267   0.952
## Type 3: with drift and trend
##       lag   ADF p.value
##  [1,]   0 -1.15   0.914
##  [2,]   1 -1.30   0.876
##  [3,]   2 -1.34   0.857
##  [4,]   3 -1.31   0.869
##  [5,]   4 -1.27   0.887
##  [6,]   5 -1.34   0.857
##  [7,]   6 -1.32   0.865
##  [8,]   7 -1.35   0.853
##  [9,]   8 -1.32   0.867
## [10,]   9 -1.31   0.872
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

首先嘗試的分析方法為一次差分。從圖中可以發現資料仍舊存在短期的劇烈變動以及穩定的增加趨勢。而且可以發現此增加趨勢似乎沒有緩和的現象，表明此時間數列可能存在單根。是以，近一步透過 ADF 檢定方法，分析趨勢、飄移及滯後期之下的資料是否存在單根。從檢定結果的報表中可以發現不管有無趨勢(Trend)或飄移（Drift）的加入，以及滯後期為 0-9 期之間，強烈的證據都指向單根存在於此序列資料中。為了符合對於穩定態（Stationary）資料的統計假設，以下使用常見的資料處理方法：取對數及差分進行探討。

```
# Step 1: 差分後的價格
gold_diff <- diff(AU_OIL$GoldPrice_interp)

# Step 2: 調整時間向量（去掉第一天，因為差分少一天）
dates_diff <- AU_OIL$Date[-1]  # 或 tail(dates, -1)

# Step 3: 繪圖（關閉預設 x 軸）
plot(dates_diff, gold_diff, type = "l",
     main = "Differenced Gold Price Time Series",
     ylab = "Differenced Gold Price",
     xlab = "Date",
     xaxt = "n")

# Step 4: 標出每年 1 月 1 日的位置
years <- seq(from = as.Date("2010-01-01"), to = as.Date("2025-01-01"),
```

```
by = "year")

# 只選擇 years 有在日期範圍內的
valid_years <- years[years %in% dates_diff]

# Step 5: 畫上 x 軸標籤（以年份顯示）
axis(side = 1, at = valid_years, labels = format(valid_years, "%Y"))
```



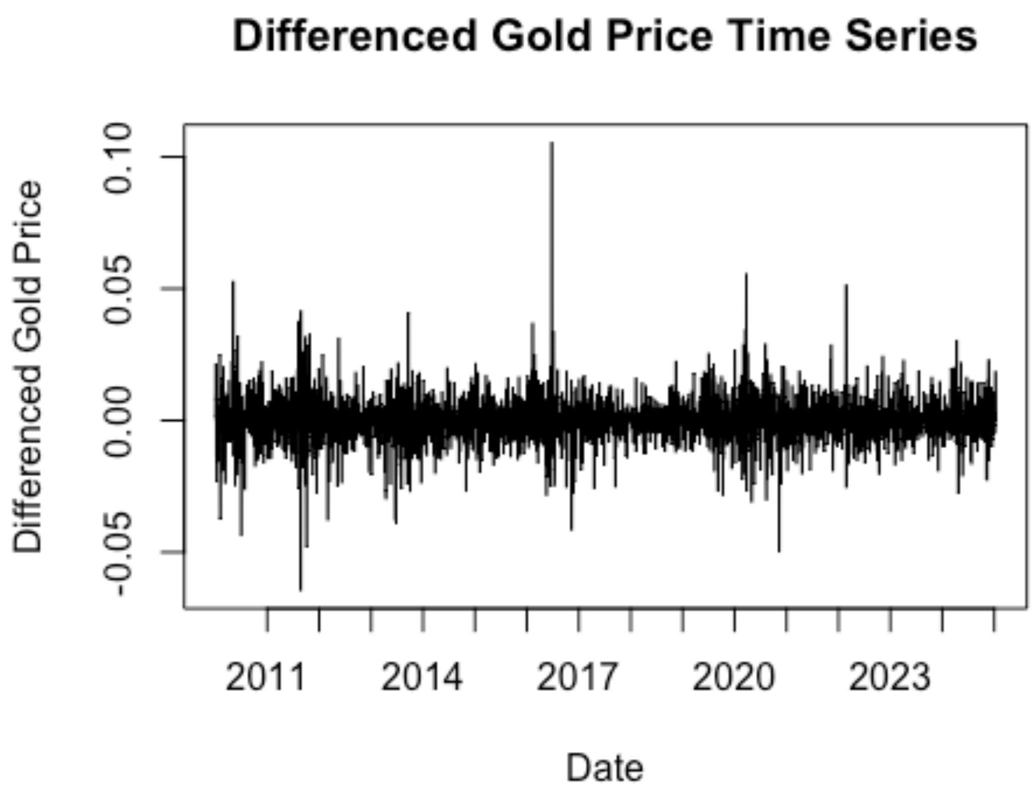

```
adf.test(diff(AU_OIL$GoldPrice_interp))

## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##       lag   ADF p.value
##  [1,]   0 -69.8    0.01
##  [2,]   1 -49.7    0.01
##  [3,]   2 -41.9    0.01
##  [4,]   3 -37.0    0.01
##  [5,]   4 -32.2    0.01
##  [6,]   5 -29.6    0.01
##  [7,]   6 -27.3    0.01
##  [8,]   7 -26.2    0.01
```

```
##  [9,]   8 -25.1    0.01
## [10,]   9 -23.4    0.01
## Type 2: with drift no trend
##       lag   ADF p.value
##  [1,]   0 -69.8    0.01
##  [2,]   1 -49.8    0.01
##  [3,]   2 -41.9    0.01
##  [4,]   3 -37.1    0.01
##  [5,]   4 -32.2    0.01
##  [6,]   5 -29.7    0.01
##  [7,]   6 -27.4    0.01
##  [8,]   7 -26.3    0.01
##  [9,]   8 -25.2    0.01
## [10,]   9 -23.5    0.01
## Type 3: with drift and trend
##       lag   ADF p.value
##  [1,]   0 -69.9    0.01
##  [2,]   1 -49.8    0.01
##  [3,]   2 -42.0    0.01
##  [4,]   3 -37.1    0.01
##  [5,]   4 -32.3    0.01
##  [6,]   5 -29.8    0.01
##  [7,]   6 -27.5    0.01
##  [8,]   7 -26.4    0.01
##  [9,]   8 -25.3    0.01
## [10,]   9 -23.6    0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

其次，嘗試的資料屬於進行差分處理的資料，從資料來看，雖然仍舊有部分時間的資料有劇烈波動，但是相較前兩個圖形，明顯少了穩定上升的現象。進一步檢測資料的單根，可以發現與前兩個方法完全不同的結果。不論趨勢是否存在以及落後期數，在以 0.05 的信心水準之下，所以組合的檢定結果都表明這個方法處理的資料不具有單根，合理推測此筆資料適合進行分析。總結來看，透過差分處理的資料不再有單根，僅剩劇烈的波動問題存在於此筆資料。

```r
# Step 1: 差分後的價格
gold_diff <- diff(log(AU_OIL$GoldPrice_interp))

# Step 2: 調整時間向量（去掉第一天，因為差分少一天）
dates_diff <- AU_OIL$Date[-1]  # 或 tail(dates, -1)

# Step 3: 繪圖（關閉預設 x 軸）
plot(dates_diff, gold_diff, type = "l",
     main = "Differenced Gold Price Time Series",
     ylab = "Differenced Gold Price",
     xlab = "Date",
     xaxt = "n")
```

```
# Step 4: 標出每年 1 月 1 日的位置
years <- seq(from = as.Date("2010-01-01"), to = as.Date("2025-01-01"),
by = "year")

# 只選擇 years 有在日期範圍內的
valid_years <- years[years %in% dates_diff]

# Step 5: 畫上 x 軸標籤（以年份顯示）
axis(side = 1, at = valid_years, labels = format(valid_years, "%Y"))
```



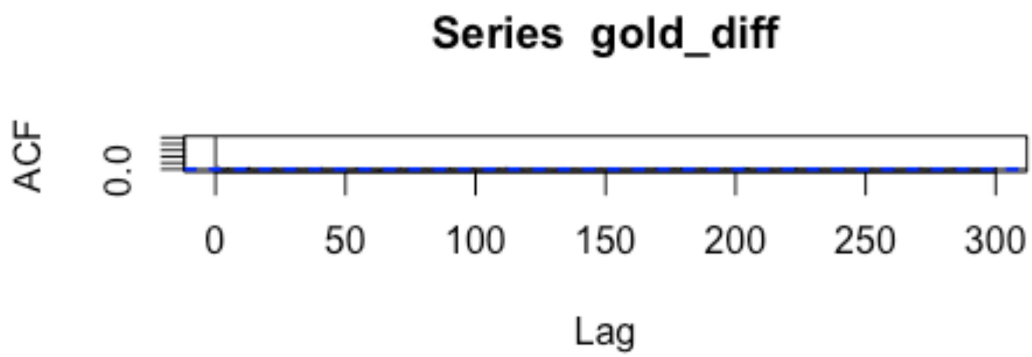

```
adf.test(diff(log(AU_OIL$GoldPrice_interp)))

## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##       lag   ADF p.value
## [1,]    0 -69.5    0.01
## [2,]    1 -49.7    0.01
## [3,]    2 -41.8    0.01
## [4,]    3 -36.8    0.01
## [5,]    4 -31.9    0.01
```

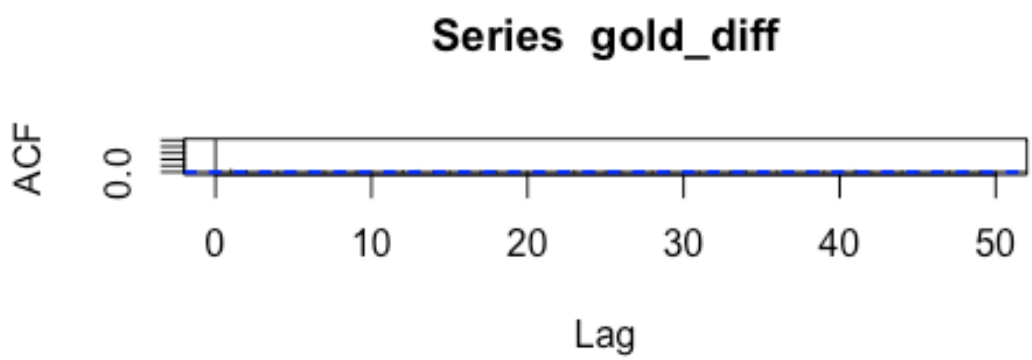
```
##  [6,]   5 -29.5    0.01
##  [7,]   6 -27.1    0.01
##  [8,]   7 -25.7    0.01
##  [9,]   8 -24.4    0.01
## [10,]   9 -22.8    0.01
## Type 2: with drift no trend
##       lag   ADF p.value
##  [1,]   0 -69.6    0.01
##  [2,]   1 -49.8    0.01
##  [3,]   2 -41.8    0.01
##  [4,]   3 -36.9    0.01
##  [5,]   4 -32.0    0.01
##  [6,]   5 -29.6    0.01
##  [7,]   6 -27.2    0.01
##  [8,]   7 -25.8    0.01
##  [9,]   8 -24.5    0.01
## [10,]   9 -22.9    0.01
## Type 3: with drift and trend
##       lag   ADF p.value
##  [1,]   0 -69.6    0.01
##  [2,]   1 -49.8    0.01
##  [3,]   2 -41.8    0.01
##  [4,]   3 -36.9    0.01
##  [5,]   4 -32.0    0.01
##  [6,]   5 -29.6    0.01
##  [7,]   6 -27.2    0.01
##  [8,]   7 -25.8    0.01
##  [9,]   8 -24.5    0.01
## [10,]   9 -22.9    0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

再來，嘗試結合差分與對數處理的方法。從圖中可以看到劇烈波動趨於穩定，且同樣少了長期穩定上升的趨勢，合理推測此種方法不只消除單根、也緩和劇烈波動的現象。進一步檢測資料的單根，可以發現與前兩個方法完全不同的結果。不論趨勢是否存在以及落後期數，在以 0.05 的信心水準之下，所以組合的檢定結果都表明這個方法處理的資料不具有單根，合理推測此筆資料適合進行分析。 總結個段資料處理方法可以發現，原始資料存在的劇烈波動及穩定上升趨勢（單根），透過結合差分及對數的處理為相對較佳的方法。故而，後續將以此方法進行深入分析與建立解釋模型。

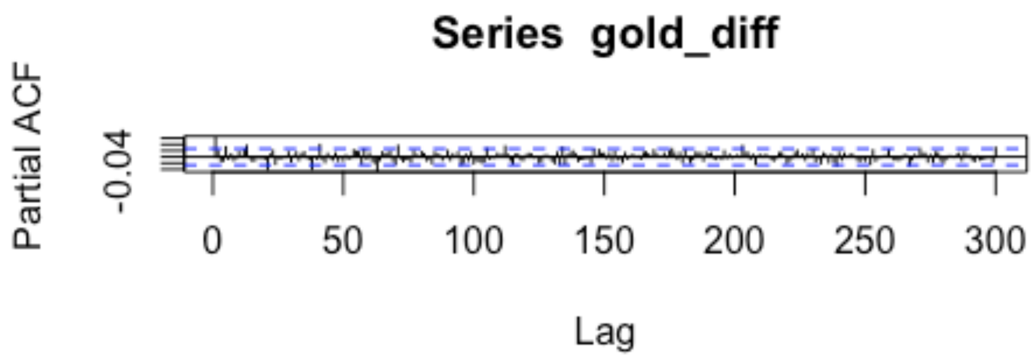
##Data mining

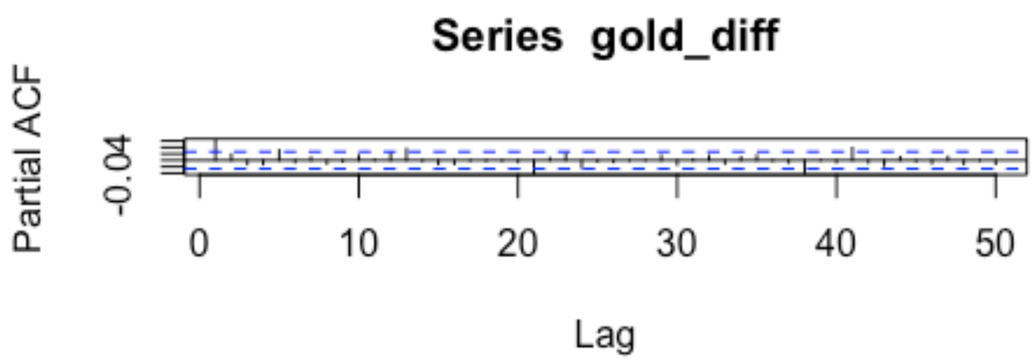
```
par(mfrow=c(2,1))
acf(gold_diff, lag.max=300)
acf(gold_diff, lag.max=50)
```




```
pacf(gold_diff, lag.max=300)
pacf(gold_diff, lag.max=50)
```



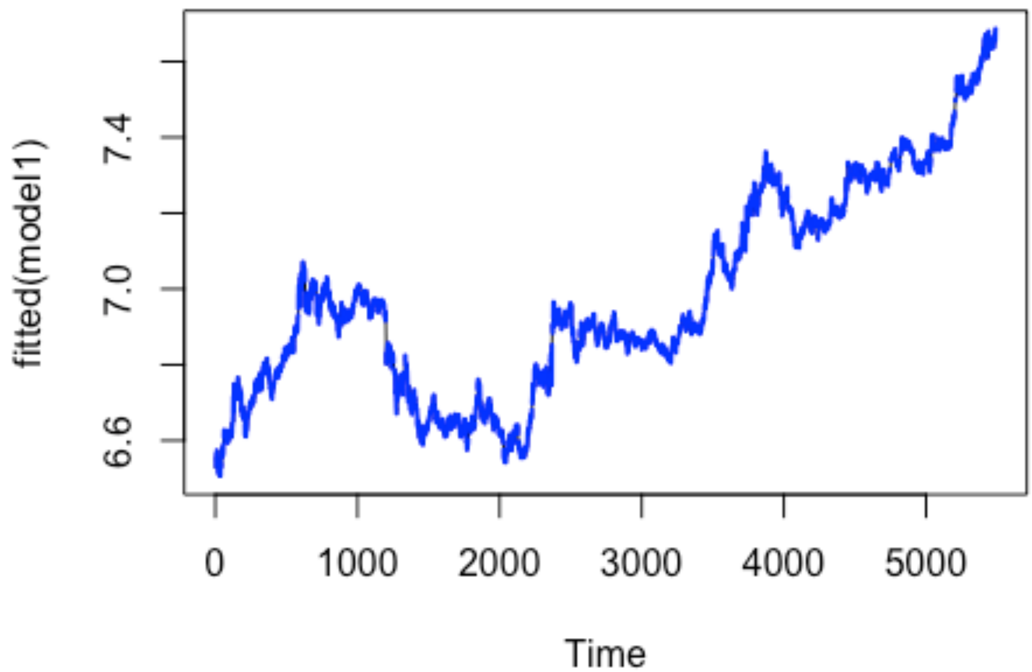




```
par(mfrow=c(1,1))
```

AR(1) ,MA(1)

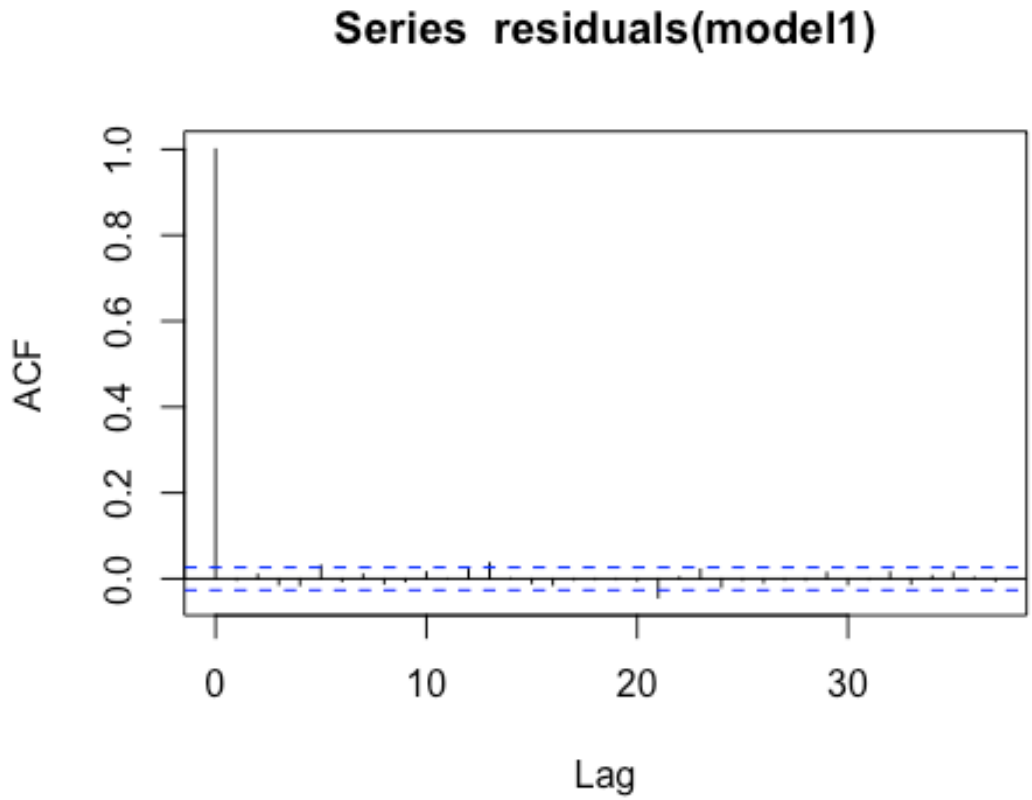
```
model1 <- arima(log(AU_OIL$GoldPrice_interp), order = c(1,1,1))
model1

##
## Call:
## arima(x = log(AU_OIL$GoldPrice_interp), order = c(1, 1, 1))
##
## Coefficients:
##          ar1      ma1
##       0.2256  -0.1624
## s.e.  0.1665   0.1678
##
## sigma^2 estimated as 5.253e-05:  log likelihood = 19256.16,  aic = -
38506.31

plot(fitted(model1))
nrow(AU_OIL)

## [1] 5490
```

```
lines(1:nrow(AU_OIL), y=log(AU_OIL$GoldPrice), type="l", lwd=2, col="bl
ue")
```

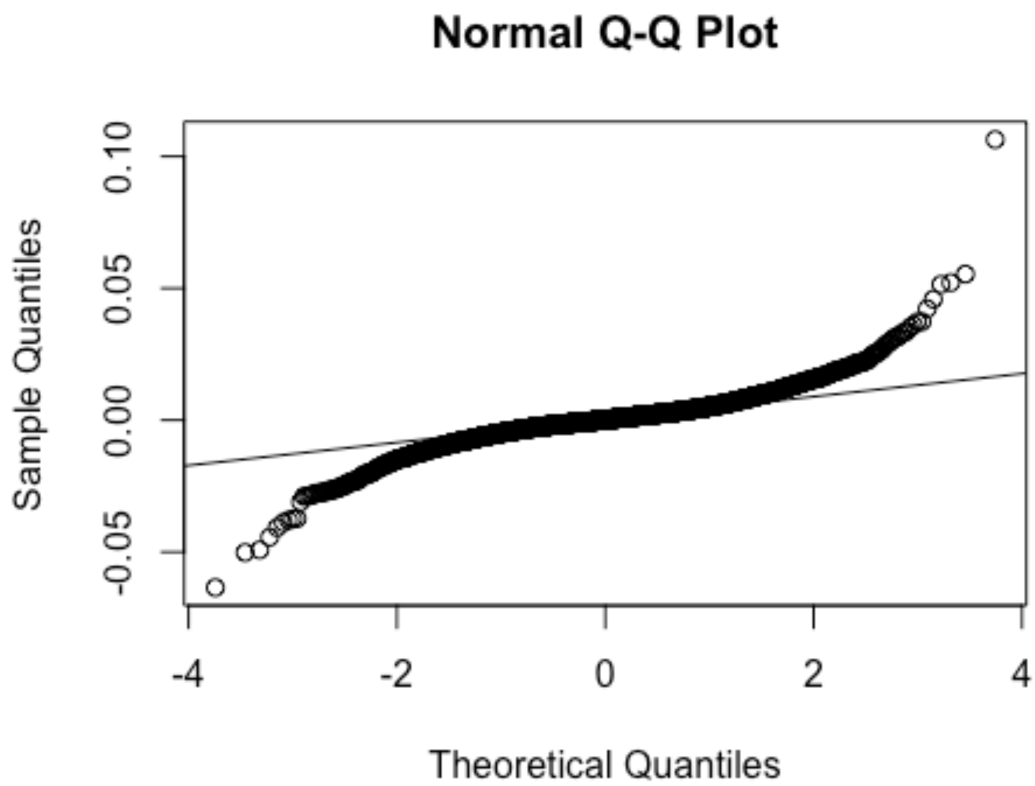



AR1, MA1 的模型看起來不錯，

```
acf(residuals(model1))
```




```
qqnorm(residuals(model1)); qqline(residuals(model1))
```




```
ks.test(residuals(model1), "pnorm")

##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  residuals(model1)
## D = 0.48664, p-value < 2.2e-16
## alternative hypothesis: two-sided

library(nortest)
ad.test(residuals(model1))

##
##  Anderson-Darling normality test
##
## data:  residuals(model1)
## A = 119.85, p-value < 2.2e-16

Box.test(residuals(model1), lag = 20, type = "Ljung-Box")

##
##  Box-Ljung test
##
```

```
## data:  residuals(model1)
## X-squared = 21.921, df = 20, p-value = 0.3448
```

對殘差作分析，以便符合模型設計。分析問題有：是否為常態、殘差是否有自相關性、變異數是否相等

```
set.seed(42)  # 固定隨機種子

# 假設你用的是模型殘差
res <- residuals(model1)

# 分組
n <- length(res)
half <- floor(n / 2)
group1 <- res[1:half]
group2 <- res[(half + 1):n]

# 真實標準差差異
obs_diff <- abs(sd(group1) - sd(group2))

# 置換檢定
n_perm <- 10000
perm_diffs <- replicate(n_perm, {
  perm <- sample(res)  # 隨機重組
  g1 <- perm[1:half]
  g2 <- perm[(half + 1):n]
  abs(sd(g1) - sd(g2))
})

# p 值
p_value <- mean(perm_diffs >= obs_diff)
cat("Permutation test p-value:", p_value, "\n")

## Permutation test p-value: 0
```

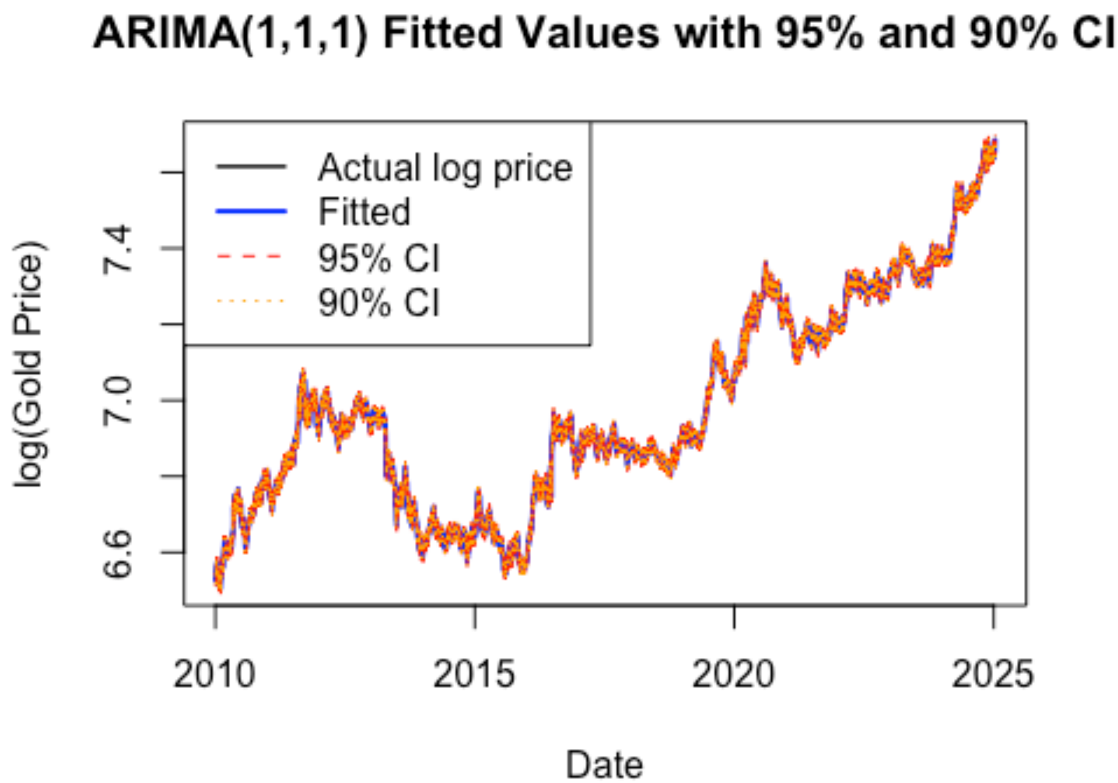
可以發現不管怎麼變動，p 值都小於 0.01，顯示該筆資料以 ARIMA 模型配置來說是相當合適得。

```
library(forecast)

# Fit ARIMA(1,1,1) model
model1 <- arima(log(AU_OIL$GoldPrice_interp), order = c(1,1,1))

# Extract fitted differences (Δlog_price)
z_fit <- fitted(model1)  # length = N - 1

# Get full time series
log_gold <- log(AU_OIL$GoldPrice_interp)
dates <- AU_OIL$Date
```

```r
# Initialize fitted log-price vector
log_price_hat <- rep(NA, length(log_gold))
log_price_hat[2:length(log_gold)] <- z_fit
```

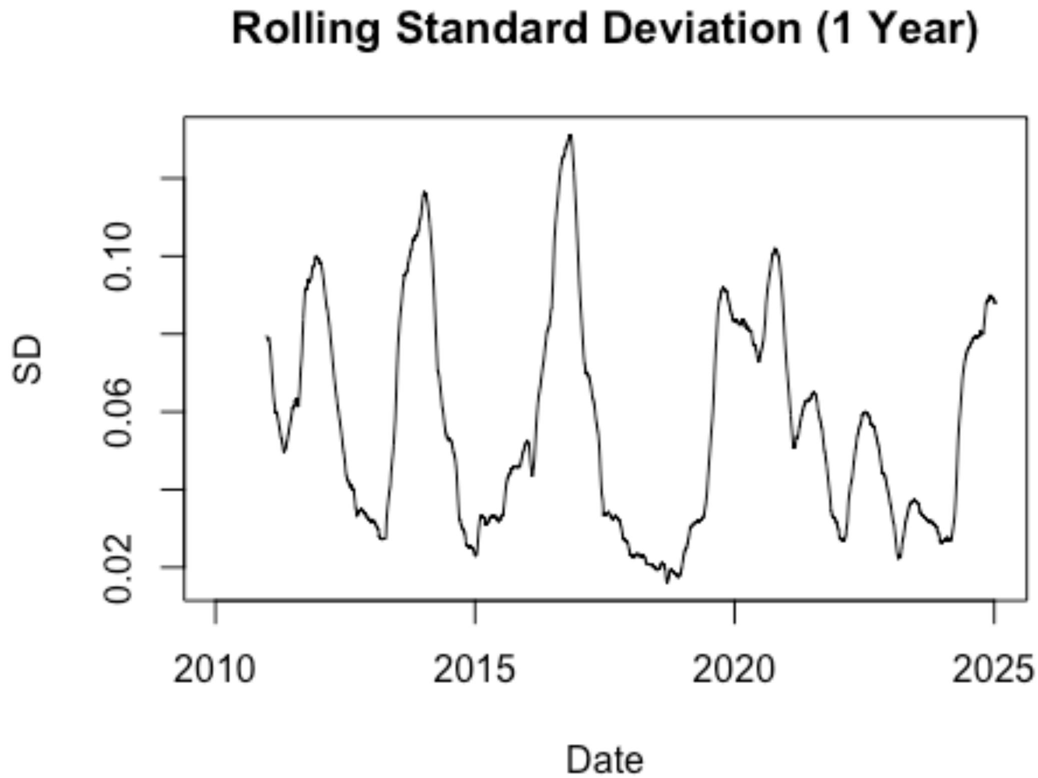
```
## Warning in log_price_hat[2:length(log_gold)] <- z_fit:
## 被替換的項目不是替換值長度的倍數
```

```r
# Residual standard deviation
resid_sd <- sd(residuals(model1), na.rm = TRUE)

# CI quantiles
z95 <- qnorm(0.975)
z90 <- qnorm(0.95)

# Confidence intervals on log scale
upper95 <- log_price_hat + z95 * resid_sd
lower95 <- log_price_hat - z95 * resid_sd
upper90 <- log_price_hat + z90 * resid_sd
lower90 <- log_price_hat - z90 * resid_sd

# Plot
plot(dates, log_gold, type = "l", col = "black", lwd = 1.5,
     main = "ARIMA(1,1,1) Fitted Values with 95% and 90% CI",
     xlab = "Date", ylab = "log(Gold Price)")

lines(dates, log_price_hat, col = "blue", lwd = 2)
lines(dates, upper95, col = "red", lty = 2)
lines(dates, lower95, col = "red", lty = 2)
lines(dates, upper90, col = "orange", lty = 3)
lines(dates, lower90, col = "orange", lty = 3)

legend("topleft", legend = c("Actual log price", "Fitted", "95% CI", "9
0% CI"),
       col = c("black", "blue", "red", "orange"),
       lty = c(1, 1, 2, 3), lwd = c(1.5, 2, 1, 1))
```




所有的觀測值都落在 fitted value 的 95%信賴區間內

```
p <-predict(model1, n.ahead = 12)
exp(p$pred)

## Time Series:
## Start = 5491
## End = 5502
## Frequency = 1
##  [1] 2178.263 2178.295 2178.302 2178.304 2178.304 2178.304 2178.304
2178.304
##  [9] 2178.304 2178.304 2178.304 2178.304

library(zoo)
roll_sd <- rollapply(log(AU_OIL$GoldPrice_interp), width = 365, FUN = s
d, align = "right", fill = NA)

plot(AU_OIL$Date, roll_sd, type = "l",
     main = "Rolling Standard Deviation (1 Year)",
     ylab = "SD", xlab = "Date")
```




所有的觀測值都落在 fitted value 的 95%信賴區間內