PyData Delhi 2019

Quantitative Finance with R

Harshit Joshi
duskybomb | harshitjoshi.in
Vedant Bonde
vedantbonde

Cluster Innovation Centre
University of Delhi

Modules

Introduction to R and Finance

The Random Walk Hypothesis and Geometric Brownian Motion

Efficient Frontier Theory

Capital Asset Pricing Model

Playing with Intraday data

Link to repository

https://github.com/duskybomb/quantitative-finance-talk


dont worry about it if you don't
understand

Introduction to R and Finance

Compute Log-return

log-return is the consecutive differences of prices in log-scale, i.e.,

$$\begin{aligned} r_t &= \log(P_t) - \log(P_{t-1}) \\ &= \log\left(\frac{P_t}{P_{t-1}}\right) \end{aligned}$$

where P_t is the price of a stock (or value of an index) at t time point

Compute simple return

Simple return can be calculated as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$
$$= \frac{P_t}{P_{t-1}} - 1$$
$$= e^{r_t} - 1,$$

Where r_t is the log-return.

K-period Simple Return

$$R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}} = \frac{P_t}{P_{t-k}} - 1$$

$$\begin{aligned} 1 + R_t(k) &= \frac{P_t}{P_{t-k}} = \left(\frac{P_t}{P_{t-1}}\right)\left(\frac{P_{t-1}}{P_{t-2}}\right)\ldots\left(\frac{P_{t-k+1}}{P_{t-k}}\right) \\ &= (1 + R_t)(1 + R_{t-1})\ldots(1 + R_{t-k+1}) \end{aligned}$$

where P_t is the price of a stock (or value of an index) at t time point

K-period log Return

$$\begin{aligned} r_t(k) &= \log\{1 + R_t(k)\} \\ &= \log\{(1 + R_t)(1 + R_{t-1})\ldots(1 + R_{t-k+1})\} \\ &= \log(1 + R_t) + \log(1 + R_{t-1}) + \ldots + \log(1 + R_{t-k+1})\} \\ &= r_t + r_{t-1} + \ldots + r_{t-k+1} \end{aligned}$$

- $1 + R_t(k)$ is k-period simple return
- k-period log-return is sum of k-single period log-return

How to compute 30-days Volatility

Suppose $r_t, r_{t-1}, \ldots, r_{t-k+1}$ are k-single period log-return of an asset, where

$$\mathbb{E}(r_{t-i+1}) = \mu \quad \forall i = 1, 2, \ldots, k$$
$$\mathbb{V}ar(r_{t-i+1}) = \sigma^2 \quad \forall i$$
$$\mathbb{C}ov(r_{t-i+1}, r_{t-j+1}) = 0 \quad \forall i \neq j = 1, 2..., k.$$

That is a covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{pmatrix}_{k \times k}$$

The k-period return can be presented in matrix notation as

$$\begin{aligned} r_t(k) &= r_t + r_{t-1} + \ldots + r_{t-k+1} \\ &= c^T \mathbf{r}, \end{aligned}$$

Where $c^T = (1, 1, \ldots, 1)_k$ is an unit vector of order k and $\mathbf{r} = \{r_t, r_{t-1}, \ldots, r_{t-k+1}\}$

The mean and variance of k-**period return** are

$$\begin{aligned} \mathbb{E}(r_t(k)) &= c^T \mu = k\mu, \\ \mathbb{V}ar(r_t(k)) &= c^T \Sigma c = k\sigma^2 \end{aligned}$$

Therefore k-**period volatility** is $\sqrt{k}\sigma$

Random Walk Hypothesis

Random walk theory suggests that changes in stock prices have the same distribution and are independent of each other. Therefore, it assumes the past movement or trend of a stock price or market cannot be used to predict its future movement.

Example: Random Walk with Fixed Moves

- Suppose price of a stock move up by 5 paisa with probability 0.5 or move down by 5 paisa with probability 0.5 every seconds.
- If the price of the stock is Re 1/-; then what will be the price of the stock after 21600 seconds.
- The model $P_t = P_{t-1} \pm M_t$ where $M_t = 5$
- Here the price of stock is the same whether it rises or falls and therefore is a fixed move random walk.
- This model is good candidate to model the stock price movement.
- However, it cannot take care of the **limited liability** feature of the stock market.
- A more general model with **random moves** can be obtained with changing the value $M_t \sim Poisson(\lambda = 5)$

Example: Random Walk with Random Return

- We use the formula of the simple Return: $P_t = P_{t-1}(1 + R_t)$
- Suppose $r_t \sim N(\mu = 0, \sigma = 0.01)$ on every second

Example: Random Walk with Random log-Return

- Now suppose we use the formula of the log-Return: $P_t = P_{t-1}e^{r_t}$
- Suppose $r_t \sim N(\mu = 0, \sigma = 0.01)$ on every second

Random Walk Model

- Suppose $r_1, r_2, \ldots$ be i.i.d with mean **μ** and standard deviation **σ**
- Let P_0 be an arbitrary starting point and

$$P_t = P_0 + r_1 + r_2 + \ldots + r_t, \quad t \geq 1$$

- The process $P_0, P_1, P_2, \ldots$ is known as **random walk** and $r_1, r_2, \ldots$ are **corresponding steps of that random walk.**
- The conditional expectation and variance of P_t given P_0 is

$$E(P_t|P_0) = P_0 + \mu t \text{ and } Var(P_t|P_0) = \sigma^2 t$$

- The parameter **μ** is the drift and set an overall trend of the random walk and parameter **σ** is the volatility and controls fluctuations around $P_0 + \mu t$

Stationary Process

- In probability, a stochastic process is known as stationary if the joint probability distribution is independent of time.
- Parameters such as mean, volatility, correlation etc., do not change over time
- A strictly stationary time series is one for which the probabilistic behavior of every collection of values $\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}$ is identical to that of the time shifted set $\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}$
- A continuous time random process P_t is stationary , if it has the following restrictions on its mean function,

$$\mathbb{E}(P_t) = \mu(t) = \mu(t+h) = \mu \quad \forall t, h \in \mathbb{R},$$

and autocovariance function

$$\mathbb{C}ov_P(t_1, t_2) = \mathbb{E}[\{P(t_1) - \mu(t_1)\}\{P(t_2) - \mu(t_2)\}] = \mathbb{C}ov_P(t_1 - t_2).$$

Unit Root of Random Walk

- The random walk is said to have unit root. To understand what this means, you should consider the **AR(1)** model (Auto-Regressive model with lag 1),

  $p_t = \phi p_{t-1} + r_t$ where $\phi = 1$and $p_t = \log(P_t)$ is the price of the asset in log-scale.

- The generic AR(1) model can be presented as

$$
\begin{aligned}
p_t &= \phi p_{t-1} + r_t \\
&= \phi(\phi p_{t-2} + r_{t-1}) + r_t \\
&\vdots \\
&= \phi^k p_{t-k} + \phi^{k-1} r_{t-(k-1)} + \ldots + \phi r_{t-1} + r_t
\end{aligned}
$$

Some properties of Autoregressive models

- If $\phi = 1$ then the process is non-stationary. Because $\sum_{i=0}^{k-1} \phi^{i-1} r_{t-(i-1)}$ accumulates the information over time. **Hence a random walk is a non-stationary process.**
- However $|\phi| < 1$, i.e., $-1 < \phi < 1$ implies that the process is stationary.
- If $\phi = 0$ that means the process is stationary and p_t and p_{t-1} are independent $\forall t$

Dickey-Fuller test for Stationarity in a Time Series

The Dickey-Fuller test (1979) compares the null hypothesis

$$H_0 : p_t = p_{t-1} + r_t$$

i.e., that the series is a random walk without drift, where r_t is a white noise with mean **0** and variance σ^2.

The alternative hypothesis is

$$H_1 : p_t = \mu + \phi p_{t-1} + r_t$$

where μ and ϕ are constant with $|\phi| < 1$.

- According to H_1, the process is stationary AR(1) with mean $\frac{\mu}{1-\phi}$.

Geometric Brownian Motion

For the risk-free assets the price grows at risk-free rate $P_T = P_0 e^{rT}$.

To introduce the trend and volatility parameters, we consider the following model:

$P_T = P_0 e^{\mu T + \lambda_T Z_1}$, where $Z_1 \sim N(0,1)$ and μ and λ_T are trend and volatility parameters.

Using Ito formula to add a damping effect to nullify the effect of random term of growth, we get $P_T = P_0 e^{\mu T} e^{\lambda_T Z_1 - \lambda_T^2/2}$.

Finally, we get the model for the spot price at time T:

$P_T = P_0 e^{(\mu - \sigma^2/2)T + \sigma W_T}$, where $W_T \sim N(0,T)$ where **μ** and **σ** are known as drift and volatility parameters respectively.

Geometric Brownian Motion

Let $P_i (i = 0, 1, \ldots, n)$ be spot prices observed over a consecutive time period of **Δt** each. Then, log return is: $r_i = \log(P_{i+1}) - \log(P_i)$.

Assuming returns follow geometric brownian motion model, according to Geometric Brownian Motion model $r_i = (\mu - \sigma^2/2)\Delta t + \sigma W_{\Delta t}$, where $W_{\Delta t} \sim N(0, \Delta t)$. Therefore, $r_i \sim N((\mu - \sigma^2/2)\Delta t, \sigma\sqrt{\Delta t})$. Using the formula of sample mean and variance, we can estimate **μ** and **σ** as:

$$\hat{\mu} \approx \frac{\bar{r} + s^2/2}{\Delta t},$$
$$\hat{\sigma} \approx \frac{s}{\Delta t}.$$

where $\bar{r}$ is sample mean and s^2 is sample variance of the returns.

Efficient Frontier

Two portfolio assets

$$\mathbb{E}(R_P) = \omega_1 \mathbb{E}(R_1) + \omega_2 \mathbb{E}(R_2),$$

- $\mathbb{E}(R_P)$ is the expected return of the portfolio
- ω_i is the weight of asset
- $\mathbb{E}(R_i)$ is the expected return of asset

Portfolio variance

$$\sigma_P^2 = \omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + 2\omega_1\omega_2\sigma_{1,2}$$

- σ_P^2 is the portfolio variance
- σ_i^2 is the variance of assets
- $\sigma_{1,2}$ is the covariance between asset 1 and 2
- $\omega_1 + \omega_2 = 1$

Portfolio volatility or standard deviation

$$\sigma_P = \left\{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + 2\omega_1\omega_2\sigma_{1,2}\right\}^{\frac{1}{2}}.$$

Example

Two securities, say X and Y

$$\mathbb{E}(R_X) = 5\% \text{ and } \mathbb{E}(R_Y) = 4\%$$
$$\sigma_X^2 = 9\%, \sigma_Y^2 = 6\%$$
$$\sigma_{XY} = 3\%$$

The following table presents the portfolio return and volatility for five different portfolio combinations.

	X1	X2	X3	X4	X5	X6
W_x	100%	80%	60%	40%	20%	0%
W_y	0%	20%	40%	60%	80%	100%
Expected Return	5%	4.8%	4.6%	4.4%	4.2%	4%
Volatility	9%	6.96%	5.64%	5.04%	5.16%	6%

Portfolio of N assets

Expected portfolio return

$$\mathbb{E}(R_P) = \omega^T \mu$$

where $\omega^T = \{\omega_1, \omega_2, \ldots, \omega_N\}$, $\mu = \{\mu_1, \mu_2, \ldots, \mu_N\}$ $\mu_i = \mathbb{E}(R_i)$ $i = 1, 2..., N$

Portfolio volatility is

$$\sigma_P = \sqrt{\omega^T \Sigma \omega},$$

where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_1^2 \end{bmatrix}$$

is the portfolio covariance matrix.

Portfolio Optimization

Markowitz's Portfolio Optimization (1952) minimizes the portfolio variance for a given level of expected return, i.e.,

$$\mathbf{min}_{\omega}\, \omega^T \Sigma \omega$$

such that

$$\omega^T \mu = \mu_0\,,$$

Note that you need to provide Σ

However, in reality we do not know what is true Σ

So we should estimate the Σ

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{r})(r_i - \bar{r})^T.$$

Capital Asset Pricing Model

In order to measure the performance of a particular security or fund against that of benchmark portfolio, the `security characteristic lines' (SCL) are very useful.

The SCL equation is

$$\mathbb{E}(R_i) - R_F = \alpha_i + \beta_i(\mathbb{E}(R_i) - R_F)$$

Where α_i is active return.

$\beta_i(\mathbb{E}(R_i) - R_F)$ is a nondiversifiable or systematic risk

Suppose there are N risky assets with return $R_i, R_2, \ldots, R_N$ with weights $\omega_1, \omega_2, \ldots, \omega_N$

The return of the portfolio is

$$R_P = \omega_1 R_1 + \omega_2 R_2 + \ldots + \omega_N R_N .$$

Let R_M be the return of the market index. According to Security Market Line,

$$\mathbb{E}(R_i) = R_F + \alpha_i + \beta_i \mathbb{E}(R_M - R_F),$$

$$\mathbb{E}(R_P) = R_F + \alpha_P + \beta_P \mathbb{E}(R_M - R_F),$$

Factor Model

Factor models generalize the CAPM by allowing more factors than simply the market risk and the unexplained unique risk of each asset.

The model is defined as,

$$R_i - R_F = \beta_{0i} + \beta_{1i} x_1 + \ldots + \beta_{pi} x_p + \epsilon_i ,$$

where $x_1, x_2, \ldots, x_p$ are the p factors.

One widely used multi-factor model is the Fama and French three-factor model. The Fama and French model has three factors: **size of firms**, **book-to-market values** and **excess return on the market**

Intraday Data!

Thank You

Harshit Joshi [harshit.joshi@outlook.com]

Vedant Bonde [vedantbonde19@gmail.com]