Alfarama Journal of Basic & Applied Sciences

https://ajbas.journals.ekb.eg
ajbas@sci.psu.edu.eg

Faculty of Science Port Said University

http://sci.psu.edu.eg/en/

July 2023, Volume 4 , Issue III

DOI: 10.21608/ajbas.2023.165619.1127

ISSN 2682-275X

Submitted: 27/09/2022
Accepted: 23/02/2023

Pages: 485 - 494

Estimation of Log-Pearson Type III Distribution Using Percentile Roots: ATM Transactions Over 24 Years in Saudi Arabia as a Case Study

Mohammed El Genidy [1,*], Wesal Megahed [2], Khaled Mahfouz [1]

[1] Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt

[2] Department of Basic Science, Obour Higher Institute for Management Informatics, Egypt

***Corresponding author:** drmmg2016@yahoo.com

ABSTRACT

Estimation methods of three-parameter distributions are essential in data-fitting distributions. In this paper, an application to a simple estimation method named Percentile Root (PR) was presented with ATM transactions as a case study. The PR method was applied to the probability distribution Log-Pearson type III distribution. The statistical properties of the distribution are exploited by PR to obtain the estimated parameters, ensuring the efficiency of the method. The Anderson-Darling and Kolmogorov-Smirnov test were performed on the results of PR method. The significant results ofPR method are compared with Maximum Likelihood Estimation (MLE), and it is clear that the PR method is simple in coding by a computer and provides precise results. The point is highly beneficial to applications of economic and commercial sciences. The Log-Pearson type III distribution was used by PR method and fitted to the ATM data set to overcome the problems of predicting the prospect numbers of financial transactions in banks.

Key Words:

Anderson-Darling Test; ATM Transactions; Log-Pearson Type III; Percentile Roots

1. INTRODUCTION

Researchers have faced severe difficulties in modelling, estimating, and obtaining the prediction equation. Mathematical modelling and forecasting are significant issues to precise decision process. Therefore, significant efforts have been devoted in developing several statistical estimation methods. A statistical method for predicting the design flood for a river at a particular site uses frequency distribution data to fit the Log-Pearson type III distribution. A frequency analysis can be created when the statistical information for the river site has been calculated [1]. The research community was interested in the log Pearson type III distributions because they are used in many scientific fields, including hydrology and seismology [2]. Additionally, the log Pearson type III distribution could be used to assess, design magnitudes and effectively describe the behavior of the maximum earthquake magnitudes for all ranges

[3]. The hydro-logic literature has investigated a number of parameter estimation methods for the log-Pearson type III distribution, including the method of moments in both log and real space, maximum likelihood estimators (MLEs), and the method of mixed moments. The comparing of the three methods side-by-side, by Griffis and Stedinger illustrate the discrepancies resulting from the conflicting conclusions [4]. Langat et al. discussed the current approaches for identifying the probability distribution functions which are best fit for estimating maximum, minimum and mean stream flow [5]. Parameter estimation can be complicated and difficult. That is why applying and studying new proposed methods, such as Percentile Root, is required.

Researchers usually use only P25 and P75 percentiles because they ensure the efficiency of the estimates. Percentile algorithm was applied as an estimation method to Log-logistic distribution [6]. Nested percentile algorithm as paired equations has been proposed and studied using Weibull distribution [7]. The percentile is quite useful for dealing with parameter estimation. Klugman et al. provide more information on the percentile methodology [8]. This article is an extension to a previous work on the PR [9]. PR method is applied to several distributions such as Log-normal, Fatigue lifetime, Erlang, Pert and Fréchet. The estimates are compared to ML estimates to verify the method. The PR estimates of Log-Pearson type III distributions to the ATM data set are obtained through the practical steps of the method. The estimates depend on combining percentile equations with central-tendency measurements, like the method of the maximum likelihood.

2. Main Results

The results show that the PR method enhances the estimated parameters to match the data in a more effective way. The Log-Pearson Type III distribution fitting data, estimated by PR and MLE are shown in Fig. 4. PR has illustrated efficient and effective results. Maximum Likelihood and is considered popular and generally used estimators, even though it is hard to apply in comparison to PR. PR also includes percentiles equations with the relevant measures such as the median, mean, and variance.

Hence, the PR simply handles the equations of percentiles. As described, PR is an acceptable and appropriate method of parameter estimating, especially for distributions with three parameters. Compared to MLE, estimated parameters seem accurate and efficient. The difficulty of estimating three parameters is decreased because PR provides a single equation that can be numerically solved using any software, such as Mathematica's FindRoot function. Unlike the traditional old MLE method, this is considered a research improvement and can be handled numerically or through mathematics-related applications.

Percentile Roots (PR) Estimation Method: The PR method is based on solving equations to obtain numerically approximated values for unknown parameters.The chosen equations must be solvable in a closed form. The equations used are percentiles of the CDF equation for all values of the variable X, with the theoretical mean and variance. The procedure combines the percentile roots with common measures of the given dataset and probability distribution properties, as mean, median, and skewness, given an accurate estimation of the parameters based on empirical measures from the given dataset. Theoretically, the steps to apply PR to any probability distribution are illustrated below.

Algorithm of PR

Step 1. Determine the dataset and the appropriate probability distribution $f(x;\alpha,\beta,\theta)$.

Step 2. Evaluate the essential measures to the data; Mean, Median, and standard deviation

Step 3. Get the parameters in terms of one, say θ *;* $R:\alpha \rightarrow f(\theta),\beta \rightarrow g(\theta)$

Step 4. Substitute parameter in cdf equation $F(x;\alpha,\beta,\theta) \rightarrow F(x;\theta) = U_i$.

Step 5. Solve the cdf equation to get the value of the parameter $\theta = F^{-1}(x;\theta)$.

Step 6. Obtain the value of other parameters.

Step 7. Check the validation of parameters through theoretical mean, variance, and

$$0 < F(x_i;\alpha,\beta,\theta) < 1;$$
$$F(x_1) < F(x_2) < \ldots < F(x_N)$$

Step 8. If step 7 is valid; Repeat steps 5,6, and 7 for N times to obtain $\alpha,\beta,$ and θ.

Step 9. If step 7 is not valid, repeat step 3 with different relation R $R_2:\alpha \rightarrow l(\theta),\beta \rightarrow w(\theta)$,

then continue with 4-7, and check again

Step 10. The estimated parameter value is the median of all obtained values.

The most challenging step in the Algorithm is to select the relation/functions among the parameters. As, some relations are not acceptable at all, others are acceptable, and others are the optimal. Therefore, possible relations must be investigated and compared to MLE to select the best relations with the best estimates.

Three-parameter Log-Pearson Distribution: Let $Y = \ln(X)$ where X is a positive random variable [10]. If Y follows a Pearson III distribution, then X will be a log-Pearson type III distribution variable with a shape parameter α, a scale parameter β, and a location parameter γ.

The probability density function, as mentioned in [11], Where $0 < \alpha,\ \beta \neq 0$, is defined as follows:

$$f\left(x;\alpha,\beta,\gamma\right) = \frac{1}{x\ \beta\ \Gamma(\alpha)}\left(\frac{\ln(x)-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\frac{\ln(x)-\gamma}{\beta}\right) \quad (1)$$

The cumulative distribution function (CDF) is:

$$F\left(x;\alpha,\beta,\gamma\right)=\frac{\Gamma_{\frac{\ln(x)-\gamma}{\beta}}(\alpha)}{\Gamma\left(\alpha\right)} \tag{2}$$

$$F(x;\alpha,\beta,\gamma)=\int_{e^{\gamma}}^{x} f(t;\alpha,\beta,\gamma)\ \mathrm{dt}$$

$$=\int_{e^{\gamma}}^{x}\frac{1}{\mathrm{t}\beta\Gamma(\alpha)}\left(\frac{\ln(\mathrm{t})-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\frac{\ln(\mathrm{t})-\gamma}{\beta}\right)\mathrm{dt}$$

Let $u=\frac{\ln(t)-\gamma}{\beta}$ $\rightarrow$ $\mathrm{du}=\frac{1}{\mathrm{t}\,\beta}\mathrm{dt}$,

then the boundaries are at $t=e^{\gamma}$ then $u=0$ $t=x$ then u= $\frac{\ln(x)-\gamma}{\beta}$ $\rightarrow$ ∞

$$=\frac{1}{\Gamma(\alpha)}\int_{0}^{u}(u)^{\alpha-1}\exp(-u)\ \mathrm{du}$$

$$=\frac{1}{\Gamma(\alpha)}\ \Gamma_{u}\left(\alpha\right)$$

Note that: $\int_{0}^{x}(\mathrm{t})^{\alpha-1}\exp(-\mathrm{t})\ \mathrm{dt}$ called the lower incomplete gamma function $\Gamma_{x}(\alpha)$

The used statistical properties are the mean, skew and variance equations shown as:

Mean: $E(x)=\gamma+\alpha\beta$ (3)

variance: $\mathrm{var}(x)=\alpha\beta^{2}$ (4)

The shape parameter (β), and location parameter (γ) are described in terms of the empirical expected value $\bar{X}$ and the scale parameter (α). The mean, and variance equation 3, 4, are employed and combined to obtain a relation between the shape parameter (β) and the scale parameter (α). The percentiles of CDF with variable x, location parameter, and scale parameter values are substituted in the equation (2) as values or relations to be solved to obtain the values of shape parameter (α).

$$\beta=\frac{s}{\sqrt{\alpha}} \tag{5}$$

$$\gamma=\mu-\alpha\beta$$

$$=\mu-\alpha\frac{s}{\sqrt{\alpha}}$$

$$\gamma=\mu-s\sqrt{\alpha} \tag{6}$$

$$F\left(x;\alpha\right)=\frac{\Gamma_{\frac{\ln(x)-\mu+s\sqrt{\alpha}}{s/\sqrt{\alpha}}}(\alpha)}{\Gamma\left(\alpha\right)}=\frac{\Gamma_{\frac{\sqrt{\alpha}}{s}(\ln(x)-\mu)+\alpha}(\alpha)}{\Gamma\left(\alpha\right)} \tag{7}$$

The estimated value of the parameter is obtained through solving the equation (7) numerically using Findroot function as shown below:

```
FindRoot[Gamma[Sqrt[α]/s * (Log[x] - m) + α, α] / Gamma[α] == y, {α, 0.1}]
```

Where (s) is the standard deviation and (y) is the percentile value, y = i/n.

Then, via substituting the value of in the equations (5),(6) then the estimates of parameters and are obtained. As mentioned before, the most challenging part is to select the relations among parameters. If the skew were used to find the shape parameter, then a relation between parameters and is specified, the estimates would not be acceptable. If the relations among parameters are defined in terms of β, the estimates would be acceptable, but not the best values.

3. Application on PR

Data Description: The given data set is the statistics of Automated Teller Machines (ATM) in Saudi Arabia from 1996 to 2021. Data set is reported by Saudi Central Bank (SAMA), to enhance energy economics research. It is provided online at datasource.kapsarc.org. (Fig.1) describes variables states of the data set as time series to show the change of years according to [12].

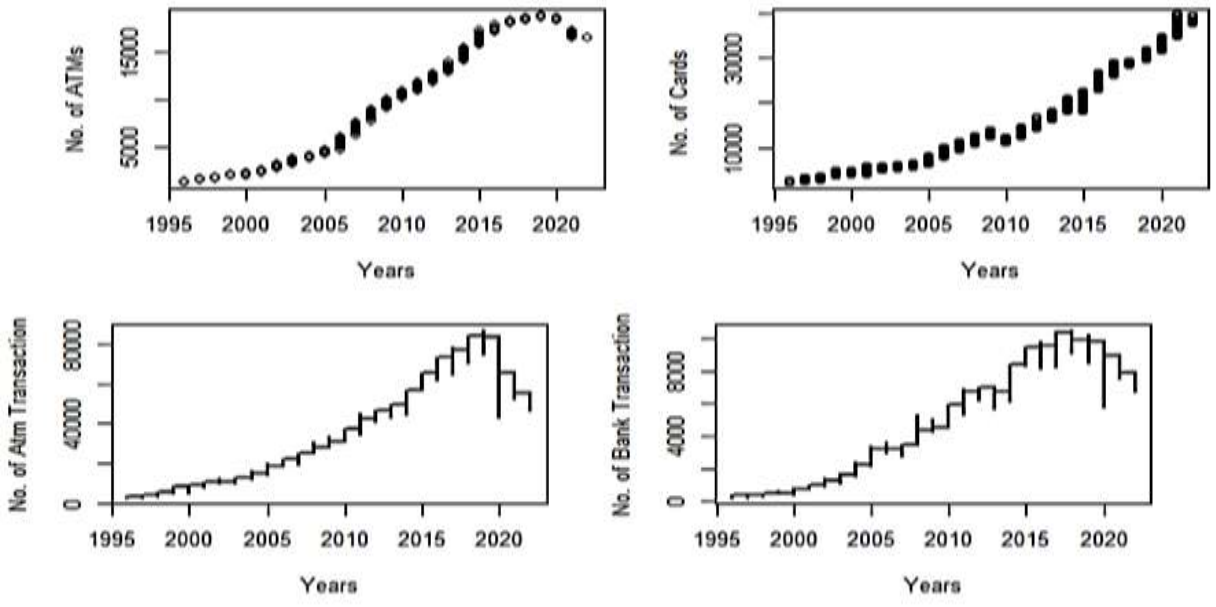

Fig (1): Time series of datasets includes ATMs and banks over 24 years, Saudi Arabia.

As the changes in the ATM statistics over the years are showed in Fig. 1. The number of ATM has rapidly increased along with the number of Cards since 2005. The number of ATM transactions has increased 10 times the number of Bank Transactions. Yet, there has been a drop in both in 2019 because of Covid-19 epidemic. Then, the online transactions have become more prevalent. The total number of transactions in both ATM and Bank will be applied in this study.

Several distributions can fit the given data set as shown in (Fig. 2) such as Weibull, Gamma, Log-normal and others. Log-Pearson is chosen in this study, although it is not the optimal choice. As the purpose of the proposed method "Nested Percentile" is to provide estimation values that provide an accurate fitting to the data and enhance the fitting by changing the relations among parameters. The validity, and efficiency of PR results is measured using (K-S) and (A-D).

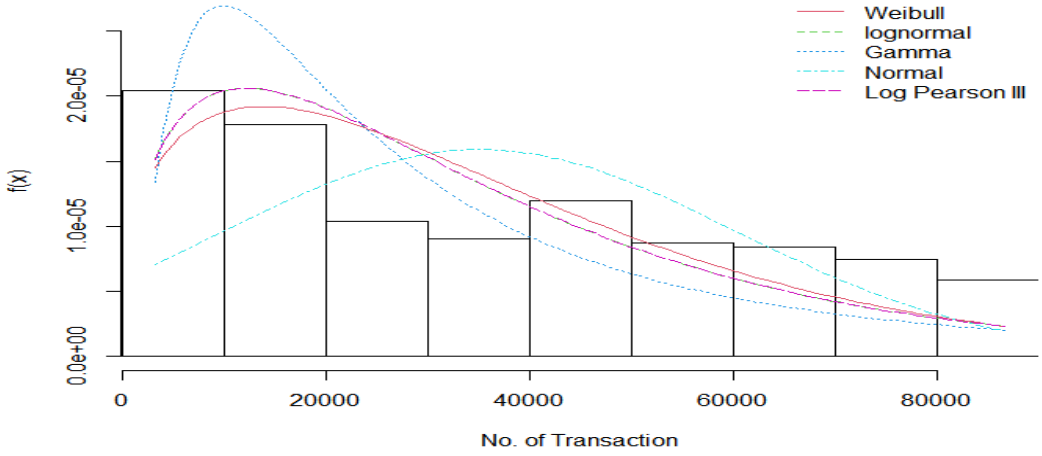

Fig (2):Several Distributions fitting to the total no. of transactions, Saudi Arabia.

Kolmogorov-Smirnov test (K-S) is based on the greatest vertical variation between the theoretical and empirical cumulative density functions. (K-S) is defined as:

$$KS = \underset{1\le i\le n}{Max}\left(F(X_i) - \frac{i-1}{n}, \frac{i}{n} - F(X_i) \right) \tag{8}$$

Anderson-Darling test (A-D) compares the fits of an observed cumulative distribution function to a specified cumulative distribution function [13], [14].

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln(F(X_i)) + \ln(1 - F(X_{n-i+1}))] \tag{9}$$

Numerical Results:

For Log Pearson type III, we solved the equation (7) to derive the value of the shape parameter. The location parameter was obtained and then the scale parameter was also obtained using the equations (5), (6). The values of estimated parameters are demonstrated in Table (1) through the proposed method PR.

The (K-S) and (A-D) test statistic values of PR are obtained by applying the equations (8),(9) and compared to values of MLE.

The estimated parameters and test statistic values of Log-Pearson type III distribution obtained by PR and MLE are shown in Table 2. The AD test statistic value of PR is less than the AD test statistic value of MLE. In addition, the KS value of KS value of PR is less than the value of MLE. The fitting of Log-Pearson type III distribution estimated by PR and MLE to the total number of transactions data set are shown in Fig. 3. The estimations by MLE were obtained via R for Log-Pearson type III distribution as follows:

```
ll <- function(alpha,beta,gamma){-sum(log(fx(x,alpha,beta,gamma)))}

fit_mle <- mle2(minuslog=ll,start=list(alpha=1,beta=1,gamma=1),
data=list(x),method="Nelder-Mead")
```

Table 1: Values of the Log Pearson type III estimated parameters (α), (β), and (γ) using PR.

	No. of Transactions (Total)	F(xi)	α	β	γ
1	3229480	0.003236246	0.00137956	11.11243	4.3736
2	3246509	0.006472492	0.0046091	6.079548	4.3609
3	3266096	0.009708738	0.00887876	4.380293	4.3500
4	3269422	0.012944984	0.013333	3.574501	4.3412
5	3339868	0.01618123	0.0176449	3.107204	4.3341
6	3445906	0.019417476	0.0216474	2.805282	4.3282
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
304	83661928	0.98381877	0.0720205	1.537982	4.2781
305	83802824	0.987055016	0.0720496	1.537672	4.2781
306	84070333	0.990291262	0.0720554	1.53761	4.2781
307	84330498	0.99350508	0.0720624	1.537535	4.2781
308	86412642	0.996763754	0.0717448	1.540934	4.2784
309	86606672	1	0.0717644	1.540724	4.2783
Average			0.071498629	1.613447	4.2792

Table 2: Estimated parameters values of Log-Pearson type III distribution.

Distribution	Parameter	PR	MLE
Log-Pearson type III	A	0.071499	13.041
	B	1.613447	0.2928
	Γ	4.27922	6.287
AD		4.754	9.453
Ks		0.08913	0.13192

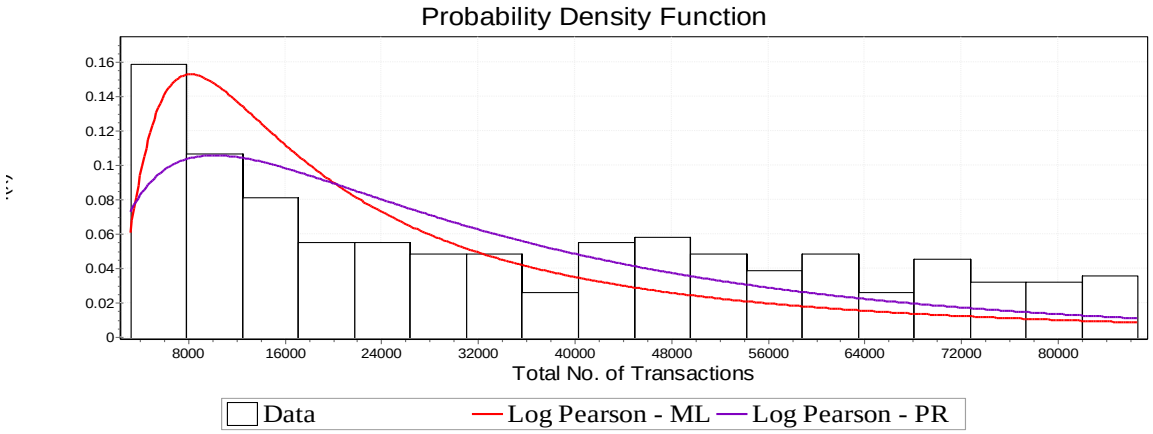

Fig (3): Log- Pearson type III distribution fitting to dataset by PR and MLE.

Fig. 3 shows the fitting of selected data set, total numbers of transactions, for theoretical Log-Pearson type III distribution estimated by PR and MLE. Both estimated parameters give a fitting curve to the frequency histogram of the data set. The fitting curve of MLE is more likely to cover the extreme values and outlays of the data set. Therefore, it has not provided the best-fitting curve. The fitting curve of PR provides an average coverage of the data, giving a better fitting. Since PR relies more on the descriptive characteristics of the data set, moreover the check-steps to ensure the accuracy of the results.

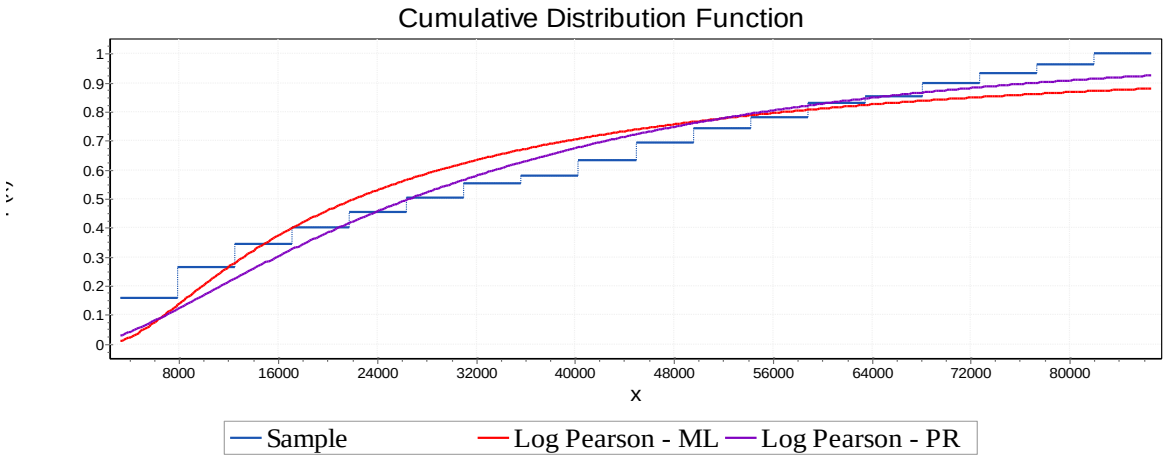

Fig (4): **Cumulative distribution function of Log-Pearson III PR and MLE.**

The difference between cumulative density function (CDF) of the estimated Log-Pearson type III distribution by PR and MLE in Fig. 4, which it shows that the difference in CDF fitting curves is slightly. Since MLE cover the extreme values and outlays of the data set, its curve takes a larger scale.

4. Discussion

Four different methods for estimating the parameters of the log Pearson type-3 distribution were provided [15]. They are given appropriate computational schemes, in addition to the methods of

maximum likelihood and moments. According to simulation results, the method based on the first two moments of the original data set and the variance of the log-transformed yields the best estimation. Whereby they are relatively better than those derived using the maximum likelihood method. Ashkar and Bobée compared the models in fitting the data set using Monte Carlo simulation [16].

Koutrouvelis and Canavos provided a comparison among the various methods of moments for estimating Log Pearson type III [17]. The maximum likelihood estimation, the variances and co-variances of a log Pearson type III distribution were derived using the logarithmic likelihood function and the inverse dispersion matrix, respectively. The parameters were estimated using the moment method, the Kolmogorov-Smirnov test, and the generalized extreme value distribution's properties, and their relevance with the used data set was verified [18].

Compared to the maximum likelihood estimation method, the PRE algorithm is more accurate and simpler to code. Through the chosen relations among parameters. The equation (7) is obtained and solved through function "FindRoot" in Mathematica program. Thus, the estimated value is substituted in the equations (5) ,(6) to obtain other parameters. For each value of x, these steps are repeated. The check-step is to ensure that the estimated values hold the nature of CDF. The MLE estimated parameter values are obtained through R Code. The (K-S) and (A-D) statistic values are obtained also to ensure the efficient of the estimated values and to compare the PR and MLE methods.

5. CONCLUSION

Percentile Root (PR) estimation method provides an accurate results of the statistics of ATM transactions of banks. The point is highly beneficial to applications of economic and commerce sciences. The Log-Pearson type III distribution was used with PR method and fitted to the data set of the total number of ATM transactions. It also helps to uncover the critical areas of banks. At the present time, researchers can work out with PR method and Log-Pearson type III distribution in other applications. Moreover, the results enable us to better understand the crises of ATM transactions in the banks.

6. REFERENCES

[1] V. P. Singh, *Entropy-Based Parameter Estimation in Hydrology*, vol. 30. Dordrecht: Springer Netherlands, 1998.

[2] V. P. Singh, "Log-Pearson Type III Distribution," Springer, Dordrecht, 1998, pp. 252–274.

[3] Coban, Kaan Hakan, and N. Sayil, "Evaluation of earthquake recurrences with different distribution models in western Anatolia," *J. Seismol.*, vol. 23, no. 6, pp. 1405–1422, Nov. 2019.

[4] V. W. Griffis and J. R. Stedinger, "Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. II: Parameter Estimation Methods," *J. Hydrol. Eng.*, vol. 12, no. 5, pp. 492–500, Sep. 2007.

[5] P. Langat, L. Kumar, and R. Koech, "Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow," *Water*, vol. 11, no. 4, p. 734, Apr. 2019.

[6] MM El Genidy and Doaa Abd El-Shafi, "Three Parameters Estimation of Log-Logistic Distribution Using Algorithm of Percentile Roots," in *The 54th Annual Conference on Statistics, Computer Science and Operation Research 9-11 Dec, 2019*, 2020, pp. 200–210.

[7] MM El Genidy and Doaa Abd El-Shafi, "A New High Accurate Estimation Method for Evaluating the Daily Solar Energy by Nested Percentiles Algorithm," *Asian J. Sci. Res.*, vol. 12, no. 4, pp. 480–487, Aug. 2019.

[8] S. Klugman, H. Panjer, and G. Willmot, *Loss models: from data to decisions*, vol. 715. John Wiley & Sons, 2012.

[9] E. genidy M.M, W. M, and M. K, "Algorithms of Solar Energy Prediction Combined with Percentile Root Estimation of Three-Parameters Distributions," *Appl. Math. Inf. Sci.*, vol. 16, no. 4,

pp. 529–547, Jul. 2022.

[10] S. A. Tegos, G. K. Karagiannidis, P. D. Diamantoulakis, and N. D. Chatzidiamantis, "New results for Pearson type III family of distributions and application in wireless power transfer," *IEEE Internet Things J.*, vol. 9, no. 23, pp. 24038–24050, 2020.

[11] N. Millington, S. Das, and S. Simonovic, "The comparison of GEV, log-Pearson type 3 and Gumbel distributions in the Upper Thames River watershed under global climate models," 2011.

[12] KAPSARC, "Automated Teller Machines Statistics — KAPSARC Data Portal." Saudi Central Bank (SAMA), Saudi Arabia.

[13] E. L. Lehmann and J. P. Romano, " Testing Goodness of Fit," in *Testing Statistical Hypotheses*, Springer, Cham: Springer International Publishing, 2022, pp. 773–829.

[14] Yadolah Dodge, "Anderson–Darling Test," in *The Concise Encyclopedia of Statistics*, Springer, New York, NY, 2008, pp. 12–14. doi: 10.1007/978-0-387-32833-1_11.

[15] H. N. Phine and M. A. Hira, "Log Pearson type-3 distribution: Parameter estimation," *J. Hydrol.*, vol. 64, no. 1–4, pp. 25–37, Jul. 1983, doi: 10.1016/0022-1694(83)90058-6.

[16] F. Ashkar and B. Bobée, "The generalized method of moments as applied to problems of flood frequency analysis: Some practical results for the log-Pearson type 3 distribution," *J. Hydrol.*, vol. 90, no. 3–4, pp. 199–217, Apr. 1987, doi: 10.1016/0022-1694(87)90067-9.

[17] I. A. Koutrouvelis and G. C. Canavos, "A comparison of moment-based methods of estimation for the log Pearson type 3 distribution," *J. Hydrol.*, vol. 234, no. 1–2, pp. 71–81, Jun. 2000, doi: 10.1016/S0022-1694(00)00241-9.

[18] MM El Genidy, "Statistical modeling of the daily global solar radiation in Queensland, Australia.," *Songklanakarin J. Sci. Technol.*, vol. 41, no. 6, 2019.