STAT 428 Final Project

Application of Monte-Carlo Method in Gold Price Analysis

Prepared by Group 14

Chongjun Peng, Qihui Hu, Tianyi Mao and Lingyi Xu
Professor: Uma Ravat

May 02, 2019

Abstract

As everyone knows, gold is the internationally recognized hard currency. Also, it is one of the most important ways for investors to maintain and increase the value of assets. In near ten years, due to the fact of economic crisis, the price of gold was comparatively unstable. Therefore, many investors suffered heavy losses. We utilized the Time Series analysis and some methods from our class materials, such as Monte Carlo, Bayesian, Jackknife, and bootstrap, etc. to achieve our goals: 1. Generalize the change of gold price based on time variation. 2. Forecast the wealth process of gold.

Introduction

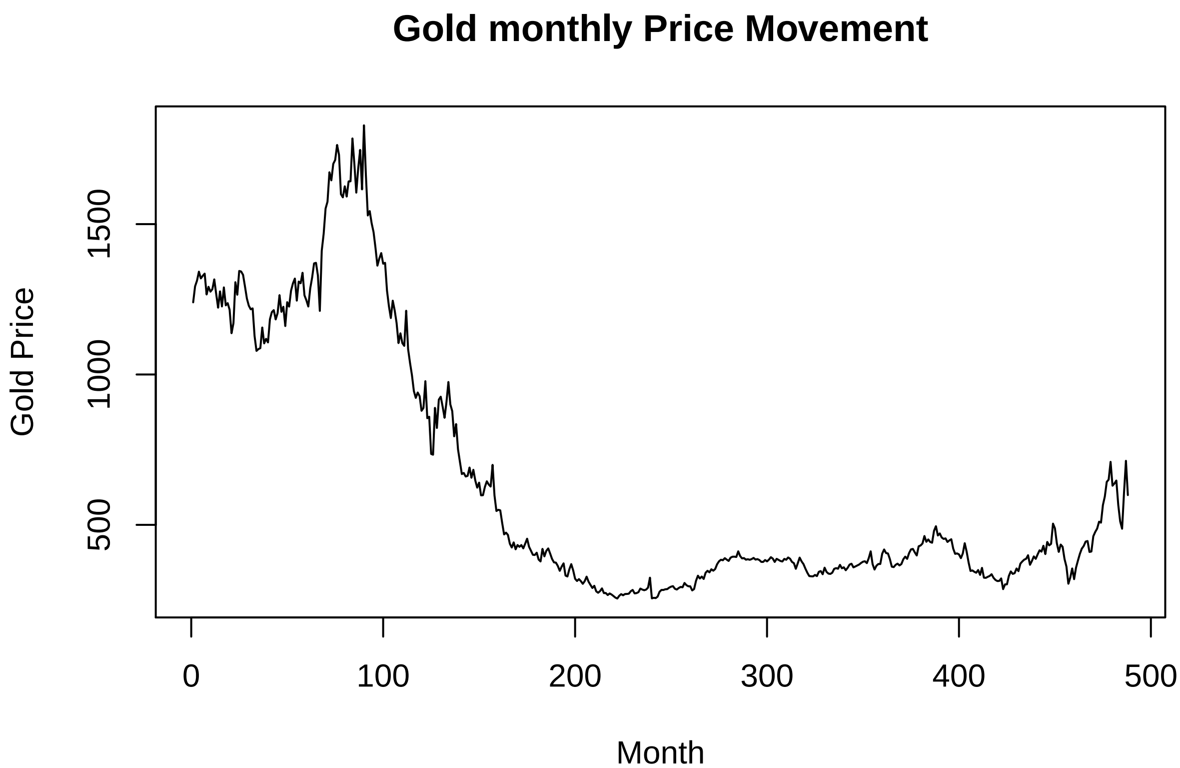
The word on the street is that investing in the gold market has a long list of benefits. Gold, as a financial hedging instrument against the US dollars, is fairly easy to trade with. However, in recent years, due to the instability of the finance industry, it is unsure whether gold maintains its functionality. By analyzing the gold price, we would like to reconfirm the function of gold and demonstrate the future trend. Specifically, we wanted to make predictions based on different empirical models (Monte Carlo and Bayesian). As for the mean and the standard deviation of the price, we applied resampling techniques including bootstrap, jackknife, and permutation.

We downloaded the daily gold price data from 1979 to 2018 from Kaggle. Immediately after we looked at the data, we started to wonder whether the price of gold has been increasing over the past few years over a relatively constant rate. Using the time-series data, we had to calculate the first order differences to reach a stationary state. We then applied jackknife and bootstrap to calculate the mean and variance of the price difference. In this way, we were able to construct the Markov Chain Monte Carlo model to forecast the future price. Because the yearly or monthly gold price is not independent and somehow depends on the price of the previous years, we then used Bayesian analysis to obtain the average price from our model for the year 1980-2019. Lastly, we compared the results of the MCMC and the Bayesian.

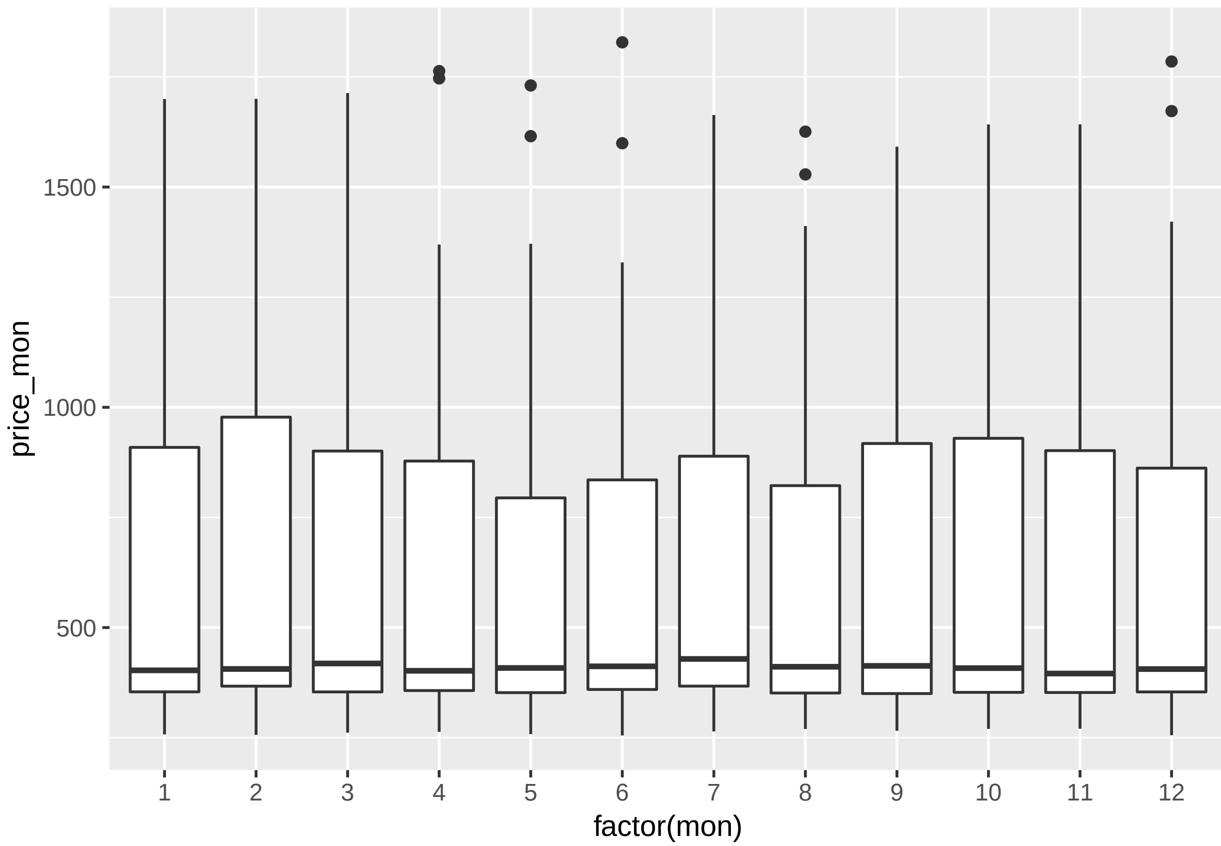
Exploratory Analysis

```
##     X      Date  Price   Open   High    Low month day year
## 1   1 2008/1/18 1216.6 1223.4 1223.4 1216.2     8   1 2018
## 2   2   7/31/18 1223.7 1220.4 1228.1 1213.0     7  31 2018
## 3   3   7/30/18 1221.3 1222.5 1223.9 1218.1     7  30 2018
## 4   4   7/27/18 1222.2 1223.5 1226.8 1216.7     7  27 2018
## 5   5   7/26/18 1225.3 1227.8 1227.8 1227.8     7  26 2018
## 6   6   7/25/18 1233.0 1225.3 1234.3 1223.2     7  25 2018
## 7   7   7/24/18 1225.3 1224.4 1229.6 1218.2     7  24 2018
## 8   8   7/23/18 1224.0 1222.0 1222.0 1222.0     7  23 2018
## 9   9   7/20/18 1229.5 1229.5 1229.5 1229.5     7  20 2018
## 10 10   7/19/18 1222.4 1227.2 1228.0 1211.7     7  19 2018
```

The data is in day-bar format, including the Date, Price, Open, High, Low of gold price.
We only use the price of gold, and here are the plots showing these time series.



There was no large-scale fluctuation in gold prices from 1979 to 2000, while in the 12 years around 2012, gold prices showed an upward trend and very rapid, and then gradually declined from 2012 to 2016. By looking up historical resource, we find that since 2000, there have been terrorist attacks and wars, such as the 911 incident, the Iranian nuclear weapons incident and so on, leading to a slow rise in gold prices. Until 2005, gold price has risen sharply from about \$400 an ounce to a record high of \$1923.20 an ounce, driven by multiple factors such as inflation and regional conflicts in the United States. From 2012 to 2013, the price of gold dropped suddenly. In fact, from 2011, the price of gold has began to stop rising or even falling gradually. This was caused from the reduction of gold holdings by central bank and investment crocodiles in 2013, the good economic figures of the recovery of the U.S. employment market and the continuing recession of the European economy.



As can be seen from 1979 to 2018, the monthly change of gold price is very small, that is, very little fluctuation. Therefore, it is basically certain that there is no cyclical change in the gold price series.

Methodology

Data Processing

We plan to model the price of gold using random walk.
Suppose X_t is the price of gold at time t, we assume the gold price as follows:

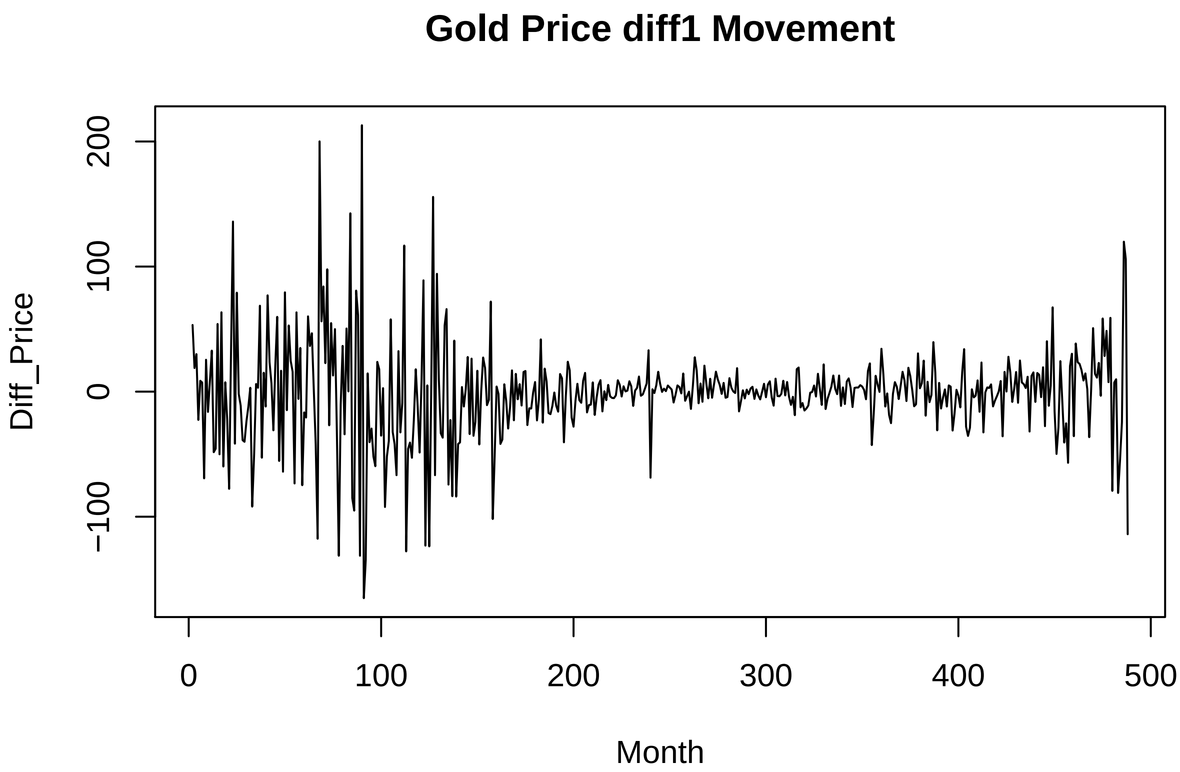
$$X_t = X_{t-1} + Z_t + \epsilon_t$$

where ϵ_t is a random disturbance term. $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[\epsilon_t \epsilon_\tau] = 0$ for $\tau \neq t$.

Before adopting the model, we test the stationality of the time series on the gold price first.

We found that the p value of adf test is 0.96 which indicates that the time series is not station. So we plan to calculate the first order difference.

```
#diff1
price_diff1 <- diff(price_mon)
```

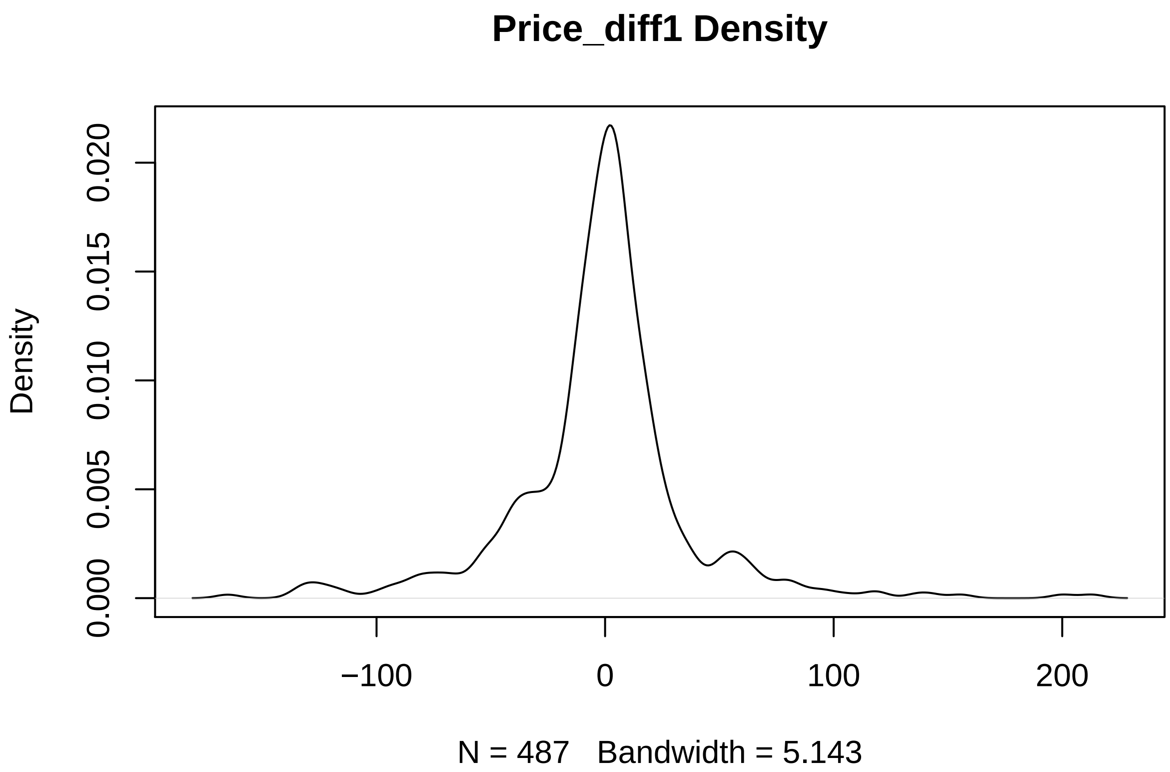



```
## Warning in adf.test(price_diff1, alternative = "stationary"): p-value
## smaller than printed p-value
```

```
## [1] 0.01
```

```
## [1] 0.02051402
```

The test results show that the differences are stationary and not correlated, so we can then explore the distribution of them.
The distribution plot is as follows:



We found that it is similar to the double exponentrial distribution, so in the next section, we implement the bootstrap and jackknife methods to estimate the variance of the differentiated price and then use permutation test to verify our assumption.

Bootstrap

Our bootstrap is implemented as follow:

1. Calculate the mean of the differentiated price from the data. Denote it as $\bar{T}^*$
2. For $b = 1, 2, ..., B$

a. Generate bth bootstrap resample $x^*_{1b}, x^*_{2b}, ..., x^*_{nb}$ from the data.

b. Compute the bth bootstrap replicate $T^*_b = T(x^*_{1b}, x^*_{2b}, ..., x^*_{nb})$

3. Finally, the bias can be calculated by

$$mean(T^*_b) - \bar{T}^*$$

the standard deviation can be calculated by

$$\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(T^*_b - \bar{T}^*)^2}$$

```
set.seed(1)
n=length(price_diff1)
B=10000
thetahat=mean(price_diff1)
boot_thetahat=numeric(B)
```

```r
for(b in 1:B){
  xboot=sample(price_diff1,n,replace=TRUE)
  boot_thetahat[b]=mean(xboot)
}
bootbias=mean(boot_thetahat)-thetahat
bootse=sd(boot_thetahat)
sprintf("The bias of price mean is %f and the standard error is %f.",bootbias,bootse)
```

```
## [1] "The bias of price mean is -0.000013 and the standard error is 1.796296."
```

Jackknife

Our Jackknife is implemented as follow:

1. Let $\bar{\hat{\theta}}_{()} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)} = mean$ of JK estimates.
2. The jackknife estimates of bias is $\hat{bias}_{jack} = (n-1)(\bar{\hat{\theta}}_{()} - \hat{\theta})$
3. The standard error of the estimator can be calculated by $\hat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \bar{\hat{\theta}}_{()})^2}$

```r
set.seed(1)
jack_thetahat=numeric(n)
for(i in 1:n){
  jack_thetahat[i]=mean(price_diff1[-i])
}
jackbias=(n-1)*(mean(jack_thetahat)-thetahat)
sumsq=sum((jack_thetahat-mean(jack_thetahat))^2)
jackse=sqrt((n-1)/n)*sqrt(sumsq)
sprintf("The bias of price mean is %f and the standard error is %f.",jackbias,jackse)
```

```
## [1] "The bias of price mean is 0.000000 and the standard error is 1.797401."
```

Permutation

Our permuation test is implemented as follow:

1. Compute the observed test statistic (t test or ks test) $\hat{\theta} = \hat{\theta}(Z, v)$. Z denotes the permutation of the data
2. For each replicated, indexed $b = 1, 2, ..., B$:

(a) Generate a random permutation π_b of v.

(b) Generate a statistic $\hat{\theta}^{(b)} = \hat{\theta}^{*}(Z, \pi_b)$

3. Use this statistic, calculate p value, and decide whether we reject the null hypothesis (the true distribution and the proposal distribution are the same)

Our Assumption is that the first order difference of gold price is similar to the double exponentrial distribution. Following is the the test result and comparision of two density functions.

```r
#permutation
s=sqrt(bootse^2/2)
double_exp=rlaplace(n,thetahat,s)
n_price = length(price_diff1)
n_exp = length(double_exp)
t.test(price_diff1, double_exp, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  price_diff1 and double_exp
## t = -0.03675, df = 972, p-value = 0.9707
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.596359  3.464138
## sample estimates:
## mean of x mean of y
## -1.315811 -1.249700
```

```r
# Compute the observed test statistic
B = 10000
z = c(price_diff1, double_exp)
nu = 1:(n_price + n_exp)
reps = numeric(B)
t0 = abs(t.test(price_diff1, double_exp, var.equal=TRUE)$statistic)

# Applying permutation test
for(i in 1:B){
  # Generate a random permutation for the data for weights of the chicks on the two diets
  perm = sample(nu, size = n_price,replace=FALSE)
  x4 = z[perm]
  y4 = z[-perm]
  # Generate a test statistic
  reps[i] = t.test(x4, y4, var.equal=TRUE)$statistic
  reps[i] = abs(reps[i])
}
# compute empirical p-value
p = mean(c(t0, reps) >= t0);p
```

```
## [1] 0.9692031
```

From the permutation test, we cannot reject null hypothesis and conclude that the two density functions are not significantly different. Thus, having the estimated density function, we move on to next step, forecasting.

Markov Chain Monte Carlo

After having the estimated density function tested from permutation test, we could then generate data from double exponential distribution and apply Markov Chain Monte Carlo in order to obtain expected gold price at the terminal time and forecast the value of gold price in the following twenty months with 100-time repeating in the plot below.

MCMC is implemented as follow:

1. Choose a proposal distribution $g(|X_t)$ from noral distribution.
2. Begin with X_0.
3. Repeat the following steps until many draws from the chain have been made after converging to its stationary distribution.

a. Generate Y from $g(|X_t)$.

b. Generate U from *Uniform(0,1)*

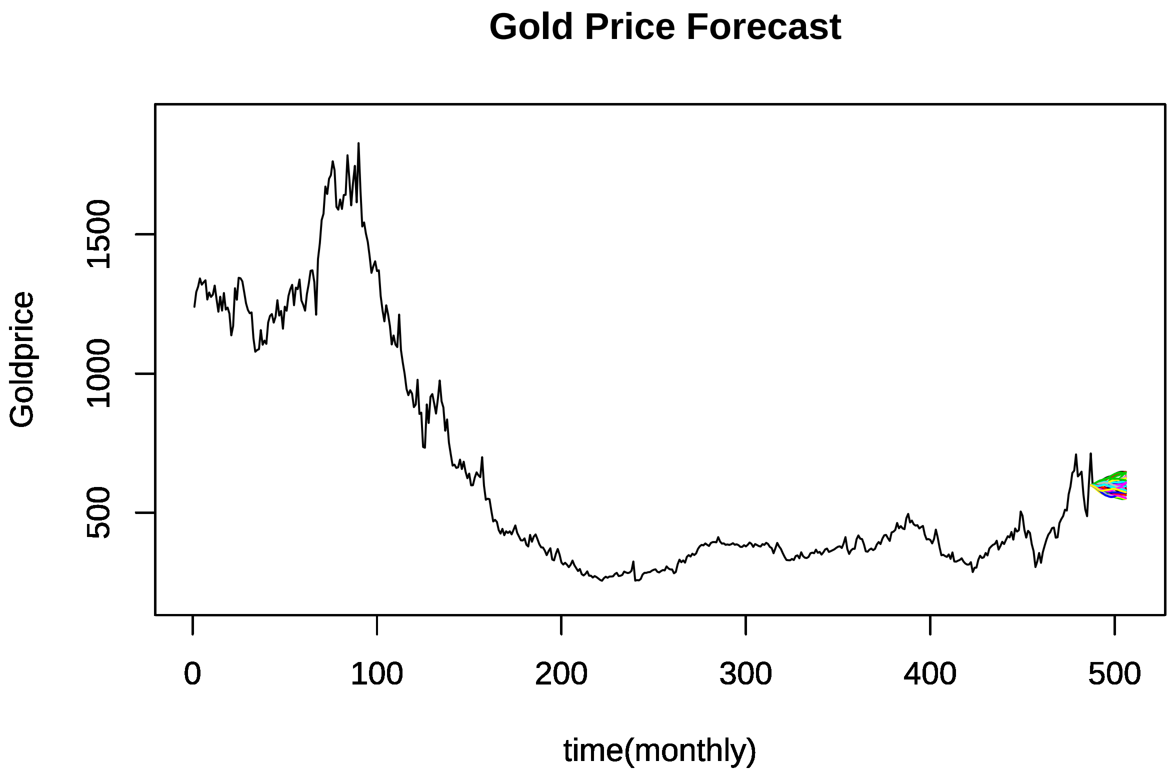
c. If

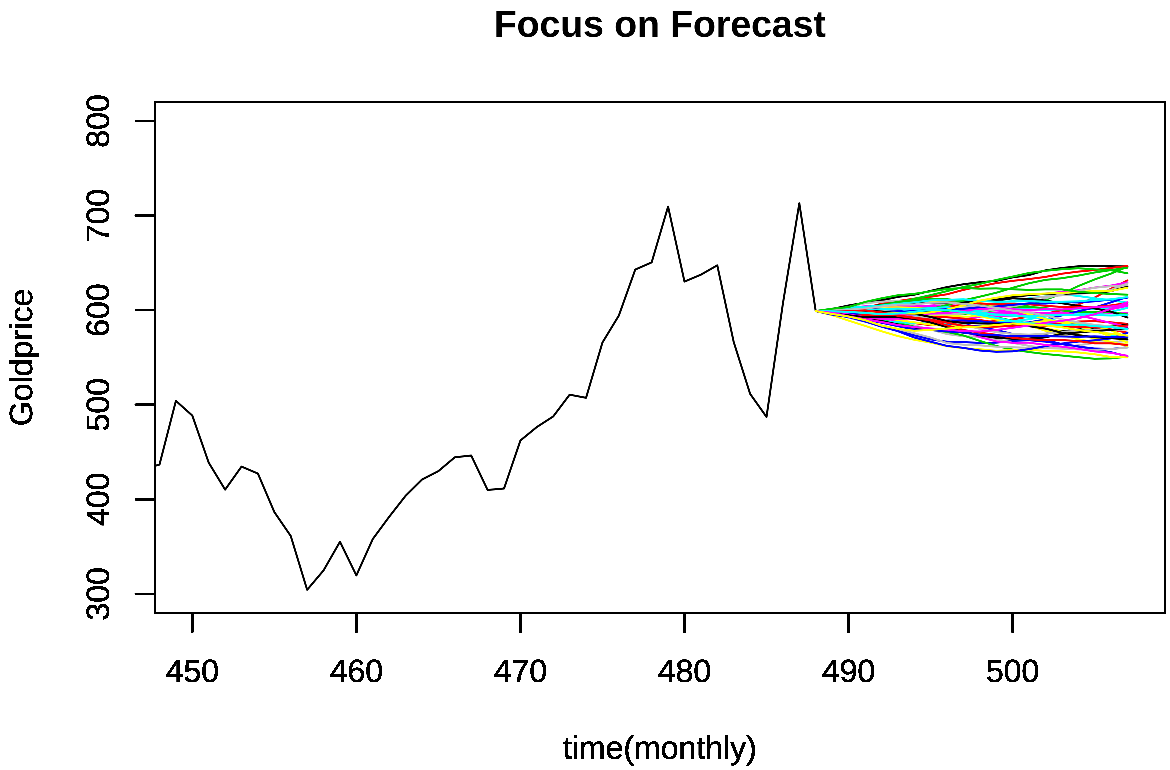
$$U \leq \frac{f(Y)g(X_t|Y)}{f(X_t)g(Y|X_t)}$$

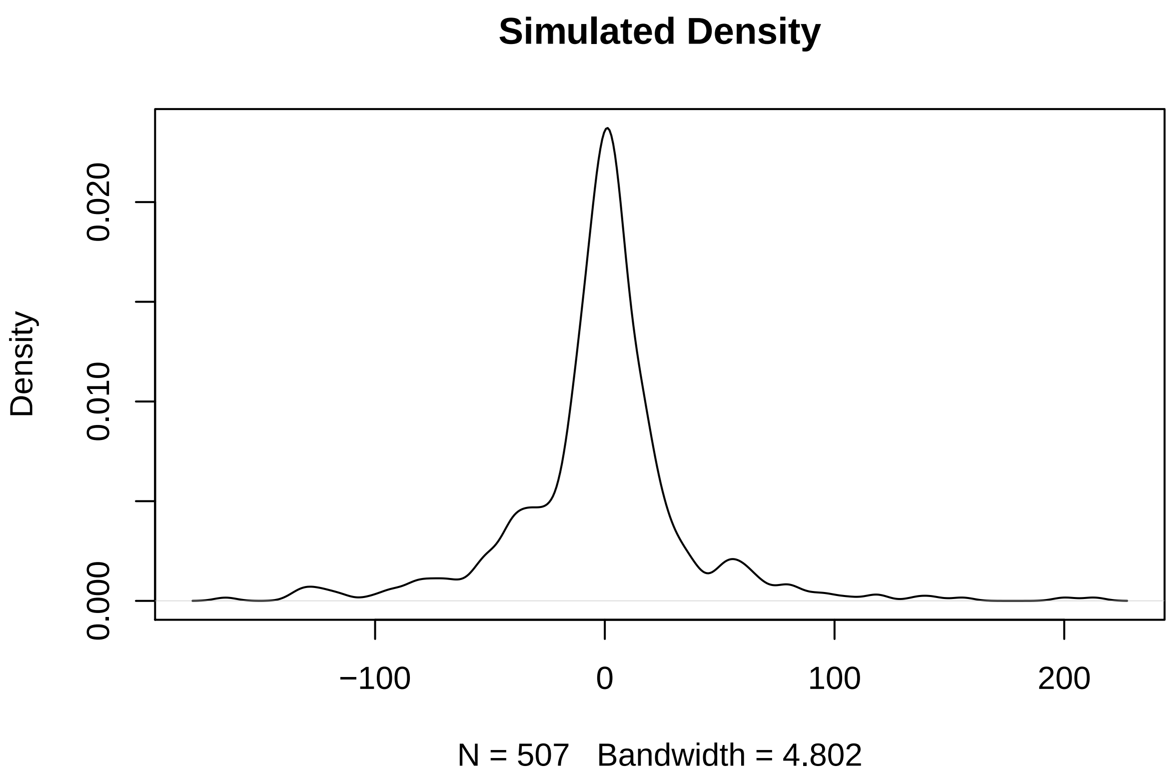
accept Y and set $X_{t+1} = Y$; otherwise set $X_{t+1} = Xt$.

```
forecast_length = 20
h=100     #100 times
gold_simulation <- matrix(NA, nrow = h, ncol = forecast_length, byrow = TRUE)
#sediff=sqrt(var(price_diff1)/2)
sediff=sqrt((bootse)^2/2)
j = 1
i=2
while(j <= h) {
  gold_simulation[j,1] <- price_mon[488]
  x <-numeric(forecast_length)
  x[1]=rnorm(1,thetahat,sediff)
  k=0
  u <-runif(forecast_length)
  for(i in 2:forecast_length) {
    xt <-x[i-1]
    y <-rnorm(1,xt,sediff)
    num <- dlaplace(y,xt,sediff)*dnorm(xt,y,sediff)
    den <- dlaplace(xt,y,sediff)*dnorm(y,xt,sediff)
    if(u[i] <= num/den)  x[i] <-y
    else {
      x[i]<-xt
      k=k+1
    }
  }

  for (i in 2:forecast_length){
    gold_simulation[j,i] <- gold_simulation[j,i-1] + x[i]
  }
    #gold_step <- rlaplace(1,thetahat,28)
  j = j + 1;
}
```

```
a<-c()
for (i in 1:100){
 if ((gold_simulation[i,20]<550) || (gold_simulation[i,20]>650))
  {a=c(a,i) }
 i=i+1
}
gold_simulation_1 <-gold_simulation[-a,]
len=dim(gold_simulation_1)[1]
```







Conclusion for MCMC :

The forecasting plot shows slight fluctuations in most of the reiterated results. Also, the density plot of values obtained from the MCMC simulation can roughly fit. Although there would always be some chance that the urgent political event or change have influence on gold price which we cannot control since gold price is highly related to dollar prices. In addition to this kind of situation, the forecasting outcome from MCMC simulation can be useful to future gold market.

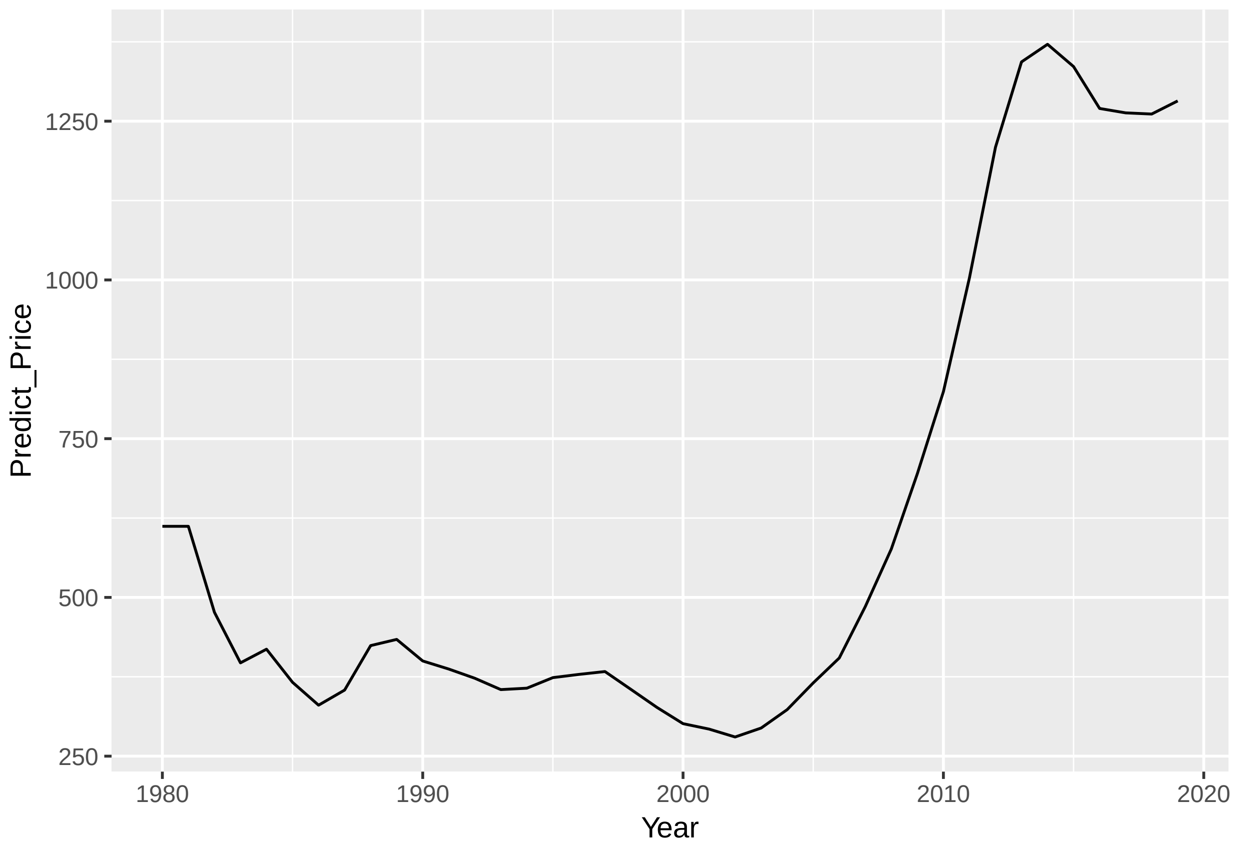
Bayesian analysis

Because we found that the price distribution of gold for each year is normal, for Bayesian analysis, we decided to use a normal conjugate prior distribution. The formula of porsterior mean is

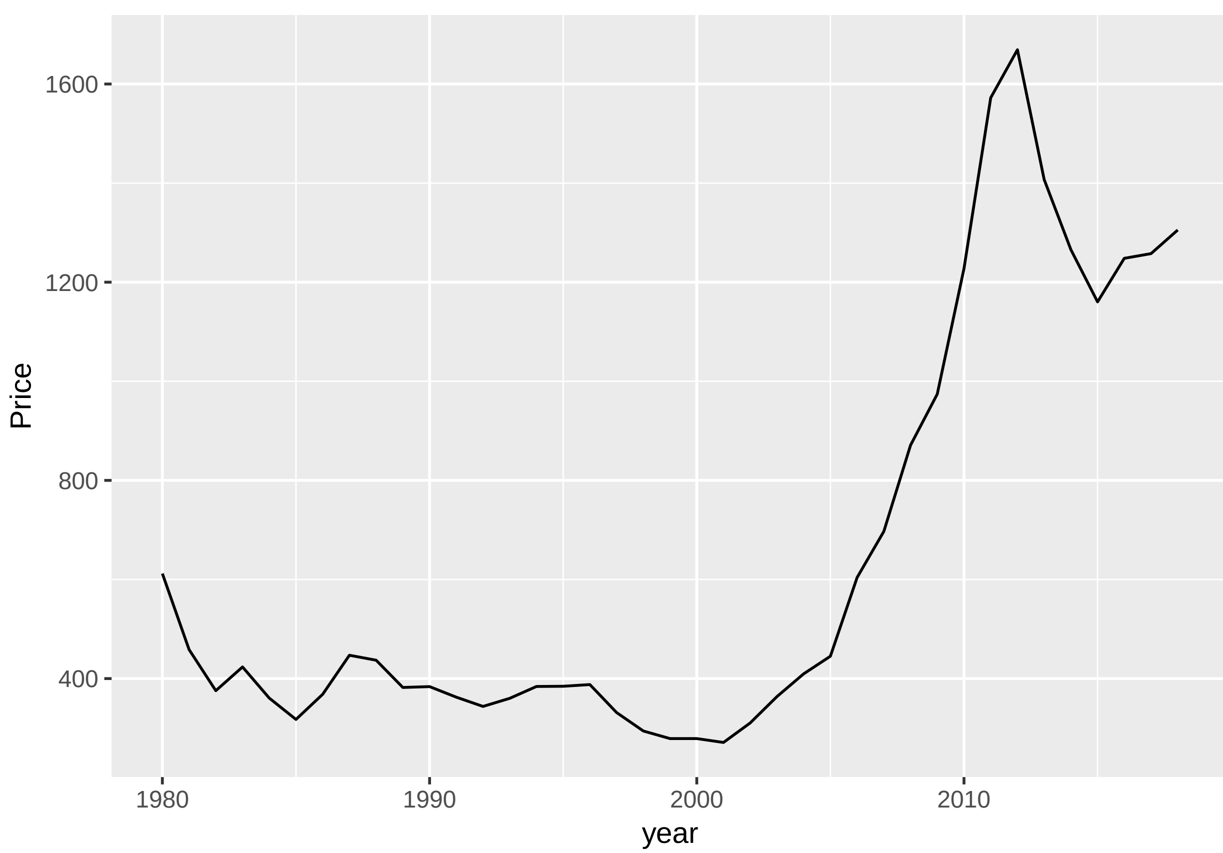
$$\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}(\frac{\mu_0}{{\sigma_0}^2}+\frac{\Sigma_i x_i}{\sigma^2}),(\frac{1}{\sigma^2}+\frac{n}{\sigma^2})^{-1}$$

The very first prior mean and prior sigma are self-set, becasue we would like to compute the posterior mean for the year 1981. We did not use the data in 1979 because there are only three observations in 1979. For the following years, the prior mean equals to the posterior mean we calculated from the previous year. Eventually, we predicted the price of 2019, and we found that the model looks moderately fair.

```
update_mean = function(prior_sigma, prior_mean, x){
  n = length(x)
  sigma = sd(x)/sqrt(n)
  1/(1/prior_sigma^2+n/sigma^2) * (prior_mean/prior_sigma^2 + sum(x)/sigma^2)
}

update_sd = function(prior_sigma, x){
  n = length(x)
  sigma = sd(x)
  sqrt((1/prior_sigma^2 + n/sigma^2) ^ (-1))
}

Year = seq(1980, 2019, 1)
post_mean = numeric(length(Year))
post_sd = numeric(length(Year))
names(post_mean) = names(post_sd) = Year
prior_mean = 600
prior_sigma = 1

x = splitdf[[1]]$Price
samples = rnorm(20, mean(x), 0.06)
post_mean[1] = update_mean(prior_sigma, prior_mean, samples)
post_sd = update_sd(prior_sigma, samples)


for (i in 2:length(post_mean)) {
  prior_sigma = post_sd[i-1]
  prior_mean = post_mean[i-1]
  x = splitdf[[i-1]]$Price

  samples = rnorm(20, mean(x), 0.06)

  post_mean[i] = update_mean(prior_sigma, prior_mean, samples)
  post_sd[i] = update_sd(prior_sigma, samples)
}
result = data.frame(Year = Year, Predict_Price = post_mean)
```



1250
1000
750
500
250
Predict_Price
1980
1990
2000
2010
2020
Year



1600
1200
800
400
Price
1980
1990
2000
2010
year

Conclusion for Bayesian analysis:

Based on the trend of the actual average price of gold price and our predicted average price over the years, we could see that our model not only successfully predicts the drop in the late 2010s but also has a very similar trend as the original data. However, the problem of overfitting may readily arise. Meanwhile, our prediction, compared to the original data, has a smaller variance, inferring that Bayesian model is not good at predicting the sudden change of the trend. That being said, one of the determinants of gold price may be the economic performance, therefore, for the project in the future, we may use GDP as one of the independent variables and gold price as the dependent variable.

Reference

[1] I. Akgüla, M. Bildiricib, S. Özdemirc, "Evaluating the Nonlinear Linkage between Gold Prices and Stock Market Index Using Markov-Switching Bayesian VAR Models", Dec. 2015, Proceedings of the 4th International Conference on Leadership, Technology, Innovation and Business Management (ICLTIBM-2014)

[2] M. Rizzo, Mar. 2019, "Statistical Computing with R", Chapman & Hall/CRC The R Series

[3] M. Rio, "Historic Gold Prices", Jun. 2018, Kaggle, retrieved from https://www.kaggle.com/omdatas/historic-gold-prices

[4] H. Kunsch, Nov. 1989, "The Jackknife and the Bootstrap for General Stationary Observations", The Annals of Statistics

[5] D. Barber, A. Cemgil, S. Chiappa, Jan. 2011, "Bayesian Time Series Models", Cambridge University Press