Special Issue in honor of John W. Neuberger
Electronic Journal of Differential Equations, Special Issue 02 (2023), pp. 193–207.
ISSN: 1072-6691. URL: https://ejde.math.txstate.edu or https://ejde.math.unt.edu

DETERMINING THE BACKGROUND DRIVING PROCESS OF THE ORNSTEIN-UHLENBECK MODEL

MARIA C. MARIANI, PETER K. ASANTE, WILLIAM KUBIN, OSEI K. TWENEBOAH, MARIA BECCAR-VARELA

Dedicated to the memory of Prof. John W. Neuberger

Abstract. In this work, we determine appropriate background driving processes for the 3-component superposed Ornstein-Uhlenbeck model by analyzing the fractal characteristics of the data sets using the rescaled range analysis (R/S), the detrended fluctuation analysis (DFA), and the diffusion entropy analysis (DEA).

1. Introduction

Ever since the Ornstein-Uhlenbeck model was proposed in 1930 by Leonard Ornstein and George Eugene Uhlenbeck, the technique has many areas of application including health care [47], nanotechnology, thermodynamics [8], geophysics [30], and finance [4, 25, 31, 41].

According to [1], the Ornstein-Uhlenbeck process is a natural model to consider in a biological context since it stabilizes around some equilibrium point. When Ornstein and Uhlenbeck proposed the model, it was an alternative to the Brownian Motion, and thus in its original presentation, the background driving process (BDP) was a standard Brownian Motion.

Unlike its original proposition which involved a Brownian motion as its background driving process, there have been many extensions or modifications to it to truly capture the behavior of the data set, which otherwise could not be modeled rightly with Brownian motions [27, 37]. Empirical results have shown evidence of non-Brownian behavior in many real world complex systems [12, 31, 32]]. In fact, according to [12] statistics of Lévy type is a ubiquitous phenomenon observed in a wide variety of areas, including physics, seismology, and engineering, to mention a few. Lévy motions constitute one of the essential and fundamental families of random movements which have stationary and independent increments. This means that the distribution of the increments is the same for any time interval, and that the increments are statistically independent from each other.

2020 *Mathematics Subject Classification.* 34k50, 34F50, 60H10.
Key words and phrases. Stochastic differential equation; Itô Calculus; Lévy Process; Ornstein-Uhlenbeck model; Superposed Ornstein-Uhlenbeck model; Gaussian process; Background driving process (BDP); Diffusion entropy analysis (DEA); long-range correlations; detrended fluctuation analysis (DFA); rescaled range analysis (R/S).

Published March 27, 2023.

One major challenge in utilizing the Ornstein-Uhlenbeck model is the ability to choose an appropriate BDP for the random process under consideration. As [31] and [14] have shown, the choice of the BDP has a significant influence on the performance of the model. In this work, our goal is to use the fractal characteristics of some stochastic processes to characterize them as Gaussian, Lévy walks, or Lévy flights.

If the characterization of the data is Gaussian, the appropriate BDP is the standard Brownian motion; if the data describes a Lévy walk or Lévy flight, the appropriate BDP is a Lévy process. This work considers a Gamma(a, b) process and an Inverse-Gaussian(a, b) process for our Lévy driven models. We compare results when a different BDP is used from our proposed BDP to see how the performance of the 3-component superposed Ornstein-Uhlenbeck model [33] is affected.

The outline of this article is as follows: In section 2 we introduce stochastic differential equations, the Ornstein-Uhlenbeck model, and the 3-component superposed Ornstein-Uhlenbeck model. In section 3, we present the parameter estimation approach for the 3-component superposed Ornstein-Uhlenbeck model. We then characterize the data in section 4. In section 5, we discuss the results obtained from our simulation using Matlab software. Finally, section 6 offers our conclusion and possible future work based on the results obtained.

2. Modeling with stochastic differential equations

2.1. **Stochastic differential equations (SDEs).** Stochastic differential equations (SDEs) have been used for modeling different phenomena in economics and finance, physics, and population dynamics among others.

Definition 2.1. A stochastic differential equation (SDE) is a deterministic differential equation perturbed by random noise.

In general, an SDE can be formulated as

$$dX_t = f(t, X_t)dt + g(t, X_t)dB_t, \tag{2.1}$$

where x_t is a stochastic process, and B denotes a Wiener process (standard Brownian motion). For this stochastic differential equation, the random noise dB_t is also termed its background driving process. From the above equation we get the corresponding integral equation

$$X_{t+s} - X_t = \int_t^{t+s} f(v, X_v)dv + \int_t^{t+s} g(v, X_v)dB_v, \tag{2.2}$$

we see that equation 2.2 defines the continuous time stochastic process X_t as the sum of an ordinary Lebesgue integral and an Itô integral.

The theory of differential equations is the origin of classical calculus and motivated the creation of differential and integral calculus. A differential equation is an equation involving an unknown function and its derivative. Typically, a differential equation is a functional relationship

$$f(t, x(t), x'(t), x''(t), \dots) = 0, \quad 0 \le t \le T \tag{2.3}$$

involving the time t, an unknown function $x(t)$ and its derivative. The solution of the differential equation is a function $x(t)$ which satisfies (2.3).

Now, consider the deterministic differential equation

$$dx(t) = a(t, x(t))dt, \quad x(0) = x_0. \tag{2.4}$$

The easiest way to introduce randomness in this equation is to randomize the initial condition. The solution $x(t)$ then becomes a stochastic process $(X_t, t \in [0,T])$ defined as

$$dX_t = a(t, X_t)dt, \quad X_0(t) = Y(t). \tag{2.5}$$

Equation (2.5) is called a random differential equation. Random differential equations can be considered deterministic equations with a perturbed initial condition. Note that this is not a full stochastic differential equation.

In general, a stochastic differential equation is a differential equation in which one or more of the terms is a stochastic process; the solution will be also a stochastic process.

2.2. **Itô calculus.** As we have observed above, in solving SDEs we encounter integrands and integrators that are stochastic processes. Thus, the classical calculus methods cannot be used in solving SDEs. Kiyosi Itô, a Japanese mathematician, advanced the solutions of SDEs by developing the theory of Itô calculus, which is an extension of the methods of calculus to stochastic processes. See [24] for further readings. Before stating the properties of the Itô integral, we first define some important terms.

Definition 2.2. A probability space is a triplet $(\Omega, \mathcal{F}, P)$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω and P is a probability measure $P : \mathcal{F} \to [0,1]$.

Definition 2.3. A continuous filtration of a set Ω is a collection $\mathcal{F}_t$ of σ-algebra subsets of Ω such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$. For $n \in \mathbb{N}$, a discrete filtration $\mathcal{F}_n$ is similarly defined.

Definition 2.4. An $\mathcal{F}_t$-measurable random variable is a random variable whose value is known at time t.

Definition 2.5. A sequence $\{X_t\}$ of random variables is an adaptive process relative to $\{\mathcal{F}_t\}$, if the random variable $\{X_t\}$ is $\{\mathcal{F}_t\}$-measurable for each t.

We recall that a stochastic process $V = V(S,T)$ is a collection of random variables $V = V_t : t \in T$, where each V_t is defined on a common probability space, and takes values in a common set S, called the state space. The collection is indexed or labeled by a set T, that in general is N or $[0,\infty)$. If $T = N$ the process is called discrete, and if $T = [0,\infty)$ it is called continuous.

Given a stochastic process $V(S,T)$ and a Brownian motion B_t, let $f, g \in V(S,T)$ and $0 \le S < U < T$. Then the following properties hold for the Itô integral:

(i) $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$;
(ii) $\int_S^T (cf + g) dB_t = c\int_S^T f dB_t + \int_S^T g dB_t$, for $c \in \mathbb{R}$;
(iii) $\mathbb{E}\big[\big(\int_s^T f dB_t\big)\big] = 0$;
(iv) $\int_S^T f dB_t$ is $\{\mathcal{F}_T\}$-measurable.

Another essential property of the Itô integral is the fact that it is a martingale. We recall that a martingale is a stochastic process or sequence of random variables such that at a given time, the conditional expectation of the next value in the sequence is equal to the present value, independently of all prior values.

2.3. **Example.** In this section, we present a classic application of SDEs in Finance.

2.3.1. *Modeling asset prices.* The SDE that models the evolution of an asset price is

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

The above equation is in the class of SDEs formulated as

$$dX_t = f(t, X_t)dt + c(t)X_t dB_t.$$

Taking $f(t, X_t) = \mu X_t$ and $c(t) = \sigma$, we have that

$$F_t = \exp(-\int_0^t \sigma dB_s + \frac{1}{2}\int_0^t \sigma^2 ds) = \exp(-\sigma B_t + \frac{1}{2}\sigma^2 t)$$

is an integrating factor for the SDE that transforms it into a deterministic differential equation. Setting $Y_t = F_t X_t$ we have

$$dY_t = d(F_t X_t) = F_t f(t, X_t)dt = \mu F_t X_t dt = \mu Y_t dt.$$

Therefore,

$$\frac{dY_t}{Y_t} = \mu dt.$$

By integrating both sides, we obtain $\log Y_t = \mu t + C$, or equivalently,

$$Y_t = e^{\mu t + C} = Ae^{\mu t}.$$

Furthermore, from this last equation, we can obtain X_t as

$$X_t = Ae^{\frac{\mu t}{F_t}} = Ae^{\mu t}e^{\sigma B_t - \frac{1}{2}\sigma^2 t}.$$

2.4. **Ornstein-Uhlenbeck model (OU).** Having introduced SDEs and some essential tools used in solving them, let us discuss the Ornstein-Uhlenbeck model and the 3-component superposition of the Ornstein-Uhlenbeck model (OU model). We begin by defining a Markov process.

Definition 2.6. A Markov process, also known as a Markov chain, is a stochastic process that has the Markov property, which means that the future state of the process depends only on its present state and not on its past states. In other words, given the present state, the probability of moving to any future state is independent of how the process arrived at its present state.

The Gaussian OU process is the solution to the stochastic differential equation

$$dX_t = \lambda(m - X_t)dt + \alpha dB_t, \quad t > 0, \tag{2.6}$$

where λ, m, and α are real constants and B_t is a standard Brownian Motion on $\mathbb{R}$. The initial value X_0 is a random variable independent of $(B_t)_{t\geq 0}$. We can show that the stochastic integral

$$X_t = m(1 - e^{-\lambda t}) + \alpha^{-\lambda t}\int_0^t e^{\lambda s}dB_s + X_0 e^{-\lambda t}, \quad t \geq 0, \tag{2.7}$$

satisfies (2.6) for any λ, m, α and choice of X_0. The solution X as defined in equation (2.7) is the unique, strong Markov solution to (2.6) see [27].

As stated earlier, empirical results have shown that many financial stock indexes deviate from normalcy. Hence, modeling with the ordinary Gaussian OU may result in poor forecasts. A modification of the Gaussian OU model through replacing the Weiner process in (2.6) with a Lévy process has been developed and applied in

various literature. To define the Lévy OU process, let (ϵ, η) be a bivariate Lévy process and define

$$X_t = m(1 - e^{-\epsilon_t}) + e^{-\epsilon_t} \int_0^t e^{-\epsilon_s} d\eta_s + X_0 e^{-\epsilon_t}, \quad t \geq 0, \tag{2.8}$$

where X_0 is independent of $(\epsilon_t, \eta_t)_{t\geq 0}$ and assumed $\mathcal{F}_0 - measurable$. Equation (2.8) is the generalized OU model (GOU) and seems to have been first considered by Carmona, Petit, and Yor (1997) as well as being implicit in Haan and Karandukar (1989) [27, 9, 10].

Now, from equation 2.7, we set $m = 0$, $\alpha = 1$ and replace B_t with a Lévy process $Z_{\lambda t}$ to get the SDE

$$dX_t = -\lambda X_t dt + dZ_{\lambda t}, \quad X_0 > 0, \quad \lambda \in \mathbb{R}^+. \tag{2.9}$$

The OU process below is the solution for equation (2.9):

$$X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dZ_{\lambda s}. \tag{2.10}$$

2.5. **Superposed Ornstein-Uhlenbeck model.** The OU process in (2.10) is redefined as a sum of m independent Ornstein-Uhlenbeck processes [33, 15, 36] as

$$X_t = \sum_{i=1}^{m} w_i e^{-\lambda_i t} X_0 + \int_0^t \sum_{i=1}^{m} w_i e^{-\lambda_i (t-s)} dZ_{\lambda_i s}, \tag{2.11}$$

where $\sum_{i=1}^{m} w_i = 1$.

For the purpose of this work, we consider a 3-component model of (2.11) which results in

$$X_t = w_1 X_{t_1} + w_2 X_{t_2} + w_3 X_{t_3},$$

with $\sum_i w_i \approx 1$.

3. Parameter estimation

This section presents the estimation method for the λ parameters and the weights for the 3-component superposed Ornstein-Uhlenbeck model. With our data, we apply an iterative approach in estimating the λ values at different lags of the autocorrelation function. We then derive a matrix system that uses the autocorrelation function and estimated λ_i values to determine the weights (w_i).

3.1. **The 3-component model.** We recall that the autocorrelation function analyzes the similarity between a time series and a lagged version of itself, see [5].

We consider the autocorrelation function at lags k, $k + h_1$ and $k + h_2$:

$$\begin{aligned} \rho(k) &= w_1 e^{-\lambda_1 |k|} + w_2 e^{-\lambda_2 |k|} + w_3 e^{-\lambda_3 |k|} \\ \rho(k + h_1) &= w_1 e^{-\lambda_1 |k+h_1|} + w_2 e^{-\lambda_2 |k+h_1|} + w_3 e^{-\lambda_3 |k+h_1|} \\ \rho(k + h_2) &= w_1 e^{-\lambda_1 |k+h_1|} + w_2 e^{-\lambda_2 |k+h_1|} + w_3 e^{-\lambda_3 |k+h_2|} \end{aligned} \tag{3.1}$$

Assume $\lambda_1 = \lambda_2 = \lambda_3$, then from the first expression in equation (3.1)

$$\rho(k) = (w_1 + w_2 + w_3) e^{-\lambda_1 |k|} \tag{3.2}$$

$$\lambda_1 = -\frac{\log(\rho(k))}{|k|}. \tag{3.3}$$

From the second expression in equation (3.1),

$$\rho(k+h_1) = (w_1 + w_2 + w_3)e^{-\lambda_2|k+h_1|}, \tag{3.4}$$

$$\lambda_2 = -\frac{\log(\rho(k+h_1))}{|k+h_1|}. \tag{3.5}$$

From the third expression in equation (3.1),

$$\rho(k+h_2) = (w_1 + w_2 + w_3)e^{-\lambda_3|k+h_2|}, \tag{3.6}$$

$$\lambda_3 = -\frac{\log(\rho(k+h_2))}{|k+h_2|}. \tag{3.7}$$

For the 3-component OU model, h_1 and h_2 are shifts from lag k. To solve for the weights, we construct the matrix equation from (3.1) as follows:

$$A = \begin{pmatrix} e^{-\lambda_1|k|} & e^{-\lambda_2|k|} & e^{-\lambda_3|k|} \\ e^{-\lambda_1|k+h_1|} & e^{-\lambda_2|k+h_1|} & e^{-\lambda_3|k+h_1|} \\ e^{-\lambda_1|k+h_2|} & e^{-\lambda_2|k+h_2|} & e^{-\lambda_3|k+h_2|} \end{pmatrix}, \tag{3.8}$$

$$b = \begin{pmatrix} \rho(k) \\ \rho(k+h_1) \\ \rho(k+h_2) \end{pmatrix}, \tag{3.9}$$

$$W = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}. \tag{3.10}$$

From equations (3.8), (3.9), and (3.10), we can solve for the weights as: $W = A^{-1}b$ with $\sum W \approx 1$ and the inverse of A, where A^{-1} exists for an appropriately chosen lag.

3.2. **Ljung-Box Statistic.** We state the null (H0) and alternate (HA) hypothesis of the Ljung-Box test statistic as follows:

(H0) The residuals are independently distributed.
(HA) The residuals are not independently distributed; they exhibit serial correlation.

Independence of residuals means that the error terms of the dependent variable at different time points are not related to each other. This assumption is usually checked by examining the autocorrelation function (ACF) of the residuals. If the residuals are independent, the ACF will be close to zero for all lags. If the ACF shows significant values at one or more lags, it indicates that the residuals are not independent, and some kind of correlation exists. Hence, the goal of the Ljung-Box test is to determine whether the residuals in a time series model exhibit autocorrelation. A p-value greater than 0.05 indicates that there is not enough evidence to reject the null hypothesis that the residuals are independently distributed. Not rejecting the null hypothesis in the Ljung-Box test statistic is a desirable property for a time series model. It indicates that our model has captured all the systematic patterns in the data and that the residuals are random fluctuations around the mean. This is important for making accurate predictions and for conducting statistical inference, as we want to ensure that any relationships we observe are not spurious and are not driven by autocorrelated residuals.

3.3. **Autoregressive integrated moving average (ARIMA).** To ensure that the matrix A is invertible, there is the need to choose appropriate lags for each λ_i. Thus, we fit an ARIMA model to the data and perform an Ljung-box statistic of the fitted ARIMA model to select the most significant lag based on the p-value at a 5% significant level. The ARIMA model assesses the significance of one dependent variable in relation to other changing variables. We iterate the ARIMA model over some lags and perform the Ljung-Box on each fit to examine the null hypothesis of independence in our time series. We then select the lags that give a p-value > 0.05 as significant lags. The three lags chosen are then used in computing the λ_i's. This process is performed with the R-software. See [5, 17] for more information on the ARIMA model and Ljung-Box statistic.

4. Data characterization

In this section, we present the data that we will analyze and we apply the characterization approach in [32] to characterize them as Gaussian, Lévy walk, or Lévy flight.

4.1. **Data.** We will use three different types of data in the analysis of our model, namely the daily closing values of the Nasdaq, retrieved from Yahoo Finance, the Japan earthquake data in 2011 after the magnitude nine occurrence (AfterM9), and finally, a simulated fractional Brownian motion (FBM) using the function fbm from the R package *somebm*. For each data set, we use two-thirds of the data for estimating the parameters in our model. The remaining one-third is predicted with our model, using the estimated parameters. Error estimates are then computed between the true values and predicted values.

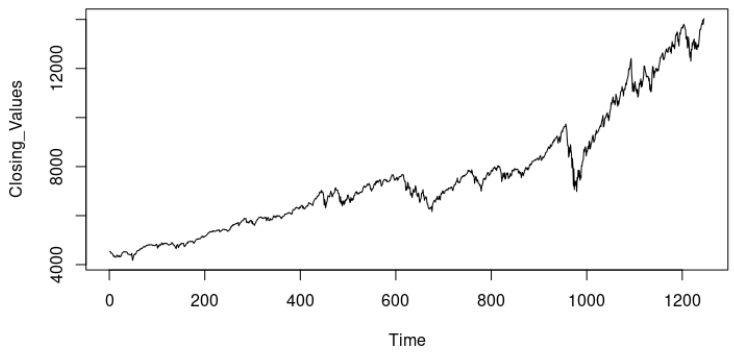

Figure 1. Time series plot of daily closing values of the Nasdaq index. Here, time is measured in days, and the closing value is in US dollars.

4.2. **Variance scaling methods.** This subsection briefly introduces the Rescaled Range Analysis, Detrended Fluctuation Analysis, and Diffusion Entropy Analysis.

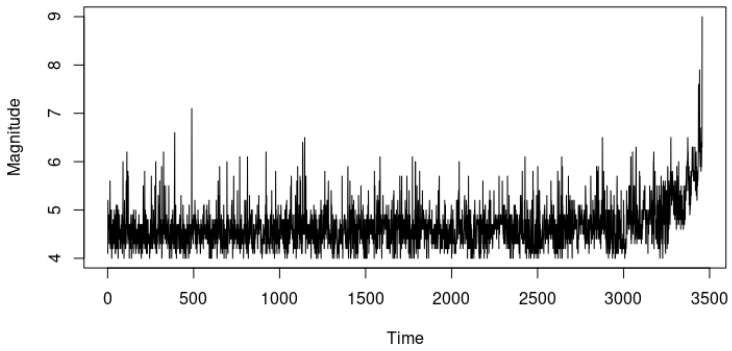

FIGURE 2. Time series plot of recorded earthquake magnitudes after magnitude nine event. Here, time is measured in minutes, and the magnitude is measured on the Richter scale (ergs).

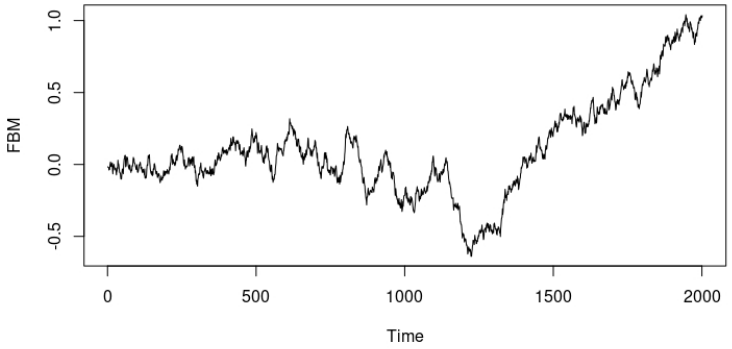

FIGURE 3. Time series plot of simulated fractional Brownian motion. Here, time is measured in days.

4.2.1. *Rescaled range analysis.* The Rescaled-Range analysis (R/S) was presented by Hurst in his study on the long-run variations of the water level of the Nile river [19].

Mandelbrot coined the name H for the parameter derived from this technique in tribute to the hydrologist Hurst and the mathematician Holder. The parameter H, also known as the index of dependence, represents the relative trend of a time series (i.e. persistence, anti-persistence or randomness) and always lies between 0 and 1; it is equal to $1/2$ in the case of processes with independent increments. Of particular interest for our work is the case in which $0.5 < H < 1$ since it is an indicator of long-range correlations [32]. However, because its sensitivity to abnormal values in the series, the rescaled range analysis method is unsuitable for analyzing long-range auto-correlation for non-stationary series [19].

4.2.2. *Detrended fluctuation analysis.* Peng et al. [38] proposed the Detrended Fluctuation Analysis (DFA) while examining a sequence of DNA nucleotides to study the self-similarity [35] and long-range dependence of time series. From the moment it was submitted to date, DFA has become a widely used method for determining fractal scaling properties and detecting long-range correlations in non-stationary time series. There are known applications in biology, meteorology, geophysics, and economics [34, 38, 46].

The principal advantage of the DFA lies in its ability to differentiate the intrinsic autocorrelations of the time series from those imposed by non-stationary external trends. The method focuses on the inherent structure of the correlations of market fluctuations at different time scales, leaving aside non-stationary trends [20].

The application of the DFA method allows obtaining a scaling exponent α from estimating the slope of the function $F(s)$ that measures the mean square deviation from an optimal linear approximation around the trend signal in segments of length s. The fluctuation function vs. s behaves as a power law. Therefore, it is possible to compute the value of the exponent α from the slope of the function in a log-log scale plot of $F(s)$ vs. s [35]. The DFA exponent α and the Hurst parameter H are related by

$$H = \begin{cases} \alpha & \text{if } 0 < \alpha < 1 \\ \alpha - 1 & \text{if } \alpha \geq 1. \end{cases} \tag{4.1}$$

4.2.3. *Diffusion entropy analysis.* Based on the direct evaluation of the Shannon entropy [16, 42, 44, 6], the Diffusion Entropy Analysis (DEA) is a probability density function (PDF) scaling method that perceives the numbers in a time series as the trajectory of a diffusion process [18]. The scaling property for the stationary time series takes the form

$$p(x,t) = \frac{1}{t^{\delta}} F(\frac{x}{t^{\delta}}), \tag{4.2}$$

where x denotes the diffusion variable, $p(x,t)$ is its PDF at time t, and $0 < \delta < 1$ is the scaling exponent. The scaling property for the non-stationary time series takes the form

$$p(x,t) = \frac{1}{t^{\delta(t)}} F(\frac{x}{t^{\delta(t)}}). \tag{4.3}$$

Note that for the stationary time series, the parameter δ in equation 4.2 is scalar, whereas for the non-stationary time series, the parameter $\delta(t)$ in 4.3 is a function of time [42]. We recall that a Lévy flight is a random walk in which the step-lengths follow a Lévy distribution. The Lévy distribution is a probability distribution that is heavy-tailed, see [29]. Lévy walks have been studied from 1937, see [26], they are special forms of random walks. Both Lévy flights and Lévy walks are named after the French mathematician Paul Lévy, see [40].

As derived in [42, 43], a diffusion process generated by Lévy walk characterization follows the relation

$$\delta = \frac{1}{3 - 2(H, \alpha)}. \tag{4.4}$$

The variables (H, α) in (4.4) indicates either H or α will be used in equation (4.4). Recall that H represents the scaling exponent from the rescaled range analysis, while α represents the scaling exponent from the detrended fluctuation analysis.

If $\delta = (H, \alpha) = 0.5$, the time series can be characterized by Fractional Brownian Motion (FBM), since the variance methods are based subtly on the Gaussian assumption [19, 20].

However, if $\delta \neq (H, \alpha)$ and equation (4.4) holds, then the time series can be characterized by Lévy statistics in particular a Lévy walk. Now, if $\delta \neq (H, \alpha)$ and equation (4.4) does not hold, then the time series can be characterized by a Lévy flight.

4.2.4. *Estimation procedure.* In this subsection, we describe the estimation technique for the scaling exponent δ. We first present a brief background on the Shannon Entropy used for estimating δ.

4.2.5. *Shannon entropy.* Rudolph Clausius developed the concept of entropy in 1865, a few years after he stated the laws of thermodynamics [21, 22]. Entropy is an indicator of the lack of information about the measure of an event that occurs with probability p, [21]. Other types of entropies are the Kolmogorov-Sinai entropy, the Renyi entropy and the Tsallis entropy [21, 42, 44]. The Shannon entropy measures information of a probability distribution as

$$S(t) = -\sum_{1}^{N} p_i \log p_i. \tag{4.5}$$

The summation is replaced by the integral in the case of continuous probability distributions. The above equation derives the log equation determining the DEA δ scaling. We present below the process for estimating δ:

- First, we transform the time series into a diffusion process. See [42, 43].
- We compute Shannon's Entropy of the diffusion process. A log-linear equation or log-quadratic equation is derived from the Shannon entropy by substituting equations $(4.2))and(4.3))$ respectively into equation (4.5). The reader is invited to read [3, 32, 42, 43] for further information on the Shannon entropy, transformation of time series into diffusion processes and the derivation of the shannon entropy for the stationary and non-stationary series. Now, simplifying the result from the substitutions, we have the relation for stationary time series given by

$$S(t) = A + \delta \log(t). \tag{4.6}$$

For the non-stationary series, the relation is

$$S(t) = A + \delta(t)\tau, \tag{4.7}$$

where $\delta(t) = \delta_0 + \eta \log(t)$, and $\tau = \log(t)$ with $\eta \log(t) < 1 - \delta_0$ and δ_0 and η are constants. Substituting $\delta(t)$ and τ gives us

$$S(t) = A + \delta_0 \log(t) + \eta \log^2(t).$$

We assume there exists some constant K multiplying $\log(t)$ such that $\eta = 1 - \delta_0$ and equation (4.7) becomes

$$S(t) = A + \delta_0 \log(t) - K \log(t) + (1 - \delta_0) \log^2(t).$$

This simplifies to

$$S(t) = A + (\delta_0 - K) \log(t) + (1 - \delta_0)(\log(t))^2. \tag{4.8}$$

Thus, by fitting a log-quadratic model in the non-stationary series and a log-linear model in the stationary series, we can determine the δ (or δ_0) scaling. At $t = 1$, it is clear that the constant A in both equations (4.6) and (4.7) is given by $S(1)$.

We derive the δ (or δ_0) by estimating the slope of the above linear-log equation or the coefficients from the quadratic-log equation. For details of the algorithm used when transforming the series into a diffusion process, we refer the reader to [35, 42].

4.3. **Scaling exponents.** Now, we apply the variance scaling methods (Rescaled Range Analysis and Detrended Fluctuation Analysis) and the PDF scaling method (Diffusion Entropy Analysis) to compute the scaling exponents of the data under consideration. We characterize the time series as Gaussian, Lévy walk, or Lévy flight using the scaling methods and equation (4.4) discussed earlier. The characterization approach of the time series has been extensively discussed in [32, 3].

TABLE 1. Scaling exponents for Financial time series obtained using the Rescaled range analysis, the Detrended fluctuation analysis and the Diffusion entropy analysis. The δ_{Levy}(R/S)= $\frac{1}{3-2H}$ and δ_{Levy}(DFA)= $\frac{1}{3-2\alpha}$ are obtained from equation 4.4

Data	R/S(H)	DFA (α)	DEA(δ)	δ_{Levy} (R/S)	δ_{Levy} (DFA)
AfterM9	0.3149	0.6518	0.77046	0.4219	0.5895
FBM	0.2928	0.4992	0.4986	0.4142	0.5
Nasdaq	0.5829	0.4298	0.4989	0.5452	0.4672

Remark 4.1. After running the characterization procedure developed in [32], we obtain the characterization of the Nasdaq, earthquake data, and simulated FBM as Lévy walk, Lévy flight, and Gaussian processes, respectively. Since all three scaling methods are obtained using numerical approximations, we may not get two scaling exponents to be the same. For example, for the simulated FBM, the α value obtained in the DFA is 0.4992, while the δ value obtained in the DEA is 0.4986. Both are approximately 0.5, but not equal; thus, we apply a $(-0.06, 0.06)$ adjustment to the scaling exponents obtained from our simulations. This adjustment interval is chosen arbitrarily. Furthermore, despite the fact that the computed $\delta_{Levy}(R/S)$ is approximately 0.5 after adjustments, the R/S method fails to compute the correct scaling exponent for the simulated FBM. The R/S method fails to correctly estimate the scaling exponent due to the non-stationarity of the simulated FBM.

5. RESULTS

This section presents the results from applying the proposed background driving process (BDP) in modeling with the 3-component superposed Ornstein-Uhlenbeck equation, i.e., the superposition of three Ornstein-Uhlenbeck processes (2.11) to solve the stochastic differential equation in (2.9). We further compare results to modeling the data with the two other processes discussed to ascertain the effect of the background driving process (BDP) on the model performance.

5.1. **Model simulation.** This section presents our simulation of the 3-component superposed Ornstein-Uhlenbeck model [33]. The solution of the proposed Ornstein-Uhlenbeck stochastic differential equation in 2.11 is used to simulate the process. We simulate our Ornstein-Uhlenbeck stochastic differential equation model solutions via the Brownian motion, Inverse-Gaussian(a, b) and the Gamma(a, b) process. For the Inverse-Gaussian(a, b), a is the mean and b is the rate parameter, while for the Gamma(a, b), a is the shape parameter and b is the rate parameter. We determine the model's performance by computing the root mean squared errors. Table 2 summarizes the numerical results of simulating each data set. We have utilized both R and Matlab software in the simulation of our model.

TABLE 2. Error estimates obtained from the model simulation of the remaining one third of each data set using the Inverse-Gaussian(a, b), the Gamma(a, b) and the Brownian motion as Background processes.

Data	Inverse-Gaussian(a, b)	Gamma(a, b)	Brownian Motion
Nasdaq	0.0286	0.25774	1.5939
AfterM9	0.0488	0.8037	1.3796
FBM	1.0525	1.4037	0.478

5.2. **Discussion.** The error estimates are compared in this section, with smaller error values implying better performance. Recall that in this study, the Lévy driven models use either a Gamma(a, b) process or an Inverse-Gaussian(a, b) process as the background driving process. The table above records the error estimates between the predicted values and true values. Each column represents the model being simulated with the respective background driving process in the first row. Now observing the results above, table 2 shows that for both the financial and earthquake data, the standard Brownian motion as a BDP leads to a poor performance (with error estimate of 1.5939 and 1.3796 respectively) of the 3-component superposed Ornstein-Uhlenbeck model, in comparison to the Lévy driven 3-component superposed Ornstein-Uhlenbeck model (with error estimates less than 1 for both IG(a, b) and Gamma(a, b)). Similarly, for the simulated FBM characterized by a Gaussian process, we see from table 2 that using the standard Brownian motion as a BDP leads to a better model (with error estimate of 0.478) performance compared to when a Lévy driven BDP (with error estimates greater than 1) is used. Finally, we observe that for the Lévy driven BDP, the Inverse-Gaussian(a, b) performs better than the Gamma(a, b). Thus, further investigation of the time series data is needed to classify the performance when modeling with a Lévy driven 3-component superposed Ornstein-Uhlenbeck equation since we do not have enough data to conclude that the IG(a, b) driven Ornstein-Uhlenbeck model will always perform better than the Gamma(a, b) driven Ornstein-Uhlenbeck model.

6. Conclusion

This work shows the importance of selecting appropriate background driving processes (BDP) when modeling time series with the 3-component superposed

Ornstein-Uhlenbeck stochastic differential equation (2.11)). We used the fractal characteristics of the time series data to classify them as Gaussian processes, Lévy walks, or Lévy flights. Next, we modeled each data set using the three background driving processes presented in this work (Standard Brownian motion, Inverse-Gaussian process, and Gamma process) which would help us determine an appropriate BDP for better model performance. Our results show that when the time series characterization is a Lévy process, a Lévy driven 3-component superposed Ornstein-Uhlenbeck model performs better. In contrast, the standard Brownian motion as a BDP performs better for a time series characterized by a Gaussian process. By observing the performance of the Inverse-Gaussian(a, b) and Gamma(a, b) Ornstein-Uhlenbeck processes, we conclude that further information about the time series data is needed to classify the performance of the Lévy driven 3-component superposed Ornstein-Uhlenbeck model, which we will be investigating in future works.

Acknowledgments. The authors are especially grateful to Prof. Alfonso Castro and the reviewers for their careful reading of the manuscript and their fruitful remarks.

This study was funded partially by the National Institute on Minority Health and Health Disparities (NIMHD) grant U54MD007592, and by the Department of Education (DoE) grant P120A220040-FY 2022 MSEIP.

References

[1] Aalen, O. O.; and Gjessing, H. K.; *Survival models based on the Ornstein-Uhlenbeck process.* Lifetime data analysis, 10 (2004), 407-423.

[2] Abadie, L. M.; *Current expectations and actual values for the clean spark spread: The case of Spain in the Covid-19 crisis.* Journal of Cleaner Production, 285 (2021), 124842.

[3] Asante, P. K.; *Lévy Processes: Characterizing Volcanic and Financial Time Series*(Doctoral dissertation, The University of Texas at El Paso), 2020.

[4] Barndorff-Nielsen, O. E.; Shephard, N.; *Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63 (2001) (2), 167-241.

[5] Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; *Time series analysis, forecasting and control.* Englewood Clifs, 1994.

[6] Brooks, C.; *A measure of persistence in daily pound exchange rates.* Applied economics letters, 2 (1995) (11), 428-431.

[7] Calafiore, G. C.; Novara, C.; Possieri, C.; *A time-varying SIRD model for the COVID-19 contagion in Italy.* Annual reviews in control, 50 (2020), 361-372.

[8] Caprini, L.; Marconi, U. M. B.; Puglisi, A.; Vulpiani, A. ; *The entropy production of Ornstein–Uhlenbeck active particles: a path integral method for correlations.* Journal of Statistical Mechanics: Theory and Experiment, 2019 (2019) (5), 053203.

[9] Carmona, P.; Petit, F.; Yor, M.; *On the distribution and asymptotic results for exponential functionals of Lévy processes.* Exponential functionals and principal values related to Brownian motion, (1997), 73-121.

[10] De Haan, L.; and Karandikar, R. L.; *Embedding a stochastic difference equation into a continuous-time process.* Stochastic processes and their applications, 32 (1989) (2), 225-235.

[11] Donado, F.; Moctezuma, R. E.; López-Flores, L.; Medina-Noyola, M.; and Arauz-Lara, J. L.; *Brownian motion in non-equilibrium systems and the Ornstein-Uhlenbeck stochastic process.* Scientific reports, 7 (2017) (1), 12614.

[12] Eliazar, I.; and Klafter, J.; *Lévy, Ornstein–Uhlenbeck, and subordination: Spectral vs. jump description.* Journal of statistical physics, 119 (2005), 165-196.

[13] Endres, S.; and St*ddot*ubinger, J.; *Optimal trading strategies for Lévy-driven Ornstein–Uhlenbeck processes.* Applied Economics, 51 (2019) (29), 3153-3169.

[14] Habtemicael, S.; SenGupta, I.; *Ornstein-Uhlenbeck processes for geophysical data analysis.* Physica A: Statistical Mechanics and its Applications, 399 (2014), 147-156.
[15] Habtemicael, S.; Ghebremichael, M.; and SenGupta, I.; *Volatility and variance swap using superposition of the Barndorff-Nielsen and Shephard type Lévy processes.* Sankhya B, 81 (2019), 75-92.
[16] Haubold, H. J.; Mathai, A. M.; and Saxena, R. K.; *Analysis of solar neutrino data from Super-Kamiokande I and II.* Entropy, 16 (2014) (3), 1414-1425.
[17] Hassani, H.; and Yeganegi, M. R.; *Selecting optimal lag order in Ljung-Box test.* Physica A: Statistical Mechanics and its Applications, 541 (2020), 123700.
[18] Huang, J.; Shang, P.; Zhao, X.; *Multifractal diffusion entropy analysis on stock volatility in financial markets.* Physica A: Statistical Mechanics and its Applications, 391 (2012) (22), 5739-5745.
[19] Hurst, H. E.; *Long-term storage capacity of reservoirs.* Transactions of the American society of civil engineers, 116 (1951) (1), 770-799.
[20] *https* : *//www.physionet.org/tutorials/fmnc/node*5.*html*
[21] *http* : *//www.ueltschi.org/teaching/chapShannon.pdf*
[22] *http* : *//galileo.phys.virginia.edu/classes/*152.*mf*1*i.spring*02*/Entropy.pdf*
[23] *https* : *//en.wikipedia.org/wiki/Ornstein*%*E*2%80%93*Uhlenbeck*$_p$*rocess*
[24] *https* : *//en.wikipedia.org/wiki/It*%*C*3%*B*4*_calculus#It*%*C*3%*B*4*_processes*
[25] Janczura, J.; Orzel, S.; Wylomańska, A.; *Subordinated α-stable Ornstein-Uhlenbeck process as a tool for financial data description.* Physica A: Statistical Mechanics and its Applications, 390 (2011) (23-24), 4379-4387.
[26] Lévy, P.; *Théorie de l'addition des variables aléatoires, Gauthier-Villars, Paris.*, 1954.
[27] Maller, R. A.; Müller, G.; Szimayer, A.; *Ornstein–Uhlenbeck processes and extensions.* Handbook of financial time series, (2009), 421-437.
[28] Mandal, M.; Jana, S.; Nandi, S. K.; Khatua, A.; Adak, S.; Kar, T. K.; *A model based study on the dynamics of COVID-19: Prediction and control.* Chaos, Solitons and Fractals, 136 (2020), 109889.
[29] Mantegna, R. N.; Stanley, H. E.; *Introduction to econophysics: correlations and complexity in finance.* Cambridge university press, 1999.
[30] Mariani, M. C.; Tweneboah, O. K.; *Stochastic differential equations applied to the study of geophysical and financial time series.* Physica A: Statistical Mechanics and its Applications, 443 (2016), 170-178.
[31] Mariani, M.C; Tweneboah, O. K.; *Modeling high frequency stock market data by using stochastic models.* Stochastic Analysis and Applications, 40 (2022) (4), 573-588.
[32] Mariani, M. C.; Asante, P. K.; Bhuiyan, M. A. M.; Beccar-Varela, M. P.; Jaroszewicz, S.; Tweneboah, O. K.; *Long-range correlations and characterization of financial and volcanic time series.* Mathematics, 8 (2020) (3), 441.
[33] Mariani, M. C.; Asante, P. K.; Tweneboah, O. K.; Kubin, W.; *A 3-component superposed Ornstein-Uhlenbeck model applied to financial stock markets.* Research in Mathematics, 9 (2022) (1), 2024339.
[34] Mariani, M. C.; Kubin, W.; Asante, P. K.; Guthrie, J. A.; Tweneboah, O. K.; *Relationship between Continuum of Hurst Exponents of Noise-like Time Series and the Cantor Set.* Entropy, 23 (2021) (11), 1505.
[35] Mariani, M. C.; Kubin, W.; Asante, P. K.; Tweneboah, O. K.; Beccar-Varela, M. P.; Jaroszewicz, S.; Gonzalez-Huizar, H.; *Self-Similar Models: Relationship between the Diffusion Entropy Analysis, Detrended Fluctuation Analysis and Lévy Models.* Mathematics, 8 (2020) (7), 1046.
[36] Nicolato, E.; Venardos, E.; *Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type.* Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 13 (2003) (4), 445-466.
[37] Obuchowski, J.; Wylomanska, A.; *Ornstein-Uhlenbeck Process with Non-Gaussian Structure.* Acta Physica Polonica B, 44(5).
[38] Peng, C. K.; Buldyrev, S. V.; Havlin, S.; Simons, M.; Stanley, H. E.; Goldberger, A. L.; *Mosaic organization of DNA nucleotides.* Physical review e, 49 (1994) (2), 1685.
[39] Punzo, A.; *A new look at the inverse Gaussian distribution with applications to insurance and economic data.* Journal of Applied Statistics, 46 (2019) (7), 1260-1287.

[40] Reynolds, A. M.; *Current status and future directions of Lévy walk research.* Biology open, 7 (2018) (1), bio030106.
[41] Salmon, N.; SenGupta, I.; *Fractional Barndorff-Nielsen and Shephard model: applications in variance and volatility swaps, and hedging.* Annals of Finance, 17 (2021) (4), 529-558.
[42] Scafetta, N.; *An Entropic Approach to the Analysis of Time Series*, University of North Texas Libraries, 2003.
[43] Scafetta, N.; Grigolini, P.; *Scaling detection in time series: diffusion entropy analysis,* Physical Review E, 66 (2002) (3).
[44] Scafetta, N.; Latora, V.; Grigolini, P.; *Lévy scaling: The diffusion entropy analysis applied to DNA sequences,* Physical review. E, Statistical, nonlinear, and soft matter physics, 66 (2002), 031906. 10.1103/PhysRevE.66.031906.
[45] Shu, Y.; Feng, Q.; Kao, E. P.; Liu, H.; *Lévy-driven non-Gaussian Ornstein–Uhlenbeck processes for degradation-based reliability analysis.* IIE Transactions, 48 (2016) (11), 993-1003.
[46] Talkner, P.; Lutz, E.; H*ddot*anggi, P.; *Fluctuation theorems: Work is not an observable,* Physical review. E, 75 (2007), 032903.
[47] Tian, M. Y.; Wang, C. J.; Yang, K. L.; Fu, P.; Xia, C. Y.; Zhuo, X. J.; Wang, L.; *Estimating the nonlinear effects of an ecological system driven by Ornstein-Uhlenbeck noise.* Chaos, Solitons and Fractals, 136 (2020), 109788.
[48] Valdivieso, L.; Schoutens, W.; Tuerlinckx, F.; *Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type.* Statistical Inference for Stochastic Processes, 12 (2009), 1-19.

Maria C. Mariani
Department of Mathematical Sciences, the University of Texas at El Paso, TX 79968, USA
Email address: mcmariani@utep.edu

Peter K. Asante
Computational Science Program, the University of Texas at El Paso, TX 79968, USA
Email address: pkasante@miners.utep.edu

William Kubin
Computational Science Program, the University of Texas at El Paso, TX 79968, USA
Email address: wkubin@miners.utep.edu

Osei K. Tweneboah
Department of Data Science, Ramapo College of New Jersey, 505 Ramapo Valley Road, Mahwah, NJ 07430, USA
Email address: otwenebo@ramapo.edu

Maria Beccar-Varela
Department of Mathematical Sciences, the University of Texas at El Paso, TX 79968, USA
Email address: mpvarela@utep.edu