Random Walk & Variance Ratio Test

Christopher Ting

http://www.mysmu.edu/faculty/christophert/

✉: christophert@smu.edu.sg

☎: 6828 0364

: LKCSB 5036

January 20, 2017

Table of Contents

Learning Objectives

ð Understand the concept of random walk and its application on the study of return time series.

ð Describe variance ratio test of random walk and calculate the test statistics for inferences.

Random Walk

- A random walk is a walk where the direction of each step is chosen at random.

- Let $Z_1, Z_2, \ldots, Z_t$ be a time series of i.i.d. random variables with mean μ and standard deviation σ. Let S_0 be any starting point.

$$S_t = S_0 + Z_1 + \cdots + Z_t \,, \qquad t \geq 1 \,.$$

- Conditional mean and variance are proportional to t.

$$\mathbb{E}\left(S_t \middle| S_0\right) = S_0 + \mu\, t$$
$$\mathbb{V}\left(S_t \middle| S_0\right) = \sigma^2\, t$$

- The parameter μ is called the drift, and the parameter σ is called the volatility, which is responsible for diffusion.

Geometric Random Walk

❀ Recall that $\ln\big(1 + R_t(q)\big) = r_t + r_{t-1} + \cdots + r_{t-q+1}$. So

$$\frac{P_t}{P_{t-q}} = 1 + R_t(q) = \exp\big(r_t + \cdots + r_{t-q+1}\big).$$

❀ Let $q = t$, we have

$$P_t = P_0 \exp(r_t + r_{t-1} + \cdots + r_1)$$

❀ If $r_1, r_2, \ldots, r_t$ are i.i.d. and for time t,

$$r_t \sim N(\mu, \sigma^2),$$

then P_t is lognormal for all t. This stochastic process is known as geometric random walk with parameters μ and σ^2.

Lognormal Geometric Random Walk

✿ Big Assumptions

1. i.i.d.
2. Each r_t is a normally distributed random variable

✿ Consequences

1. Log returns are uncorrelated.
2. Log returns cannot be forecasted.

Mean and Variance of Lognormal Random Variable

❈ Mean

The expected value of price is

$$\mathbb{E}\big(P_t \big| P_0\big) = P_0 \exp\left(\left(\mu + \frac{\sigma^2}{2}\right) t\right)$$

❈ Variance

The price variance is

$$\mathbb{V}\big(P_t \big| P_0\big) = P_0^2\, e^{(2\mu+\sigma^2)t}\big(e^{\sigma^2 t} - 1\big)$$

Information Set and Random Walk

$\rightsquigarrow$ A better forecast of next-period price P_{t+1} is obtainable as the conditional expectation based on information ϕ_t available at t:

$$\mathbb{E}_t(P_{t+1}) \equiv \mathbb{E}\big(P_{t+1}\big|\phi_t\big)$$

$\rightsquigarrow$ A random walk is a process that exhibits no preference in the direction it is taking for the next time step. Thus, no “pattern” can be deciphered from a time series of random walks.

$\rightsquigarrow$ If the price process is a random walk, i.e, the probability of up move is the same as the probability of down move:

$$P_{t+1} = P_t + e_{t+1}$$

where e_{t+1} is a noise process with zero mean.

$\rightsquigarrow$ The noise e_{t+1} is not correlated with P_t, i.e., $\mathbb{E}\big(e_{t+1}\big|P_t\big) = 0$. Then

$$\mathbb{E}_t\big(P_{t+1}\big) = \mathbb{E}_t\big(P_t + e_{t+1}\big) = P_t.$$

Log Price and Random Walk

- Consider a random variable $\xi_{t+1} > 0$ such that $\mathbb{E}_t\big(\xi_{t+1}\big) = 1$ and

$$P_{t+1} = P_t \xi_{t+1}$$

Then

$$\mathbb{E}_t\big(P_{t+1}\big) = P_t$$

- In this model, log return is noise!

$$\begin{aligned} r_{t+1} &= \ln P_{t+1} - \ln P_t = \ln P_t + \ln \xi_{t+1} - \ln P_t \\ &= \ln \xi_{t+1} \end{aligned}$$

Assumptions

❑ If the daily log return r_t is treated as a random variable, the variance of a sum of q daily log returns *in sequel* is

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

❑ Two assumptions are made

1. Zero covariance: $\mathbb{C}\big(r_s, r_t\big) = 0$ for any $s \neq t$
2. Homoskedasticity: $\mathbb{V}\big(r_t\big) = \sigma^2$

❑ Under these two assumptions,

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

Variance Ratio

❒ Definition of variance ratio

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

❒ $\mathrm{VR}(q)$ should be equal to one 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,

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

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

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

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 Variance Estimate

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

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

$$\mathbb{V}\big(\widehat{\sigma}_1^2\big) = \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{V}\Big(\big(r_t - \widehat{r}_1\big)^2\Big) \longrightarrow \frac{1}{T}\mathbb{V}\big(\sigma^2 x^2\big) = \frac{\sigma^4}{T}\mathbb{V}\big(x^2\big),$$

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

- By the central limit theorem, as $T \longrightarrow \infty$,

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

Estimation of q-Daily Log Return and Variance

- The q-daily return is

$$r_{qj}(q) = \ln P_{qj} - \ln P_{q(j-1)},$$

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.

- The sample average of $r_{qj}(q)$ is simply q times of $\widehat{r}_1$, i.e., $q\widehat{r}_1$.

- The sample variance is estimated as

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

Asymptotic Limits

▲ The asymptotic limits of the expected value and variance of $\widehat{\sigma}_q^2$ are as follows:

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

$$\mathbb{V}\left(\frac{\widehat{\sigma}_q^2}{q}\right) = \frac{1}{M^2q^2}\sum_{j=1}^{M}\mathbb{V}\Big(\big(r_{qj}(q) - q\widehat{r}_1\big)^2\Big) \longrightarrow \frac{1}{Mq^2}\mathbb{V}\left(q\sigma^2x^2\right) = \frac{1}{M}\sigma^4\mathbb{V}(x^2).$$

▲ As in Slide 13,

▲ By the central limit theorem, as $M \longrightarrow \infty$,

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

Test Statistics

- To perform the test, we define the test statistics

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

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

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

Asymptotic Distributions

Theorem 3.1

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

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

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

- In light of this theorem, for $q > 2$, the z score is computed as

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

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
$\widehat{\mathrm{VR}}(q)$	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?

Takeaways

- Asset returns are likely to be not normally distributed

- The variance of return increases with the holding period

- Statistical arbitrage is difficult but possible because prices are not strictly random walks.