Modeling Chaotic Time Series using Stochastic Differential Equation

S.O.N. Agwuegbo*; E.E. Onugha; and A.A. Akintunde

Department of Statistics, University of Agriculture, Abeokuta.

E-mail: agwuegbo_son@yahoo.com

ABSTRACT

In this paper, attempts were made to build an appropriate model for the prediction of chaotic time series by using Langevin equation. Langevin equation is a linear stochastic differential equation related to the world of time series and is called Ornstein-Uhlenbeck process. The Ornstein-Uhlenbeck process is a Gaussian process with autocovariance and which can be transformed into state dependent time series model. The state dependent model can be reduced to an autoregressive integrated moving average (ARIMA) process or an autoregressive moving average (ARMA) process. The study of chaotic models is fascinating and this paper may contribute to the understanding of random behavior in time series modelling.

(Keywords: Langevin equation, stochastic differentia equation, Ornstein-Uhlenbeck process, dependent time series model)

INTRODUCTION

Many systems exhibit large changes in output corresponding to very small changes in initial conditions. Such systems are said to display chaotic behavior (Guttorp, 1995). Chaotic behavior exists in many natural systems, such as weather and climate and also occurs spontaneously in some systems with artificial components such as road traffic.

Chaos theory has provided a new tool in the understanding of behaviors that arises from physical systems and has applications in meteorology, physics, environmental science, computer science, engineering, economics and finance. The paradigm of chaos was introduced by Lorenz (1963), with several attempts having been made in several fields to demonstrate basic characteristics of chaotic behavior, such as the irregularity of motion, unpredictability and sensitivity to initial conditions (Takens, 1981).

Chaos theory as a branch of mathematics focuses on the behavior of dynamical systems that are highly sensitive to initial conditions. Chaos theory is an interdisciplinary theory stating that within the apparent randomness of chaotic complex systems, there are underlying patterns, constant feedback loops, repetition, self-similarity, fractals, self-organization, and reliance on programming at the initial point known as sensitive dependence on initial conditions.

Chaos theory has shown that complex erratic behavior in physical systems sensitive to initial conditions will generally magnify through time rather than die out. This is exemplified in the so-called butterfly effect, whereby a butterfly flapping its wings could produce an effect that is eventually transformed into a tropical storm. The sensitivity to initial conditions (that is, the rate at which a small perturbation is magnified) is measured by the Lyapunov exponent (Chatfield, 2004). These behaviors can be studied through analysis of a chaotic mathematical model, or through analytical techniques such as recurrence plots and Poincare maps.

Taken (1981) has shown that a chaotic dynamical system can be accurately reconstructed from a sequence of observations of the state of a dynamical system. Differential equations are used in sciences to model dynamic processes. They provided the basic simple model of any phenomenon in which one or more variables depend continuously on time without any random influences (Glendinning, 1994). One of the most exciting developments in recent theory of differential equations is the discovery that relatively simple differential equations can have solutions which are much more complicated than periodic and quasi periodic solutions. A differential equation is said to be chaotic if there are bounded solutions which are neither periodic nor quasi-periodic and which diverge from each other locally (Glendinning, 1994).

The existence of chaotic solutions has had a profound effect on thinking in many disciplines. One immediate corollary of the local divergence of nearby solutions is that one losses predictive power in practical situations. The solutions of differential equations are deterministic in the sense that if the initial conditions are precisely specified then the solution is completely determined and so, in principle one should be able to predict the value of the solution at some later time (Glendinning, 1994). Of course in practice, the initial condition can only be known to some finite precision and if the equation is chaotic, information about the system rapidly may be lost, since solution through the approximate initial condition does not stay close to the desired solution.

In Hooker (2009), recent research has seen a significant increase in interest in fitting nonlinear differential equation to data. Many systems of differential equations used to describe real world phenomena are developed from first-principles approaches by a combination of conservation laws and rough guesses. Such a priori modelling has led to proposed models that mimic the qualitative behavior of observed systems, but have poor quantitative agreement with empirical measurements. The problem of modeling dynamics of nonlinear systems has become an active field due to its potential applications especially in finance and engineering and is viewed as a realization from stochastic process of a nonlinear dynamical system. In the stochastic formulation, the dynamic behavior is modeled as a stochastic differential equation.

Stochastic differential equations can be understood as deterministic differential equations which are perturbed by random noise. The term stochastic differential equation was actually introduced by Bernštein in the limiting study of a sequence of Markov chains arising in a stochastic difference scheme (Karatzas and Shreve , 1994). Since the early work of Itǒ and Gihman, the interest in the methodology and the mathematical theory of stochastic differential equations has enjoyed remarkable successes. The constructive and intuitive nature of the concept, as well as its strong physical appeal has been responsible for its popularity among applied sciences.

MATERIALS AND METHODS

Chaotic time series is a complex nonlinear dynamical system that is specified by a state vector and a function which describes how the system evolves over time. The state vector is a list of numbers which may change as time progresses and is a numerical description of the current configuration of the system. The function is a rule which shows how the system changes over time. The mathematical representation is usually a set of dynamical differential equations, with unique solutions. The main idea is evident in the simplest possible system, the doubling map, sometimes called the one-sided Bernoulli shift:

$$X_{n+1} = \begin{cases} 2X_n & 0 \le X_n \le \frac{1}{2} \\ 2X_n - 1, & \frac{1}{2} \le X_n \le 1 \end{cases} \quad \textbf{(1)}$$

or by the logistic map, sometimes called the quadratic map:

$$\mathrm{X}_{n+1} = \lambda\, \mathrm{X}_n (1 - \mathrm{X}_n) \qquad 0 \le \lambda \le 4 \quad \textbf{(2)}$$

Equations (1) and (2) are examples of what mathematicians would call a difference equation, but which statisticians would probably regard as a deterministic time series. Of course in practice, the initial condition can only be known to some finite precision and if the equation is chaotic, information about the system rapidly may be lost, since solution through the approximate initial condition does not stay close to the desired solution. In Hong (1996), a random time series is a stochastic process in discrete time index.

Roughly speaking, a random time series under time reversal can be a deterministic sequence, that is, iterates of a chaotic map. On the other hand, some iterates of a chaotic map can be a sample sequence from a random time series under time reversal. These links may be helpful in understanding the relationship between the random behavior of time series in probability and in chaos. There may be some advantage in considering the general linear stochastic differential equation.

Stochastic process in continuous times is defined as solution of stochastic differential equation. It is a special class of stochastic differential equations which have an explicit solution in terms of the coefficient functions of the underlying Brownian motion. Chaotic processes are random processes that can be described mathematically as a set of dynamical differential equations. Consider the deterministic differential equation:

$$dx(t) = a(t, x(t))dt, \qquad x(0) = x_0 \qquad \textbf{(3)}$$

The easiest way to introduce randomness in this equation is to randomize the initial condition. The solution $x(t)$ then becomes a stochastic process $(\mathrm{X}_t, t \in [0,T]$:

$$d\mathrm{X}_t = a(t, \mathrm{X}_t)dt \qquad \mathrm{X}_0(\omega) = \mathrm{Y}(\omega) \quad \textbf{(4)}$$

Equation (4) is a random differential equation. Random differential equations (Mikosch, 2008) can be considered as deterministic differential equations with a perturbed initial condition. For our purpose, the randomness in the differential equation is introduced through an additional random noise term:

$$d\mathrm{X}_t = a(t, \mathrm{X}_t)dt + b(t, \mathrm{X}_t)d\mathrm{B}_t,$$
$$\mathrm{X}_0(\omega) = \mathrm{Y}(\omega) \qquad \textbf{(5)}$$

In (5), $\mathrm{B} = (\mathrm{B}_t, t \geq 0)$ denotes Brownian motion, and $a(t,x)$ and $b(t,x)$ are deterministic functions. The solution X, if it exists, is then a stochastic process.

The randomness of $\mathrm{X} = (\mathrm{X}_t, t \in [0,T]$ results, on one hands, from the initial condition, and on the other hand, from the noise generated by Brownian motion. Since Brownian motion does not have differentiable sample paths, we can propose (5) as an $\mathrm{It\hat{o}}$ calculus by interpreting it as the stochastic integral equation:

$$\mathrm{X}_t = \mathrm{X}_0 + \int_0^t a(s, \mathrm{X}_s)ds + \int_0^t b(s, \mathrm{X}_s)d\mathrm{B}_s$$
$$0 \leq t \leq T \qquad \textbf{(6)}$$

Equation (6) is an $\mathrm{It\hat{o}}$ stochastic differential equation. Brownian motion is the driving process of the $\mathrm{It\hat{o}}$ stochastic differential equation (Miskosch, 2008). There are two kinds of solution to (6) called strong and weak solutions. A strong solution to (6) is a stochastic process $\mathrm{X} = (\mathrm{X}_t, t \in [0,T]$, based on the path of the underlying Brownian motion. Weak solutions, Chung and Williams (1990) are sufficient in order to determine the distributional characteristics of $\mathrm{X} = (\mathrm{X}_t, t \in [0,T])$, such as the expectation, variance and covariance functions of the process.

A strong or weak solution X of the $\mathrm{It\hat{o}}$ stochastic differential equation (6) is called diffusion. In particular, (6) is a diffusion process if we take $a(t,x) = 0$ and $b(t,x) = 1$.

The solution to an $\mathrm{It\hat{o}}$ stochastic differential equation can also be derived as the solution of a deterministic partial differential equation. By considering the Ornstein-Uhlebeck process, we have another linear stochastic differential equation

$$t \in [0,T] \qquad \textbf{(7)}$$

Langevin (1908) studied this kind of stochastic differential equation to model the velocity of a Brownian particle. In the physical literature, the random forcing in (7) is called additive noise which is an adequate description of this phenomenon. This time series model can be considered as a discrete analogue of the solution to the Langevin equation (7) and the Langevin equation is a linear itô stochastic differential equation. To solve (7), the following transformation of X is convenient:

$$\mathrm{Y}_t = e^{-ct}\mathrm{X}_t \qquad \textbf{(8)}$$

We have to note that both process X and Y satisfy the same initial condition that:

$$X_0 = Y_0 \quad \textbf{(9)}$$

Hence the Langevin stochastic differential equations in (7) is an Ornstein-Uhlenbeck process given as:

$$X_t = e^{ct}X_0 + \sigma e^{ct}\int_0^t e^{-cs} dB_s \quad \textbf{(10)}$$

Equation (10) is actually a solution to (7). The Ornstein-Uhlenbeck process in (10) is a Gaussian process. The stochastic process X_t follows a random walk and can be represented as:

$$X_{t+1} = \phi X_t + Z_t \quad \textbf{(11)}$$

with a constant c and white noise a_t. If c is not zero then the variables, $X_t - X_{t-1} = c + a_t$ have a non-zero mean and is called a random walk with a drift. The state space model sets out to capture the salient features of the time series and these are apparent from the nature of the series. The state space model can be reduced to an autoregressive integrated moving average (ARIMA) process or an autoregressive moving average (ARMA) process. ARIMA (p,d,q) models are typically parsimonious model. ARIMA model selection is based on the premise that the ACF and the related statistics can be accurately estimated and are stable over time. By adopting Box ~ Jenkins ARIMA (p,d,q) model approach to time series analysis, model identification, parameter estimation and diagnostic check were feasible.

RESULTS

The empirical data used in this study is the claims submitted to the secretariat of the Nigeria Insurance Association (NIA) for the period 2007 and is as shown in Table 1. The examination of the time plot of Table 1 revealed greater variability of claims as shown in Figure 1.

Table 1: Insurance Claims Portfolio

Jan	4469654, 991698, 1243344, 513000, 522473, 3800000, 744538, 610536, 900000, 573750, 2025000, 570978, 542400, 2592000, 574536, 705682, 719059, 933038, 696173, 665766, 581487, 700000, 750000, 661000
Feb	684068, 750988, 3401510, 1389917, 1113524, 623396, 3036377, 2334479, 2208789, 1107032, 1050000, 691572, 612716, 537750, 1072512, 1312500, 838080, 12831900, 10159375, 1516854, 500000, 7000000, 1682558, 1012500
Mar	1700000, 899965, 513000, 3000000, 1066400, 5772690, 629803, 819500, 1923000, 720000, 1624500, 2183298, 1133930, 1680120, 1938070, 1163191
Apr	1091250, 1327784, 3066729, 17093431, 645188, 612000, 586528, 6277603, 864374, 3278926, 1260000, 738000
May	678000, 577776, 869660, 501785, 777600, 965037, 1260000, 1304394, 1800000, 1083598, 1177808, 1310143, 1298097, 1968740, 760089, 2700000, 4302997, 1224172, 643204, 1344823, 2588500, 3165761.
Jun	532010, 1495237, 742808, 724693, 1176641, 7000000, 2109000, 810000, 699510, 531644, 966845, 2618136, 1230618, 3162224, 1081947, 666750, 1500000, 652500, 529376, 4850000, 878500, 520600, 3298464, 1980000, 679115, 632591, 540000.
Jul	627838, 723050, 504900, 2263523, 606000, 1729917, 1950000, 704660, 1000000, 2479351, 1417898, 500000, 810000, 1408603, 1256584, 1620000, 540000, 522969, 717039, 982816, 4250000, 700000.
Aug	760000, 2353302, 546826, 531451, 823125, 515735, 1364993, 1030228, 1393273, 3244220, 1080000, 1044000, 24400000, 1075284, 1070244, 1197000, 995663, 761846, 4254817, 1162800, 24579269.
Sep	1338750, 1338750, 6300000, 3179432, 800650, 891950, 4454095, 2307436, 559558, 559545, 2063844, 2685354, 1246300, 1245983, 4884223, 857719, 566820, 631125, 1648438, 832733, 3254900, 2061216, 1085797.
Oct	5235988, 688500, 1411242, 2607147, 1530000, 800000, 1620000, 1067600, 826350, 1982973, 576000, 1381026, 6697192, 3265331, 3222164, 1238226, 828800, 1657500, 14552619, 1121850, 842387, 728946, 3734997, 1341743, 546950, 1134488, 544266, 1351500, 562002, 1851600, 1823018, 3054268.
Nov	1048478, 1625570, 3886258, 3910305, 1313125, 2900000, 600750, 800000, 1026667, 14490580, 563170, 705093, 1792500, 2153730, 2920459, 643357, 7435587, 542500, 565213, 5178084, 5161160, 2207540, 513359, 746971, 1882850, 2089548, 680400, 553248, 914973, 1080000, 1346386, 27311939.
Dec	500000, 1042321, 765000, 1344823, 746971, 1339595, 3134790, 540510, 661500, 671700, 1768000, 3587542, 1051200, 1303154, 1298996, 544000, 1744652, 3017240, 3865360, 711461, 992062, 515800, 870795, 665000, 675000, 1080000, 3451391, 524846.

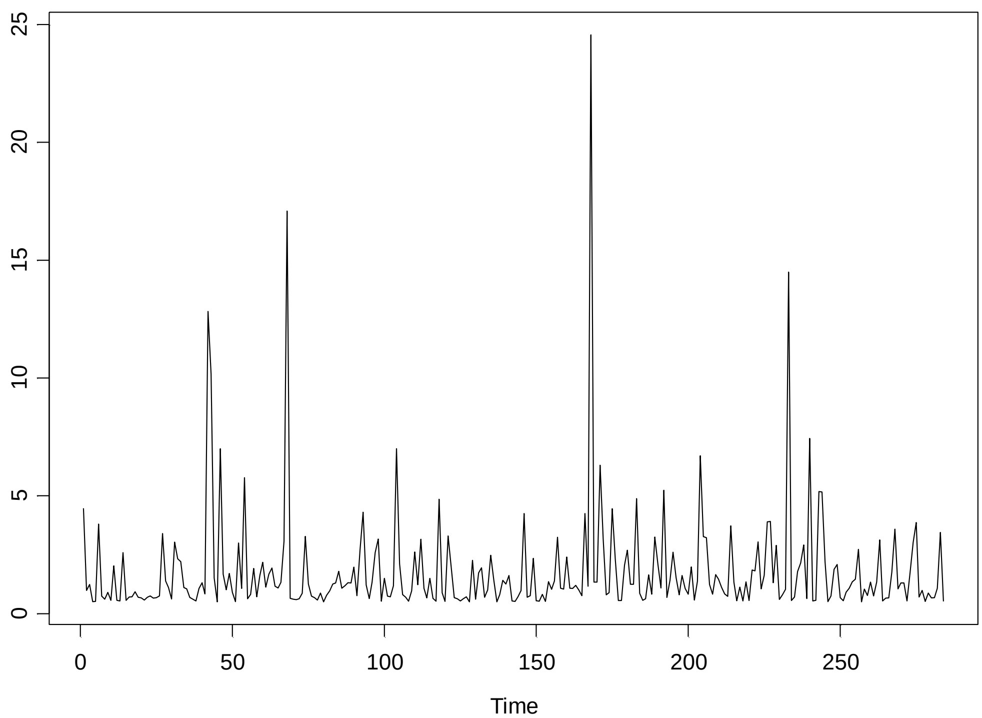

Figure 1: Sample Paths for the Claim Portfolio.

The path of the model as depicted in Figure 1 is more of a chaotic situation. There were lots of irregularities making it difficult to define the path of the process. The series can be understood as discretization of stochastic differential equations. The sample paths of the process fluctuate wildly in contrast to the limiting process of Brownian motion which has a continuous sample path. In order to avoid such irregular behavior, deconvolution of the data in Table 1 was carried out using Brownian sample paths, as in Figure 2.

By the examination of the ACF, the PACF, and the AIC for claim processes suggests an autoregressive model of order 1 as in Table 3.

The S-PLUS package used the Akaike information criterion to provide the best fit for an autoregressive model to a set of data. The correlogram for the ACF and the PACF is as in Figures 3 and 4.

The corresponding fitted autoregressive model is:

$$X_t = 2.36 + 0.578X_{t-1} + a_t$$

(0.066) (0.0031)

An overall test of model adequacy is provided by Ljung-Box chi-squared statistics. These statistics also known as the Box-Pierce chi-square statistics contain what are known as the portmanteau statistics with their associated p-values.

In fitting the AR (1) model to the structured claims data, none of the chi-square values is significant at the 5% level. The ARIMA model diagnostic is as shown in Figure 5 with various plots produced such as the standardized residuals, the ACF of the residuals, the PACF of the residuals, and the the p-values of Ljung-Box Chi-squared statistics.

Table 3: Sample ACF, PACF and AIC for Claims Portfolio.

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
ACF	0.482	0.299	0.207	0.104	0.140	0.060	0.049	0.066	0.058	0.015	0.123	0.053
PACF	0.482	0.087	0.045	-0.04	0.106	-0.06	0.020	0.033	0.024	-0.06	0.017	0.059
AIC	0.000	0.340	1.902	3.595	3.120	4.298	6.212	7.978	9.851	11.12	13.06	14.28

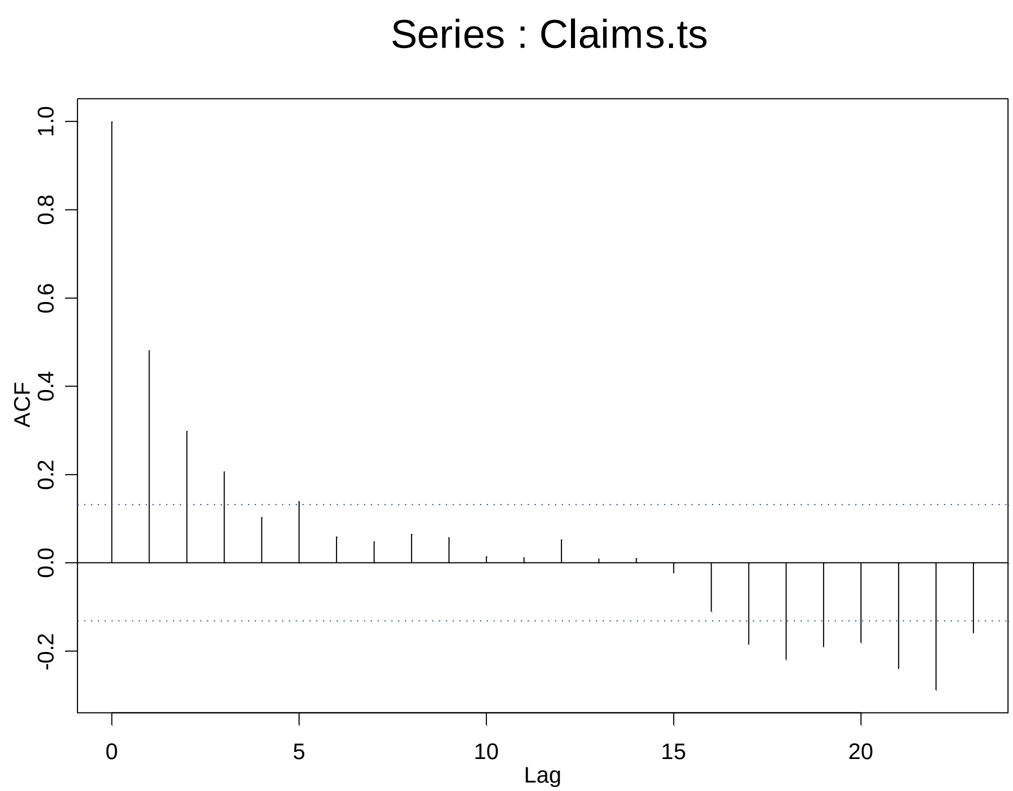

Figure 3: The Estimated Autocorrelations Function (ACF).

Almost all the plots are based on the examination of the residuals, $\hat{e}_t = y_t - \hat{y}_t$, where $\hat{y}_t$ is the fitted value, or some function of the residuals.

The rationale is that if the model is satisfactory, the residuals, which are estimates of the error components $\hat{e}_t$,should be uncorrelated at any lag and should be approximately normally distributed with mean zero and a variance estimated by the residual mean square. Thus the AR (1) model appears quite adequate.

DISCUSSION

The numerical solution of stochastic differential equations is a relatively new area of applied probability theory with an overviews in Kloeden and Platen (1992).

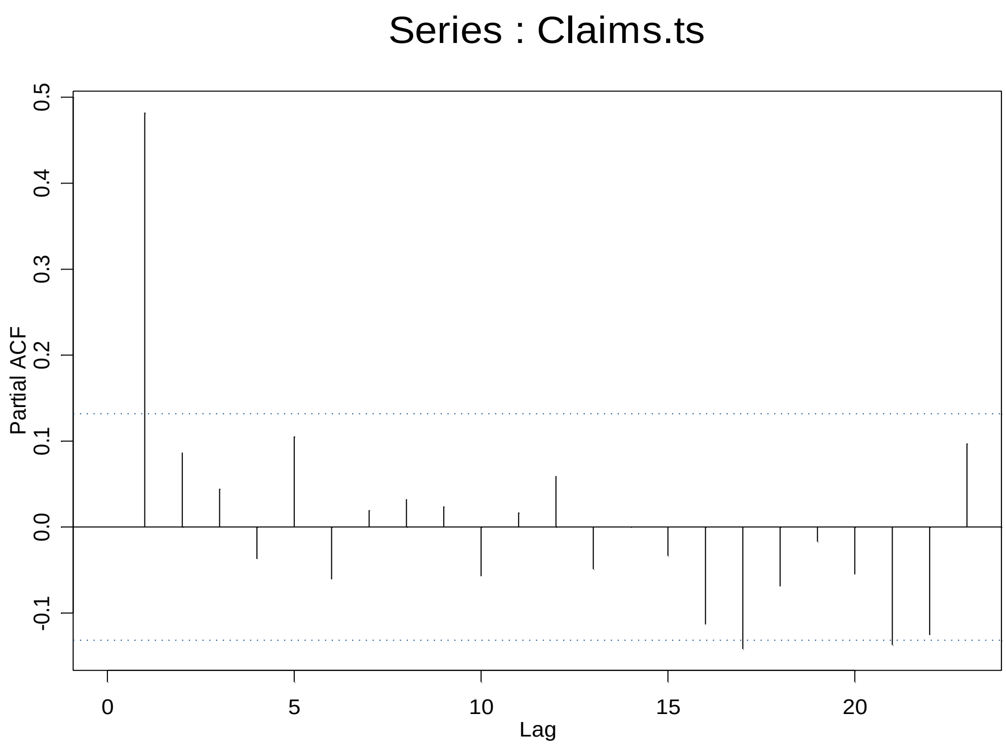

Figure 4: The Estimated Partial Autocorrelation Function (PACF).

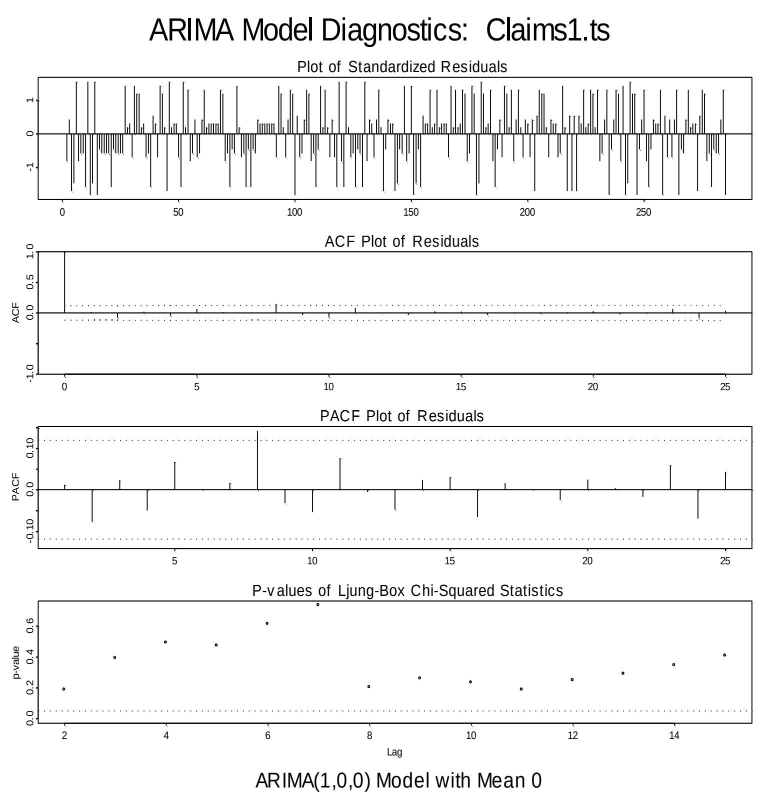

Figure 5: ARIMA Model Diagnostic for the Claims Portfolio

This study described chaotic evolution or Insurance claim process over a given time period as a Langevin equation. Langevin equations are one of the major classes of models in the response time domain. The Langevin equation is a linear stochastic differential equation related to the world of time series and is called Ornstein-Uhlenbeck process.

The Ornstein-Uhlenbeck process is a Gaussian process with autocovariance and which can be transformed into state space time series model. The state space model sets out to capture the salient features of the time series and these are apparent from the nature of the series. The state space model can be reduced to an autoregressive integrated moving average (ARIMA) process or an autoregressive moving average (ARMA) process. For the purpose, the study restricted to the numerical solution of the stochastic differential equation using stochastic differential equation with multiplicative noise.

REFERENCES

1. Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis, Forecasting and Control (rev.ed),* Holden-Day: San Francisco, CA.
2. Chatfield, C. 2004. *The Analysis of Times Series: An Introduction,6th ed*, Chapman and Hall/CRC: Boca Raton, FL.
3. Chung, K.L and R.J. Williams. 1990. *Introduction to Stochastic Integration, 2nd edition*. Birkhauser; Basel, Switzerland.
4. Glendinning, P. 1994. *Stability, Instability and Chaos: An Introduction to the Theory of Non-linear Differential Equations*. Cambridge University Press: Boston, MA.
5. Guttorp, P. 1995. *Stochastic Modelling of Scientific Data*. Chapman and Hall: Boca Raton, FL.
6. Hong, Z.A. 1961. “A Note on Chaotic Maps and Time Series”. Athens Conference on Applied Probability and Time Series Analysis. pp 15-26. Springer.
7. Hooker, G. 2009. “Forcing Function Diagnostics for Nonlinear Dynamics”. *Biometrics*. 65(3).
8. Lalley, S.P. and A.B. Nobel. 2006. “Denoising Deterministic Time Series”. *Dynamics of PDE*. 3(4): 259-279.
9. Lorenz, E.N. 1963. “Deterministic Nonperiodic Flow”. *Journal of the Atmospheric Sciences*. 20(2):130-141.
10. Karatzas, I. and E.S. Shreve. 1994. *Graduate Text in Mathematics, Brownian Motion and Stochastic Calculus,2nd ed.* Springer-Verlag: New York, NY.
11. Mikosch, T. 2008. “Elementary Stochastic Calculus With Finance in View”, *Advanced Series on Statistical Science and Applied Probability, World Scientific*, Princeton, NY.
12. Takens, F. 1981. “Detecting Strange Attractors in Turbulence”. *Lecture notes in Mathematics*. 898

SUGGESTED CITATION

Agwuegbo, S.O.N, E.E. Onugha, and A.A. Akintunde. 2020. “Modeling Chaotic Time Series using Stochastic Differential Equation”. *Pacific Journal of Science and Technology*. 21(1):93-100.

Pacific Journal of Science and Technology