A test of financial time-series data to discriminate among lognormal, Gaussian and square-root random walks

Yuri Heymann

3 rue Chandieu, 1202 Geneva, Switzerland

(Dated: March 11, 2014)

Abstract

In the present study, Monte Carlo simulations show how a simple test applied to financial time-series data can discriminate among the lognormal random walk used in the Black-Scholes-Merton model, the Gaussian random walk used in the Ornstein-Uhlenbeck stochastic process, and the square-root random walk used in the Cox, Ingersoll and Ross process. Alpha-level hypothesis testing is provided. As a conclusion, this test appears to be helpful for selecting the best stochastic processes for pricing contingent claims and risk management.

I. INTRODUCTION

One approach to testing for lognormality of financial time series is to analyze the distribution of the returns. If a time series follows a lognormal random walk, then the continuously compounded returns $\ln(S_t/S_{t-1})$ must be normally distributed. This approach has been used by [1] for interest rate changes. There are many ways to test the normality of a distribution such as the Kolmogorov-Smirnov test and the Anderson-Darling test [2]. Such tests may be used to accept or reject the hypothesis of normally distributed returns given a certain significance level.

Inferring lognormality from a normality test of the distribution of returns has some weaknesses. For example, in many cases a Gaussian random walk or square-root random walk may exhibit close to normally distributed returns. Accordingly, testing for normality of returns is not enough to infer lognormality of a time series. In addition, returns of financial time series exhibit departures from normality such as anomalies of the skewness and kurtosis, outliers and, for equity and market indexes, fat tails [3, 4]. Therefore, a normality test may give a negative result at the same time that we accept that the underlying process is lognormal in overall. These weaknesses motivate the development of tests that do not rely on the hypothesis that returns are normally distributed.

The aim of the present study is to offer a testing framework for the structural properties of the Brownian motion of the underlying stochastic process of a time series. For example the Ornstein-Uhlenbeck stochastic process is expressed as $dS_t = \lambda\left(\mu - S_t\right)dt + \sigma dW_t$, where λ, μ, and σ are model parameters, and W_t the Wiener process. In this stochastic differential equation, the Brownian motion is a Gaussian random walk of the form σdW_t. In the Cox, Ingersoll and Ross process, which is expressed as $dS_t = \lambda\left(\mu - S_t\right)dt + \sigma\sqrt{S_t}dW_t$, the Brownian motion is a square-root random walk of the form $\sigma\sqrt{S_t}dW_t$. The lognormal random walk, which is expressed as $dS_t = \mu S_t dt + \sigma S_t dW_t$, has Brownian motion of the form $\sigma S_t dW_t$. For the general form of stochastic processes, $dS_t = \mu(S_t, t)dt + \sigma_n S_t^n dW_t$, the present study proposes a method to estimate n which is the order of the Brownian motion.

II. MODEL

Because daily time increments are generally used for the purpose of analyzing time-series data, the drift term of the stochastic differential equation is considered to be negligible in our analyses. Hence, for a lognormal random walk the stochastic differential equation is as follows:

$$dS_t = \sigma S_t dW_t \,, \tag{1}$$

where S_t is the asset price, σ is the volatility, and dW_t the Brownian term which is a $\mathcal{N}(0,1)$ variable.

Let us consider the expected value of the absolute value of dS_t/S_t in eq. (1). We get:

$$\mathbb{E}\left(\left|\frac{dS_t}{S_t}\right|\right) = \sigma\mathbb{E}\left(|dW_t|\right) \,. \tag{2}$$

To evaluate $\mathbb{E}\left(|dW_t|\right)$, we need to integrate the density function of $\mathcal{N}(0,1)$ multiplied by the absolute value of the integration variable between $-\infty$ and ∞, which is equal to twice the integral of the density function multiplied by the integration variable between 0 and ∞:

$$\mathbb{E}\left(|dW_t|\right) = 2\int_0^\infty x\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)dx \,. \tag{3}$$

Hence, the expected value of the absolute value of dW_t is as follows:

$$\mathbb{E}\left(|dW_t|\right) = \sqrt{\frac{2}{\pi}} \,. \tag{4}$$

The following estimator of the volatility is obtained from eqs. (2) and (4):

$$\hat{\sigma_1} = \sqrt{\frac{\pi}{2}}\,\frac{1}{n}\sum_{i=1}^{n}|r_i| \,, \tag{5}$$

where $\hat{\sigma_1}$ is the volatility estimator for n returns $r_i = \ln\frac{S_i}{S_{i-1}}$.

Eq. (5) estimates volatility based on the average value of absolute returns assuming that the returns are normally distributed. In contrast, the usual approach to computing volatility calculates a standard deviation from the sum of squared returns:

$$\hat{\sigma_2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} r_i^2}\,. \qquad (6)$$

Eq. (6) makes no assumptions about return distribution.

To test the lognormality of a time series, let us do a linear regression of the absolute value of the differences $|dS_t| = |S_t - S_{t-1}|$ versus S_t. The regression model is $|S_i - S_{i-1}| = \alpha + \beta S_i + \epsilon_i$, where α is the intercept, β the slope, and ϵ_i the error term. Then let us compute the following parameter which is also an estimator of the volatility for the lognormal random walk:

$$\hat{\sigma_3} = \sqrt{\frac{\pi}{2}}\,\beta\,, \qquad (7)$$

where β is the slope of the linear regression.

For a lognormal random walk, $\hat{\sigma_3}$ will converge toward $\hat{\sigma_1}$ or $\hat{\sigma_2}$; whereas for a Gaussian random walk of the form $dS_t = \sigma_{Gauss} dW_t$, $\hat{\sigma_3}$ will converge toward 0. This approach can also be used to test the square-root random walk of the form $dS_t = \sigma_{sqrt}\sqrt{S_t}dW_t$.

III. SIMULATIONS

To assess the present test, let us run some Monte Carlo simulations with different stochastic processes, respectively the lognormal random walk, the Gaussian random walk, and the square-root random walk. For comparability, we need to scale the σ_{Gauss} and σ_{sqrt} of the Gaussian and square-root random walk. Let us take $\sigma_{Gauss} = S_0\sigma$, and $\sigma_{sqrt} = \sqrt{S_0}\sigma$, where S_0 is the initial asset price of the simulation. For the Monte Carlo simulation, let us take $S_0 = 100$, $\sigma = 0.15$, over a time horizon T of 2 years with daily time increments.

TABLE I. Simulated test statistical parameters over 200,000 paths

	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\sigma_3}$
lognormal rand. walk	0.1500	0.1499	0.1510
Gaussian rand. walk	0.1537	0.1543	-0.000278
square-root rand. walk	0.1513	0.1514	0.0758

In the simulations of table I, the volatility estimators have been scaled to annual basis using the $\sqrt{dt}$ scaling factor. The volatility estimators converge quickly to their expected values from the first simulation path. For the purpose of running the simulations over longer time horizons (e.g. T of 20 years), we must add a constraint on the minimal value that S can reach; otherwise, the statistics $\hat{\sigma_1}$ and $\hat{\sigma_2}$ become unstable when S is close to zero for the Gaussian and square-root random walk. A floor price of 1 unity when S_0 is equal to 100 would be adequate.

To show that testing for normality of the returns is not sufficient to make inferences about the lognormality of a time series, let us run some Monte Carlo simulations under different scenarios and compute the percentage rejection of the null hypothesis (i.e. normally distributed returns) using the Kolmogorov-Smirnov test at a significance level of 5% over 10,000 simulation paths (see table II).

TABLE II. Rejection rate of the normally distributed return hypothesis with the Kolmogorov-Smirnov test at 5% significance level

	T of 2 yrs	T of 20 yrs
lognormal rand. walk	0.0%	0.0%
Gaussian rand. walk	0.01%	42.5%
square-root rand. walk	0.0%	9.9%

For the simulations with the 2-year horizon, the test fails to reject the normally distributed return hypothesis for the Gaussian and square-root random walks; whereas, for the 20-year horizon, the rejection rate is 9.9% for the square-root random walk and 42.5% for the Gaussian random walk. The reason for this result is that, for the 2-year horizon, the variations of S are small. Hence, returns from the stochastic differential equation are almost normally distributed for both the Gaussian and square-root random walks. For the 20-year horizon, the variations of S are larger (S spans a larger range of values); therefore, we start to observe some deviations from normality for the returns.

IV. GENERALIZATION

Let us show that for the general form of stochastic processes $dS_t = S_t^n \sigma_n dW_t$, the ratio $\hat{\sigma_3}/\hat{\sigma_1}$ converges toward n.

Let us set $y = |dS|$, hence the slope of $|dS|$ versus S for arbitrary t is as follows:

$$\beta_n = \frac{dy}{dS} = nS^{n-1}\sigma_n|dW|\,. \tag{8}$$

From the scaling relationship $S_i^n \sigma_n = S_i \sigma_i$, we get $\sigma_n = \frac{\sigma}{\mathbb{E}(S^{n-1})}$, hence:

$$\beta_n = nS^{n-1}\frac{\sigma}{\mathbb{E}(S^{n-1})}|dW|\,. \tag{9}$$

Finally, the expected value of β_n is as follows:

$$\mathbb{E}(\beta_n) = n\sqrt{\frac{2}{\pi}}\sigma\,. \tag{10}$$

V. HYPOTHESIS TESTING OF LOGNORMALITY

Finally, the parametric test of lognormality of a given time series is described below. The null hypothesis is that the underlying stochastic process of the time series is lognormal. Let us assume that the error terms of the linear regression of $|dS_t|$ versus S_t used for the estimation of $\hat{\sigma_3}$ are Gaussian centered in zero, and of finite variance. If a time series is lognormal, then the slope of the linear regression β must converge toward $\sqrt{\frac{2}{\pi}}\sigma$.

Hence, we can compute the value of the test statistic as follows:

$$t^* = \frac{\beta - \sqrt{\frac{2}{\pi}}\sigma}{\mathrm{se}\left(\beta\right)}\,, \tag{11}$$

where the standard error of the slope of the linear regression β is as follows:

$$\mathrm{se}\left(\beta\right) = \frac{\sqrt{MSE}}{\sqrt{\sum_{i=1}^{n}\left(S_i - \bar{S}\right)^2}}\,, \tag{12}$$

where MSE is the mean square error of the linear regression which is the sum of square errors divided by $n-2$:

$$MSE = \frac{\sum_{i=1}^{n}\epsilon_i^2}{n-2}\,, \tag{13}$$

where ϵ_i are the error terms of the linear regression $y_i = \alpha + \beta x_i + \epsilon_i$, with n observations.

To compute t^* in eq. (11), we suggest using $\hat{\sigma_2}$. The p-value is determined by referring to a Student's t-distribution with t_{n-2} degrees of freedom. If the p-value is smaller than the significance level α, we reject the null hypothesis; if it is larger than α, we can conclude that the time series is lognormal at the significance level α.

VI. CONCLUSION

The present study presents a testing framework for the structural properties of the underlying stochastic process of a time series. This test is aimed at discriminating among stochastic processes, in particular among the lognormal random walk used in the Black-Scholes-Merton model, the Gaussian random walk used in the Ornstein-Uhlenbeck stochastic process, and the square-root random walk used the Cox, Ingersoll and Ross process. The test is based on three parameters: $\hat{\sigma_1}$, $\hat{\sigma_2}$ and $\hat{\sigma_3}$, where $\hat{\sigma_1}$ and $\hat{\sigma_2}$ are volatility estimators. For a lognormal random walk, $\hat{\sigma_3}$ converges toward $\hat{\sigma_1}$ and $\hat{\sigma_2}$. For a Gaussian random walk, $\hat{\sigma_3}$ converges toward 0; and for a square-root random walk $\hat{\sigma_3}$ converges toward half of $\hat{\sigma_1}$ or $\hat{\sigma_2}$. Finally, an α-level hypothesis test is provided to test for the lognormality of a time series. In conclusion, practitioners may find the present test useful for selecting the stochastic processes they use for contigent claim valuation and risk management. The test can be applied to any asset class.

VII. ACKNOWLEDGEMENTS

The author is grateful to Professor Anthony Davison from the Institute of Mathematics at the EPFL and Professor Thomas Mikosch from the Department of Mathematics at the University of Copenhagen for their suggestions and improvements to the manuscript.

[1] D. Becker, Transactions of Society of Actuaries **43**, 7 (1991).

[2] E. Lehmann and J. Romano, "Testing statistical hypotheses," (Springer, 2010) pp. 584–599.

[3] S. Rachev, S. Stoyanov, A. Biglova, and F. Fabozzi, "Data analysis and decision support," (Springer Berlin Heidelbergr, 2005) pp. 269–281.

[4] S. Rachev, S. Stoyanov, C. Wu, and F. Fabozzi, Annals of Economics and Finance **8**, 21 (2007).