MAXIMUM LIKELIHOOD ESTIMATION OF VASICEK MODEL AND CALIBRATION ON MERTON JUMP DIFFUSION USING METHOD OF MOMENTS

RAN ZHAO

ABSTRACT.

1. MAXIMUM LIKELIHOOD AND VASICEK MODEL ESTIMATION

1.1. **Introduction.** The Vasicek model [5] is a mathematical finance model that describes the dynamics of interest rates. This model allows the short-term interest rate, the spot rate, to follow a random walk, which leads to a parabolic partial differential equation for the prices of bonds and other interest rate derivative products. The Vasicek model referred in this paper belongs to one factor interest rate model, where there is only one source of randomness, the spot interest rate.

The spot rate evolution is modeled in the following stochastic differential equation:

$$dr_t = a(b - r_t)dt + \sigma dW_t \tag{1}$$

where W_t is the Wiener process. a can be interpreted as the speed of mean-reverting, and b is the long term level of the spot rate. σ, here as a constant, is the instantaneous volatility.

In this paper, I use the 3-month Treasury Bills rates to approximate the spot rate. The data source is Federal Reserve Bank of St. Louis website [1]. Using monthly data, the 3-month Treasury Bill rates plots in the Figure 1. From the graph, the 3-month T-bill rates display cyclical pattern, where the rate fluctuates between its long-term trend indicating by rate's moving average. However, these long term trends vary among different time period, over 1954 to 2015. Intuitively, the short term rate is tightly connected with monetary policy, and is adjusted according to the economic status. The economic cycle enhances the pattern of the short term.

1.2. **Discretization of Vasicek Model.** In discrete time, use ϵ_t to represent the white noise with expected value of 0 and variance of 1, evolving at time t. Then the discrete version of Vasicek model expresses as

$$\begin{aligned}
\Delta r_t &= a(b - r_t)\Delta t + \sigma \Delta W_t \\
r_{t+1} - r_t &= a(b - r_t)\Delta t + \sigma(W_{t+1} - W_t) \\
&= a(b - r_t)\Delta t + \sigma \epsilon_{t+1} \\
r_{t+1} &= ab\Delta t + (1 - a\Delta t) r_t + \sigma \epsilon_{t+1}
\end{aligned}$$

Change the notation from b to $\bar{X}$ to represent the mean level of rates, from $(1 - a\Delta t)$ to ϕ simplifying the parameter. Use X_t to denote r_t, we yield

$$X_{t+1} = \bar{X}(1 - \phi) + \phi X_t + \sigma \epsilon_{t+1} \tag{2}$$

which has the form of an AR(1) process.

[1] `https://fred.stlouisfed.org/series/DTB3`

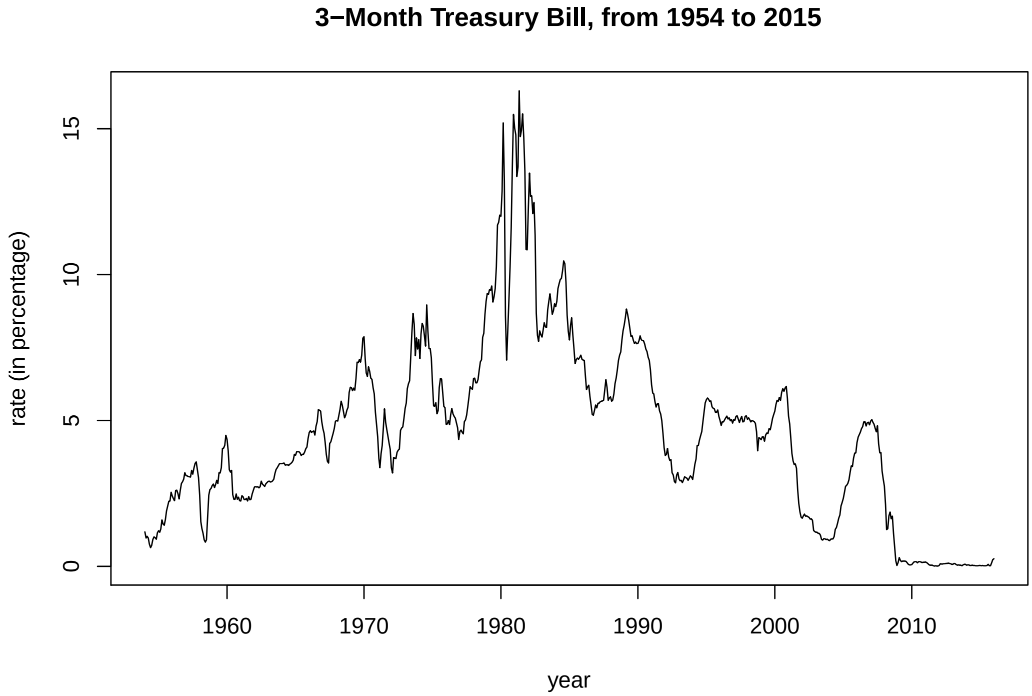

Figure 1 Monthly 3-month Treasury Bill rate at secondary market. Time period is from 1954 to 2015.

1.3. **Likelihood of the AR(1) Process.** From Equation 2, the conditional distribution of X_{t+1} is straightforward:

$$X_{t+1}|X_t \sim N(\bar{X}(1-\phi)+\phi X_t, \sigma^2) \tag{3}$$

since the ϵ_{t+1} is a white noise term with mean 0 and variance 1. The information available is enough to determine X_t, but ϵ_{t+1} provides the source of randomness. The parameters remain to be estimated, $\theta = (\bar{X}, \phi, \sigma^2)'$.

The distribution of the initial value would be $X_0 \sim N(\bar{X}, \frac{\sigma}{1-\phi^2})$. Hence,

$$\begin{aligned} f_{X_0}(x_0;\theta) &= f_{X_0}(x_0;\bar{X},\phi,\sigma^2) \\ &= \frac{1}{\sqrt{2\pi}/\sqrt{\sigma^2/(1-\phi^2)}}\exp\left[-\frac{1}{2}\frac{(x_0-\bar{X})^2}{\sigma^2/(1-\phi^2)}\right] \end{aligned}$$

Next consider the conditional distribution of the second observation. According to Equation 3, we have

$$f_{X_1|X_0}(x_1,x_0;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x_1-\bar{X}(1-\phi)-\phi x_0)^2}{\sigma^2}\right]$$

The joint density of observations 1 and 2 is then just

$$f_{X_2,X_1}(x_2.x_1;\theta) = f_{X_1|X_0}(x_1,x_0;\theta)f_{X_0}(x_0;\theta)$$

In general, the value of $X_0, X_1, \ldots, X_{T-1}$ matters for X_T only through the value X_{T-1}, and the density of observation T conditional on the preceding $T-1$ observations is given by

$$\begin{aligned} f_{X_T|X_{T-1},X_{T-2},\ldots,X_0}(x_T|x_{T-1},\ldots,x_0;\theta) &= f_{X_T|X_{T-1}}(x_T|x_{T-1};\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x_T-\bar{X}(1-\phi)-\phi x_{T-1})^2}{\sigma^2}\right] \end{aligned}$$

The likelihood of the complete sample can thus be calculated as

$$f_{X_T,X_{T-1},\ldots,X_0}(x_T,x_{T-1},\ldots,x_0;\theta)=f_{X_0}(x_0;\theta)\cdot\prod_{t=0}^{T-1}f_{X_{t+1}|X_t}(x_{t+1}|x_t;\theta)$$

Then the log-likelihood function (denoted by $\ell(\theta)$) is

$$\begin{aligned}
\ell((\theta)) &= \log f_{X_0}(x_0;\theta)+\sum_{t=0}^{T-1}\log f_{X_{t+1}|X_t}(x_{t+1}|x_t;\theta)\\
&= -\frac{1}{2}\log 2\pi+\frac{1}{2}\log(1-\phi^2)-\frac{1}{2}\log\sigma^2-\frac{1}{2}\frac{(1-\phi^2)(x_0-\bar{X})^2}{\sigma^2}\\
&\quad -\frac{T}{2}\log 2\pi-\frac{T}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{t=0}^{T-1}[x_{t+1}-\bar{X}(1-\phi)-\phi x_t]^2\\
&= -\frac{(T+1)}{2}\log 2\pi+\frac{1}{2}\log(1-\phi^2)-\frac{(T+1)}{2}\log\sigma^2-\frac{1}{2}\frac{(1-\phi^2)(x_0-\bar{X})^2}{\sigma^2}\\
&\quad -\frac{1}{2\sigma^2}\sum_{t=0}^{T-1}[x_{t+1}-\bar{X}(1-\phi)-\phi x_t]^2
\end{aligned}$$

1.4. **Maximum Likelihood Estimation.** Maximum likelihood estimation (MLE) is the method of estimating model parameters given observations, by finding the parameters (here θ) that maximize the likelihood of the model. To optimize the likelihood function derived in Equation 4, use the derivative of the likelihood function with respect to the parameter set θ:

$$\frac{\partial\ell(\theta)}{\partial\theta^*}=0$$

where again $\theta=(\bar{X},\phi,\sigma^2)'$. Here the parameter σ^2 represents the conditional variance of the short rate. The optimization requires

$$\begin{cases}
\frac{\ell(\theta)}{\partial\bar{X}} &= 0\\
\frac{\ell(\theta)}{\partial\phi} &= 0\\
\frac{\ell(\theta)}{\partial\sigma^2} &= 0
\end{cases}$$

Substitute the log-likelihood function in Equation 4 into the optimization condition above yields

$$\begin{aligned}
\frac{\ell(\theta)}{\partial\bar{X}} &= -\frac{(1-\phi^2)(\bar{X}-x_0)}{2\sigma^2}+\frac{1-\phi}{\sigma^2}\sum_{t=0}^{T-1}[x_{t+1}-\bar{X}(1-\phi)-\phi x_t]=0\\
\frac{\ell(\theta)}{\partial\phi} &= -\frac{\phi}{1-\phi^2}+\frac{\phi(x_0-\bar{X})^2}{\sigma^2}-\frac{1}{\sigma^2}\sum_{t=0}^{T-1}(\bar{X}-x_t)[x_{t+1}-\bar{X}(1-\phi)-\phi x_t]=0\\
\frac{\ell(\theta)}{\partial\sigma^2} &= -\frac{T+1}{2\sigma^2}+\frac{(1-\phi^2)(x_0-\bar{X})^2}{2\sigma^4}+\frac{1}{2\sigma^4}\sum_{t=0}^{T-1}[x_{t+1}-\bar{X}(1-\phi)-\phi x_t]^2=0
\end{aligned}$$

The exact maximum likelihood estimators requires to solve the above equations. In other words, the solutions of the above equations become the exact ML estimator. However, these equations are nonlinear and difficult to solve analytically. Numerical methods and algorithms would be used in this case, and in this paper I use R to obtain the

2. Merton Jump Diffusion Model Calibration

2.1. **Introduction.** A vast of literatures have extended the Black-Scholes model [3] in option pricing by making more reasonable assumptions on market factors, such as the distribution of the underlying. Merton's 1976 JFE article [4] was the first to explore jump diffusion models. The jump

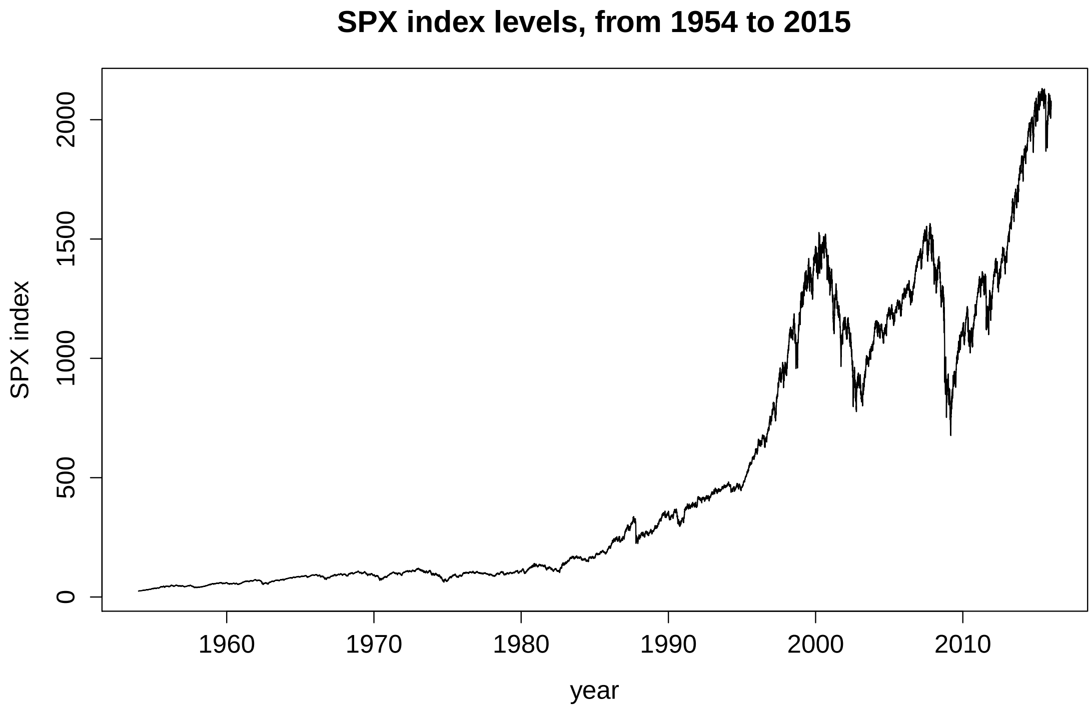

Figure 2 The SPX index levels. Time period is from 1954 to 2015.

diffusion model is designed to address the issue of fat tails, which is observed in dynamics many asset classes. However, when the underlying can jump to any level, the market is not complete, for the reason that there are more states than assets. Merton's innovative solution lies on extra randomness due to jumps can be diversified away.

The Merton jump diffusion (MJD) model was introduced to model the asset price S_t, mainly the equity (stock) prices. Figure 2 plots the SPX index levels from 1954 to 2015. The data source is Bloomberg. The long term return on the equity market is tremendous. The SPX index level begins at 25 in the beginning of 1954, and reached over 1500 points around 2000 and 2006. in 2015, the SPX index level exceeds 2000 points.

2.2. **Jump Diffusion Process.** Consider the dynamics of the model follows the Black-Scholes dynamics, which supposes that the behavior of the stock price, S_t, is determined by the stochastic process

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where W_t is the Wiener process. μ and σ are assumed to be constant and represent the drift and diffusion, respectively.

Now consider the asset price S_t with log-normal jumps $V_1, \ldots, V_j$ at random times $\tau_1, \ldots, \tau_j$ representing the moments of jumps in a Poisson process.

The MJD model assumes the S_t to follow the stochastic process:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + d\left(\sum_{j=0}^{N_t} (V_j - 1)\right) \tag{4}$$

The discontinuities of the price process are described by the Poisson process N_t with intensity λ (mean arrival rate of jumps per unit time) and jump V_j. And $\log V_j \sim N(\theta, \delta^2)$. The jump interprets by random variable V which transforms the price S_t to VS_t. The difference $V-1$ is the relative price change when a Poisson jump occurs.

Using Ito's Lemma, the strong solution of the Equation 4 is

$$S_t = S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right]\prod_{j=0}^{N_t} V_j$$

where S_0 is the initial value of the stock price. Let $Y_j = \log V_j$ and rewrite

$$X_t = \log\frac{S_t}{S_0} = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{j=0}^{N_t} Y_j \tag{5}$$

and we assume in this paper that the processes W_t, N_t and Y_j are independent. The parameter set of this MJD model is $\theta = (\mu, \sigma^2, \theta, \delta^2, \lambda)'$.

Discretize Equation 5 over time period $[t, t+1]$ and yield

$$X_{t+1} = X_t + \left(\mu - \frac{1}{2}\sigma^2\right) + \sigma\epsilon_{t+1} + \sum_{j=0}^{N_{t+1}-N_t} Y_j$$

where ϵ_{t+1} is the white noise with mean 0 and variance 1.

Hence the probability density of ΔX_t can be expressed as

$$f(x) = \sum_{n=0}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}\frac{1}{\sqrt{2\pi(\sigma^2+n\delta^2)}}\exp\left[-\frac{(x-\mu+\sigma^2/2-n\theta)^2}{2(\sigma^2+n\delta^2)}\right]$$

2.3. **Parameter Estimation with Method of Moments.** From the density function of ΔX_t (the return process), we can have

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=0}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}\int_{-\infty}^{+\infty}\frac{x}{\sqrt{2\pi(\sigma^2+n\delta^2)}}\exp\left[-\frac{(x-\mu+\sigma^2/2-n\theta)^2}{2(\sigma^2+n\delta^2)}\right]dx \\
\mathbb{E}[(X-\mathbb{E}(X))^k] &= \sum_{n=0}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}\int_{-\infty}^{+\infty}\frac{(x-\mu+\sigma^2/2-n\theta)^k}{\sqrt{2\pi(\sigma^2+n\delta^2)}}\exp\left[-\frac{(x-\mu+\sigma^2/2-n\theta)^2}{2(\sigma^2+n\delta^2)}\right]dx
\end{aligned}$$

for $k \geq 1$. Note that the improper integral is the central moment of order k of the normal random variable with $N(\mu - (\sigma^2/2) - n\theta, \sigma^2 + n\delta^2)$. Intuitively, the Poisson mixture of normals contributes to but only mean but also the variance (or volatility) term of the return process. Since the central moment with odd order becomes null for normal random variables, we use the central moments with even order to estimate the parameters. The central moments of even order is

$$\mathbb{E}[(X-\mathbb{E}(X))^{2k}] = \frac{(2k)!}{2^k k!}\sum_{n=0}^{+\infty} e^{-\lambda}\frac{\lambda^n}{n!}(\sigma^2+n\delta^2)^k$$

Follow Beckers [2] and Ball and Torous [1], set θ to be 0, as to assume symmetric jumps. Using the following (lowest available) four central moments to estimate the four parameters:

$$\begin{aligned}
\mathbb{E}(X) &= \mu - \frac{\sigma^2}{2} \\
\mathbb{E}[(X-\mathbb{E}(X))^2] &= \sigma^2 + \lambda\delta^2 \\
\mathbb{E}[(X-\mathbb{E}(X))^4] &= 3[(\sigma^2+\lambda\delta^2)^2 + \lambda\delta^4] \\
\mathbb{E}[(X-\mathbb{E}(X))^6] &= 15[(\sigma^2+\lambda\delta^2)^3 + 3\lambda\delta^4(\sigma^2+\lambda\delta^2) + \lambda\delta^6]
\end{aligned}$$

Then optimize the following expression to get parameter estimations:

$$\underset{\theta}{\operatorname{argmin}}\, g'(\theta) g(\theta) \tag{6}$$

where $g(\theta)$ is the difference between theoretical moments and moments from market data. To be specific,

$$g(\theta) = \begin{pmatrix} \frac{1}{N}\sum_{i=0}^{N} x_i - \mathbb{E}(X) \\ \frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{X})^2 - \mathbb{E}[(X - \mathbb{E}(X))^2] \\ \frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{X})^4 - \mathbb{E}[(X - \mathbb{E}(X))^4] \\ \frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{X})^6 - \mathbb{E}[(X - \mathbb{E}(X))^6] \end{pmatrix}$$

where x_i is the ith observation from the market data. N is the total sample size. $\bar{X}$ is the sample mean. The optimization assumes equal weights among moments.

2.4. **Calibration Results.** Optimizing (minimizing) the objective function shown in 6 with help of statistical package yields the calibrated parameters as listed in Table 1.

Table 1 Table Caption

Parameters	Calibrated Value
μ	0.0003547
σ^2	0.00004836
θ	0
δ^2	0.03310
λ	0.001458

References

[1] Ball C. A., Torous W. N. On jumps in common stock prices and their impact on call option pricing. *Journal of Finance*, 40(1):155–173, 1985.

[2] Becker S. A note on estimating the parameters of the diffusion-jump model of stock returns. *Journal of Financial and Quantitative Analysis*, 16(1):127–140, 1981.

[3] Black F., Scholes M. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[4] Merton R. C. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[5] Vasicek O. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

Appendix: Code for Vasicek model ML estimation

```r
setwd('C:\\Users\\ranzhao\\Documents\\Empirical Asset Pricing\\Assignment 1')
setwd('D:\\PhD FE\\Empirical-Asset-Pricing\\Assignment 1')

# Data loading
require(ggplot2)
spx_index_values = read.csv('spx_index_values.csv', header = TRUE)
t_bill_3M_values = read.csv('TB3MS.csv', header = TRUE)
plot(as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y"), t_bill_3M_values$TB3MS,
     type='l',
     main='3-Month Treasury Bill, from 1954 to 2015',
     xlab='year', ylab='rate (in percentage)')
# add the moving average of the rates to the plot, ggplot?


# Data segments
ir_full = t_bill_3M_values
ir_1954_1975 = t_bill_3M_values[
  as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") >= as.Date('1954-01-01') &
  as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") <= as.Date('1975-12-31'), ]
ir_1976_1981 = t_bill_3M_values[
  as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") >= as.Date('1976-01-01') &
    as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") <= as.Date('1981-12-31'),
        ]
ir_1982_2005 = t_bill_3M_values[
  as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") >= as.Date('1982-01-01') &
    as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") <= as.Date('2005-12-31'),
        ]
ir_2006_2015 = t_bill_3M_values[
  as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") >= as.Date('2006-01-01') &
    as.Date(as.character(t_bill_3M_values$DATE), "%m/%d/%Y") <= as.Date('2015-12-31'),
        ]
```

assignment1–1.R

Appendix: Code for Merton jump diffusion model

```r
setwd('C:\\Users\\ranzhao\\Documents\\Empirical-Asset-Pricing\\Assignment 1')
setwd('D:\\PhD FE\\Empirical-Asset-Pricing\\Assignment 1')


# Data loading
#require(ggplot2)
spx_index_values = read.csv('spx_index_values.csv', header = TRUE)
plot(as.Date(as.character(spx_index_values$Date), "%m/%d/%Y"), spx_index_values$SPX.Index, type='l',
     main='SPX index levels, from 1954 to 2015',
     xlab='year', ylab='SPX index')

# calculate the return series
spx_index_values$Return = rep(0, dim(spx_index_values)[1])
spx_index_values$Return[2:length(spx_index_values$Return)] =
  spx_index_values$SPX.Index[2:length(spx_index_values$SPX.Index)] /
  spx_index_values$SPX.Index[1:(length(spx_index_values$SPX.Index)-1)] - 1

# calculation empirical moments
return.data = spx_index_values$Return
n.length = length(return.data)
emp.moment.1 = mean(return.data)
emp.moment.2 = 1/(n.length-1)*sum((return.data - emp.moment.1)^2)
emp.moment.4 = 1/(n.length-1)*sum((return.data - emp.moment.1)^4)
emp.moment.6 = 1/(n.length-1)*sum((return.data - emp.moment.1)^6)

# optimization function
moment.diff = function(data.input){
  mu = data.input[1]
  sigma.square = data.input[2]
  lambda = data.input[3]
  delta.square = data.input[4]

  theo.moment.1 = mu - sigma.square / 2
  theo.moment.2 = sigma.square + lambda * delta.square
  theo.moment.4 = 3 * ((sigma.square+lambda*delta.square)^2 + lambda*delta.square^2)
  theo.moment.6 = 15 * ((sigma.square+lambda*delta.square)^2 + 3*lambda*delta.square*(sigma.square+lambda*delta.square)+lambda*delta.square^3)

  least.square.obj = (theo.moment.1 - emp.moment.1)^2 + (theo.moment.2 - emp.moment.2)^2 + (theo.moment.4 - emp.moment.4)^2 + (theo.moment.6 - emp.moment.6)^2
}

# parameter calibration
output = optim(c(0, 0.006, 0.03, 0.0005), moment.diff)
output$par
```

assignment1-2.R