

Time series

Vũ Quang Hiếu, Đào Tất Thắng

2023-11-16

Chủ đề : Dự đoán giá vàng bằng mô hình ARIMA

Khai báo thư viện và dữ liệu

```
rm(list=ls())  
library(tseries)  
library(astsa)  
library(forecast)  
library(readxl)  
library(lubridate)
```

Dữ liệu giá vàng - USD tính theo tháng từ 1950-01 đến 2020-07

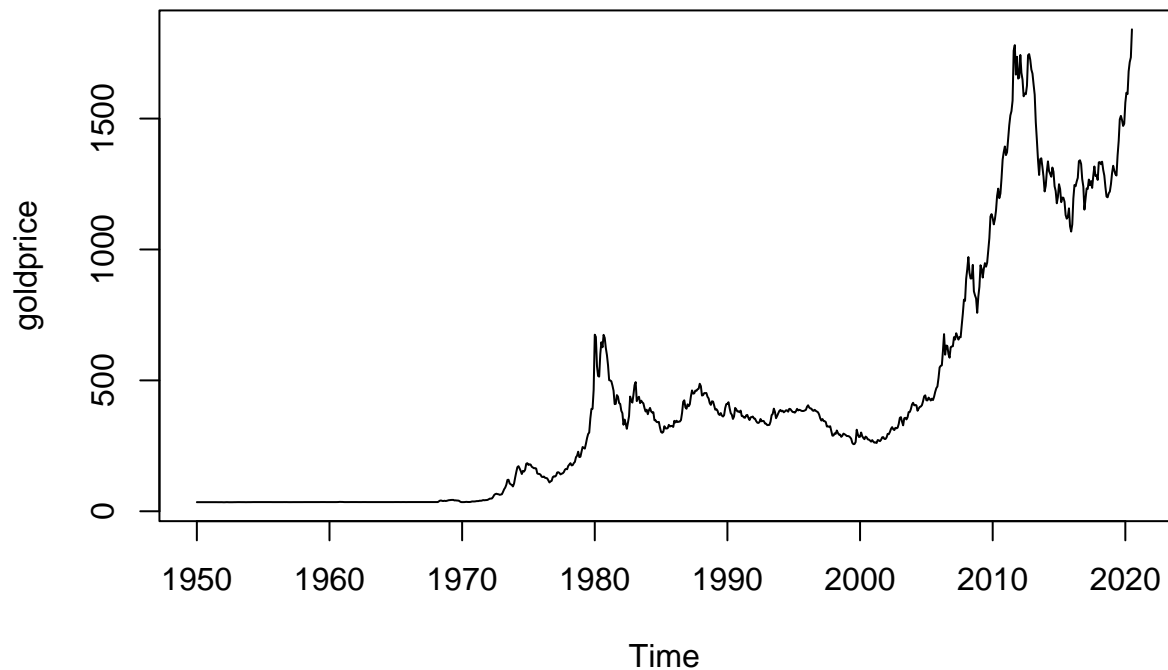
Data source: monthly.csv in Data Files

```
data = read_excel("D:\\Time Series\\data\\monthly_xlsx.xlsx")  
data$Date = ym(data$Date)  
head(data, 5)
```

```
## # A tibble: 5 x 2  
##   Date      Price  
##   <date>    <dbl>  
## 1 1950-01-01  34.7  
## 2 1950-02-01  34.7  
## 3 1950-03-01  34.7  
## 4 1950-04-01  34.7  
## 5 1950-05-01  34.7
```

Plot dữ liệu và kiểm định tính dừng

```
goldprice = ts(data$Price, start=c(1950,1), end=c(2020,7), frequency=12)  
plot.ts(goldprice)
```



Dựa vào biểu đồ có thể thấy chuỗi thời gian không có tính dừng.

Sử dụng Augmented Dickey-Fuller test để kiểm định tính dừng:

```
adf.test(goldprice)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: goldprice
## Dickey-Fuller = -0.75633, Lag order = 9, p-value = 0.9655
## alternative hypothesis: stationary
```

Bài toán kiểm định:

H_0 : Chuỗi thời gian không có tính dừng

H_1 : Chuỗi thời gian có tính dừng

Với p-value > 0.05, không có cơ sở để bác bỏ H_0

Vậy với mức ý nghĩa 5%, có thể nói chuỗi thời gian không có tính dừng

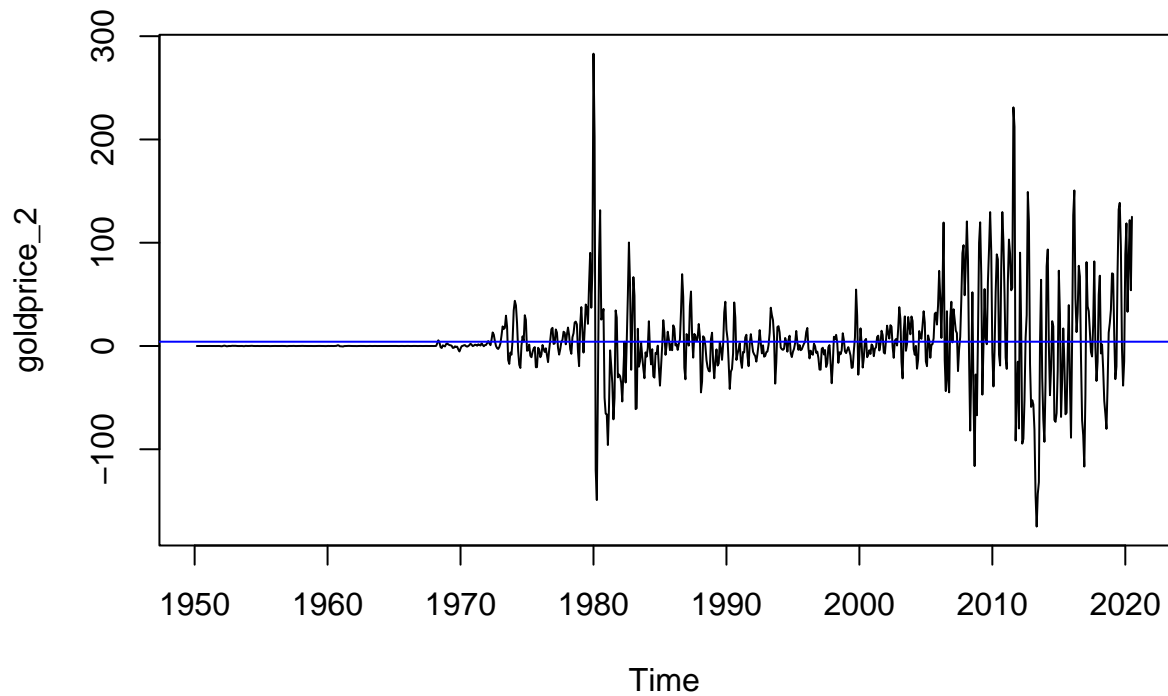
Biến đổi dữ liệu để đạt được chuỗi dừng

Sử dụng sai phân bậc 2 cho dữ liệu

```
goldprice_2 = diff(goldprice, 2)
#goldprice_2 = diff(goldprice, 3)
```

Vẽ đồ thị

```
plot(goldprice_2); abline(h=mean(goldprice_2), col="blue")
```



Kiểm định tính dừng cho chuỗi thời gian:

```
adf.test(goldprice_2)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: goldprice_2
## Dickey-Fuller = -7.0316, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

Bài toán kiểm định:

H_0 : Chuỗi thời gian không có tính dừng

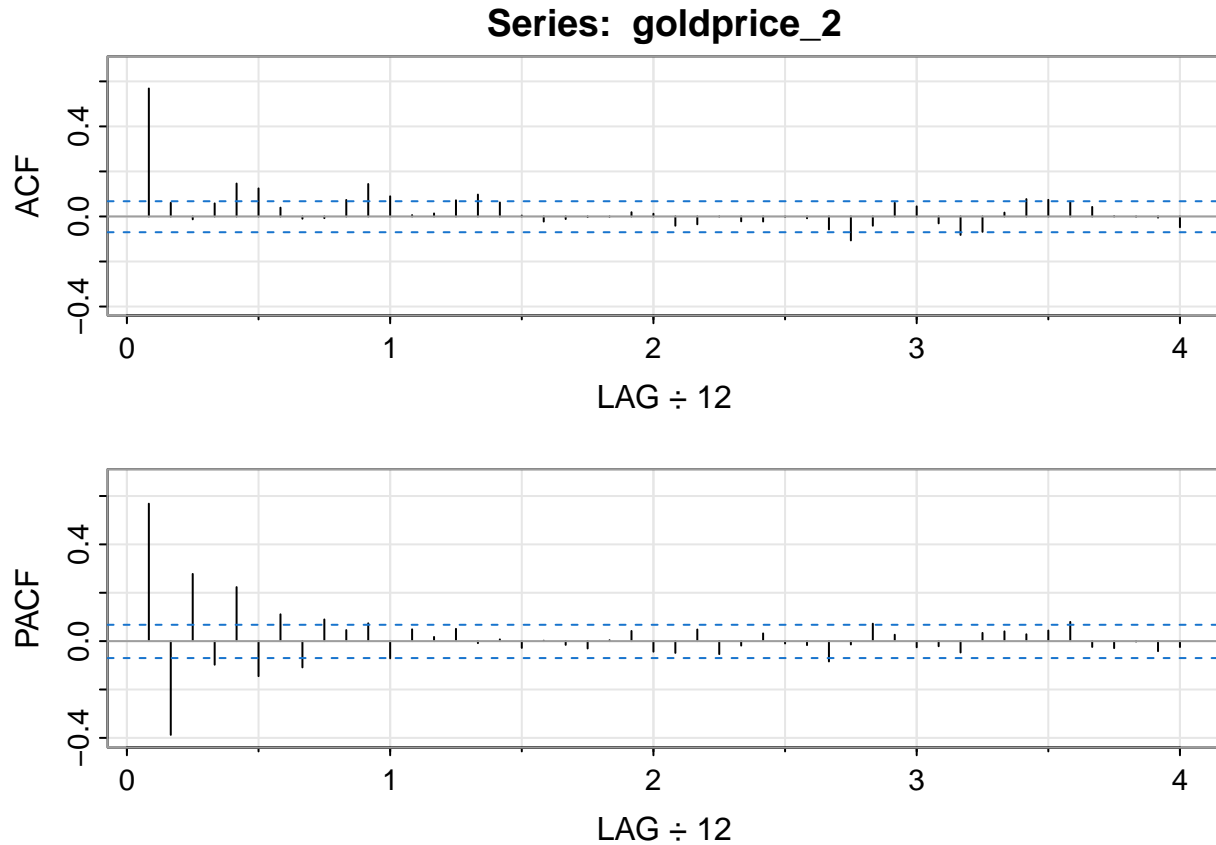
H_1 : Chuỗi thời gian có tính dừng

Với p-value < 0.05, có cơ sở để bác bỏ H_0

Vậy với mức ý nghĩa 5%, có thể nói chuỗi thời gian có tính dừng.

Xây dựng mô hình dựa vào ACF. PACF

```
acf2(goldprice_2)
```



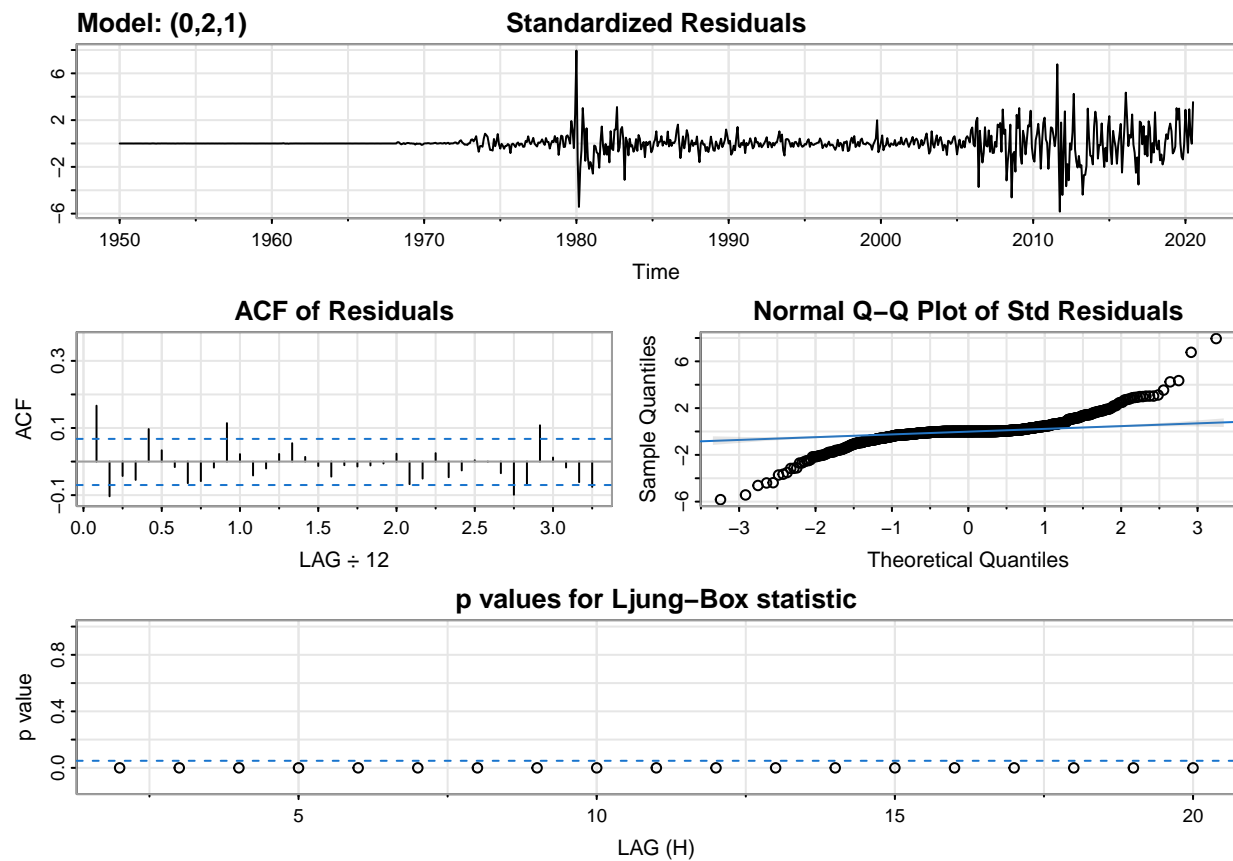
```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.57  0.06 -0.01  0.06  0.15  0.12  0.04 -0.01 -0.01  0.07  0.14  0.09  0.01
## PACF  0.57 -0.39  0.28 -0.10  0.22 -0.14  0.11 -0.11  0.09  0.05  0.07 -0.07  0.05
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.01  0.07  0.10  0.06  0.00 -0.02 -0.01  0.00    0  0.02  0.01 -0.04
## PACF   0.02  0.05 -0.01  0.01 -0.03  0.00 -0.02 -0.03    0  0.04 -0.04 -0.05
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF  -0.03  0.00 -0.02 -0.02  0.00 -0.01 -0.06 -0.11 -0.04  0.06  0.04 -0.03
## PACF   0.05 -0.05 -0.02  0.03 -0.01 -0.02 -0.08 -0.01  0.07  0.03 -0.03 -0.02
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48]
## ACF  -0.08 -0.07  0.02  0.08  0.07  0.07  0.04  0.00    0 -0.01 -0.05
## PACF -0.05  0.03  0.04  0.03  0.04  0.08 -0.02 -0.03    0 -0.04 -0.02
```

PACF từ độ trễ 1 giảm dần về 0, ACF xấp xỉ 0 sau độ trễ 1 \rightarrow MA(1) \rightarrow Hệ số (p, d, q) = (0, 2, 1)

```
sarima(goldprice, 0, 2, 1)
```

```
## initial value 3.434122
## iter 2 value 3.339969
```

```
## iter 3 value 3.289452
## iter 4 value 3.266083
## iter 5 value 3.219596
## iter 6 value 3.214347
## iter 7 value 3.212220
## iter 8 value 3.212210
## iter 9 value 3.212209
## iter 9 value 3.212209
## iter 9 value 3.212209
## final value 3.212209
## converged
## initial value 3.213517
## iter 2 value 3.213504
## iter 3 value 3.213504
## iter 3 value 3.213504
## iter 3 value 3.213504
## final value 3.213504
## converged
```



```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##       include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##       REPORT = 1, reltol = tol))
```

```
##
## Coefficients:
##          ma1
##        -0.9409
## s.e.      0.0177
##
## sigma^2 estimated as 616.7:  log likelihood = -3914.41,  aic = 7832.83
##
## $degrees_of_freedom
## [1] 844
##
## $ttable
##      Estimate      SE  t.value p.value
## ma1  -0.9409 0.0177 -53.2814      0
##
## $AIC
## [1] 9.269618
##
## $AICc
## [1] 9.269624
##
## $BIC
## [1] 9.280835
```

Kiểm định hệ số cho mô hình:

Xét $H_0 : \theta_1 = 0$; $H_1 : \theta_1 \neq 0$

Vì $P_value = 0 < 0.05$, có cơ sở để bác bỏ H_0

Với mức ý nghĩa 5%, có thể nói θ_1 khác 0.

Kiểm định phần dư:

H_0 : Phần dư tuân theo phân phối chuẩn; H_1 : Phần dư không tuân theo phân phối chuẩn

Normal Q-Q plot: các điểm dữ liệu nằm lệch khỏi đường thẳng

Ljung-Box statistic: các giá trị p_value xấp xỉ 0, có cơ sở để bác bỏ H_0

Với mức ý nghĩa 5%, có thể nói phần dư không tuân theo phân phối chuẩn.

Mô hình sau kiểm định:

$$\nabla^2 Y_t = e_t - 0.9409 * e_{t-1}$$

```
model_1 = arima(goldprice, order=c(0,2,1))
```

Xây dựng mô hình sử dụng auto.arima

```
auto.arima(goldprice,ic='aic',trace=TRUE)
```

```
##
## Fitting models using approximations to speed things up...
##
## ARIMA(2,2,2)(1,0,1)[12] : Inf
```

```
## ARIMA(0,2,0) : 8189.935
## ARIMA(1,2,0)(1,0,0)[12] : 8107.958
## ARIMA(0,2,1)(0,0,1)[12] : 7818.433
## ARIMA(0,2,1) : 7816.903
## ARIMA(0,2,1)(1,0,0)[12] : 7830.496
## ARIMA(0,2,1)(1,0,1)[12] : 7831.98
## ARIMA(1,2,1) : Inf
## ARIMA(0,2,2) : Inf
## ARIMA(1,2,0) : 8093.936
## ARIMA(1,2,2) : Inf
##
## Now re-fitting the best model(s) without approximations...
##
## ARIMA(0,2,1) : 7832.827
##
## Best model: ARIMA(0,2,1)
```

```
## Series: goldprice
## ARIMA(0,2,1)
##
## Coefficients:
##          ma1
##        -0.9409
## s.e.    0.0177
##
## sigma^2 = 617.5: log likelihood = -3914.41
## AIC=7832.83 AICc=7832.84 BIC=7842.31
```

$ARIMA(0,2,1) : \nabla^2 Y_t = e_t - 0.9409 * e_{t-1}$
→ auto.arima đưa ra mô hình giống model_1.

Dự báo

Dự đoán giá vàng cho 3 năm tiếp theo:

```
pred = forecast(model_1, level=0.95, h=12*3)
pred
```

```
##          Point Forecast    Lo 95    Hi 95
## Aug 2020      1864.860 1816.186 1913.534
## Sep 2020      1888.913 1818.015 1959.811
## Oct 2020      1912.966 1823.587 2002.345
## Nov 2020      1937.019 1830.849 2043.188
## Dec 2020      1961.072 1839.034 2083.109
## Jan 2021      1985.124 1847.761 2122.488
## Feb 2021      2009.177 1856.813 2161.542
## Mar 2021      2033.230 1866.051 2200.409
## Apr 2021      2057.283 1875.386 2239.180
## May 2021      2081.336 1884.753 2277.919
## Jun 2021      2105.389 1894.107 2316.671
## Jul 2021      2129.442 1903.414 2355.470
## Aug 2021      2153.495 1912.647 2394.343
```

## Sep 2021	2177.548	1921.788	2433.308
## Oct 2021	2201.601	1930.820	2472.381
## Nov 2021	2225.653	1939.733	2511.574
## Dec 2021	2249.706	1948.517	2550.896
## Jan 2022	2273.759	1957.164	2590.354
## Feb 2022	2297.812	1965.669	2629.955
## Mar 2022	2321.865	1974.027	2669.704
## Apr 2022	2345.918	1982.233	2709.603
## May 2022	2369.971	1990.286	2749.656
## Jun 2022	2394.024	1998.183	2789.865
## Jul 2022	2418.077	2005.922	2830.231
## Aug 2022	2442.130	2013.503	2870.757
## Sep 2022	2466.183	2020.923	2911.442
## Oct 2022	2490.235	2028.183	2952.288
## Nov 2022	2514.288	2035.282	2993.295
## Dec 2022	2538.341	2042.220	3034.463
## Jan 2023	2562.394	2048.996	3075.792
## Feb 2023	2586.447	2055.612	3117.282
## Mar 2023	2610.500	2062.067	3158.933
## Apr 2023	2634.553	2068.362	3200.744
## May 2023	2658.606	2074.497	3242.714
## Jun 2023	2682.659	2080.473	3284.844
## Jul 2023	2706.712	2086.290	3327.133

```
autoplot(pred)
```

