

Research Article

The Role of Twitter Medium in Business with Regression Analysis and Statistical Modelling

Jia Cong,¹ Zubair Ahmad ,² Basim S.O. Alsaedi ,³ Osama Abdulaziz Alamri ,³ Ibrahim Alkhairy,⁴ and Hassan Alsuhabi ⁴

¹Marketing Department, Management School, Lancaster University, Lancaster, UK

²Department of Statistics, Yazd University, P.O. Box 89175-741, Yazd, Iran

³Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia

⁴Department of Mathematics, Al-Qunfudah University College, Umm Al-Qura University, Mecca, Saudi Arabia

Correspondence should be addressed to Zubair Ahmad; zubair@stu.yazd.ac.ir

Received 29 July 2021; Accepted 11 October 2021; Published 20 October 2021

Academic Editor: Ahmed Mostafa Khalil

Copyright © 2021 Jia Cong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Marketing refers to the strategies a company undertakes to promote its brands to its potential audience. Advertising provides useful venues for marketing to promote a company's survives/goods to the audience. It has a positive impact on the sale of services or products. In this study, we consider a well-known online medium called Twitter (the fourth most popular social media platform used by marketers) to check its impact on sales. For this purpose, the simple linear regression modeling approach is implemented to test the significance and usefulness of Twitter advertising on sale. Statistical tests such as *t*-test and correlation test are adopted to test the hypothesis of the "impact of Twitter advertising on sales." Based on the findings of this study, it is observed that Twitter advertising has a positive impact on sales. Furthermore, a new statistical model called the exponential T-X exponentiated exponential is introduced. The proposed model is very interesting and possesses heavy-tailed characteristics which are useful in finance and other related sectors. Finally, the applicability of the new model is illustrated by considering the sales data.

1. Introduction

Advertisement is a way of marketing to communicate with the potential consumers of certain goods, products, or services. Advertisements are the information usually paid by a company to influence consumer's decision to buy their products. Numerous mediums are available for advertisements and marketing. The available mediums can be divided into two main groups, such as printed mediums and online mediums. Online media is the most effective medium to reach more consumers [1, 2].

Online advertising is a way of marketing to convey messages to potential users via using the Internet. It is also called online marketing and helps in finding the potential and right audience. Online advertisements appear in a search engine, on social media, in a browser, or even in e-mail. Online advertising can be classified into different

groups such as display ads, e-mail ads, native ads, video ads, and social media ads [3, 4].

Twitter is one of the most prominent online/social media platforms that can be used quite effectively for advertising a particular brand. It is a fruitful tool for marketing the business and engaging the audience. It permits its users to create accounts and interact with others through short messages called "tweets." Twitter is one of the marketer's favorite social media platforms. For instance, 53% of marketers worldwide used it as a channel for their marketing activities in 2020. Social media generally provides brands with personally consumer-related opportunities. Twitter has a huge impact on consumer behavior. According to a report by the Digital Marketing Institute, it turns out that about 40% of Twitter users bought something after seeing it on Twitter. This reflects the steadily increasing impact of social media on consumers' buying habits [5–8].

Not only is it enough to get the company brand involved on Twitter but also brands need to consider the facilities of social media influencers. The influence of social media like Twitter is also competing with friends in building consumer trust. When trying to find a product recommendation, 56% of consumers asked their friends for instructions, and 49% of respondents said that they bought something after consulting their friends; see [9]. On Twitter, recommendations from influencers resulted in a 20% increase in the sale of products [10]. It is ranked fourth in the marketing arena; see Figure 1.

In this paper, we analyze a real-life data set related to Twitter advertising to see its impact on sales. We will use a well-known statistical methodology called the regression technique to see how much the sales can be increased by spending money on Twitter advertising. To carry out the regression analysis, we will test the hypothesis of “the impact of Twitter advertising on sales.” In this regard, the null hypothesis (NH) expressed by H_0 and alternative hypothesis (AH) denoted by H_A will be formulated as follows: $H_0 =$ Twitter advertising has no significant impact on sales versus $H_A =$ Twitter advertising has a significant impact on sales.

Furthermore, to provide an adequate fit to the sales data, a new heavy-tailed (HT) statistical model is introduced. The new model is introduced by combining the exponential distribution with the exponential TX strategy. The new model is entitled the exponential T-X exponentiated exponential (ETXE-exponential) model. The comparison of the ETXE-exponential distribution is done with the other models. The numerical results show that the ETXE-exponential is a suitable model for dealing with data in business, finance, and management sciences.

2. Regression Analysis

The regression method helps to carry out the regression analysis to forecast the demand for a particular product. It determines the nature of the relationship between the product’s demand and its corresponding determinants. The regression analysis can be carried out via three main approaches, depending upon the nature of the study. Figure 2 illustrates the main divisions of the regression analysis.

2.1. Simple Linear Regression Approach. In this paper, we adopt the simple linear regression (SLReg) approach as we have one independent variable (Twitter advertisement). This approach helps to predict the response variable, usually represented by Y based on the independent variable(s) usually represented by X . The simple linear regression can be expressed mathematically as follows:

$$Y = \theta_0 + \theta_1 X + \varepsilon, \quad (1)$$

where θ_0 is representing the intercept of the model and θ_1 is called the slope of the model. A value 0 of θ_1 indicates no effect of X on Y , and in such situation, the model is given by $Y = \theta_0$. The quantity ε is the residual error term having a mean value of 0.

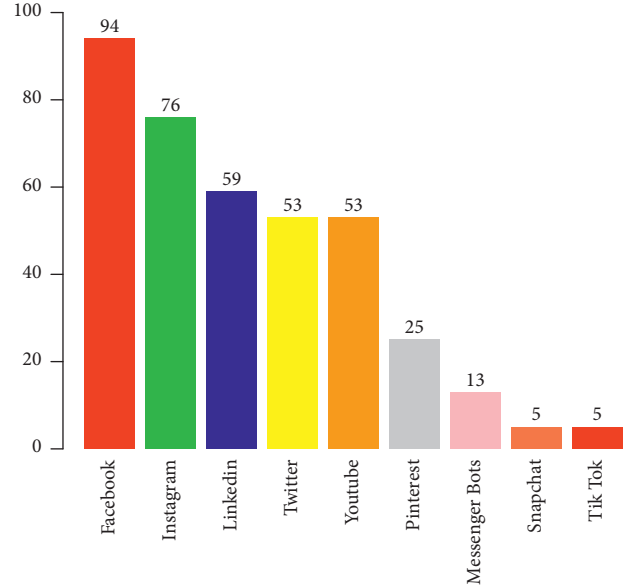


FIGURE 1: Leading social media platforms used by the business community around the globe in 2020.

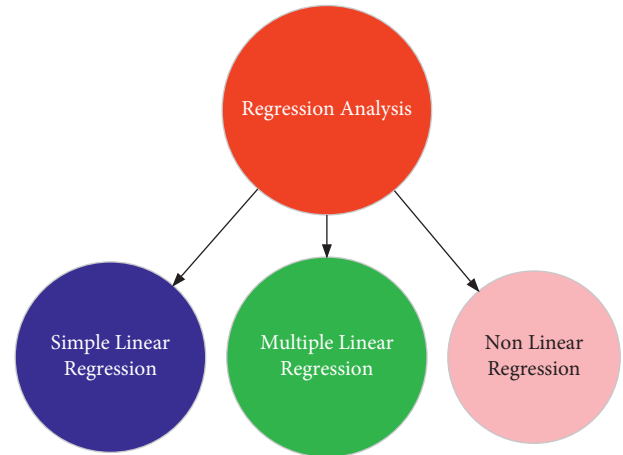


FIGURE 2: Main division of the regression modeling.

In case of explaining the relationship between sales Y and Twitter advertising X , the simple linear regression becomes

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \varepsilon. \quad (2)$$

After carrying out the analysis using the regression technique, we found that $\theta_0 = 5.621$ representing the predicted sales in thousands of USD (United States dollars), when there is no investment on the Twitter medium. Therefore, under the circumstances of no investment on Twitter medium, the expected sale would be $5.621 * 1000 = 5621$ dollars.

The slope (also called the regression coefficient) of the regression model θ_1 , provided in equation (2), is 0.193 representing 193 ($0.193 * 100$) units increment in the sales. Henceforth, using the Twitter medium as a marketing tool, the predicted/estimated sale is $5.621 + 0.193 * 1000 =$

198.621, representing a sale of 198621 dollars. Corresponding to Twitter and sales data, the fitted regression model is

$$\text{Sales} = 5.621 + 0.193\text{Twitter}. \quad (3)$$

The visual display of the relationship between Twitter advertising and the sale of goods is provided in Figure 3. From the plot in Figure 3, it is obvious that there is a positive relationship (PR) between Twitter advertising and sales. The PR means that the more we spend on marketing via the Twitter medium, the more will be the sale.

2.2. Statistical Testing. Is there a PR between sales and Twitter advertising medium? Certainly, there is a relationship between them, since the estimated slope θ_1 of the regression line is 5.621, not 0, showing that there is a relationship between sales and Twitter advertisement in the sample of 130 observations. However, we want to know if there is a relationship between the observations of all the sales and Twitter advertisement; that is, we want to know that if θ_1 is unlikely to be 0.

For these purposes, we adopt a standard hypothesis approach to conduct a hypothesis test for θ_1 . To carry out the procedure, first, we need to accomplish and specify H_0 and H_A as follows: $H_0: \theta_1 = 0$ (no significant relationship exists between sales and Twitter advertisement) versus $H_A: \theta_1 \neq 0$ (a significant relationship exists between sales and Twitter advertisement). Second, we obtain the value of *t-statistics* using the following formula:

$$t = \frac{\hat{\theta}_1 - \theta_1}{S.E(\hat{\theta}_1)}, \quad (4)$$

$$t = \frac{\hat{\theta}_1 - \theta_1}{\sqrt{MSE} / \sqrt{\sum_{i=0}^n (x_i - \bar{x})^2}}.$$

Third, we compute the value of *t-statistics* using statistical software. By default, the *t-statistics* is calculated by assuming that $\theta_1 = 0$. Divide the estimated coefficient $\hat{\theta}_1 = 5.621$ by its standard error $S.E\hat{\theta}_1 = 0.44428$ to get the value of *t-statistics* as follows:

$$t = \frac{5.621 - 0}{0.44428},$$

$$t = \frac{5.621}{0.44428}, \quad (5)$$

$$t = 12.652.$$

Fourth, we calculate the *p* value. By default, the *p* value is calculated by assuming H_A is a "not-equal-to." If the *p* value is less than 0.05, then we reject H_A . After performing the analysis, we observe that the *p* value is $2e - 16$.

Since the *p* value is very small (less than 0.001), we reject the hypothesis of $\theta_1 = 0$ and accept the hypothesis of $\theta_1 \neq 0$. Henceforth, there is significant evidence that there is a linear relationship between sales and Twitter advertisement.

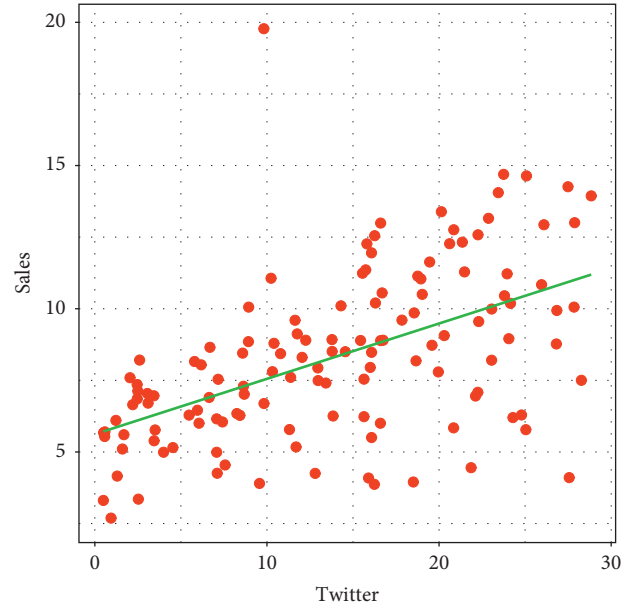


FIGURE 3: Graphical display of the relationship between Twitter advertising and sales.

2.3. Residuals. The regression line is used to quantify a linear trend in the data. The residual represents the distance between the true value and observed values of Y . Mathematically, it is given by

$$\varepsilon = y - \hat{y}. \quad (6)$$

The residual standard error is used to quantify the quality of the regression fit. In this study, it represents the average amount of the sales variable (Y) varies from the true regression line. In the context of this study, a visual display of the behavior of the residuals is provided in Figure 4.

From the plots presented in Figure 4, we have the following observations:

- (1) From the residual versus fitted plot, we can see that the red line is quite close to the residual value of 0. Therefore, we assume that the residuals possess the linear property. By linearity, we mean that the X in the regression model has a straight-line relationship with Y .
- (2) One of the fundamental assumptions of regression modeling is the normality of the data. The QQ (quantile-quantile) plot makes an angle of 45 showing that the residuals are approximately normally distributed.
- (3) Homoscedasticity is one of the basic assumptions of linear regression modeling. If this assumption of homoscedasticity does not hold, then the problem of heteroscedasticity arises. The scale-location plot shows that the residuals satisfy the homoscedasticity property.
- (4) In regression modeling, the influential observations have a serious effect on the parameters estimates. The residual versus leverage plot shows that there are fewer influential observations in the data.

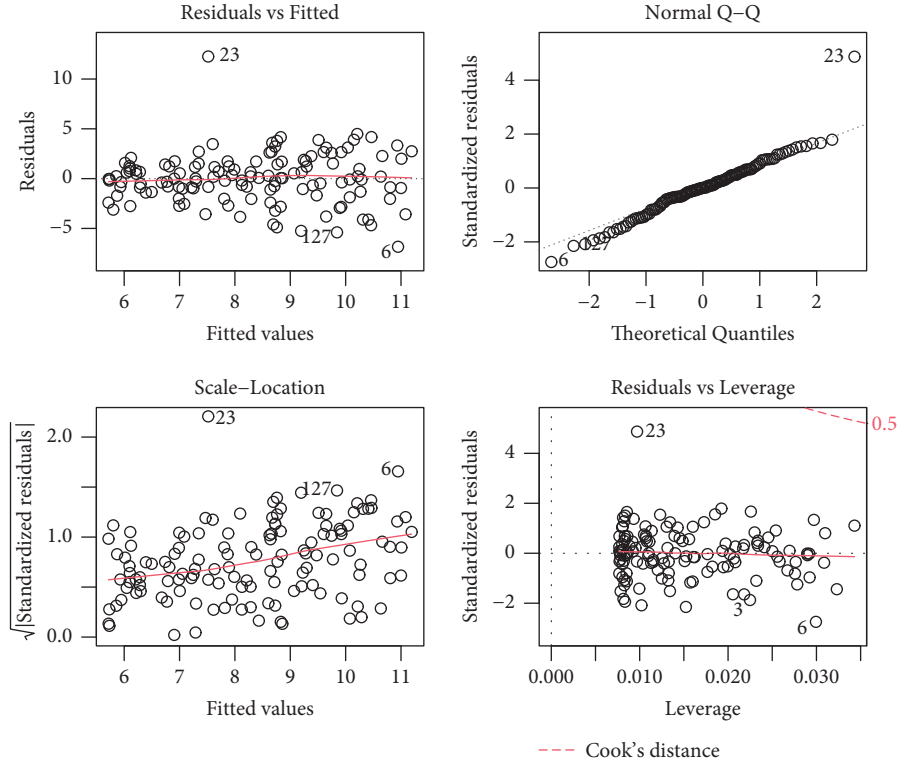


FIGURE 4: Graphical sketching of the residuals behaviors.

2.4. Correlation Test. A correlation test is a statistical tool used to quantify the relationship between two or more variables. In this study, we have only two variables (sales and Twitter advertising). Henceforth, we adopt the Pearson correlation approach, which is used to measure a linear relationship between two variables. The coefficient of the Pearson correlation is usually denoted by r and derived as

$$r = \frac{\sum_{k=1}^{130} (\text{Twitter} - \mu_{\text{Twitter}})(\text{Sales} - \mu_{\text{Sales}})}{\sqrt{\sum_{k=1}^{130} (\text{Twitter} - \mu_{\text{Twitter}})^2 (\text{Sales} - \mu_{\text{Sales}})^2}} \quad (7)$$

where μ_{Twitter} and μ_{Sales} represent the mean values of Twitter and sales, respectively. The significance of r can be tested by using the t -test and is given by

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}. \quad (8)$$

To the test r , we use the hypothesis procedure as $H_0 = 0$ (there is no relationship between Twitter advertising and sale) versus $H_A \neq 0$. After performing the required steps of analysis, we observe that $r = 0.5253$, showing a positive relationship between sales and Twitter advertising. Furthermore, we observe that the Spearman rank correlation test represented by Sp , is 161 with p value given by $4.634e-12$. As we see that p -value < 0.05 , we reject H_0 and conclude that there is a significant relationship between Twitter advertising and sale. This relationship has also been shown graphically in Figure 5.

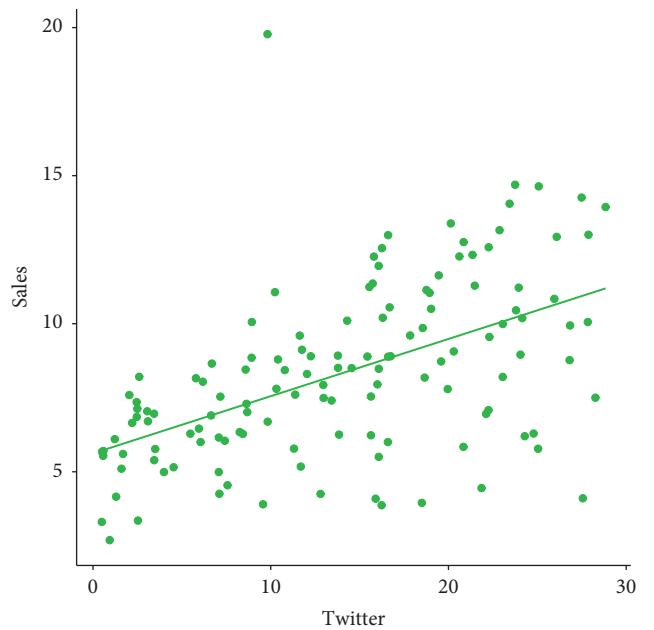


FIGURE 5: Visual display of the correlation test results.

2.5. Testing of Normality. As we stated above that in the domain of regression modeling, the assessment of the normality of data is one of the basic requirements. Basically, there are two approaches for testing the normality of the data: (i) numerically and (ii) graphically. For numerical analysis, we use the Shapiro-Wilk (SW) test for checking the normality of the sales and Twitter advertising data. This test

is carried out in the following manner, H_0 = the sale data is not normally distributed versus H_A = the sale data is normally distributed. After applying the SW test, we found that (i) for the Twitter advertising data, $SW = 0.96011$ with p value = 0.000739 and (ii) for the sale data, $SW = 0.967900$ with p value = 0.00356.

Since the p values associated with the Twitter advertising data and sales data are less than 0.05, we reject H_0 and accept the assumption of normality.

In addition to the numerical test (SW test), we also use a visual approach based on the QQ plot; see Figure 6. For assuming the normality, the QQ plot makes an angle of 45 degrees. Since all the data points are scattered around the referenced line, we can say that the sales and Twitter advertising data sets are normally distributed.

3. Statistical Modeling

This section consists of four sections: (i) in the very first section, we describe how the new model is obtained, (ii) in the second section, the proposed model is introduced and its different PDF (probability density function) plots are sketched, (iii) in the third section, we discuss the regularly varying tail behavior of the proposed model, and (iv) finally, a real application is discussed in the last section.

3.1. Development of the New Model. Before investing in a particular domain, it is always of interest to have the best description of that domain. The statistical distributions play a great role in describing the phenomena under consideration. Among the statistical distributions in the available literature, the EE (exponentiated exponential) model applied as an alternative to the two-parameter Weibull and Gamma distributions, has attracted many researchers. The DF (distribution function) of the EE model expressed by $H(y; \Phi)$ is

$$H(y; \Phi) = (1 - \exp\{-\delta_2 y\})^{\delta_1}, \quad y \geq 0, \delta_1, \delta_2 > 0, \quad (9)$$

where δ_1 is a shape and δ_2 is the scale parameter of the EE distribution. The PDF (probability density function) of the EE model represented by $h(y; \Phi)$ is

$$h(y; \Phi) = \delta_1 \delta_2 \exp\{-\delta_2 y\} (1 - \exp\{-\delta_2 y\})^{\delta_1 - 1}, \quad y > 0. \quad (10)$$

As we know in the finance sector, the data sets, particularly the financial returns, are frequently right-skewed with heavy tails. Among the class of statistical distributions, the HT distributions have shown better performance in modeling data in the finance sector where the financial returns experience HT, right-skewed, and unimodal behavior. According to Beirlant et al. [11], a model is called HT distribution, its SF (survival function) $\bar{H}(y; \Phi) = 1 - H(y; \Phi)$, satisfies

$$\lim_{y \rightarrow \infty} \frac{1 - H(uy; \Phi)}{1 - H(y; \Phi)} = \exp\{ay\} H(y; \Phi) = \infty, \quad (11)$$

where $a > 0$.

An important characteristic of the HT distributions is the RVP (regular variational property); see Resnick [12]. A model is called RVD (regular varying distribution) if it satisfies

$$\lim_{y \rightarrow \infty} \frac{1 - H(uy; \Phi)}{1 - H(y; \Phi)} = u^{-a}, \quad a \in \{0, \infty\}. \quad (12)$$

The distributions which possess the RVP are very useful models for dealing with the financial returns and HT data; see [13], [14], [15], and [16]. Numerous methods to obtain new HT and flexible distributions have been introduced. Among the available methods in the literature, the exponential TX family is a prominent method to generate HT distributions [17]. For $\sigma > 1$ and $y \in \mathbb{R}$, the DF $G(y; \Phi)$, and PDF $g(y; \Phi)$ of the exponential TX family are given by

$$G(y; \Phi) = 1 - \frac{\sigma \bar{H}(y; \Phi)}{\sigma - H(y; \Phi)}, \quad (13)$$

$$g(y; \Phi) = \frac{\sigma(\sigma - 1)h(y; \Phi)}{(\sigma - H(y; \Phi))^2}. \quad (14)$$

Keeping in view the role of the HT distributions in the finance sectors, we are motivated to study a new HT model called ETXE-exponential distribution. The ETXE-exponential distribution is obtained by incorporating equation (1) in equation (13).

The ETXE-exponential distribution is very interesting and offers to model the data that are skewed to right with a heavy tail. This fact is shown in the coming three sections. For example, the graphical behavior of its PDF is shown in section 3.2, where we can see that the ETXE-exponential distribution is a HT model. Secondly, the HT properties of the ETXE-exponential distribution are proved in section 3.3. Finally, positively skewed data related to the sales are analyzed in section 3.4, where we show that the ETXE-exponential distribution can be a better model for the HT financial data sets as compared to other existing distributions.

3.2. The ETXE-Exponential Distribution. A random variable Y has the ETXE-exponential model with scale parameters $\delta_2 > 0, \sigma > 1$ and shape parameter $\delta_1 > 0$, if its DF is

$$G(y; \Phi) = 1 - \frac{\sigma(1 - (1 - \exp\{-\delta_2 y\})^{\delta_1})}{\sigma - (1 - \exp\{-\delta_2 y\})^{\delta_1}}, \quad y \geq 0, \quad (15)$$

and PDF is

$$g(y; \Phi) = \frac{\delta_1 \delta_2 \sigma (\sigma - 1) \exp\{-\delta_2 y\} (1 - \exp\{-\delta_2 y\})^{\delta_1 - 1}}{(\sigma - (1 - \exp\{-\delta_2 y\})^{\delta_1})^2}, \quad y > 0. \quad (16)$$

Some possible behaviors of $g(y; \Phi)$ are sketched in Figure 7. The plots of $g(y; \Phi)$ in Figure 7 (left panel), are sketched for $\delta_1 = 1.2, \delta_2 = 0.8, \sigma = 1.8$ (red line); $\delta_1 = 2.8, \delta_2 = 0.8, \sigma = 1.8$ (green line); and $\delta_1 = 4.8, \delta_2 = 0.8, \sigma = 1.8$ (black line), whereas the plots of $g(y; \Phi)$ in Figure 7

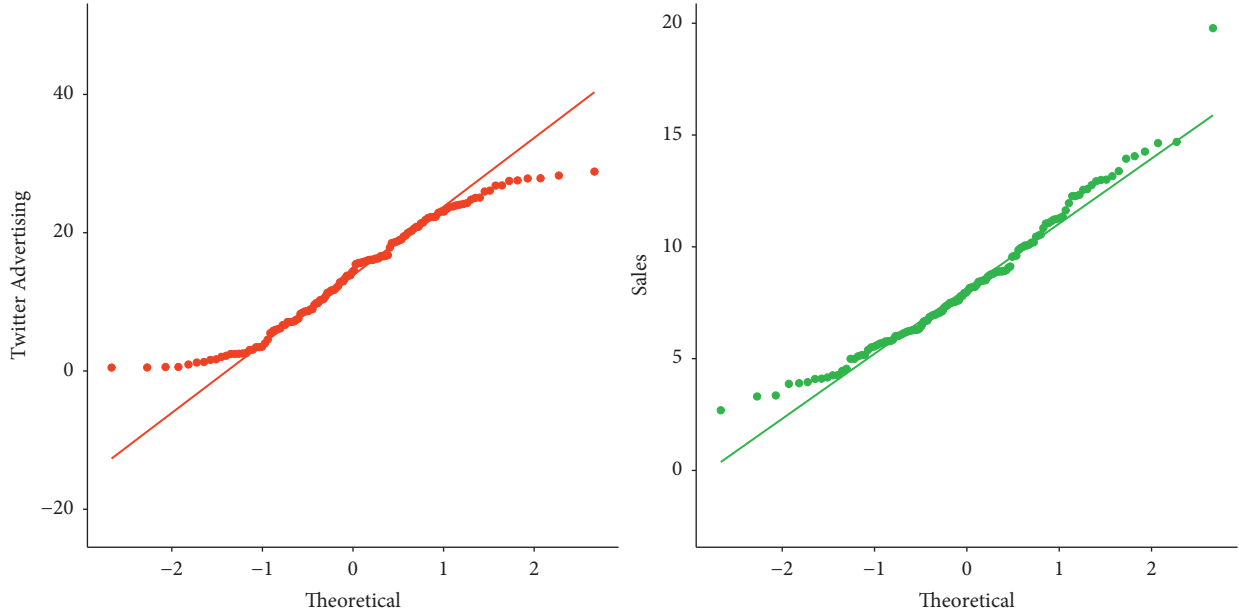


FIGURE 6: Graphical display of the SW tests result for the Twitter advertising and sale data.

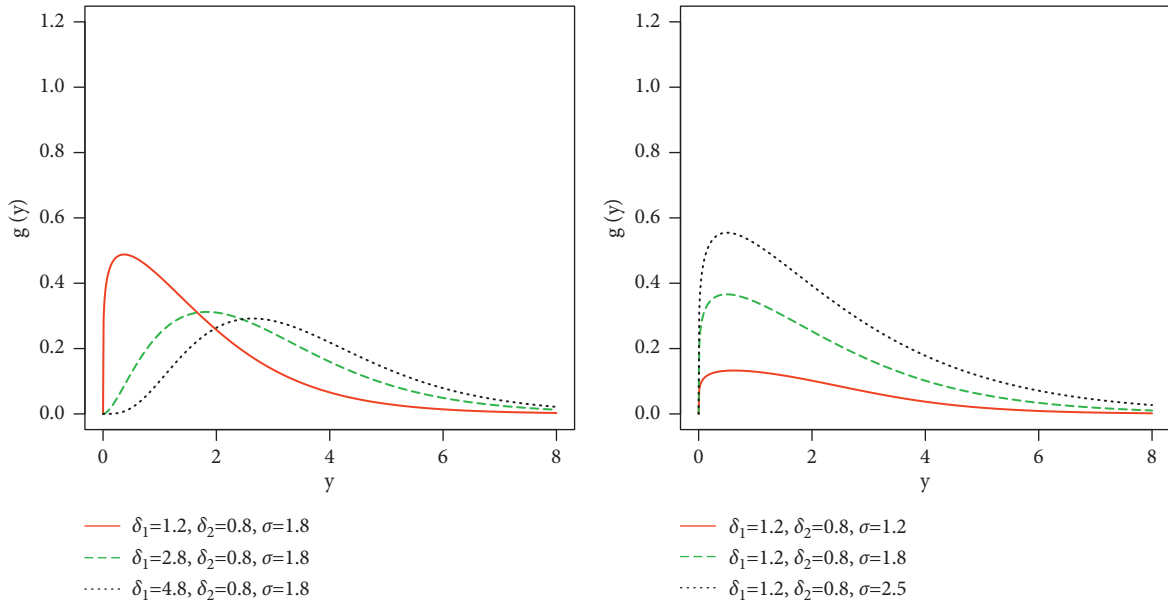


FIGURE 7: Different PDF plots of the ETXE-exponential distribution.

(right panel) are sketched for $\delta_1 = 1.2, \delta_2 = 0.8, \sigma = 1.2$ (red line); $\delta_1 = 1.2, \delta_2 = 0.8, \sigma = 1.8$ (green line); and $\delta_1 = 1.2, \delta_2 = 0.8, \sigma = 2.5$ (black line).

From the PDF plots sketched in Figure 7 (left panel), it is clear that when the value of δ_1 increases, the ETXE-exponential distribution tends to a HT distribution. Also, from the PDF plots sketched in Figure 7 (right panel), as the value of σ increases the proposed EXTE-exponential distribution possesses HT behavior.

3.3. The Heavy-Tailed Behavior. This section deals with the regularly varying tail behavior of the ETXE-exponential distribution which is very important to characterize the HT property of a model.

3.3.1. The Regularly Varying Tail Behavior. According to Karamata's theorem [18], in terms of SF (survival function) $\bar{G}(y; \Phi)$, we have the following.

Theorem 1. *If $\bar{H}(y; \Phi)$ is the SF of the RVD, then $\bar{G}(y; \sigma, \Phi)$ is a RVD.*

Proof. Suppose $\lim_{y \rightarrow \infty} \bar{H}(ky; \Phi) / \bar{H}(y; \Phi) = g(k)$ is finite but nonzero for every $k > 0$. Using equation (3), we have

$$\lim_{y \rightarrow \infty} \frac{\bar{G}(ky; \sigma, \Phi)}{\bar{G}(y; \sigma, \Phi)} = \lim_{y \rightarrow \infty} \frac{\sigma \bar{H}(ky; \Phi)}{\sigma \bar{H}(y; \Phi)} \cdot \frac{\sigma - H(y; \Phi)}{\sigma - H(ky; \Phi)} \tag{17}$$

Using equation (1) in equation (17), we get

$$\begin{aligned}
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)} \cdot \frac{\sigma - (1 - \exp\{-\delta_2 y\})^{\delta_1}}{\sigma - (1 - \exp\{-\delta_2 y\})^{\delta_1}}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)} \cdot \frac{\sigma - (1 - \exp\{-\delta_2 \infty\})^{\delta_1}}{\sigma - (1 - \exp\{-\delta_2 \infty\})^{\delta_1}}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)} \cdot \frac{\sigma - (1 - \exp\{\infty\})^{\delta_1}}{\sigma - (1 - \exp\{-\infty\})^{\delta_1}}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)} \cdot \frac{\sigma - (1 - 0)^{\delta_1}}{\sigma - (1 - 0)^{\delta_1}}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)} \cdot \frac{\sigma - 1}{\sigma - 1}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= \lim_{y \rightarrow \infty} \frac{\sigma \overline{H}(ky; \Phi)}{\sigma \overline{H}(y; \Phi)}, \\
\lim_{y \rightarrow \infty} \frac{\overline{G}(ky; \sigma, \Phi)}{\overline{G}(y; \sigma, \Phi)} &= g(k),
\end{aligned} \tag{18}$$

which is finite but nonzero for every $k > 0$; thus, $\overline{G}(y; \sigma, \Phi)$ is a RVD. \square

3.3.2. An Application of the Regularly Varying Tail Behavior. Let us assume that the distribution of Y has PLB (power law behavior); then, we have

$$1 - H(y; \Phi) = \overline{H}(y; \Phi) = \mathbb{P}(Y > y) \sim y^{-\beta}. \tag{19}$$

Utilizing Karamata's theorem [18], we can write $\overline{G}(y; \Phi)$ as

$$\overline{G}(y; \Phi) = y^{-\beta} L(y; \Phi), \tag{20}$$

where $L(y; \Phi)$ is slowly varying. Note that

$$\overline{G}(y; \Phi) = \frac{\sigma \overline{H}(y; \Phi)}{\sigma - 1 + \overline{H}(y; \Phi)}. \tag{21}$$

Since $\overline{H}(y; \Phi) \sim y^{-\beta}$, we can write

$$\overline{G}(y; \Phi) = \frac{\sigma y^{-\beta}}{\sigma - 1 + y^{-\beta}} = y^{-\beta} L(y; \Phi), \tag{22}$$

where $L(y; \Phi) = \sigma / (\sigma - 1 + y^{-\beta})$. If $L(y; \Phi)$ is a SVF (slowly varying function), then the variational result obtained is true. As per Resnick [12], for all $\beta > 0$, we have to show that

$$\lim_{y \rightarrow \infty} \frac{L(\beta y; \Phi)}{L(y; \Phi)} = 1. \tag{23}$$

After some computation, we get

$$\frac{L(\beta y; \Phi)}{L(y; \Phi)} = \frac{\sigma(\sigma - 1 + (1/y^\beta))}{\sigma(\sigma - 1 + (1/y^\beta b^\beta))}. \tag{24}$$

If $y \rightarrow \infty$, then $\lim_{y \rightarrow \infty} (1/y^\beta) = 0$, and $\lim_{y \rightarrow \infty} (1/y^\beta b^\beta) = 0$. Therefore, from equation (14), we have

$$\frac{L(\beta y; \Phi)}{L(y; \Phi)} = \frac{\sigma(\sigma - 1)}{\sigma(\sigma - 1)}. \tag{25}$$

Finally, we get the required result

$$\lim_{y \rightarrow \infty} \frac{L(\beta y; \Phi)}{L(y; \Phi)} = 1. \tag{26}$$

3.4. Analyzing the Sales Data. As we mentioned earlier that the HT statistical models are very useful in describing the financial phenomena. In this section, we show the best fitting power of the ETXE-exponential model via analyzing the sales data which consists of 130 observations. The Twitter advertising and sales data sets are available at <https://data.world/datasets/twitter>. Table 1 offers the BMs (basic measures) of the sales data.

The total time test (TTT), histogram, and box plot of the sales data are provided in Figure 8. From Figure 8, we can easily observe that the data set is unimodal and skewed to the right. The data set possessing such characteristics can be better modeled via using the ETXE-exponential distribution.

The comparison of the ETXE-exponential is made with three well-known distributions available in the literature. The DFs of the competing models are as follows:

TABLE 1: The SMs of the sales data.

Observations	Min.	1 st Qu.	Median	Mean	3 rd Qu.	Max.
130	2.691	6.167	7.997	8.310	10.091	19.777

(1) MOW (Marshall–Olkin Weibull) model [19]:

$$H(y; \Phi) = 1 - \frac{\sigma e^{-\delta_2 y^{\delta_1}}}{\left(1 - (1 - \sigma)e^{-\delta_2 y^{\delta_1}}\right)}, \quad y \geq 0, \sigma, \delta_1, \delta_2 > 0. \quad (27)$$

(2) FW (flexible Weibull) model [20]:

$$H(y; \Phi) = 1 - e^{-e^{(\gamma y^{\beta})^{\gamma}}}, \quad y \geq 0, \gamma, \beta > 0. \quad (28)$$

(3) APTW (alpha power transformed Weibull) model [21]:

$$H(y; \Phi) = \frac{\alpha^{1 - e^{-\delta_2 y^{\delta_1}}} - 1}{\alpha - 1}, \quad y \geq 0, \alpha, \delta_1, \delta_2 > 0, \alpha \neq 1. \quad (29)$$

The term “better modeling” is used in the sense that the ETXE-exponential distribution has the smaller values of the selected IC (information criterion) considered for comparison. The expressions of the IC are given by the following:

(4) The AIC (Akaike IC) [22] is

$$AIC = 2k - 2\ell. \quad (30)$$

(5) The BIC (Bayesian IC) [23] is

$$BIC = k \log(n) - 2\ell. \quad (31)$$

(6) The HQIC (Hannan–Quinn IC) [24] is

$$HQIC = 2k \log(\log(n)) - 2\ell. \quad (32)$$

(7) The CAIC (Corrected Akaike IC) [25] is

$$CAIC = \frac{2nk}{n - k - 1} - 2\ell. \quad (33)$$

In addition to the IC, we further considered three goodness of fit measures (g-o-f) including the following:

(1) The AD (Anderson–Darling) test statistic:

$$AD = -k - \frac{1}{k} \sum_{u=1}^k (2u - 1) [\log H(y_u) + \log\{1 - H(y_{k-u+1})\}]. \quad (34)$$

(2) The CM (Cramer–von Mises) test statistic:

$$CM = \frac{1}{12k} + \sum_{u=1}^k \left[\frac{2u - 1}{2k} - H(y_u) \right]^2. \quad (35)$$

(3) The KS (Kolmogorov–Smirnov (KS) test statistic:

$$KS = \sup_y [H_k(y) - H(y)]. \quad (36)$$

The optimization technique is adopted to get the MLEs (maximum likelihood estimators) of the competing models. In this section, we implement the Newton–Raphson (NR) iteration approach to get the MLEs of the fitted models using the sales data. The MLEs of the fitted models are provided in Table 2. The values of the IC are reported in Table 3, and the g-o-f measures with p value are presented in Table 4.

Based on the sales data, the results reported in Tables 3 and 4, we can see that the ETXE-exponential distribution has the smallest values of the IC and g-o-f measures. The values of the selected measures for the ETXE-exponential distribution are IC = 0.0237, CAIC = 0.2027, BIC = 0.0336, HQIC = 0.9985, AD = 0.0237, CM = 0.2027, KS = 0.0336, and p value = 0.9985. For the second-best competitive model (FW distribution), these measures are AIC = 648.4654, CAIC = 648.5599, BIC = 654.2005, HQIC = 650.7957, AD = 0.0719, CM = 0.4315, KS = 0.0861, and p value = 0.2896. We can see that the FW is the second-best model, but not in terms of KS and p value. From Table 4, it is clear that the APTW distribution is the second-best model as it has the second smallest KS value which is 0.0559, and the second-highest p value which is 0.8106.

Furthermore, for the best description of the results of the sales data, the estimated PDF and CDF plots of the fitted models (EXTE-exponential (red line), MOW (green line), APTW (pink line), and FW (blue line)) are provided in Figure 9. The PP (probability-probability), and QQ plots are sketched in Figures 10 and 11, respectively.

4. Future Research Directions

In this work, we restricted our study to simple linear regression analysis by using the Twitter medium as a predictor variable. However, many other online mediums can be used for advertising purposes such as YouTube, Facebook, and Twitter. Therefore, in the future, we are intended to implement the multiple linear regression modeling approach to see the impact of different advertising mediums on sales. Generally, the multiple linear regression model (MLRM) with k predictors is given by

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots + \theta_k X_k + \varepsilon. \quad (37)$$

The future research study will be carried out in three different phases. For example, (i) in phase one, the MLRM will be implemented with two (X_1 = medium 1, X_2 = medium 2) advertising mediums; (ii) in phase two, the MLRM will be used by considering three (X_1 = medium 1, X_2 = medium 2, X_3 = medium 3) advertising mediums; and (iii) in phase three, the MLRM will be applied by taking four (X_1 = medium 1, X_2 = medium 2, X_3 = medium 3, X_4 = medium 4) advertising mediums.

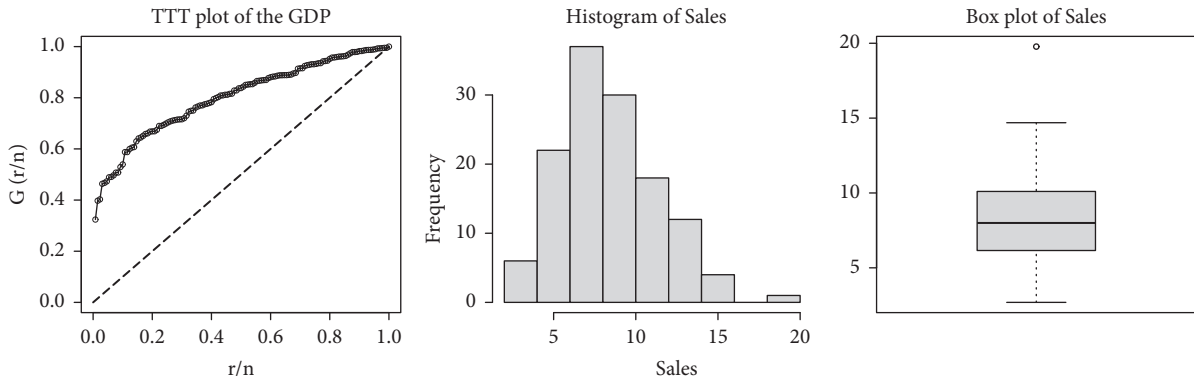


FIGURE 8: The TTT, histogram, and box plots for the sales data.

TABLE 2: The MLEs of the fitted models.

Models	δ_1	δ_2	σ	α	γ	β
ETXE-exponential	13.6215	0.4655	1.7019	—	—	—
MOW	2.6988	0.0030	1.4851	—	—	—
APTW	1.8410	0.0367	—	44.6860	—	—
FW	—	—	—	—	0.1482	13.0846

TABLE 3: The IC of the fitted models.

Models	AIC	CAIC	BIC	HQIC
ETXE-exponential	644.9836	645.1741	653.5862	648.4791
MOW	655.2653	655.4558	663.8679	658.7608
APTW	651.5548	651.7453	660.1574	655.0503
FW	648.4654	648.5599	654.2005	650.7957

TABLE 4: The goodness of fit measures of the fitted models.

Models	AD	CM	KS	p value
ETXE-exponential	0.0237	0.2027	0.0336	0.9985
MOW	0.1433	0.8511	0.0773	0.4188
APTW	0.0912	0.5847	0.0559	0.8106
FW	0.0719	0.4315	0.0861	0.2896

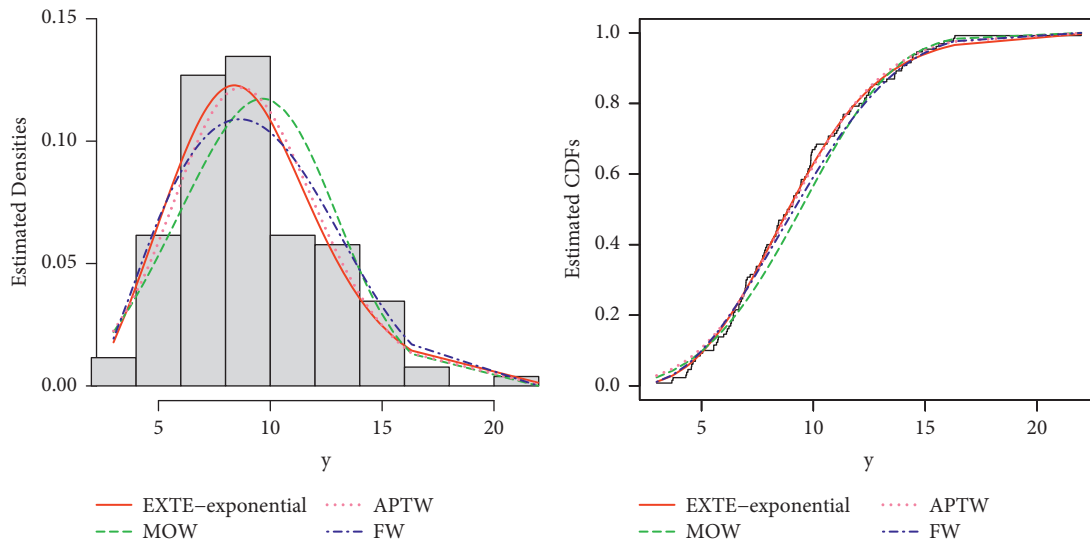


FIGURE 9: The estimated PDF, CDF, and KMS plots of the fitted distributions.

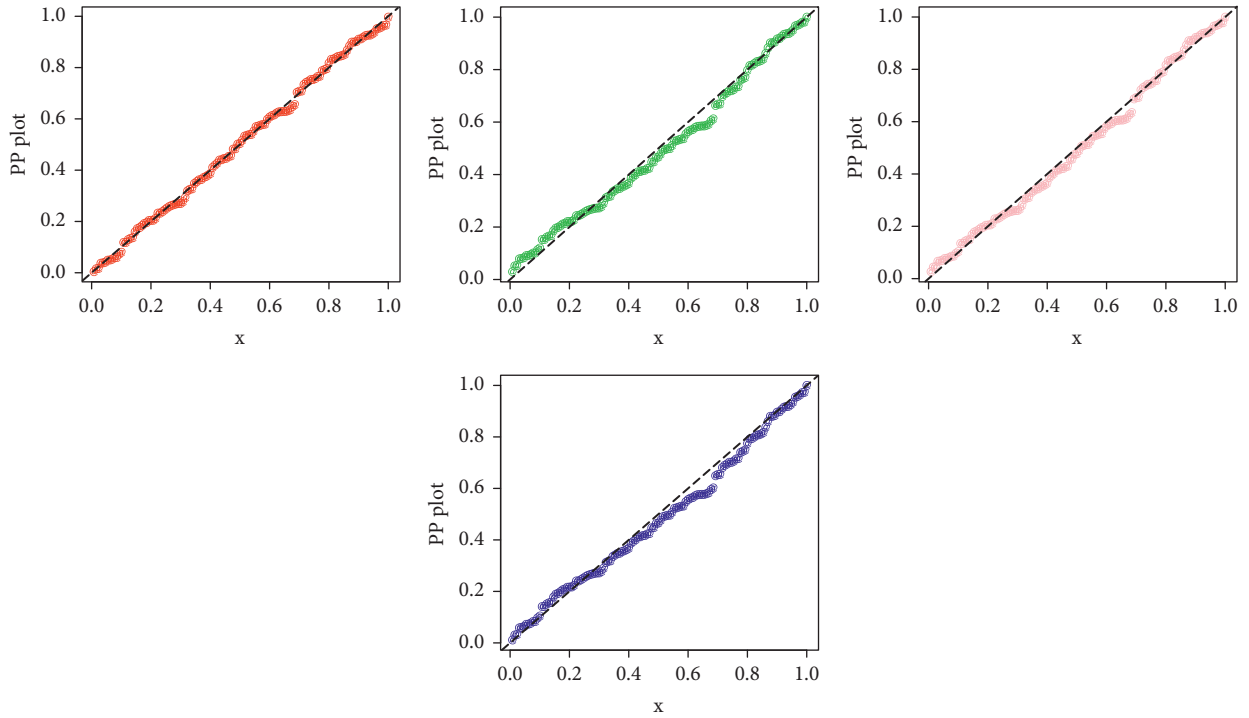


FIGURE 10: The PP plots of the fitted distributions.

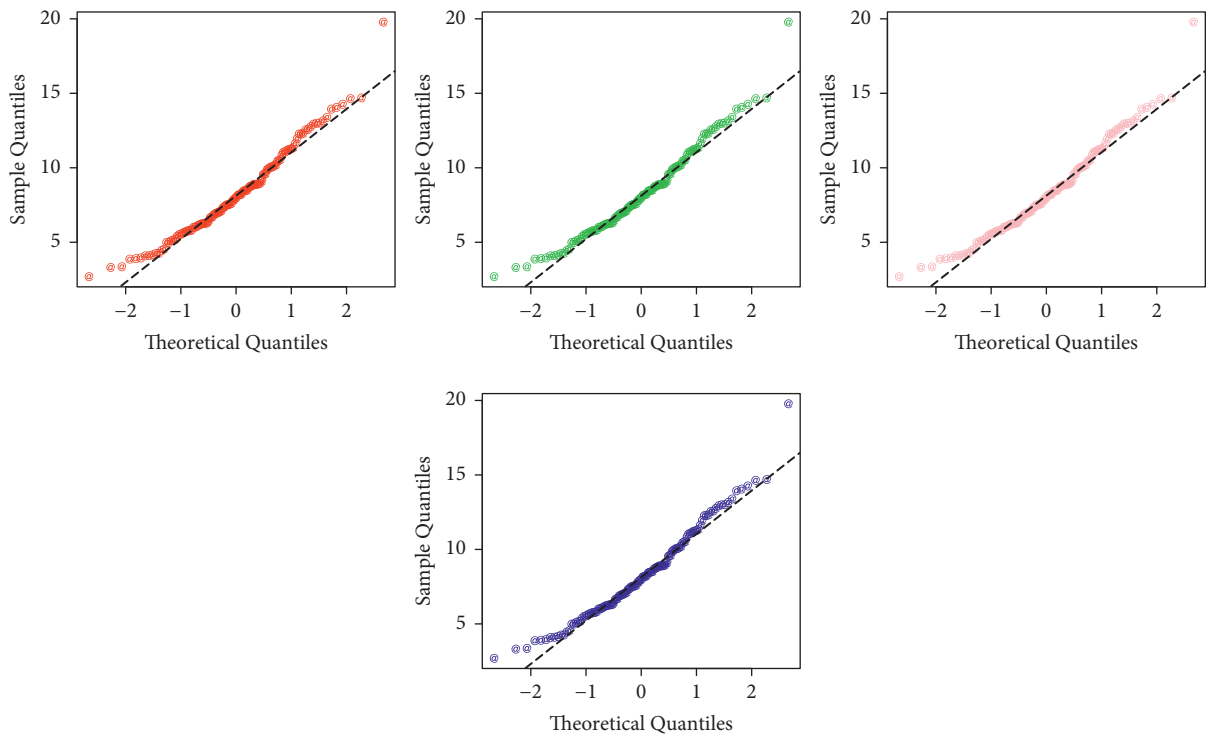


FIGURE 11: The QQ plots of the fitted distributions.

4.1. *The MLRM with Two Predictors.* The MLRM with two predictors is given by

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \varepsilon. \tag{38}$$

In the future, we are planning to use the MLRM defined in equation (38) to see the impact of two different advertising mediums on sales. The MLRM models with possible combinations of two different advertising mediums are given by the following:

- (1) Effect of Twitter and YouTube advertising mediums on sales

The MLRM with two advertising mediums (X_1 =Twitter and X_2 =YouTube) is given by

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{YouTube} + \varepsilon. \quad (39)$$

- (2) Effect of Twitter and Facebook advertising mediums on sales

The MLRM with two advertising mediums (X_1 =Twitter and X_2 =Facebook) is given by

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{Facebook} + \varepsilon. \quad (40)$$

- (3) Effect of Twitter and Instagram advertising mediums on sales.

The MLRM with two advertising mediums (X_1 =Twitter and X_2 =Instagram) is given by

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{Instagram} + \varepsilon. \quad (41)$$

4.2. The MLRM with Three Predictors. The MLRM with three predictors is given by

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \varepsilon. \quad (42)$$

In the future, we are also interested in using the MLRM defined in equation (42) to see the impact of three different advertising mediums on sales. The MLRM models with possible combinations of three different advertising mediums are given by the following:

- (1) Effect of Twitter, YouTube, and Facebook advertising mediums on sales.

The MLRM with three different advertising mediums (X_1 =Twitter, X_2 =YouTube, X_3 =Facebook) is given by

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{YouTube} + \theta_3 \text{Facebook} + \varepsilon. \quad (43)$$

- (2) Effect of Twitter, YouTube, and Instagram advertising mediums on sales.

The MLRM with three different advertising mediums (X_1 =Twitter, X_2 =YouTube, X_3 =Instagram) is given by

$$\text{Sales} = \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{YouTube} + \theta_3 \text{Instagram} + \varepsilon. \quad (44)$$

4.3. The MLRM with Four Predictors. The MLRM with four predictors is given by

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \varepsilon. \quad (45)$$

In the future, we are also motivated to use the MLRM defined in equation (45) to see the impact of four different advertising mediums on sales.

- (1) Effect of Twitter, YouTube, Facebook, and Instagram advertising mediums on sales

The MLRM with four different advertising mediums (X_1 =Twitter, X_2 =YouTube, X_3 =Facebook, X_4 =Instagram) is given by

$$\begin{aligned} \text{Sales} = & \theta_0 + \theta_1 \text{Twitter} + \theta_2 \text{YouTube} + \theta_3 \text{Facebook} \\ & + \theta_4 \text{Instagram} + \varepsilon. \end{aligned} \quad (46)$$

5. Concluding Remarks

In this research, we studied the relationship between Twitter advertising and sales by adopting the simple linear regression approach. For testing the impact of Twitter medium as an advertising tool, two main statistical approaches such as t -test and correlation test are utilized. Based on the results of the t -test and correlation test, it is observed that Twitter advertising plays a useful role in increasing sales. Furthermore, a new HT model called the ETXE-exponential distribution is developed by combining the exponentiated-exponential model with the ETX family. The ETXE-exponential distribution was introduced to provide a close fit to sales data, which is an important area of the finance sector. For proving the applicability of the proposed model, the sales data is analyzed. The comparison of the ETXE-exponential distribution was made with MOW, APTW, and FW distributions. By taking into account certain analytical tools, we showed that the ETXE-exponential distribution is the best model for modeling data in the finance sector.

Data Availability

The data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Jia Cong and Zubair Ahmad conceptualized the study. Zubair Ahmad developed the methodology. Jia Cong, Zubair Ahmad, and Basim S.O. Alsaedi wrote the original draft. Zubair Ahmad and Osama Abdulaziz Alamri performed formal analysis. Jia Cong performed supervision. Zubair Ahmad, Ibrahim Alkhairy, and Hassan Alsuhabi provided software. Jia Cong and Osama Abdulaziz Alamri performed investigation. Zubair Ahmad, Ibrahim Alkhairy, Basim S.O. Alsaedi, and Hassan Alsuhabi reviewed and edited the article.

Acknowledgments

The authors acknowledge the support of Dr. Zhou Yang (Xiamen University, China) and Dr. M. Nagy (Fayoum

University, Egypt) for their support and guidance in preparing this paper.

References

- [1] A. Lambrecht, C. Tucker, and C. Wiertz, "Advertising to early trend propagators: evidence from Twitter," *Marketing Science*, vol. 37, no. 2, pp. 177–199, 2018.
- [2] R. P. Sari, "Marketing mix implementation in small medium enterprises: a study of Galeristorey online business," *Etikonomi*, vol. 16, no. 1, pp. 115–126, 2017.
- [3] M. De Rosa, G. Fenza, A. Gallo, M. Gallo, and V. Loia, "Pharmacovigilance in the era of social media: discovering adverse drug events cross-relating Twitter and PubMed," *Future Generation Computer Systems*, vol. 114, pp. 394–402, 2021.
- [4] W. Janssens and P. De Pelsmacker, "Advertising for new and existing brands: the impact of media context and type of advertisement," *Journal of Marketing Communications*, vol. 11, no. 2, pp. 113–128, 2005.
- [5] A. Anagnostopoulos, F. Petroni, and M. Sorella, "Targeted interest-driven advertising in cities using Twitter," *Data Mining and Knowledge Discovery*, vol. 32, no. 3, pp. 737–763, 2018.
- [6] J. L. Hayes, G. Golan, B. Britt, and J. Applequist, "How advertising relevance and consumer-Brand relationship strength limit disclosure effects of native ads on Twitter," *International Journal of Advertising*, vol. 39, no. 1, pp. 131–165, 2020.
- [7] E. Murillo, M. Merino, and A. Núñez, "The advertising value of Twitter Ads: a study among Mexican Millennials," *Review of Business Management*, vol. 18, no. 61, pp. 436–456, 2016.
- [8] T. Yigitcanlar, N. Kankanamge, and K. Vella, "How are smart city concepts and technologies perceived and utilized? A systematic geo-twitter analysis of smart cities in Australia," *Journal of Urban Technology*, vol. 28, no. 1–2, pp. 135–154, 2021.
- [9] S. Lou, "Applying data analytics to social media advertising: a twitter advertising campaign case study," *Journal of Advertising Education*, vol. 21, no. 1, pp. 26–32, 2017.
- [10] V. Israel-Turim, J. L. Micó-Sanz, and E. Ordeix-Rigo, "Who did the top media from Spain started following on twitter? An exploratory data analysis case study," *American Behavioral Scientist*, vol. 65, no. 3, pp. 512–539, 2021.
- [11] J. Beirlant, G. Matthys, and G. Dierckx, "Heavy-tailed distributions and rating," *ASTIN Bulletin*, vol. 31, no. 1, pp. 37–58, 2001.
- [12] S. I. Resnick, "Discussion of the Danish data on large fire insurance losses," *ASTIN Bulletin*, vol. 27, no. 1, pp. 139–151, 1997.
- [13] A. Punzo, A. Mazza, and A. Maruotti, "Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated gamma distributions," *Journal of Applied Statistics*, vol. 45, no. 14, pp. 2563–2584, 2018.
- [14] D. Bhati and S. Ravi, "On generalized log-Moyal distribution: a new heavy tailed size distribution," *Insurance: Mathematics and Economics*, vol. 79, pp. 247–259, 2018.
- [15] A. Punzo, "A new look at the inverse Gaussian distribution with applications to insurance and economic data," *Journal of Applied Statistics*, vol. 46, no. 7, pp. 1260–1287, 2019.
- [16] A. Punzo and L. Bagnato, "Modeling the cryptocurrency return distribution via Laplace scale mixtures," *Physica A: Statistical Mechanics and Its Applications*, vol. 563, Article ID 125354, 2021.
- [17] Z. Ahmad, E. Mahmoudi, M. Alizadeh, R. Roozegar, and A. Z. Afify, "The exponential TX family of distributions: properties and an application to insurance data," *Journal of Mathematics*, vol. 2021, Article ID 3058170, 18 pages, 2021.
- [18] E. Seneta, "Karamata's characterization theorem, feller and regular variation in probability theory," *Publications de l'Institut Mathématique*, vol. 71, no. 85, pp. 79–89, 2002.
- [19] A. Marshall and I. Olkin, "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families," *Biometrika*, vol. 84, no. 3, pp. 641–652, 1997.
- [20] M. Bebbington, C.-D. Lai, and R. Zitnikis, "A flexible Weibull extension," *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 719–726, 2007.
- [21] S. Dey, V. K. Sharma, and M. Mesfioui, "A new extension of Weibull distribution with application to lifetime data," *Annals of Data Science*, vol. 4, no. 1, pp. 31–61, 2017.
- [22] H. Akaike, "A new look at the statistical model identification," in *Selected Papers of Hirotugu Akaike*, pp. 215–222, Springer, New York, NY, USA, 1974.
- [23] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [24] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society: Series B*, vol. 41, no. 2, pp. 190–195, 1979.
- [25] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.