Lesson 5
Random Walk and Variance Ratio Test

Christopher Ting

Hiroshima University

✉: cting@hiroshima-u.ac.jp

☺: https://cting.neocities.org/

☎: +81 082-424-6451

A1棟 131-1

Broad Lesson Plan

Four Big Questions

(1) **Are stock returns predictable?**
Common sense answer: No. Otherwise, everyone will make money without risk.

(2) **Are asset returns 100% random?**
Investment firms: No. Otherwise, we will have no client!

(3) **Will stock market indexes move up over 30 years?**
Common sense answer: Yes. Otherwise why would anyone invest in risky stocks?

(4) **Can mutual funds beat their benchmark stock market indexes?**
Mutual fund managers: Yes. It's our job to beat the market.
Academic professors: No. On average, mutual funds are equally likely to beat the market; after costs, the alpha is negative.

Types of Market Efficiency

- **weak form efficiency**: information in past returns or prices
- **semi-strong form efficiency**: all public information
- **strong form efficiency**: all private information (fundamental research on companies, economies, etc)

A Paradox

If the market is (strong form) **efficient**, no one will bother to gather **private information** because they can't beat the market. But if no one gathers private information, then the market cannot possibly reflect any private information!

John Y. Campbell

On the Impossibility of Informationally Efficient Markets

Strong form efficient market hypothesis is theoretically implausible, unless **information** is free. If information is costly to gather (i.e., costly to do fundamental research on companies), then investors will only gather information if the benefit outweighs the cost.

Grossman and Stiglitz

Implications in Practice

- Reality: Research is still ongoing to obtain and analyze new **private information**.
- Consequence: Research results will push prices to some extent, i.e., prices will reflect some but not all of the **new information**.
- In **equilibrium**, the **marginal benefit** of obtaining new information will be equal to the **marginal cost** of doing research.
- What are the practical implications?
  - You can beat the market, as long as your cost of information acquisition and analysis is less than other people's.
  - If you have no particular advantage, then you should hold primarily low-cost **ETF**s (and **free ride** on the information obtained by others), possibly with a tilt towards value.

Martingale

Definition 1

Suppose $\{P_t\}_{t=1}^{T}$ is a **stochastic process**. It is said to be a **martingale** if

$$\mathbb{E}(P_t | P_{t-1}, P_{t-2}, \ldots) = P_{t-1}.$$

Equivalently, since P_{t-1} is a known constant at time $t-1$,

$$\mathbb{E}(P_t - P_{t-1} | P_{t-1}, P_{t-2}, \ldots) = 0.$$

Implications

A. Given the information $\{P_{t-i}\}_{i=1}^{t}$ up to time $t-1$, the **best forecast** of P_t is P_{t-1}.
B. Obviously, in most cases, $P_t - P_{t-1} \neq 0$ and in that sense, the forecast is wrong and thus P_t and return are **unpredictable**.

Fundamental Value

Definition 2

The **conditional expectation** of a **random variable** X is a probability-weighted mean of the different possible values of X

$$\mathbb{E}\left(X|\mathcal{I}\right) = \sum_{i=1}^{n} \pi_i X_i,$$

where the probability π_i for each i is conditional on the **information set** $\mathcal{I}$.

Remark

The expected value conditional on the information set $\mathcal{I}$ is the **rational expectation** of an asset's **payoff**, i.e., $\mathbb{E}\left(X|\mathcal{I}\right)$ is the **fundamental value** of the asset.

A Very Important Question

- If returns are **unpredictable**, does it mean that the market price is equal to the fundamental value?

Law of Iterated Expectations (LIE)

Proposition 1

For two random variables X and Y, the **law of iterated expectations** says that

$$\mathbb{E}_Y\Big(\mathbb{E}_{X|Y}\left(X \mid Y\right)\Big) = \mathbb{E}\left(X\right).$$

Example

Suppose a portfolio has two stocks, A and B. Stock A (B) has a mean return of 6% (15%) per annum. The investment in Stock A (B) is 2 (4) billion dollars. What is the portfolio's expected return?

`Answer`: Based on the investment amounts, the weight of A is 1/3 and that of B is 2/3. The return r_P of the portfolio is

$$\begin{aligned}\mathbb{E}(r_P) &= \mathbb{E}\left(r_P|A\right)\mathbb{P}(A) + \mathbb{E}\left(r_P|B\right)\mathbb{P}(B)\\ &= 6\% \times \frac{1}{3} + 15\% \times \frac{2}{3} = 12\%.\end{aligned}$$

Proof of LIE

$$
\begin{aligned}
\mathbb{E}_Y\Big(\mathbb{E}_{X|Y}\left(X \mid Y\right)\Big) &= \mathbb{E}_Y\left[\sum_x x\cdot\mathbb{P}(X=x\mid Y)\right] \\
&= \sum_y\left[\sum_x x\cdot\mathbb{P}(X=x\mid Y=y)\right]\cdot\mathbb{P}(Y=y) \\
&= \sum_y\sum_x x\cdot\mathbb{P}(X=x\mid Y=y)\cdot\mathbb{P}(Y=y) \\
&= \sum_x x\sum_y \mathbb{P}(X=x\mid Y=y)\cdot\mathbb{P}(Y=y) \\
&= \sum_x x\sum_y \mathbb{P}(X=x, Y=y) = \sum_x x\cdot\mathbb{P}(X=x) \\
&= \mathbb{E}(X).
\end{aligned}
$$

General LIE

Proposition 2

For a **probability space** $(\Omega, \mathcal{F}_\infty, \mathbb{P})$, suppose the σ **algebras** are such that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_\infty$. A **random variable** X is defined on this space. The **law of iterated expectations** states that

$$\mathbb{E}\left(\mathbb{E}\left(X \middle| \mathcal{F}_2\right) \middle| \mathcal{F}_1\right) = \mathbb{E}\left(X \middle| \mathcal{F}_1\right).$$

- Note that

$$\int_{F_1} \mathbb{E}\left(\mathbb{E}\left(X \middle| \mathcal{F}_2\right) \middle| \mathcal{F}_1\right) \mathrm{d}\mathbb{P} = \int_{F_1} \mathbb{E}\left(X \middle| \mathcal{F}_2\right) \mathrm{d}\mathbb{P}$$
$$= \int_{F_1} X \, \mathrm{d}\mathbb{P} \text{ holds for all } F_1 \in \mathcal{F}_1 \subseteq \mathcal{F}_2$$

- In the special case of $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_2 = \sigma(Y)$, the σ algebra generated by random variable Y, we have

$$\mathbb{E}\big(\mathbb{E}(X|Y)\big) = \mathbb{E}(X).$$

LIE over Time

- Suppose time t **information** is less than time $t+1$ information, which is less than time $t+2$ information.
- Let

$$\mathbb{E}_t(X) := \mathbb{E}\left(X \middle| \text{Time } t \text{ information}\right)$$

- By the **law of Iterated Expectations**,

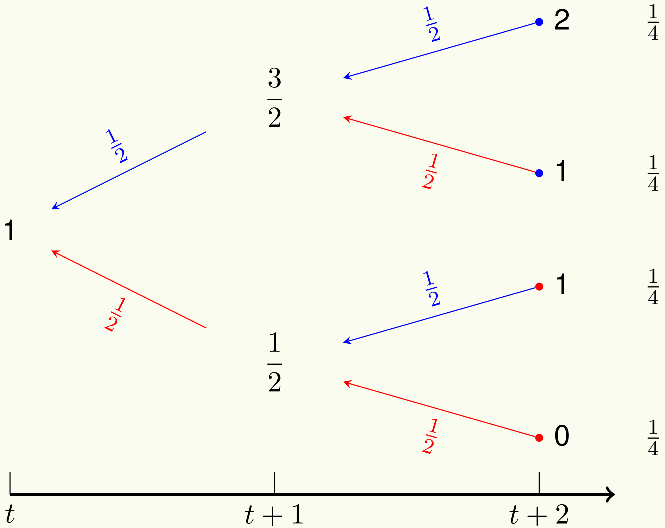
$$\mathbb{E}_t(X) = \mathbb{E}_t\big(\mathbb{E}_{t+1}(X)\big)$$

- Alternatively,

$$\mathbb{E}_t\big(\mathbb{E}_{t+1}(X) - X\big) = 0. \tag{1}$$

- The expectation at t of the **mistake** $\mathbb{E}_{t+1}(X) - X$ you are going to make at $t+1$ is zero.

Example: Tossing Two Fair Coins



Numerical Illustration

- Let X be the number of heads. Then, the direct method obtains

$$\mathbb{E}_t(X) = \frac{1}{4} \times 2 + \frac{1}{2} \times 1 + \frac{1}{4} \times 0 = 1$$

- The LIE's method obtains

$$\mathbb{E}_{t+1}(X) = \frac{3}{2} \quad \text{or} \quad \frac{1}{2}$$

$$\mathbb{E}_t\left(\mathbb{E}_{t+1}(X)\right) = \frac{1}{2} \times \frac{3}{2} + \frac{1}{2} \times \frac{1}{2} = 1.$$

- Hence, we have illustrated that

$$\mathbb{E}_t(X) = \mathbb{E}_t\Big(\mathbb{E}_{t+1}(X)\Big).$$

Verification of Alternative Formalism (1)

- With respect to t, the future forecast is $\mathbb{E}_{t+1}(X)$ for X's value at $t+2$.
- The forecast made at $t+1$ is either 1/2 or 3/2.
- In the case of 1/2, the correct value is either 0 or 1 at $t+2$ with equal probability. So the error is either (1/2 - 0) or (1/2 - 1).
- In the case of 3/2, the correct value is either 1 or 2 at $t+2$ with equal probability. So the error is either (3/2 - 1) or (3/2 - 2).
- Based on the probability measure at t,

$$\mathbb{E}_t\big(\mathbb{E}_{t+1}(X) - X\big) = \left(\frac{1}{4} \times \left(\frac{1}{2} - 0\right)\right) + \left(\frac{1}{4} \times \left(\frac{1}{2} - 1\right)\right) + \left(\frac{1}{4} \times \left(\frac{3}{2} - 1\right)\right) + \left(\frac{1}{4} \times \left(\frac{3}{2} - 2\right)\right) = 0.$$

Fundamental Value and Unpredictable Returns

- Suppose prices are based on the **fundamental value** of the future **payoff** X, i.e., $P_t = \mathbb{E}_t(X)$ and $P_{t+1} = \mathbb{E}_{t+1}(X)$

- By the LIE, since $\mathbb{E}_t(X) = \mathbb{E}_t\big(\mathbb{E}_{t+1}(X)\big)$, the price change is

$$
\begin{aligned}
\Delta P_{t+1} := P_{t+1} - P_t \\
&= P_{t+1} - \mathbb{E}_t(X) \\
&= P_{t+1} - \mathbb{E}_t\left(\mathbb{E}_{t+1}(X)\right) \\
&= P_{t+1} - \mathbb{E}_t(P_{t+1}).
\end{aligned}
$$

- Therefore, **price change** (and hence return) is equal to the “**mistake**” or error $P_{t+1} - \mathbb{E}_t(P_{t+1})$, which is **unpredictable**.

Fundamental Value and Rational Prediction

- Now, by definition,

$$\begin{aligned}\Delta P_{t+1} &= \mathbb{E}_{t+1}(X) - \mathbb{E}_t(X).\\ &= \text{future rational expectation} - \text{current } \textbf{rational expectation}\end{aligned}$$

- Implications
  - Price change or return is equally likely to be positive and negative, just like tossing a fair coin.
  - Therefore, if you are rational, you cannot predict how you will change your mind in the future!

- Insight

  LIE connects **fundamental value** with **unpredictable** returns.

Random Walk 1

Definition 3

Random walk $\{P_t\}_{t=1}^{T}$, as a **stochastic process**, is expressed as

$$P_t = \mu + P_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{IID}(0, \sigma^2),$$

where

μ	rate of drift
ϵ_t	random "noise"
σ^2	variance
$\text{IID}(0, \sigma^2)$	**i**ndependently and **i**dentically **d**istributed with mean 0 and variance σ^2.

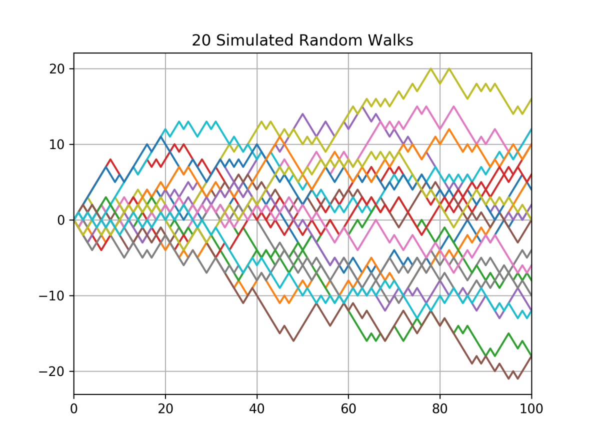
✷ Note that no explicit distribution is specified for ϵ_t.

Properties of Random Walk 1

- Let P_0 be the initial position of the random walk at time 0. At time t,

$$P_t = P_0 + \mu\, t + \sum_{s=0}^{t-1} \epsilon_{t-s}. \qquad (2)$$

- Suppose ϵ_t is either 1 or -1, with $\mathbb{P}(\epsilon_t = 1) = 0.5 = \mathbb{P}(\epsilon_t = -1)$.



```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(12)
n, paths = 100, 20
p, r = np.zeros((n+1,paths),dtype=float), np.zeros((n+1,paths),dtype=float)
for z in range(paths):
    r[:, z] = np.random.random(n+1)

up_probability, t = 0.5, range(n+1)
for z in range(paths):
    for i in t[1:n+1]:
        if r[i, z] > up_probability:
            p[i,z] = p[i-1,z]+1
        else:
            p[i,z] = p[i-1,z]-1
    plt.plot(t, p[:,z])

plt.grid()
title = str(paths) + ' Simulated Random Walks'
plt.title(title)
plt.autoscale(enable=True, axis='x', tight=True)
plt.savefig('rw.png', format = 'png', dpi=3900)
plt.show()
```

Concept Checker

- For the **random walk** $P_t = \mu + P_{t-1} + \epsilon_t$ to be a martingale, you need

A. $\mu = 0$

B. $\mathbb{E}(\epsilon_t) = 0$

C. $P_0 = 0$

D. None of the above, because random walk cannot be a **martingale**

Linear Scaling Law

- Rewriting (2), we obtain

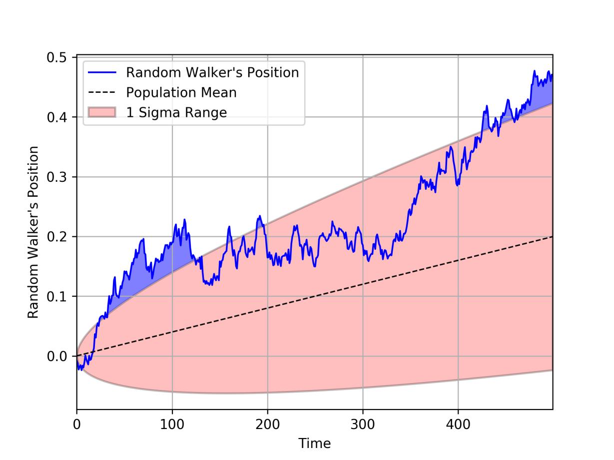
$$P_t - P_0 = \mu t + \sum_{s=0}^{t-1} \epsilon_{t-s}.$$

- It is easy to find that $\mathbb{V}\left(\epsilon_t\right) = 1$ for the assumption in Slide 20.
- Therefore, since $\mathbb{C}\left(\epsilon_t, \epsilon_s\right) = 0$ for $t \neq s$,

$$\mathbb{V}\left(P_t - P_0\right) = \mathbb{V}\left(\sum_{s=0}^{t-1} \epsilon_{t-s}\right) = \sum_{s=0}^{t-1} \mathbb{V}\left(\epsilon_{t-s}\right) = t$$

- The variance from the origin P_0 scales linearly with time t as it increases.

Simulation of Linear Scaling Law



More General Random Walks

Random Walk 2

The random noise ϵ_t is still independent but not identically distributed. In particular, the variance σ_t^2 of the noise is different at different t.

Random Walk 3

The **random noise** ϵ_t is still identically distributed but the independence assumption is relaxed to zero covariance, i.e.,

$$\mathbb{C}\left(\epsilon_t, \epsilon_{t-s}\right) = 0.$$

for $s \neq 0$. In other words, the random noise does not have any linear correlation structure. However, nonlinear correlation, for example, $\mathbb{C}\left(\epsilon_t^2, \epsilon_{t-s}^2\right) \neq 0$ is not assumed.

Random Walk 4

The most general: neither independent nor identically distributed.

A Problem with Arithmetic Random Walk

- The price of an asset such as stock is positive. But the arithmetic **random walk** 1 defined in Slide 19 can lead to a negative price if ϵ_t is very negative.
- Therefore, consider a random variable $\xi_{t+1} > 0$ such that $\mathbb{E}_t\big(\xi_{t+1}\big) = 1$ and

$$P_t = P_{t-1}\xi_t.$$

- Then

$$\mathbb{E}_{t-1}\big(P_t\big) = P_{t-1}.$$

- **In this model, log return r_t is noise!**

$$\begin{aligned} r_t &= \ln P_t - \ln P_{t-1} \\ &= \ln P_{t-1} + \ln \xi_t - \ln P_{t-1} \\ &= \ln \xi_t =: u_t \end{aligned}$$

Geometric Random Walk

Definition 4
Geometric random walk with **drift** μ is defined as

$$\ln P_t = \mu + \ln P_{t-1} + u_t,$$

where u_t is not correlated over time and identically distributed with mean 0 and $\mathbb{V}(u_t) = \sigma^2$.

- In other word, two assumptions are made or the log return r_t
  - **Zero covariance**: $\mathbb{C}(r_s, r_t) = 0$ for any $s \neq t$
  - **Homoskedasticity**: $\mathbb{V}(r_t) = \sigma^2$

Implications of Geometric Random Walk

- Definition 4 implies that

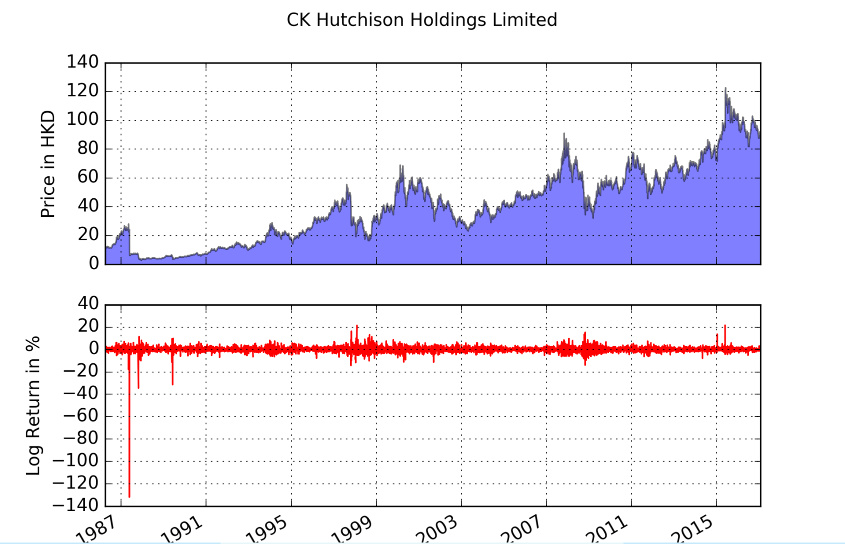
$$r_t := \ln P_t - \ln P_{t-1} = \mu + u_t. \tag{3}$$

- **Homoskedasticity** $\mathbb{V}(r_t) = \sigma^2$ implies that

$$u_t = \sigma Z_t,$$

where Z_t is a **random variable** with mean 0 and variance 1.

Real-World Example



Variance of a Sum of q Daily Log Returns

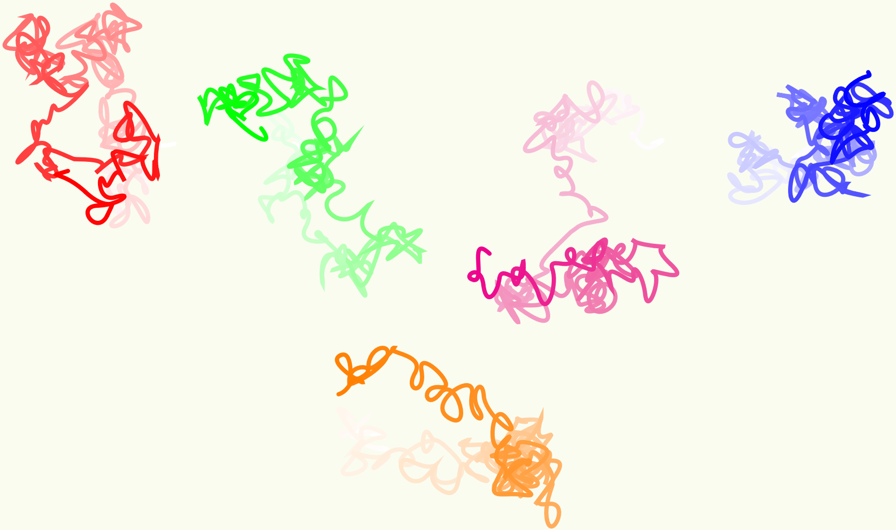
- When the daily log return r_t is treated as a random variable, the variance of a sum of q daily log returns is

$$\mathbb{V}\left(\sum_{t=1}^{q} r_t\right) = \sum_{t=1}^{q} \mathbb{V}(r_t) + 2\sum_{t=1}^{q}\sum_{s<t} \mathbb{C}(r_s, r_t).$$

- Under the two assumptions in Slide 27,

$$\mathbb{V}\big(r_t(q)\big) := \mathbb{V}\left(\sum_{t=1}^{q} r_t\right) = q\sigma^2.$$

Random Walk $\longrightarrow$ Diffusion



Starting Point of Chi-Square Random Variable

- Suppose Z is a **standard normal random variable**, i.e., $Z \sim N(0,1)$. Then Z^2 is a **chi-square random variable** with one **degree of freedom**.
- The expected value of Z^2 is, since $Z \sim N(0,1)$,

$$\mathbb{E}\left(Z^2\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz = 1.$$

- By the definition of variance,

$$\mathbb{V}\left(Z^2\right) = \mathbb{E}(Z^4) - \left(\mathbb{E}(Z^2)\right)^2.$$

- We need to show that

$$\mathbb{E}\left(Z^4\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^4 e^{-\frac{z^2}{2}} dz = 3.$$

Kurtosis

$$\int_{-\infty}^{\infty} z^4 e^{-z^2/2}\, dz = -\int_{-\infty}^{\infty} z^3\, de^{-z^2/2}$$
$$= -\left. z^3 e^{-z^2/2}\right|_{-\infty}^{+\infty} + 3\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = -3\int_{-\infty}^{\infty} z\, de^{-z^2/2}$$
$$= -\left. 3z\, e^{-z^2/2}\right|_{-\infty}^{+\infty} + 3\int_{-\infty}^{\infty} e^{-z^2/2}\, dz = 3\sqrt{2\pi}$$

Therefore,

$$\boxed{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^4 e^{-z^2/2}\, dz = 3}$$

- It follows that the variance of Z^2 is

$$\mathbb{V}\left(Z^2\right) = 3 - 1^2 = 2.$$

Chi-Square Random Variable with q Degrees

- In general, if $Z_1, \ldots, Z_q$ are independent **standard normal random variables**, then the sum of their squares,

$$Y = \sum_{i=1}^{q} Z_i^2,$$

is distributed according to the **chi-square distribution** with q **degrees of freedom**, i.e., $Y \sim \chi_q^2$.

- By the assumption of independence,

$$\mathbb{V}\left(Y\right) = \sum_{i=1}^{q} \mathbb{V}\left(Z_i^2\right) = \sum_{i=1}^{q} 2 = 2q.$$

Variance Ratio

Definition 5

The **variance ratio** is defined as

$$\mathrm{VR}(q) := \frac{\mathbb{V}\big(r_t(q)\big)}{q\sigma^2},$$

- $\mathrm{VR}(q)$ should be equal to 1 when the conditions of log returns being **serially uncorrelated** and **homoskedastic** are satisfied.
- The **variance ratio test** is a test of

$$H_0 : \mathrm{VR}(q) - 1 = 0 \qquad \text{versus} \qquad H_1 : \mathrm{VR}(q) - 1 \neq 0$$

- If the null hypothesis cannot be rejected, then it means that the two assumptions are consistent with the reality. Conversely, a rejection of H_0 implies that at least one of the two assumptions is inconsistent with reality.

Sample Mean and Variance of Daily Log Returns

- To set up the framework for inference, we recall a few definitions and facts. The sample mean of daily log returns is estimated as usual,

$$\widehat{r}_1 = \frac{1}{T}\sum_{t=1}^{T} r_t. \tag{4}$$

- But the sample variance of daily log returns is instead estimated as

$$\widehat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T} \left(r_t - \widehat{r}_1\right)^2. \tag{5}$$

- The subscript of 1 in $\widehat{r}_1$ and $\widehat{\sigma}_1^2$ is meant to indicate that these estimates are for **daily log returns**.

Distribution of Daily Variance Estimate

- By the **law of large numbers**, as $T \longrightarrow \infty$,

$$\mathbb{E}\left(\widehat{\sigma}_1^2\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\left(r_t - \widehat{r}_1\right)^2\right) \longrightarrow \sigma^2$$

- When T is very large, asymptotically, and by the central limit theorem, and since $r_t = \mu + u_t$ as in Equation (3)

$$\mathbb{V}\left(\widehat{\sigma}_1^2\right) = \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{V}\left(\left(r_t - \widehat{r}_1\right)^2\right) = \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{V}\left(\left(\mu + u_t - \widehat{r}_1\right)^2\right)$$

$$\longrightarrow \frac{1}{T}\mathbb{V}\left(u_t^2\right) = \frac{1}{T}\mathbb{V}\left(\sigma^2 Z^2\right) = \frac{\sigma^4}{T}\mathbb{V}\left(Z^2\right) = \frac{2\sigma^4}{T},$$

where $Z \sim N(0,1)$, and $\mathbb{V}\left(Z^2\right)$ is the variance of the **chi-square random variable** with 1 **degree of freedom** (χ_1^2), which equals 2.

Summary for Daily Sample Variance Estimate

- By the **central limit theorem**, as T becomes larger and larger,

$$\widehat{\sigma}_1^2 \sim N\left(\sigma^2, \frac{2\sigma^4}{T}\right)$$

- Equivalently,

$$\sqrt{T}\left(\widehat{\sigma}_1^2 - \sigma^2\right) \sim N\left(0, 2\sigma^4\right),$$

or

$$\sqrt{T}\left(\frac{\widehat{\sigma}_1^2}{\sigma^2} - 1\right) \sim N\left(0, 2\right).$$

Multi-Period Mean and Variance

- An important property of log return is the **telescoping sum**:

$$\ln\left(\frac{P_t}{P_0}\right) = \ln(P_t) - \ln(P_{t-1}) + \ln P_{t-1} - \ln P_{t-2} + \ln P_{t-2} + \cdots + \ln P_2 - \ln P_1 + \ln P_1 - \ln P_0.$$

- Daily log return $r_s := \ln P_s - \ln P_{s-1}$ for $s = 1, 2, \ldots, t$,

$$\ln P_t - \ln P_0 = \sum_{s=1}^{t} r_s.$$

- Hence for log return

$$\mathbb{E}\left(\ln\left(\frac{P_t}{P_0}\right)\right) = \sum_{s=1}^{t} \mathbb{E}(r_s) = t\mu, \qquad \mathbb{V}\left(\ln\left(\frac{P_t}{P_0}\right)\right) = \sum_{s=1}^{t} \mathbb{V}(r_s) = t\sigma^2.$$

Concept Checker

- **Sharpe ratio** is a measure of risk-adjusted return:

$$R_{\text{Sharpe}} = \frac{\mathbb{E}\left(r_{i,t} - r_{\text{rf},t}\right)}{\sqrt{\mathbb{V}\left(r_{i,t}\right)}}.$$

- You use the daily frequency to estimate the average excess log return of portfolio i and the standard deviation of the log return on portfolio i. **How should you annualize the Sharpe ratio?**

- Specifically, suppose the average excess return is 0.02% and the daily volatility is 1.1%. Suppose each year has 252 trading days. What is the annualized Sharpe ratio?

Non-Overlapping Log Return

Definition 6

The **non-overlapping q-daily log return** is defined as

$$r_q(j) := \ln\left(P_{qj}\right) - \ln\left(P_{(q-1)j}\right). \tag{6}$$

- The prices indexed from $(q-1)j+1$ through $qj-1$ are ignored.

Multi-Period Return from Sum of Daily Returns

Proposition 3
Let $r_1(qj-i)$ denote the **daily log return** constructed from $T+1$ log prices, with $j=1,2,\ldots,M$ and $i=0,1,\ldots q-1$. The **non-overlapping q-daily log return** can be expressed as

$$r_q(j)=\sum_{i=0}^{q-1} r_1(qj-i).$$

Proof: The **telescoping sum** of q daily log returns $r_1(qj-i)$ provides the proof as follows:

$$\begin{aligned}\sum_{i=0}^{q-1} r_1(qj-i) &= \big(\ln P_{qj}-\ln P_{qj-1}\big)+\big(\ln P_{qj-1}-\ln P_{qj-2}\big)+\cdots\\ &\quad\cdots+\big(\ln P_{qj-q+2}-\ln P_{qj-q+1}\big)+\big(\ln P_{qj-q+1}-\ln P_{qj-q}\big)\\ &=\ln P_{qj}-\ln P_{q(j-1)}=r_q(j).\end{aligned}$$

Average of q-Daily Log Return

- The sample mean of **daily log returns** is estimated by assuming that $T = Mq$,

$$\widehat{r}_1 = \frac{1}{T}\sum_{t=1}^{T} r_t = \frac{1}{Mq}\sum_{j=1}^{M}\sum_{i=0}^{q-1} r_1(qj-i). \tag{7}$$

Proposition 4
The sample average of q-daily log return $r_{qj}(q)$ is simply $q\widehat{r}_1$.

Proof: After multiplying both sides, we obtain $q\widehat{r}_1 = \frac{1}{M}\sum_{j=1}^{M}\sum_{i=0}^{q-1} r_1(qj-i)$. It follows that $q\widehat{r}_1 = \frac{1}{M}\sum_{j=1}^{M} r_q(j)$. The right-hand side is anything but the average of M q-daily log returns.

A Nice Result

- The q-daily log return $r_q(j)$ for each j is a sum of q daily returns.

$$r_q(j) = r_{qj} + r_{qj-1} + \cdots + r_{q(j-1)+1}.$$

- Now, $q\widehat{r}_1$ can be expanded out as a summation. Then, since $u_t = r_t - \widehat{r}_1$ for each term, we obtain

$$\begin{aligned} r_q(j) - q\widehat{r}_1 &= (r_{qj} - \widehat{r}_1) + (r_{qj-1} - \widehat{r}_1) + \cdots + (r_{q(j-1)+1} - \widehat{r}_1) \\ &= u_{qj} + u_{qj-1} + \cdots + u_{q(j-1)+1}. \end{aligned} \tag{8}$$

- Since all the u_t has zero covariance with each other, for every j,

$$\mathbb{E}\Big(\big(r_q(j) - q\widehat{r}_1\big)^2\Big) = \mathbb{E}\left(\sum_{i=0}^{q-1} u_{qj-i}^2\right) = \sum_{i=0}^{q-1} \widehat{\sigma}_1^2(j) = q\widehat{\sigma}_1^2(j).$$

Estimation of q-Daily Sample Mean and Variance

❒ The q-daily return is a sum of daily returns:

$$r(qj) := \ln P_{qj} - \ln P_{q(j-1)} = r_{qj} + r_{qj-1} + \cdots + r_{qj-(q-1)}, \tag{9}$$

for $j = 1, 2, \ldots, M$, where M is the maximum number of **non-overlapping q-daily returns** that are obtainable from $T+1$ prices starting from P_0.

❒ As a matter of fact, $M = \left\lfloor \dfrac{T}{q} \right\rfloor$.

❒ The sample average of $r(qj)$ is q times of $\widehat{r}_1$ (sample estimate of μ), i.e., $q\widehat{r}_1$.

❒ The sample variance is estimated as

$$\widehat{\sigma}_q^2 = \frac{1}{M} \sum_{j=1}^{M} \big(r(qj) - q\widehat{r}_1 \big)^2. \tag{10}$$

Asymptotic Distribution of q-Daily Variance

- The **asymptotic limit** of the expected value of $\widehat{\sigma}_q^2$ is

$$\mathbb{E}\left(\frac{\widehat{\sigma}_q^2}{q}\right) = \frac{1}{Mq}\sum_{j=1}^{M}\mathbb{E}\Big(\big(r(qj) - q\widehat{r}_1\big)^2\Big) \longrightarrow \sigma^2;$$

- The asymptotic limit of the variance of $\widehat{\sigma}_q^2$ is

$$\begin{aligned}
\mathbb{V}\left(\frac{\widehat{\sigma}_q^2}{q}\right) &= \frac{1}{M^2q^2}\mathbb{V}\left(\sum_{j=1}^{M}\big(r(qj) - q\widehat{r}_1\big)^2\right) \\
&= \frac{1}{Mq^2}\mathbb{V}\left(\sum_{j=1}^{q}u_j^2\right) = \frac{1}{(Mq)q}\mathbb{V}\left(\sum_{j=1}^{q}u_j^2\right) \\
&= \frac{1}{Tq}\mathbb{V}\left(q\sigma^2Z^2\right) = \frac{q}{T}\sigma^4\mathbb{V}(Z^2) = \frac{2q\sigma^4}{T}.
\end{aligned}$$

Null Hypothesis of Variance Ratio

- By the **central limit theorem**, as T becomes very large,

$$\frac{\widehat{\sigma}_q^2}{q} \sim N\left(\sigma^2, \frac{2q\sigma^4}{T}\right)$$

- Equivalently

$$\sqrt{T}\left(\frac{\widehat{\sigma}_q^2}{q} - \sigma^2\right) \sim N(0, 2q\sigma^4),$$

or

$$\sqrt{T}\left(\frac{\widehat{\sigma}_q^2}{q\sigma^2} - 1\right) \sim N(0, 2q).$$

- Note that the variance for the statistic $\sqrt{T}\left(\frac{\widehat{\sigma}_q^2}{q\sigma^2} - 1\right)$, i.e., $2q$, is known!

Test Statistics

- To perform the test, we define the test statistics

$$J_d(q) := \frac{\widehat{\sigma}_q^2}{q} - \widehat{\sigma}_1^2;$$

$$J_r(q) := \frac{\widehat{\sigma}_q^2}{q\widehat{\sigma}_1^2} - 1 = \widehat{\mathrm{VR}}(q) - 1. \tag{11}$$

- Note that $J_r(q) = \dfrac{J_d(q)}{\widehat{\sigma}_1^2}$.

Asymptotic Distributions

Theorem

The asymptotic distributions of $\sqrt{T}J_d(q)$ and $\sqrt{T}J_r(q)$ are normal with mean 0 and variances of, respectively, $2(q-1)\sigma^4$ and $2(q-1)$:

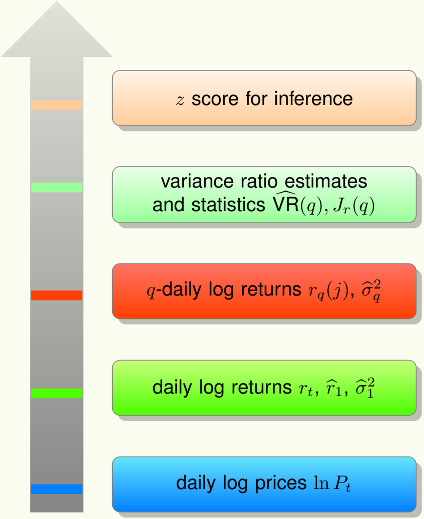
$$\sqrt{T}J_d(q) \sim N\big(0, 2(q-1)\sigma^4\big);$$

$$\sqrt{T}J_r(q) \sim N\big(0, 2(q-1)\big).$$

- In light of this theorem, for $q > 1$, the z score is computed as

$$\boxed{Z_q = \sqrt{T}\frac{J_r(q)}{\sqrt{2(q-1)}} \sim N(0,1).} \tag{12}$$

A Variance Ratio Test Algorithm



Case Study: Variance Tests on GE

q	1	2	3	4	5	6	7	8	9	10
Obs	22,776	11,388	7,592	5,694	4,555	3,796	3,253	2,847	2,530	2,277
$\widehat{\text{VR}}(q)$	1	1.002	0.946	0.939	0.916	0.926	0.968	0.933	0.871	0.920
Z_q	—	0.20	-4.08	-3.74	-4.46	-3.53	-1.40	-2.69	-4.85	-2.86

Table: Results of variance ratio tests based on GE's daily log returns.

1. Looking at Z_2, what can you infer?
2. Looking at Z_5, what can you infer?

Problem of Original Test

- Empirical studies almost always examine multiple values of q
- Without controlling the overall (joint) test size, the probability of falsely rejecting the RWH (Type I error) becomes very large.
- Chow and Denning (1993) document that naive application inflates the rejection rate to roughly 6$\times$, 3–4$\times$, and 2$\times$ the nominal size for 1%, 5%, and 10% tests, respectively.
- This inflation exacerbates **data-snooping bias** (focusing on the most **extreme statistic**).

The Multiple Comparison Procedure

- Define the maximum absolute statistic:

$$Z^* := \max_{1 \le i \le l} |Z_q|,$$

  where l is the number of aggregation parameters tested.

- Under the joint null $H_0 : \mathrm{VR}(q_1) = \mathrm{VR}(q_2) = \cdots = \mathrm{VR}(q_l) = 1$, use the **Studentized Maximum Modulus** (SMM) distribution to obtain critical values that control the family-wise error rate (overall test size) at level α.
- Compare Z^* to $\mathrm{SMM}(\alpha; l; \infty)$, which is upper $\alpha/2$ point of standard normal adjusted via Sidak inequality.
- This value also yields **joint confidence intervals** for the set of variance ratios.
- Bottom line: just replace the usual $\pm$1.96 (or $\pm$2.58) with the larger SMM **critical value**.

Studentized Maximum Modulus Distribution

- Let $z_1, z_2, \ldots, z_l$ be l (possibly correlated) standard normal random variables (or, more generally, studentized statistics with v degrees of freedom). Then

$$M = \max_{1 \leq i \leq l} |z_i|$$

- The distribution of M is the **maximum modulus distribution**.
- When the z_i are **studentized** (i.e., divided by their estimated **standard errors**), it is called the **Studentized Maximum Modulus**.
- $\mathrm{SMM}(\alpha; l; v)$ is defined as the upper α critical value (100$(1-\alpha)$% quantile) of this distribution with parameters:
  - $l =$ number of statistics (dimension)
  - $v =$ degrees of freedom

Properties

- The following holds even when the z_i are correlated (with arbitrary correlation matrix), making it very robust for the variance ratio application.

$$\mathbb{P}\left[\max_{1\leq i\leq l}|z_i| \leq \text{SMM}(\alpha; l; \upsilon)\right] \geq 1-\alpha.$$

- Asymptotically, when $\upsilon \to \infty$,

$$\text{SMM}(\alpha; l; \infty) = Z^*_{\alpha^+/2},$$

where $Z^*_{\alpha^+/2}$ is the adjusted normal critical value coming from the **Sidak inequality**:

$$\alpha^+ = 1-(1-\alpha)^{1/l}$$

Chow-Denning Procedure

- Compute Z_{q_i} for chosen set of q_i
- Take the maximum absolute value
- Compare Z^* to the $\mathrm{SMM}(\alpha; l; v)$ critical value.
- If Z^* is larger than the **critical value**, reject the **joint null hypothesis**, meaning that the **random walk** hypothesis is rejected.

SMM Critical Values (Selected Values)

l	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
2	2.326	2.576	3.090
3	2.569	2.795	3.282
4	2.728	2.935	3.402
5	2.850	3.048	3.499
6	2.949	3.140	3.581
7	3.030	3.219	3.652
8	3.100	3.290	3.715
9	3.162	3.354	3.772
10	3.218	3.411	3.823

Takeaways

- The Law of Iterated Expectations provides the connection between the unpredictability of return and the fundamental value of the asset.
- The variance of random walk's sample paths increases with time.
- Variance ratio test is an algorithm to test whether the asset price is a random walk.
- Empirical evidence suggests that asset prices are generally not random walks. They are not 100% random.
- Statistical arbitrage is difficult but possible because prices are not strictly random walks.
- If you invest in obtaining and analyzing financial information, it is possible to "beat the market" (though difficult).
- Smart beta based on VR(q), skewness, and kurtosis estimates?

Additional Reading

1. Quantitative Investing is Fundamental
2. A Non-Random Walk Down Wall Street (Page 3 to Page 83)