The Extreme Value Method for Estimating the Variance of the Rate of Return
Author(s): Michael Parkinson
Source: *The Journal of Business*, Vol. 53, No. 1 (Jan., 1980), pp. 61-65
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/2352357

Michael Parkinson*
University of Florida and
Gulf Research and Development Company

The Extreme Value Method for Estimating the Variance of the Rate of Return

If S is the price of a common stock, it is now generally accepted that ln (S) follows a random walk, at least to a very good approximation. The diffusion constant characterizing that walk, which is the same as the variance of the rate of return, thus becomes an important quantity to calculate and is traditionally estimated using closing prices only. It is shown that the use of extreme values (the high and low prices) provides a far superior estimate.

I. Introduction and Motivation

The random walk problem has a long history. In fact, its application to the movement of security prices predates the application to Brownian motion.[1] And now it is generally accepted that, at least to a good approximation, ln (S), where S is the price of a common stock, follows a random walk.[2] The diffusion constant characterizing that walk for each stock thus becomes an important quantity to calculate. In Section II, we describe the general random walk problem and show how the diffusion constant is traditionally estimated. In Section III, we discuss another way to estimate the diffusion constant, the extreme value method. In Section IV, we compare the traditional and extreme value methods and conclude that the extreme value method is about 2½–5 times better, depending on how you choose to measure the difference. In Section V, we discuss the use of this method for the estimation of the variance of the rate of return of a common stock.

* The author would like to acknowledge Robert Pisani for useful comments on an earlier version of this paper, the Northeast Florida Regional Data Center for computing support, and Nancy Parkinson for her assistance in preparing the manuscript.

1. See Bachelier's paper in Cootner (1964).
2. See Cootner (1964) for an excellent discussion of the evidence.

(*Journal of Business*, 1980, vol. 53, no. 1)

0021-9398/80/5301-0001$00.75

II. Statement of the Problem

The problem to be solved may be stated as follows: Suppose a point particle undergoes a one-dimensional, continuous random walk with a diffusion constant D. Then, the probability of finding the particle in the interval $(x, x + dx)$ at time t, if it started at point x_0 at time $t = 0$, is $(dx/\sqrt{2\pi Dt}) \exp[-(x - x_0)^2/2Dt]$. By comparison with the normal distribution, we see that D is the variance of the displacement $x - x_0$ after a unit time interval. This suggests the traditional way to estimate D: we measure $x(t)$ for $t = 0, 1, 2, \ldots, n$. Then, defining d_i = displacement during the ith interval, $d_i = x(i) - x(i - 1), i = 1, 2, \ldots, n$, we have

$$D_x = \frac{1}{n - 1} \sum_{i=1}^{n} (d_i - \bar{d})^2$$

as an estimate for D;

$$\bar{d} = \frac{1}{n} \sum_{m=1}^{n} d_m = \text{mean displacement.}$$

However, instead of measuring $x(n)$, for $n = 0, 1, 2, \ldots$, suppose we have measured only the difference l between the maximum and minimum position during each time interval. These differences should be capable of giving a good estimate for D, for it is intuitively clear that the average difference will get larger or smaller as D gets larger or smaller.

III. The Relation between the Range and the Diffusion Constant D

A derivation of the probability distribution for l has already been published.[3] Defining $P(l, t)$ to be the probability that $(x_{\max} - x_{\min}) \leq l$ during time interval t, we have

$$P(l, t) = \sum_{n=1}^{\infty} (-1)^{n+1} n\{erfc[(n + 1)l/\sqrt{2Dt}] - 2erfc(nl/\sqrt{2Dt}) + erfc[(n - 1)l/\sqrt{2Dt}]\},$$

where $erfc(x) = 1 - erf(x)$ and $erf(x)$ is the error function.

We can now show straightforwardly that

$$E[l^p] = \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{p + 1}{2}\right)\left(1 - \frac{4}{2^p}\right) \zeta(p - 1)(2Dt)^{p/2} \quad (p \text{ real and } \geq 1),$$

where $\zeta(x)$ is the Riemann zeta function.[4] In particular, we have $E[l] = \sqrt{8Dt/\pi}$ and $E[l^2] = (4 \ln 2)Dt$.

3. See Feller 1951.
4. The HP-65 Users' Library (Hewlett-Packard Company, Corvallis, Oregon) program for $\zeta(x)$ (01104A by Stuart Augustin) is both fast and convenient; $\zeta(x)$ is tabulated in Abramowitz and Stegun (1964).

Thus, for a unit time interval, we have $D = .393\ (E[l])^2 = .361\ E[l^2]$. The relationship $E[l^2] = 1.09\ (E[l])^2$ provides a very useful test for checking that a set of observed l's really come from a random walk of the kind assumed above.

Thus, given a set $(l_1, l_2, \ldots, l_n)$ of observed l values over n unit time intervals, we can write the extreme value estimate for D as follows:

$$D_l = \frac{.361}{n} \sum_{i=1}^{n} l_i^2.$$

IV. Comparison of D_x and D_l Estimates for D

Computing the variance of D_x and D_l, we find that

$$E[(D_x - D)^2] = \left[\frac{E(x^4)}{E(x^2)} - 1\right] \frac{D^2}{N_x} = \frac{2D^2}{N_x}$$

and

$$E[(D_l - D)^2] = \left[\frac{E(l^4)}{E(l^2)} - 1\right] \frac{D^2}{N_l} = \frac{.41D^2}{N_l},$$

where N is the number of observations. Thus, to obtain the same amount of variance using the two methods, we need $N_x \approx 5N_l$. Clearly, the extreme value method is far superior to the traditional method and will be much more sensitive to variations in D.

In order to check the validity of the above analysis, 100 random walks of 10,000 steps each were generated on the computer.[5] The steps were uniformly distributed over the interval $(-\frac{1}{2}, +\frac{1}{2})$. For each walk, the displacement x from the origin after 10,000 steps was recorded and also the difference between the maximum and the minimum. Since the variance of each step is 1/12, the theoretical prediction for D is $10{,}000/12 = 833$. The following quantities were then calculated from the recorded data:

$$\overline{x^2} = \frac{1}{100} \sum_{i=1}^{100} (x^2)_i = 812.63,$$

$$\sigma_{x^2} = \sqrt{\frac{1}{99} \sum_{i=1}^{100} (x_i^2 - \overline{x^2})^2} = 1164.60,$$

$$\bar{l} = \frac{1}{100} \sum_{i=1}^{100} l_i = 45.00,$$

5. Since the presence of correlations in the random generator could seriously affect a test involving so many random numbers, a specification of the particular generator used is in order. The method described by MacLaren and Marsaglia (1965) was used. The method works as follows: A table of 128 numbers is filled by $I_{n+1} = 65549 I_n \pmod{2^{31}}$. Selections are randomly made from it using $J_{n+1} = 65539 J_n \pmod{2^{31}}$. As each selection is made, the next I_n is generated and inserted into the table to replace the one just selected. I thank Robert L. Coldwell for providing me with the code for this generator and for calling my attention to the severe limitations of the standard generator which is given by the I_n's alone.

and

$$\sigma_l = \sqrt{\frac{1}{99}\sum_{i=1}^{100}(l_i - \bar{l})^2} = 14.3.$$

Then, $\sigma_{\overline{x^2}}$ = the standard deviation of $\overline{x^2}$ = 116.46 and $\sigma_{\bar{l}}$ = the standard deviation of $\bar{l}$ = 1.43. We thus find that $D_x = \overline{x^2} = 813 \pm 116$ and $D_l = .393\bar{l}^2 = 795 \pm 51$. Both estimates agree with the theoretical value within their errors, but notice how much smaller the calculated standard deviation is for D_l, as predicted above.

V. Application to the Stock Market

Let $(S_1, S_2, \ldots, S_{n+1})$ be a set of stock prices quoted at equal time intervals. Then V, the variance of the rate of return on the stock, is traditionally estimated as follows: Letting $r_i = \ln(S_{i+1}/S_i)$, $i = 1, 2, \ldots, n$ = rate of return over the ith time interval, then

$$V = \frac{1}{n}\sum_{i=1}^{n} r_i^2,$$

assuming the mean rate of return is zero.

Now, since $x = \ln(S)$ follows, to a very good approximation, a continuous random walk (see Cootner 1964) of the kind described in the first paragraph of Section II, we see that, when the time interval separating the S_i above is the unit time interval, the V calculated above is the traditional estimate of the random walk diffusion constant D.

We can conclude that the true variance of the rate of return of a common stock over a unit time interval is precisely the diffusion constant of the underlying random walk. Now, as we have seen, the extreme value method provides a much better estimate of that diffusion constant than the traditional method does.[6] So, to estimate the variance of the rate of return, we certainly ought to prefer the extreme value method.

In fact, the extreme value method is very easy to apply in practice, since daily, weekly, and in some cases, monthly highs and lows are published for every stock. If H = the high and L = the low in the time interval under consideration, then $l = \ln(H/L)$ gives one of the difference values we need for calculating D_l. The improvement using D_l as an estimate of V could be of particular importance in studies of the time and price dependence (if any) of V, for, to get a given accuracy in V, about 80% less data (and thus, an 80% smaller time interval) is needed for the extreme value method than for the traditional method.

6. The extreme value method for estimating V was used by the author in a paper on put options (Parkinson 1977).

References

Abramowitz, M., and Stegun, I. 1964. *Handbook of Mathematical Function*. Washington, D.C.: National Bureau of Standards.

Cootner, Paul, ed. 1964. *The Random Character of Stock Prices*. Cambridge, Mass.: M.I.T. Press.

Feller, W. 1951. The Asymptotic Distribution of the Range of Sums of Random Variables. *Annals of Mathematical Statistics* 22:427–32.

MacLaren, M. D., and Marsaglia, G. 1965. Uniform Random Number Generators. *Journal of the Association of Computing Machinery* 12:83.

Parkinson, M. 1977. Option Pricing: The American Put. *Journal of Business* 50 (January): 21–36.