Annals of Mathematics and Computer Science ISSN: 2789-7206
Vol 29 (2025) 1-19 https://doi.org/10.56947/amcs.v29.573

A FIRST-ORDER AUTOREGRESSIVE PROCESS WITH WEIGHTED LINDLEY INNOVATIONS AND ITS APPLICATIONS TO ENERGY AND FINANCIAL DATA

M. M. GABR[1], HASSAN S. BAKOUCH[2,3] AND HADEER M. EL-TAWEEL[4*]

Abstract. This paper presents an autoregressive process of order one in which the innovation term follows a weighted Lindley distribution. The fundamental characteristics of this process are carefully examined. The proposed process is applied to analyze time series data, using real-world datasets. The parameters of the considered model are estimated using the Gaussian estimation approach, and the study offers a comparison of the suggested model with some alternative models in the literature. Moreover, this study investigates data forecasting for the proposed model by applying the classical conditional expectation method alongside some machine learning algorithms to evaluate the prediction accuracy.

Keywords. Non-Gaussian innovations, Autoregressive model, Gaussian estimation, Forecasting
2020 Mathematics Subject Classification. Primary 62M10, 62M15

1. Introduction

In time series analysis, the linear autoregressive process of order one (AR(1)) model, which has the formula, $X_t = \rho X_{t-1} + \varepsilon_t$, is one of the important models for analyzing time series data. The choice of innovation distribution plays a critical role in capturing the underlying structure and behavior of the data. While traditional autoregressive models often assume normally distributed innovations, this assumption may not hold in many practical applications where the data exhibit skewness, heavy tails, or are constrained to positive values. To address these limitations, researchers have explored the use of more flexible innovation distributions, among which are exponential and gamma (Gaver and Lewis [5]), exponential (Andel [1]), student-t (Tiku et al. [16]; Christmas and Everson [3]; Nduka [11]), skew-normal (Sharafi and Nematollahi [14]), generalized hyperbolic (Ghasami et al. [6]), normal inverse Gaussian (Dhull et al. [4]), Lindley (Nitha and Krishnarani [12]). One such non-Gaussian distribution is the weighted Lindley (WL) distribution (Ghitany et al. [7]), a flexible and positively skewed

Date: Received: Jun 16, 2025; Accepted: Jul 12, 2025.
* Corresponding author

distribution introduced as a generalization of the classical Lindley distribution. The weighted Lindley distribution has gained popularity in different fields. Its versatility and analytical tractability make it a compelling choice for modeling innovations in time series models where normality is not appropriate.
Motivated by the versatile features of the WL, this paper aims to introduce an AR(1) process based on the weighted Lindley-distributed innovations, study its theoretical properties, estimate its parameters using Gaussian estimation techniques, and compare its performance with other established models using real-life datasets.

The probability density function (PDF) of the innovation random variable ε_t following the WL distribution with parameters $\alpha > 0$ and $c > 0$ is given by:

$$f(\varepsilon;\theta,c) = \frac{\theta^{c+1}}{(\theta+c)\Gamma(c)}\varepsilon^{c-1}(1+\varepsilon)e^{-\theta\varepsilon}, \quad \varepsilon > 0,$$

where $\Gamma(c)$ is the gamma function, c is the shape parameter, and θ is the scale parameter.
The mean, variance, and Laplace transform of the WL distribution are expressed, respectively, as

$$E(\varepsilon) = \frac{c(\theta+c+1)}{\theta(\theta+c)},$$

$$V(\varepsilon) = \frac{(c+1)(\theta+c)^2-\theta^2}{\theta^2(\theta+c)^2},$$

and

$$\Phi_\varepsilon(s) = \frac{\theta^{c+1}(\theta+c+s)}{(\theta+c)(\theta+s)^{c+1}}.$$

The remainder of this paper is structured as follows: In Section 2, AR(1) process with weighted Lindley innovations (AR-WL(1)) is constructed. Section 3 examines a few structural characteristics of the suggested AR-WL(1) model, such as the autocorrelation function, spectral density function, conditional mean and variance, and the multi-step conditional Laplace transform. The Gaussian estimation method is presented in Section 4. Section 5 focuses on applying the model to two real-life datasets and presents the residual analysis of the considered datasets. Section 6 presents the data forecasting for the chosen model using the classical conditional expectation method and some machine learning techniques.

2. The AR-WL(1) Model

The first-order autoregressive model, AR(1), is defined as:

$$X_t = \rho X_{t-1} + \varepsilon_t \ , \qquad \rho \in [0,1) \tag{2.1}$$

where ε_t are independent and identically distributed (i.i.d.) innovation terms that are assumed to follow the WL probability distribution. The restriction $\rho \in [0,1)$ ensures the strict stationarity and positivity of the process. The AR(1) process

defined by Equation (2.1) can be expressed in terms of the innovation sequence ε_t, as follows:

$$\begin{aligned} X_t &= \rho^t X_0 + \rho^{t-1}\varepsilon_1 + \cdots + \rho\ \varepsilon_{t-1} + \varepsilon_t && (2.2)\\ &= \rho^t X_0 + \sum_{r=0}^{t-1} \rho^r \varepsilon_{t-r}, && (2.3) \end{aligned}$$

which can be expressed as an infinite sum:

$$X_t = \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r}. \tag{2.4}$$

For the derivation of this representation, refer to A.
This infinite sum is valid and converges if and only if $|\rho| < 1$. Under this condition, the effect of the initial value disappears as t increases, and consequently, the process is stationary; its mean and variance remain constant over time.
Since ε_t follows the WL distribution, and according to Equations (2.3) and (2.4), the LT corresponding to X_t is given by,

$$\begin{aligned} \Phi_{X_t}(s) &= E\left[e^{-sX_t}\right] = E\left[e^{-s(\rho^t X_0 + \sum_{r=0}^{t-1}\rho^r\varepsilon_{t-r})}\right]\\ &= E\left[e^{-s\rho^t X_0}\right] E\left[e^{-s\sum_{r=0}^{t-1}\rho^r\varepsilon_{t-r})}\right]\\ &= \Phi_{X_0}(s\rho^t)\prod_{r=0}^{t-1}\Phi_{\varepsilon}(s\rho^r)\\ &= \prod_{r=0}^{\infty}\Phi_{\varepsilon}(s\rho^r)\\ &= \prod_{r=0}^{\infty}\frac{\theta^{c+1}(\theta+c+s\rho^r)}{(\theta+c)(\theta+s\rho^r)^{c+1}}. \end{aligned} \tag{2.5}$$

It is difficult to determine the distribution of X_t analytically due to the intricacy of the structure of (2.5). However, the distributional features of ε_t are used to obtain the fundamental analytical properties of X_t as specified by (2.1). From Equation (2.1), we conclude that the mean and variance of the process X_t are as follows, respectively,

$$\begin{aligned} E(X_t) &= \frac{1}{(1-\rho)}E(\varepsilon_t)\\ &= \frac{1}{(1-\rho)}\frac{c(\theta+c+1)}{\theta(\theta+c)}, \end{aligned} \tag{2.6}$$

$$\begin{aligned} V(X_t) &= \frac{1}{(1-\rho^2)}V(\varepsilon_t)\\ &= \frac{1}{(1-\rho^2)}\frac{(c+1)(\theta+c)^2-\theta^2}{\theta^2(\theta+c)^2}. \end{aligned} \tag{2.7}$$

3. Structural Characteristics of the AR-WL(1) Process

This section presents the derivation of some statistical properties of the AR-WL(1) model.

3.1. **Statistical and conditional properties.** Regression properties are useful for estimating unknown parameters and forecasting future values of the time series data. Among these properties, the one-step ahead conditional mean of the process is given by:

$$\begin{aligned} E(X_t \mid X_{t-1}) &= \rho x_{t-1} + E(\varepsilon_t) \\ &= \rho x_{t-1} + \frac{c(\theta + c + 1)}{\theta(\theta + c)}. \end{aligned} \tag{3.1}$$

As a result, the $(k+1)$-step ahead conditional mean formula can be expressed as:

$$\begin{aligned} E(X_{t+k} \mid X_{t-1}) &= \rho^{k+1} x_{t-1} + \frac{1-\rho^{k+1}}{1-\rho} E(\varepsilon_t) \\ &= \rho^{k+1} x_{t-1} + \frac{1-\rho^{k+1}}{1-\rho} \ \frac{c\ (\theta + c + 1)}{\theta(\theta + c)}. \end{aligned} \tag{3.2}$$

The expression above will converge to the unconditional mean of the main process when $k \to \infty$ as follows:

$$E(X_{t+k} \mid X_{t-1}) \to \frac{1}{1-\rho} \ \frac{c\ (\theta + c + 1)}{\theta(\theta + c)}.$$

The one-step ahead conditional variance of the process can be written as follows:

$$\begin{aligned} V(X_t \mid X_{t-1}) &= V(\varepsilon_t) \\ &= \frac{(c+1)(\theta + c)^2 - \theta^2}{\theta^2(\theta + c)^2}. \end{aligned} \tag{3.3}$$

Thus, the $(k+1)$-step ahead conditional variance is given by:

$$V(X_{t+k} \mid X_{t-1}) = \frac{1-\rho^{2(k+1)}}{1-\rho^2} V(\varepsilon_t) \tag{3.4}$$

$$= \frac{1-\rho^{2(k+1)}}{1-\rho^2} \ \frac{(c+1)(\theta + c)^2 - \theta^2}{\theta^2(\theta + c)^2}, \tag{3.5}$$

which tends to the unconditional variance at $k \to \infty$.

The multi-step ahead conditional LT of the process is derived as:

$$\begin{aligned}
\Phi_{X_{t+k}|X_{t-1}}(s) &= E\left(e^{-sX_{t+k}} \mid X_{t-1}\right) \\
&= E\left(e^{-s\left(\rho^{k+1}X_{t-1}+\sum_{i=0}^{k}\rho^{k-i}\varepsilon_{t+i}\right)} \mid X_{t-1}\right) \\
&= E\left(e^{-s\rho^{k+1}x_{t-1}}\right) E\left(e^{-s\sum_{i=0}^{k}\rho^{k-i}\varepsilon_{t+i}}\right) \\
&= e^{-s\rho^{k+1}x_{t-1}} \prod_{i=0}^{k} \Phi_{\varepsilon}(\rho^{k-i}s) \\
&= e^{-s\rho^{k+1}x_{t-1}} \prod_{i=0}^{k} \frac{\theta^{c+1}(\theta+c+\rho^{k-i}s)}{(\theta+c)(\theta+\rho^{k-i}s)^{c+1}}.
\end{aligned}$$

At $k \to \infty$, then

$$\Phi_{X_{t+k}|X_{t-1}}(s) \to \Phi_{X_t}(s)$$

which is the unconditional Laplace transform of X_t.

3.2. **Joint distribution, autocorrelation, and spectral density function.** Based on Equation (2.1), the expression for the joint LT of (X_{t-1}, X_t) is as follows:

$$\begin{aligned}
\Phi_{X_{t-1},X_t}(s_1,s_2) &= E(e^{-s_1X_{t-1}-s_2X_t}) \\
&= E(e^{-s_1X_{t-1}-s_2(\rho X_{t-1}+\varepsilon_t)}) \\
&= E(e^{-(s_1+s_2\rho)X_{t-1}})E(e^{-s_2\varepsilon_t}) \\
&= \Phi_{X_{t-1}}(s_1+s_2\rho)\ \Phi_{\varepsilon_t}(s_2) \\
&= \prod_{r=0}^{\infty} \frac{\theta^{c+1}(\theta+c+\rho^r(s_1+s_2\rho))}{(\theta+c)(\theta+\rho^r(s_1+s_2\rho))^{c+1}}\ \frac{\theta^{c+1}(\theta+c+s_2)}{(\theta+c)(\theta+s_2)^{c+1}}.
\end{aligned}$$

The AR-WL(1) process is not time-reversible because the joint Laplace transform $\Phi_{X_{t-1},X_t}(s_1,s_2)$ is not symmetric in s_1 and s_2.

The autocovariance (ACVF) and autocorrelation (ACF) functions of the AR-WL(1) process X_t at lag k are, respectively, expressed as:

$$\gamma(k) = \rho^{|k|}\operatorname{Var}(X_t) = \rho^{|k|}\frac{1}{1-\rho^2}\frac{(c+1)(\theta+c)^2-\theta^2}{\theta^2(\theta+c)^2},$$

$$\eta(k) = \rho^{|k|}.$$

The spectral density function $f_X(w)$ of the AR-WL(1) process is specified as follows:

$$f_X(w) = \frac{1}{2\pi}\sum_{z=-\infty}^{\infty}\gamma(z)e^{-iwz}, \quad \text{for } w \in [-\pi,\pi] \tag{3.6}$$

$$= \frac{(c+1)(\theta+c)^2-\theta^2}{2\pi\theta^2(\theta+c)^2(1+\rho^2-2\rho\cos w)}. \tag{3.7}$$

In the next section, we discuss the Gaussian estimation approach for parameter estimation in the AR-WL(1) process.

4. Parameter Estimation

In this section, the Gaussian estimation method (GE) is used to estimate the unknown parameters. The conditional maximum likelihood function can be written as follows:

$$L(\rho,\theta,c) = f(x_1)\prod_{t=2}^{n} f(x_t|x_{t-1}).$$

Here, the conditional and marginal probability functions of $X_t|X_{t-1}$ and X_t, respectively, are denoted by $f(x_t|x_{t-1})$ and $f(x_1)$.
The log-likelihood function can therefore be expressed as follows:

$$l(\rho,\theta,c) \;=\; \log(L(\rho,\theta,c)) = \frac{-n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\left(\frac{\left(x_t-\mu_{x_{t-1}}\right)^2}{\sigma^2_{x_{t-1}}} + \log\sigma^2_{x_{t-1}}\right),$$

where $\mu_{x_{t-1}}$ and $\sigma^2_{x_{t-1}}$ are the one-step conditional mean and variance defined by Equations (3.1) and (3.3), respectively.

For the AR-WL(1) process, the Gaussian log-likelihood function is thus as follows:

$$\begin{aligned} l(\rho,\theta,c) \;&=\; \frac{-n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\left\{\log\left(\frac{(c+1)(\theta+c)^2-\theta^2}{\theta^2(\theta+c)^2}\right)\right. \\ &+\; \left.\left(x_t-\rho x_{t-1}-\frac{c(\theta+c+1)}{\theta(\theta+c)}\right)^2\left(\frac{\theta^2(\theta+c)^2}{(c+1)(\theta+c)^2-\theta^2}\right)\right\}. \quad (4.1) \end{aligned}$$

The Gaussian estimators, $\hat{\rho}_{GE}$, $\hat{\theta}_{GE}$ and $\hat{c}_{GE}$, can therefore be obtained by solving the set of equations $\frac{\partial l(\rho,\theta,c)}{\partial\rho}=0$, $\frac{\partial l(\rho,\theta,c)}{\partial\theta}=0$, and $\frac{\partial l(\rho,\theta,c)}{\partial c}=0$.
Since the system of partial derivative equations is analytically intractable, numerical optimization methods must be employed to obtain the estimates of the parameters. The **optim** function in the `R` statistical programming language is used to numerically maximize the log-likelihood function, enabling the estimation of the unknown parameters through this optimization process.

5. Real-Life Applications and Model Selection

In this section, we examine the practical performance of the proposed model through the use of two real-life datasets and their description is:

- The first dataset includes 550 observations, representing the daily CBOE Energy Sector ETF Volatility Index (VXXLECLS) from December 1, 2017 to February 10, 2020. This data is available in https://fred.stlouisfed.org/series/VXXLECLS.
- The second dataset includes 246 daily data items of the stock price of Bharat Petroleum Corporation Limited (BPCL.BO), from April 26, 2022 to April 21, 2023. This data can be obtained from https://finance.yahoo.com/quote/BPCL.BO/history

Figures 1 and 2 present the time series, autocorrelation (ACF), and partial autocorrelation (PACF) functions for the two datasets. We can infer from these figures that the datasets are suitable for an AR(1) model since the PACF cuts off after lag one and ACF dies down quickly. The two datasets appear to be stationary based on the two figures. We further validate this conclusion by using the Augmented Dickey-Fuller (ADF) test, which is applied by using the `R` function **adf.test** from **tseries** package. It is clear from Table 1 that the p-value of the ADF test is less than the designated significance level (0.05), therefore, we reject the null hypothesis of non-stationarity, proving the stationary nature of the time series.

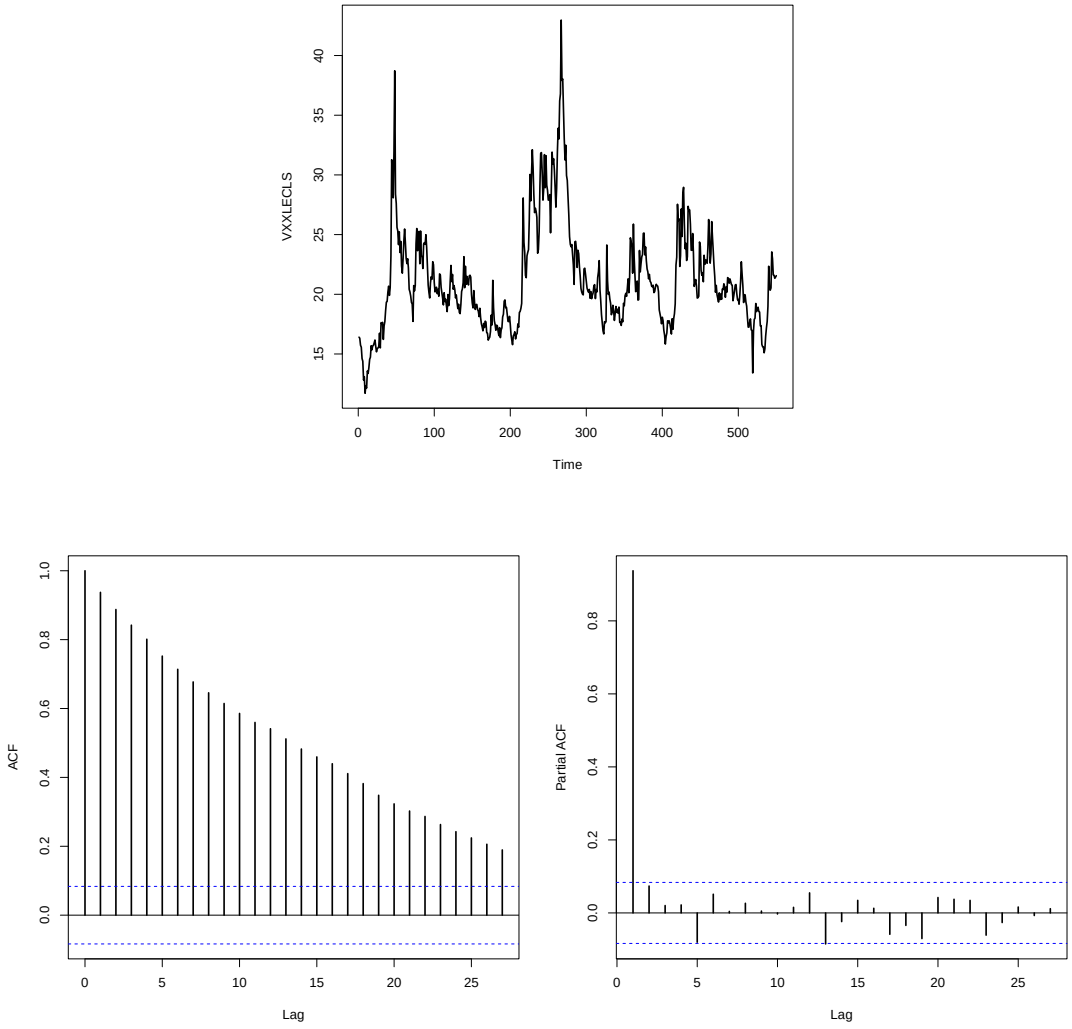

FIGURE 1. The time series, ACF, and PACF plots of the daily CBOE Energy Sector ETF Volatility Index (VXXLECLS).

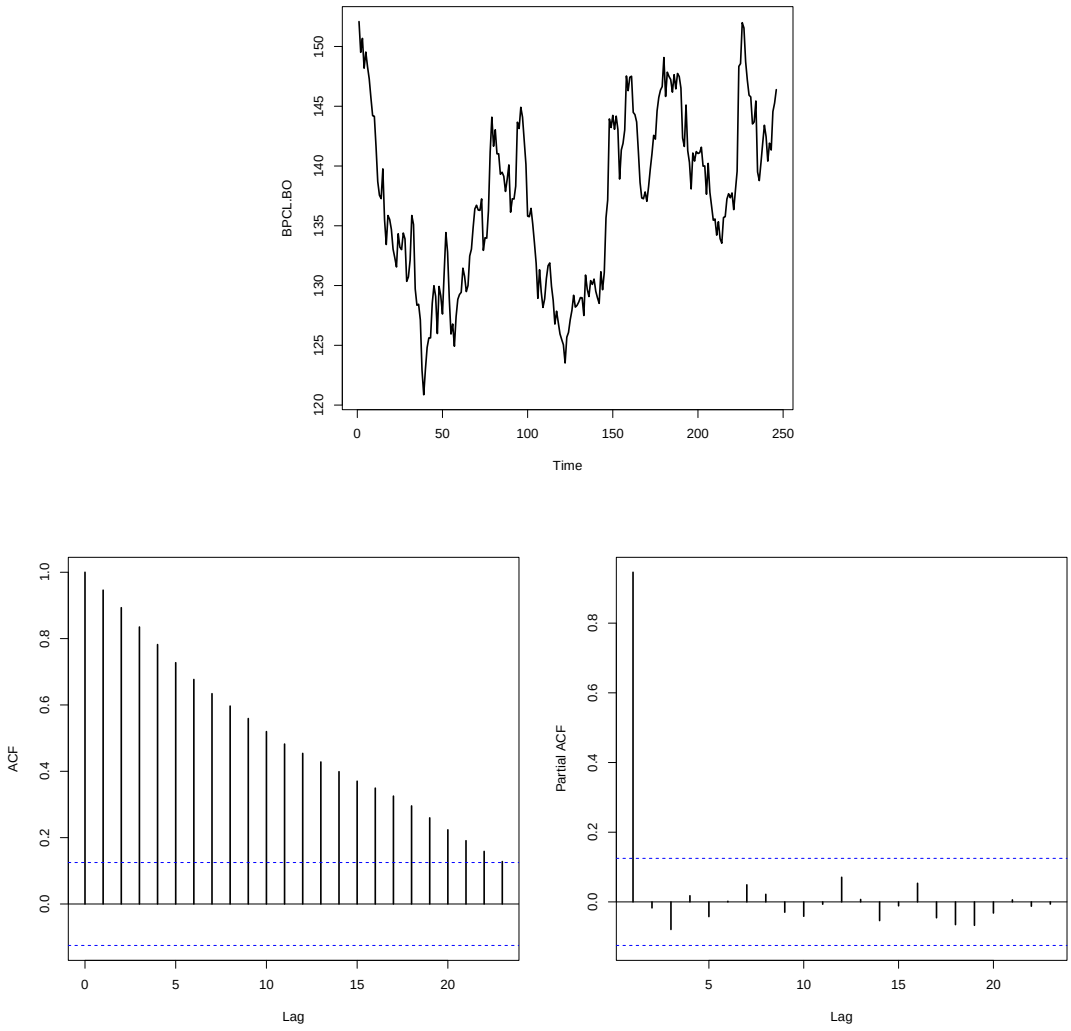

FIGURE 2. The time series, ACF, and PACF plots of the daily data items of the stock price of Bharat Petroleum Corporation Limited (BPCL.BO).

TABLE 1. Descriptive statistics of VXXLECLS and BPCL.BO datasets.

Dataset	n	**Min.**	Q_1	Q_2	**Mean**	Q_3	**Max.**	**Var**	p**-value (ADF)**
VXXLECLS	550	11.71	18.38	20.45	21.25	23.22	42.97	18.927	0.03767
BPCL.BO	246	120.9	130.6	137.2	137.0	142.5	152.1	50.151	0.04575

For comparison purposes, the proposed model AR-WL(1) is compared with some other non-Gaussian innovations, such as Lindley (LER(1)) (Nitha and Krishnarani [12]) and skew normal (ARSN(1)) (Sharafi and Nematollahi [14]), also

it is compared with non-Gaussian AR(1) process, where the process distribution is proposed and then the innovation distribution is obtained, such as exponential (E-AR(1)) (Gaver and Lewis [5]), gamma Lindley (GaL-AR(1)) (Mello et al. [10]), and gamma (G-AR(1)) (Gaver and Lewis [5]).
For each of the models under consideration, the unknown parameters will be estimated using the GE approach. Gaussian likelihood estimates, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Hannan-Quinn information criterion (HQIC) are calculated for every dataset and model. We use information criteria statistics to assess model performance. The model with the smallest values for these statistics is the most effective. Tables 2 and 3 provide a summary of the goodness-of-fit statistics and GE estimates, along with associated standard errors (SE). It is clear from these tables that the AR-WL(1) model obtains the lowest AIC, BIC, and HQIC values. We may therefore infer that the AR-WL(1) model offers the greatest fit among the other models for both datasets. Additionally, based on Raftery's guideline (Raftery [13]), a significant difference in BIC between models should exceed 2. The AR-WL(1) model meets this threshold when compared to other AR(1) models.

TABLE 2. Estimated parameters, AIC, BIC, and HQIC for VXXLECLS dataset.

Model	**GE estimates (SE)**	**AIC**	**BIC**	**HQIC**
AR-WL	$\hat{\rho} = 0.9320483$ (0.01524189) $\hat{\theta} = 0.7883768$ (0.13433453) $\hat{c} = 0.6842501$ (0.36174023)	2012.785	2025.715	2017.838
LER	$\hat{\rho}$ =0.9671705 (0.003607927) $\hat{\theta} = 2.8856176$ (0.181230147)	2037.183	2045.803	2040.552
ARSN	$\hat{\rho} = 0.9725503$ (0.01484289) $\hat{\mu}$ = -1.3974722 (0.75381243) $\hat{\sigma} = 2.4725016$ (0.54242154) $\hat{\delta} = 0.9973088$ (0.12589308)	2020.544	2037.784	2027.281
E-AR	$\hat{\rho} = 0.99053724$ (0.00210726) $\hat{\lambda} = 0.09009285$ (0.01033839)	2026.182	2034.801	2029.550
GaL-AR	$\hat{\rho} = 0.99027540$ (0.002201986) $\hat{\theta} = 0.09906001$ (0.023808570) $\hat{\beta} = 0.09933028$ (0.049133442)	2028.226	2041.156	2033.279
G-AR	$\hat{\rho} = 0.9877012$ (0.003685655) $\hat{\lambda} = 0.1138214$ (0.053515770) $\hat{k} = 1.2287354$ (0.872986806)	2028.738	2041.667	2033.790

TABLE 3. Estimated parameters, AIC, BIC, and HQIC for BPLC.BO dataset.

Model	GE estimates (SE)	AIC	BIC	HQIC
AR-WL	$\hat{\rho}$ = 0.9613549 (0.01150333) $\hat{\theta}$ = 1.3226693 (0.38068946) $\hat{c}$ = 6.1285924 (3.98009762)	1043.739	1054.255	1047.973
LER	$\hat{\rho}$ = 0.9974604 (0.0002922562) $\hat{\theta}$ = 7.5284570 (0.5214815843)	1049.385	1056.396	1052.208
ARSN	$\hat{\rho}$ = 0.9746756 (0.01521548) $\hat{\mu}$ = 4.3901815 (2.09114509) $\hat{\sigma}$ = 1.9285827 (0.09406422) $\hat{\delta}$ = -0.6249413 (0.07418625)	1065.730	1079.751	1071.375
E-AR	$\hat{\rho}$ = 0.99651492 (0.000363583) $\hat{\lambda}$ = 0.04176516 (0.002915109)	1055.763	1062.774	1058.586
GaL-AR	$\hat{\rho}$ = 0.99727907 (0.0001895084) $\hat{\theta}$ = 0.03278107 (0.0030251827) $\hat{\beta}$ = 0.02898761 (0.0038144708)	1053.997	1064.512	1058.231
G-AR	$\hat{\rho}$ = 0.99624020 (0.000448135) $\hat{\lambda}$ = 0.04504749 (0.011746707) $\hat{k}$ = 1.06998828 (0.471005345)	1059.342	1069.858	1063.576

5.1. **The residual analysis of the considered datasets.** The residual analysis of the VXXLECLS and BPCL datasets using the AR-WL(1) model is covered in this section in order to assess and validate the adequacy of the suggested model. The standardized Pearson residuals (Harvey and Fernandes [8]) are formulated as:

$$e_t = \frac{X_t - E(X_t \mid X_{t-1})}{\sqrt{Var(X_t \mid X_{t-1})}},$$

where $E(X_t \mid X_{t-1})$ and $Var(X_t \mid X_{t-1})$ are given by Equations (3.1) and (3.3), respectively. To calculate the residuals, the parameters are substituted in $E(X_t \mid X_{t-1})$ and $Var(X_t \mid X_{t-1})$ with their estimations while taking the AR-WL(1) process into account. The Pearson residuals for VXXLECLS and BPCL datasets have mean and variance values of (-0.00286, 1.0056) and (0.00715, 0.99392), respectively. The obtained values indicate that the means are approximately 0 and the variances are close to 1. Figure 3 presents the sample ACF of the Pearson residuals for the datasets, indicating no significant autocorrelation. This is corroborated by the Ljung-Box test results for the two datasets, with p-values of 0.1208 and 0.4502, confirming the residuals are uncorrelated. The cumulative periodogram of Pearson residuals, shown in Figure 4, verifies that the residuals are dispersed randomly and do not exhibit any discernible trends.

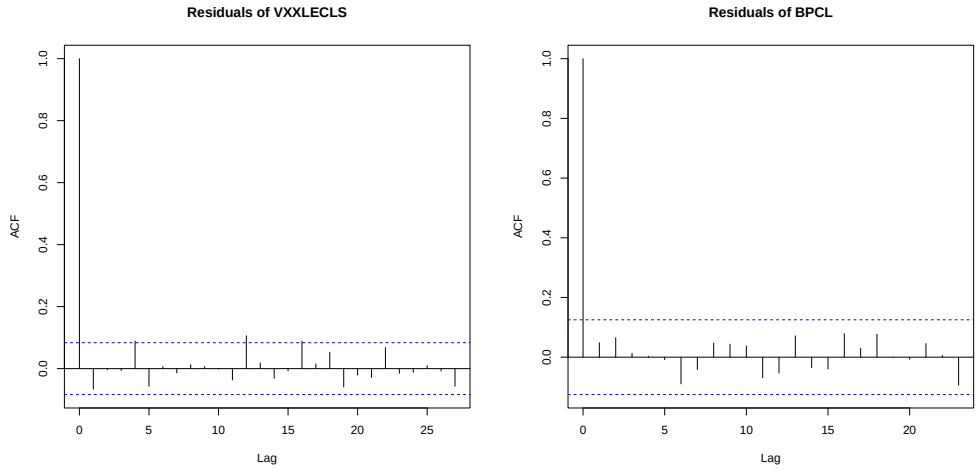

FIGURE 3. The standardized Pearson residuals ACF of VXXLECLS and BPCL datasets.

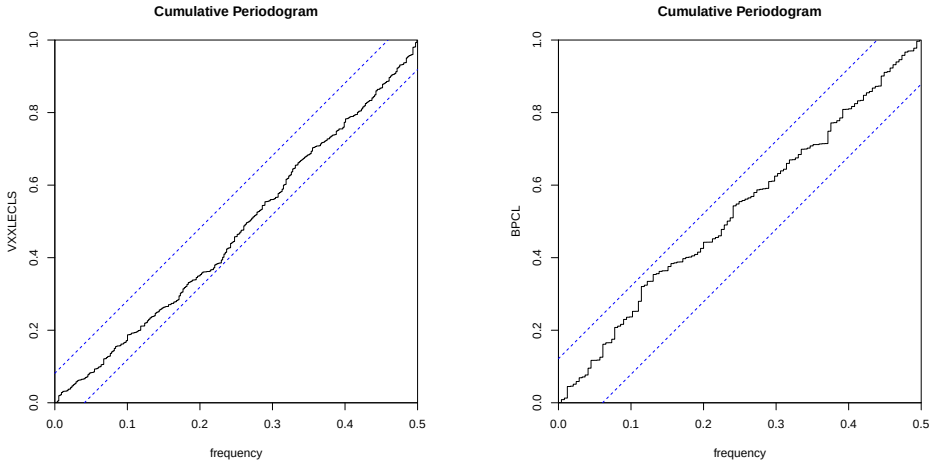

FIGURE 4. The cumulative periodogram of the standardized Pearson residuals for VXXLECLS and BPCL datasets.

6. FORECASTING

Forecasting is an essential step in time series analysis that confirms the appropriateness and predictability of the selected model. In this section, the AR-WL(1) model combines the classical conditional expectation method with machine learning algorithms, measuring the prediction accuracy for examined real-life datasets.

6.1. **Classical conditional expectation method.** One of the most popular forecasting methods is the conditional expectation method. The expected value (or conditional expectation) of X_{t+k} given the data observed up to time t is usually referred to as the forecast $\hat{X}_{t+k}$. Based on Equations (3.1) and (3.2), the

one- and k-step ahead forecasts for X_{t+1} and X_{t+k}, respectively, are as follows:

$$\hat{X}_{t+1} = E(X_{t+1}|X_t = x_t) = \hat{\rho}\ x_t + \frac{\hat{c}\ (\hat{\theta} + \hat{c} + 1)}{\hat{\theta}(\hat{\theta} + \hat{c})}\ ,$$

and

$$\hat{X}_{t+k} = E(X_{t+k}|X_t = x_t) = \hat{\rho}^k\ x_t + \frac{1 - \hat{\rho}^k}{1 - \hat{\rho}}\ .\ \frac{\hat{c}\ (\hat{\theta} + \hat{c} + 1)}{\hat{\theta}(\hat{\theta} + \hat{c})},$$

where $\hat{\rho}$, $\hat{\theta}$, and $\hat{c}$ are the estimated parameters presented at Tables 2 and 3.

6.2. **Machine learning forecasting methods.** In addition to classical statistical forecasting methods, machine learning represents another important category of forecasting approaches. Machine learning (ML) methodologies are data-driven and non-parametric, discovering patterns directly from observed data to adapt flexibly to complicated structures, in contrast to traditional statistical methods that depend on predetermined theoretical assumptions. These methods are usually trained on past data to identify patterns and relationships, which are then used to predict future values. In order to use ML techniques to describe the autoregressive behavior of the time series, we used an AR(1) structure, which predicts the current observation x_t by using its first-order lag, x_{t-1}. The data was initially reformatted into a supervised learning structure by introducing a lag 1 feature. Following this transformation, the dataset was split into training and testing subsets, typically with 80% allocated for training and 20% reserved for testing. In the next subsections, we will use for forecasting some ML methods such as k-Nearest Neighbors, Extreme Gradient Boosting, and Support Vector Regression.

6.2.1. *K-Nearest Neighbors.* The K-Nearest Neighbors (KNN) method works by calculating the distance between a new input and existing data points in the training set. It then identifies the K closest neighboring points and uses their outcomes to make a prediction for the new data point. The result is derived by averaging the outcomes of the neighboring data points as:

$$\hat{y} = \frac{1}{K}\sum_{j=1}^{K} y_{(j)},$$

where $y_{(j)}, j = 1, 2, .., K$, represents the output value associated with the j^{th} closest neighbor (Kramer [9]). The KNN method was implemented using the **FNN** package in `R`. Predictions were obtained using the **knn.reg** function, which averaged the nearest data points.

6.2.2. *Extreme Gradient Boosting.* Extreme Gradient Boosting (XGBoost) is a ML technique based on gradient-boosted decision trees (Chen and Guestrin [2]). It builds trees one after another, with each new tree improving on the errors of the previous ones. The model is trained by minimizing an objective function that

balances prediction accuracy with model complexity to avoid overfitting. The objective function takes the form:

$$\text{Obj}(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{k=1}^{K}\Omega(f_k),$$

where f_k denotes each individual regression tree, and $\Omega(f_k)$ is a regularization term that controls the complexity of the corresponding tree. XGBoost was applied using the **xgboost** package in `R`, with data organized in **xgb.DMatrix** format to improve processing speed. The model was trained with **xgb.train**, specifying the "**reg:squarederror**" objective function for regression tasks.

6.2.3. *Support Vector Regression.* Support Vector Regression (SVR) aims to identify a predictive function that estimates target values within a specified error tolerance (ϵ) (Smola and Schölkopf [15]). Given a training dataset $\{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i denotes input features (such as lagged observations) and y_i are the corresponding target values, the objective is to determine a linear function $f(x) = \langle w, x\rangle + b$ that minimizes:

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \quad & \begin{cases} y_i - \langle w, x_i\rangle - b \le \epsilon + \xi_i, \\ \langle w, x_i\rangle + b - y_i \le \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, \end{cases} \end{aligned}$$

where $\langle ., . \rangle$ indicating the dot product operation, w represents the weight vector, b is the bias term, C regulates the balance between the model's complexity and the penalty for errors on the training data, and ξ_i, ξ_i^* are slack variables that permit prediction errors exceeding the specified ϵ-insensitive margin. We utilized the **svm** function from the **e1071** package in `R`, configuring it with a linear kernel and an ϵ-insensitive loss function (type = "**eps-regression**"), ensuring consistency with the underlying assumptions of the AR(1) model.

6.3. **Forecasting Evaluation.** The accuracy of the forecasting model was evaluated using several standard error metrics commonly employed in predictive analytics. The key measures are listed below:

- **Root Mean Squared Error (RMSE)**
  RMSE is calculated as the square root of the mean squared error (MSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}.$$

- **Mean Absolute Error (MAE)**
  MAE represents the average of the absolute differences between the predicted values and the actual observations:
$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|.$$
- **Symmetric Mean Absolute Percentage Error (SMAPE)**
  SMAPE measures error as a percentage of the average of actual and predicted values:
$$\text{SMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{2\,|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \times 100\%.$$

For each machine learning method, we used grid search to tune hyperparameters and identify the parameter combination that minimized the RMSE.

The one-step ahead forecasting accuracy of both the classical conditional expectation method and ML techniques is assessed in Tables 4 and 5. As shown in these tables, the classical method has strong predictive performance. In VXXLE-CLS dataset, the performance of the classical method is slightly lower than that of SVR and better than the other methods, while for the BPCL dataset, it outperforms most competing methods and achieves accuracy on par with SVR. Although the ML techniques have notable predictive strengths, the results confirm that the classical method demonstrated highly competitive performance and even outperformed ML techniques in some cases, highlighting the effectiveness of the AR-WL(1) model.

Figures 5 and 6 display the actual versus predicted values for each dataset across all forecasting techniques. The predictions made by the classical AR-WL(1) method in these figures exhibit smoother alignment than those produced by some machine learning techniques, closely matching the trends in the actual data. This strong fit reflects the AR-WL(1) model's parametric framework, which is designed for the skewed and positive-valued nature of the data.

TABLE 4. Model performance measures for forecasting VXXLE-CLS dataset values.

Measure	**RMSE**	**MAE**	**SMAPE**
Classical	1.169084	0.8215819	4.052766 %
KNN	1.184506	0.8468396	4.152476 %
XGBoost	1.17828	0.8379002	4.143604 %
SVR	1.156722	0.7831515	3.882160 %

TABLE 5. Model performance measures for forecasting BPCL dataset values.

Measure	RMSE	MAE	SMAPE
Classical	2.107053	1.47297	1.040915 %
KNN	2.253783	1.654489	1.167888 %
XGBoost	2.285977	1.618397	1.142517 %
SVR	2.107359	1.473663	1.041394 %

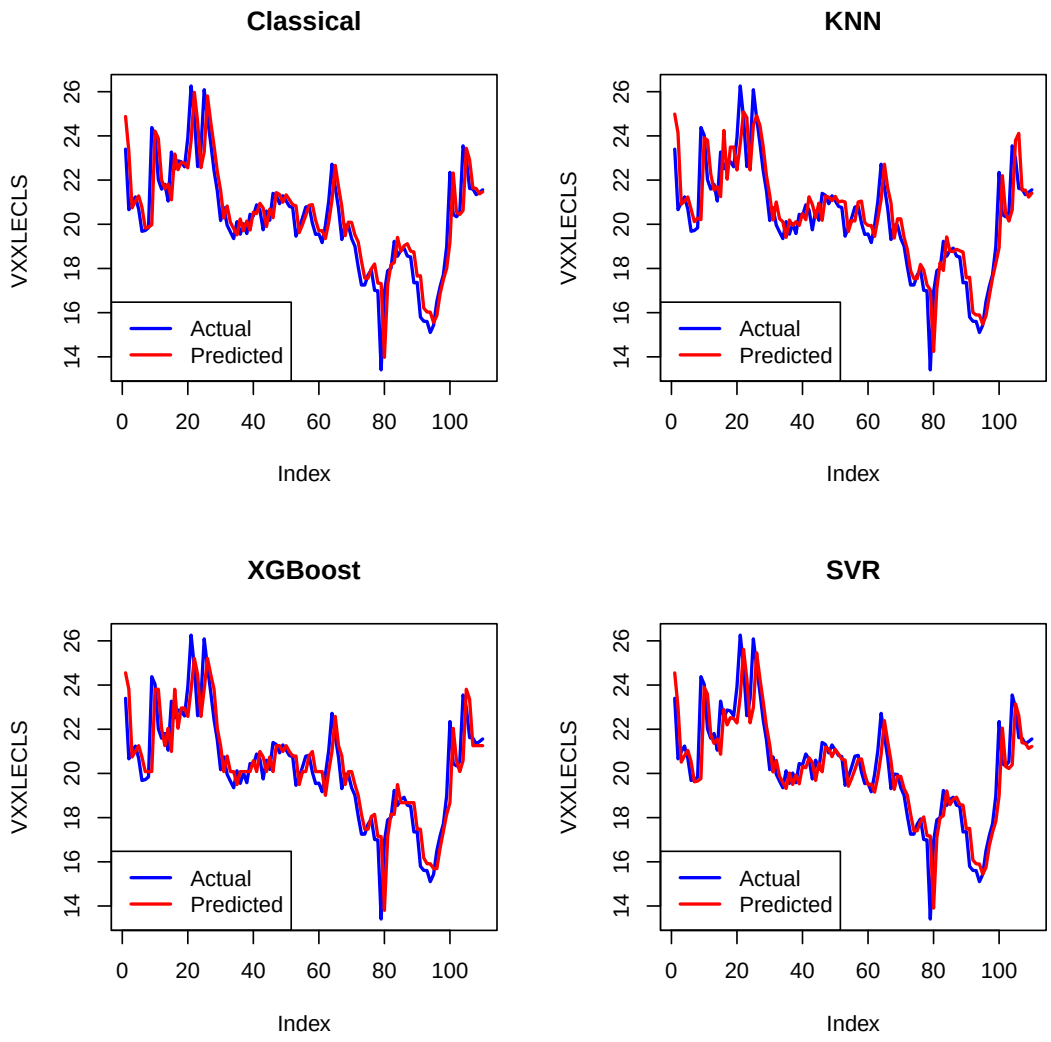

FIGURE 5. Actual and predicted values of VXXLECLS.

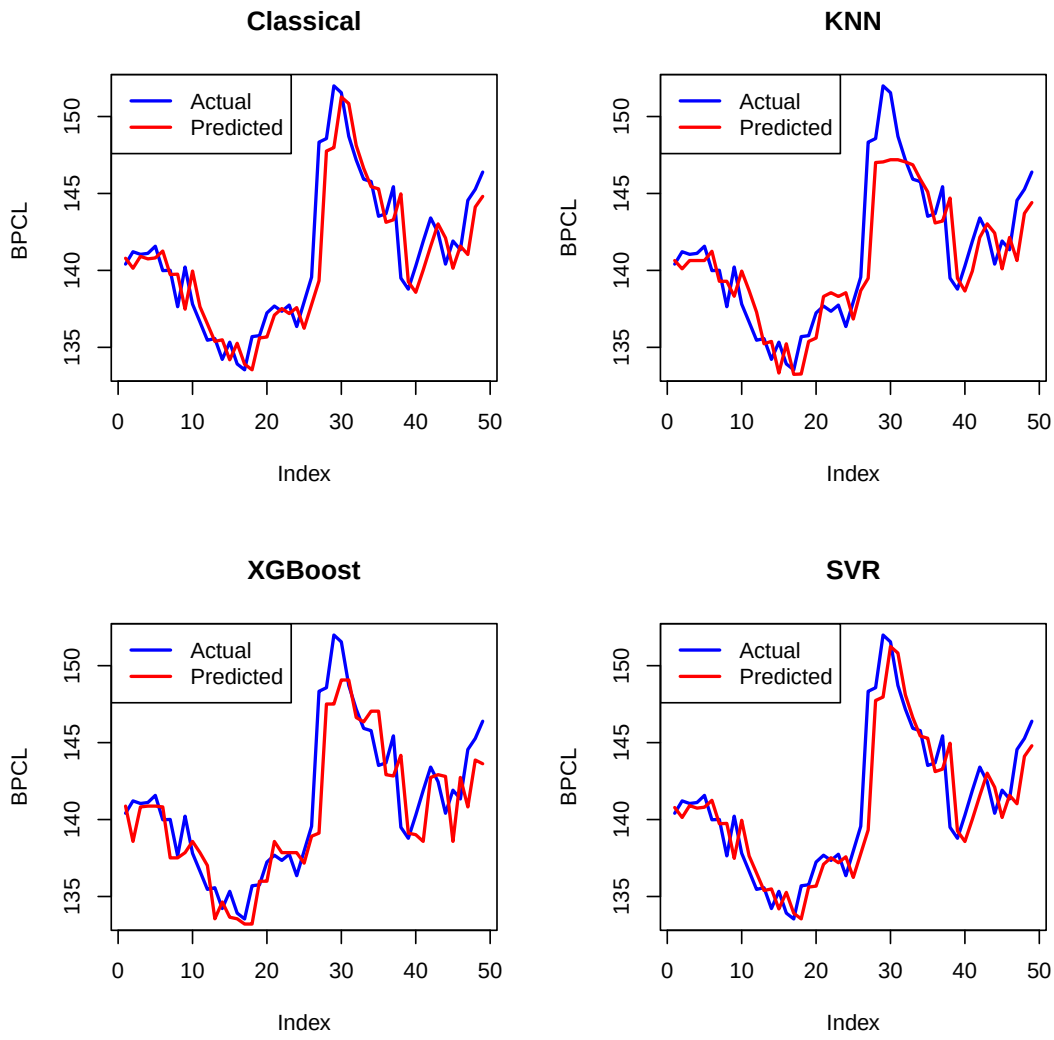

FIGURE 6. Actual and predicted values of BPCL.

7. Discussion and Conclusion

In this study, an autoregressive process of order one with innovations following a weighted Lindley distribution was proposed and its fundamental properties were investigated. The model demonstrated practical relevance through applications to real-world time series data, where parameter estimation was effectively carried out using the Gaussian estimation method, and comparative analysis highlighted the model's performance against existing alternatives. Furthermore, forecasting was addressed using both the classical conditional expectation method and various machine learning techniques, offering a comprehensive evaluation of the model's predictive capabilities. The results suggest that the proposed process provides a flexible and competitive approach for modeling and forecasting time series data.

Future research could extend the analysis by incorporating more advanced machine learning models, such as Multi-Layer Perceptrons (MLPs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures, which have demonstrated strong performance in recent time series forecasting applications. Additionally, the feature set for the machine learning models could be expanded by including multiple lagged observations as input features, effectively transforming them into higher-order autoregressive forecasting models. This would allow the ML benchmarks to capture more complex temporal dependencies by incorporating information from several past time points, similar to higher-order AR models.

Appendix A. Derivation of the Model Representation in (2.4)

$$\begin{aligned} X_t &= \rho X_{t-1} + \varepsilon_t \\ &= \rho^2 X_{t-2} + \rho\, \varepsilon_{t-1} + \varepsilon_t \\ &= \rho^3 X_{t-3} + \rho^2 \varepsilon_{t-2} + \rho\, \varepsilon_{t-1} + \varepsilon_t \\ &= \ldots \\ &= \rho^k X_{t-k} + \rho^{k-1} \varepsilon_{t-k+1} + \ldots + \rho\, \varepsilon_{t-1} + \varepsilon_t \\ &= \rho^k X_{t-k} + \sum_{r=0}^{k-1} \rho^r \varepsilon_{t-r}. \end{aligned} \tag{A.1}$$

Consequently, the above expression can be rewritten as:

$$X_t - \sum_{r=0}^{k-1} \rho^r \varepsilon_{t-r} = \rho^k X_{t-k}.$$

For $|\rho| < 1$, then

$$E\left[\left(X_t - \sum_{r=0}^{k-1} \rho^r \varepsilon_{t-r}\right)^2\right] = E\left(\rho^k X_{t-k}\right)^2.$$

As $k \to \infty$, then $E\left(\rho^k X_{t-k}\right)^2 \to 0$. Hence

$$\lim_{k\to\infty} E\left[\left(X_t - \sum_{r=0}^{k-1} \rho^r \varepsilon_{t-r}\right)^2\right] = 0, \quad \text{as } k \to \infty.$$

This implies that the sum is convergent with respect to the mean square criterion.

References

1. Andel, J., (1988). On AR(1) processes with exponential white noise. Communications in Statistics - Theory and Methods, 17(5), 1481–1495. https://doi.org/10.1080/03610928808829693
2. Chen, T., Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794. https://doi.org/10.1145/2939672.2939785
3. Christmas, J., Everson, R. (2010). Robust autoregression: Student-t innovations using variational Bayes. IEEE Transactions on Signal Processing, 59(1), 48-57. https://doi.org/10.1109/TSP.2010.2080271
4. Dhull, M. S., Kumar, A., Wyłomańska, A. (2023). The expectation-maximization algorithm for autoregressive models with normal inverse Gaussian innovations. Communications in Statistics - Simulation and Computation, 53(11), 5421–5441. https://doi.org/10.1080/03610918.2023.2186334
5. Gaver, D. P., Lewis, P. A. W., (1980). First order autoregressive gamma sequences and point processes. Advances and Applications in Probability, 12, 727–745. https://doi.org/10.2307/1426429
6. Ghasami, S., Khodadadi, Z., Maleki, M., (2020). Autoregressive processes with generalized hyperbolic innovations. Communications in Statistics-Simulation and Computation, 49(12), 3080–3092. https://doi.org/10.1080/03610918.2018.1535066
7. Ghitany, M. E., Alqallaf, F., Al-Mutairi, D. K., Husain, H. A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data. Mathematics and Computers in simulation, 81(6), 1190-1201. https://doi.org/10.1016/j.matcom.2010.11.005
8. Harvey, A. C., Fernandes, C. (1989). Time series models for count or qualitative observations. Journal of Business & Economic Statistics, 7(4), 407-417. https://doi.org/10.2307/1391639
9. Kramer, O. (2011). Dimensionality reduction by unsupervised k-nearest neighbor regression. In Proceedings of the 10th international conference on machine learning and applications and workshops, 275-278. https://doi.org/10.1109/ICMLA.2011.55
10. Mello, A. B., Lima, M. C., Nascimento, A. D. (2022). A notable gamma-Lindley first order autoregressive process: An application to hydrological data. Environmetrics, 33(4), e2724. https://doi.org/10.1002/env.2724
11. Nduka, U. C. (2018). EM-based algorithms for autoregressive models with t-distributed innovations. Communications in Statistics-Simulation and Computation, 47(1), 206-228. https://doi.org/10.1080/03610918.2017.1280164
12. Nitha, K. U., Krishnarani, S. D. (2024). On autoregressive processes with Lindley-distributed innovations: modeling and simulation. Statistics in Transition new series, 25(3), 31-47. https://doi.org/10.59170/stattrans-2024-026
13. Raftery, A. E. (1995). Bayesian model selection in social research. Sociological methodology, 111-163. https://doi.org/10.2307/271063
14. Sharafi, M., Nematollahi, A. R., (2016). AR(1) model with skew-normal innovations. Metrika, 79(8), 1011–1029. https://doi.org/10.1007/s00184-016-0587-7
15. Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
16. Tiku, M. L., Wong, W. K., Vaughan, D. C., Bian, G. (2000). Time series models in non-normal situations: Symmetric innovations. Journal of Time Series Analysis, 21(5), 571-596. https://doi.org/10.1111/1467-9892.00199

[1] Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt.
Email address: mahgabr@alexu.edu.eg

[2] Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.

[3] Department of Mathematics, College of Science, Qassim University, Buraydah, Saudi Arabia.
Email address: h.bakouch@qu.edu.sa

[4] Department of Mathematics, Faculty of Science, Damietta University, Damietta, Egypt.
Email address: hadeer_mostafa@du.edu.eg