

Appropriateness of Reduced Modified Three-Parameter Weibull Distribution Function for Predicting Gold Production in Ghana

Samuel Kwaku Obeng^{1,2*}, Christiana C. Nyarko², Lewis Brew², Kaku Sagary Nokoe³

¹Research and Statistics Unit, Minerals Commission of Ghana, Accra, Ghana

²Department of Mathematics and Statistics, University of Mines and Technology, Tarkwa, Ghana

³Department of Economics, Catholic University of Ghana, Sunyani, Ghana

Email: *Samobeng2002@yahoo.com, *samobeng2002@gmail.com

How to cite this paper: Obeng, S.K., Nyarko, C.C., Brew, L. and Nokoe, K.S. (2023) Appropriateness of Reduced Modified Three-Parameter Weibull Distribution Function for Predicting Gold Production in Ghana. *Open Journal of Statistics*, 13, 534-567.

<https://doi.org/10.4236/ojs.2023.134027>

Received: June 29, 2023

Accepted: August 11, 2023

Published: August 14, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Forecasting mine production is pertinent to gold mining as it serves as production goals for investors. It is therefore important to identify the exact distribution that gold production as a response variable naturally follows. It is even more appropriate to have a model(s) with few predictor variables. This paper seeks to identify appropriate statistical distribution functions for fitting gold production in Ghana. The empirical paper relied mainly on quarterly secondary datasets on gold production between the years 2009 and 2022 secured from the Minerals Commission of Ghana, Accra. Several known statistical distributions including Weibull, Log-Normal, Generalized Extreme Value (GEV) were explored with Maximum Likelihood Estimation (MLE) and evaluated using model selection criteria as AIC, AICc and BIC. Goodness of Fits were evaluated using Kolmogorov-Smirnov Test (K-S), Cramer-Von Mises Statistic and Anderson-Darling Statistic. Based on the analysis conducted, the reduced modified 3-parameter Weibull distribution provided the best fit for gold production in Ghana. Though the reduced modified Weibull function is proposed, it is important however to recognize that other external factors can influence production levels. Also, the average quarterly fitted gold production is $1000334.8918 \pm 75,327.080$ ($\pm 7.5\%$) [*i.e.*, 925,007.812 – 1,075,661.972]. This indicates that the average annually fitted gold production lies between 3700031.248 and 4302647.888 ounces at 99.9% confidence level. Therefore, the predicted gold production for the year 2022 is 3.7million ounces at 99.9% confidence level.

Keywords

Gold Production, Statistical Distribution Functions, Goodness of Fit Statistics

1. Introduction

Forecasting gold production is a critical task with significant economic implications for mining companies, investors, and governments [1]. Accurate predictions of gold production enable stakeholders to make informed decisions regarding investment strategies, resource allocation, and market positioning [2]. An essential aspect of gold production forecasting is identifying an appropriate statistical distribution to characterize the uncertainty and variability inherent in production data [3].

The relevance of selecting an appropriate statistical distribution for gold production forecasting is supported by research and industry practices. Academic studies, such as the research conducted by Panagiotelis *et al.* [4], emphasize the importance of statistical modeling and distribution selection for accurate gold production forecasting. They highlight the need to consider the complex nature of production data and its inherent variability to improve forecasting accuracy [4].

Moreover, industry reports, such as the “World Gold Council’s Gold Demand Trends” publication, emphasize the significance of reliable production forecasts for understanding the global gold supply and demand dynamics [5]. These reports highlight the critical role of statistical modeling techniques, including appropriate distribution selection, in generating accurate and insightful production forecasts [4] [5].

By identifying an appropriate statistical distribution, analysts can better understand the probabilistic nature of gold production, estimate production volumes, assess project feasibility, and develop robust risk management strategies [6]. Furthermore, it aids in optimizing production processes, evaluating financial performance, and facilitating effective communication with stakeholders [7].

One of the major export commodities for Ghana is gold and it remains the core of Ghana’s mining and quarrying activities [8]. Although, indigenous mining commenced in Ghana in the 4th century, the mining industry officially started in the year 1874 and it accounts for 5.5% of the country’s GDP, 14% of total tax revenue and contributes over 90% of the 48.4% minerals receipts as a share of total exports in Ghana in 2020 [9] and makes up 49% of the country’s total export value [10].

According to World Gold Council [11], Ghana is Africa’s largest gold producer, 6th largest in the world (as of December 2021) and produces more ounces of gold per square kilometer than Nevada. Ghana’s overall gold concession as at first quarter of 2022 is 2,331.729 sqkm with a total gold reserve amounting to 204,725,804.24 Metric Tons [12].

Additionally, Ghana’s gold production stood at 150 metric tons in 2020 irrespective of the COVID-19 pandemic [13] [14].

However, methods used by the mining companies as well as the environmental or weather conditions, agitations from local communities for access to mining sites and other unregulated activities have had negative consequence on gold production [15] [16].

Gold marketing has also been poorly done, especially among the small-scale local miners [17]. Due to these and many other factors, gold production and its market value have been fluctuating over the years [18]. However, a steady and progressive goal production encourages investors to pump more resources into the sector [19].

Studies, such as Matroushi [20], Yazdani-Chamzini *et al.* [21], Kaba *et al.* [22], Appiah *et al.* [23], Hafezi and Akhavan [24], and Chai *et al.* [25] modelled production or price of gold with various statistical functions such as Beta, Chi-square, Erlang, Exponential, Fisher-Tippett, Gamma, Gumbel, Log-normal, Logistic, Normal, Student t, Neural Network and Bayesian structural time series model. However, none of these findings explored most of the statistical functions, especially, those in the extreme value distribution families such as Generalized Extreme Value (GEV) and Weibull two and three parameter distributions to ascertain or identify the actual or natural distributions that production of gold follow. Meanwhile, it is always best to model data around its own natural distribution that it follows to get a parsimonious model [26] [27]. That is, identifying the exact distribution that a response or dependent variable naturally follows assists the model to accomplish the desired level of explanation or prediction with fewer predictor variables thereby making the model to reach its highest parsimony [28]. Also, since these studies did not explore the available extreme family distributions to identify the real distributions they followed before the modelling, they could not attain better precision [29]. Several fields, including structural engineering, finance, earth sciences, traffic prediction, and geological engineering, frequently employ extreme value analysis [30] [31] [32]. Likewise, the production of gold falls within the jurisdictions of geological engineering, earth sciences and finance respectively [33].

Extreme Value Distributions, such as the Weibull distribution, are commonly used for modeling production data due to their relevance in capturing extreme events and tail behavior. These distributions provide a robust framework for analyzing rare, high-impact events that can significantly impact gold production [34].

The Weibull distribution is particularly well-suited for modeling production data because it offers flexibility in capturing a wide range of shapes, including skewed and heavy-tailed distributions. This is crucial as production data often exhibit non-normal characteristics, with a propensity for occasional large deviations from the mean [35].

One of the primary reasons for the adoption of Weibull and other extreme value distributions is their ability to model the occurrence of extreme events, such as production spikes or declines. These distributions have tail properties that allow for the estimation of extreme quantiles, enabling analysts to assess the likelihood of rare events and plan for potential risks associated with them [36].

The relevance of using extreme value distributions like Weibull for modeling production data is supported by academic research and industry practices. Nu-

merous studies have demonstrated the effectiveness of Weibull and other extreme value distributions in modeling and forecasting production data in various fields, including mining and resource extraction [37].

For example, in the paper “Application of the Weibull Distribution Function for Characterizing and Modeling Gold Grades in the Spent Ore Stockpile at Bogoso Gold Limited, Ghana” by Asante *et al.* [38], the authors utilized the Weibull distribution to model gold grades in a gold mining operation. The study demonstrated that the Weibull distribution provided a good fit to the data and enabled the estimation of extreme quantiles for planning purposes.

Another relevant reference is the book “Statistical Methods for Forecasting” by Bovas Abraham and Johannes Ledolter [39], which discusses the application of extreme value distributions, including Weibull, in forecasting time series data. The authors highlight the significance of extreme value distributions in capturing tail behavior and managing risks associated with rare events.

Forecasting mine production is pertinent to gold mining since it serves as production goals for investors [40]. Hence the introduction of models and techniques that predict gold production with the focus of incorporating the impact of uncertainties by means of quantitative stochastic methods is necessary. This led the authors to consider and modify the Weibull distribution function and justify the same by comparing the results with the models reviewed above and used elsewhere.

2. Materials and Methods

The study mainly employed quarterly secondary datasets on gold production between the years 2009 and 2022 secured from the Minerals Commission of Ghana, Accra. Distributions such as Weibull, Log-Normal, Generalized Extreme Value (GEV) were considered. Parameter Estimation used was Maximum Likelihood Estimation (MLE). The Model/Distribution Selection Criteria used were AIC, AICc, BIC. The Goodness of Fit tests considered for this study are Kolmogorov-Smirnov Test (K-S), Cramer-Von Mises Statistic and Anderson-Darling Statistic.

Goodness-of-fit statistics, such as the Kolmogorov-Smirnov (KS) test, Cramer-Von Mises (CVM) statistic, and Anderson-Darling (AD) statistic, are commonly used to assess how well a statistical distribution fits a given set of data. These statistics provide quantitative measures to evaluate the agreement between the observed data and the expected distribution. Each test has its own characteristics, interpretations, strengths, and limitations.

1) Kolmogorov-Smirnov Test: The KS test compares the cumulative distribution function (CDF) of the observed data with the CDF of the expected distribution. It calculates the maximum vertical distance (D) between the two functions, representing the test statistic. The KS test assesses whether the observed data follows a specific distribution or if it significantly deviates from it. The test produces a p-value, which indicates the probability of obtaining a discrepancy as large as or

larger than the observed one if the data truly follows the expected distribution.

Strengths:

- Simple and widely used goodness-of-fit test.
- Applicable to a wide range of distributions.
- Nonparametric and distribution-free.

Limitations:

- Sensitive to discrepancies in the tails of the distribution.
- Less powerful for small sample sizes.

2) Cramer-Von Mises Statistic: The CVM statistic measures the integral of the squared difference between the observed cumulative distribution function and the expected distribution's cumulative distribution function. It quantifies the overall discrepancy between the observed data and the expected distribution.

Strengths:

- Similar to the KS test but gives more weight to the tails of the distribution.
- Suitable for comparing distributions with different shapes.

Limitations:

- May not work well with small sample sizes.
- Requires cumulative distribution function estimation.

3) Anderson-Darling Statistic: The AD statistic, similar to the CVM statistic, assesses the integral of the squared difference between the observed cumulative distribution function and the expected distribution's cumulative distribution function. However, the AD test places greater emphasis on the tails of the distribution, making it more sensitive to discrepancies in those regions.

Strengths:

- Particularly useful for assessing goodness-of-fit in the tails of the distribution.
- Applicable to a wide range of distributions.

Limitations:

- Can be sensitive to estimation errors.
- Sample size dependency, with larger sample sizes leading to higher power.

The choice of these goodness-of-fit statistics depends on the specific requirements and characteristics of the data. The KS test is commonly used as a general-purpose test, while the CVM and AD statistics are preferred when there is a particular interest in tail behavior. It is often recommended to employ multiple goodness-of-fit tests to gain a comprehensive understanding of how well the expected distribution fits the data.

In a nutshell, goodness-of-fit statistics such as the KS test, CVM statistic, and AD statistic provide quantitative measures to assess the agreement between observed data and expected distributions. While they have their respective strengths and limitations, they play a valuable role in evaluating the appropriateness of a chosen statistical distribution for modeling purposes.

The study covered gold production between 2009 and 2021 gold production. Several distribution functions were evaluated as initial step towards identifying the likely candidate. The initial distribution fitting for gold production using the

XLSTAT software revealed that the three-parameter Weibull distribution perfectly fit the gold production data with Kolmogorov-Smirnoff test p-value of 0.9438. The closest distribution was Beta four-parameter (K-S: 0.9420). However, in view of the versatility of the Weibull three-parameter distribution was selected and modified for the study (**Table 1**).

3. Model Formulation and Modification

3.1. Concepts of the Extreme Value Distributions

The limiting distributions for the lowest or maximum of sizable collections of independent random variables drawn from the same arbitrary distribution are called extreme value distributions. The topic of extreme value theory is, by definition, restricting distributions (which are distinct from the normal distribution). Statistics' severe deviations from the median of probability distributions are the focus of extreme value theory (EVT) or extreme value analysis (EVA) [41]. With an ordered sample of a particular random variable, it aims to determine the likelihood of events that are more extreme than anything previously recorded [41]. Several fields, including structural engineering, finance, earth sciences, traffic prediction, and geological engineering, frequently employ extreme value analysis [30] [31] [32]. It should therefore be noted that production and price of gold fall within the jurisdictions of geological engineering, earth sciences and finance respectively [33]. Therefore, modifying the distributions

Table 1. Initial distribution fitting for gold production using XLSTAT software.

Distribution	K-S-(p-value)
Beta4	0.9420
Chi-square	<0.0001
Erlang	<0.0001
Exponential	<0.0001
Fisher-Tippett (1)	<0.0001
Fisher-Tippett (2)	0.0005
Gamma (1)	<0.0001
Gamma (2)	0.7297
Gumbel	<0.0001
Log-normal	0.6700
Normal	0.9253
Normal (Standard)	<0.0001
Student	<0.0001
Weibull (1)	<0.0001
Weibull (3)	0.9438

Source: Authors own estimation, 2023

that fall within the extreme value families to fit production and price of gold is laudable.

For instance, the Leadbetter *et al.* [42] [43] extreme value theory queried that; given X_1, \dots, X_k as a set of independent identically distributed random variables, what possible limiting distributions will $M_k = a_k [\max(X_1, \dots, X_k) - b_k]$ follow as $k \rightarrow \infty$? That is,

$$F^k \left(\frac{x - b_k}{a_k} \right) \xrightarrow{k \rightarrow \infty} G(x)$$

However, to respond to this question, if a nondegenerate limiting cumulative distribution exists for some sequences of constants a_k and b_k , it should fall within the following three categories.

- 1) $F(x) = \exp[-e^{-x}]$, $-\infty < x < \infty$;
- 2) $F(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0 \end{cases}$;
- 3) $F(x) = \begin{cases} \exp[-(-x)^\alpha], & x < 0, \alpha > 0 \\ 1, & x \geq 0 \end{cases}$.

These three categories represent the Gumbel, Fréchet and Weibull distributions respectively. However, in a more modernized approach, these distributions have been combined to form the Generalized Extreme Value distribution with a cumulative density function as

$$F(y) = \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{\frac{-1}{\xi}} \right\}, -\infty < \mu, \xi < \infty, \sigma > 0$$

Defined for values of y for which $1 + \xi \left(\frac{y - \mu}{\sigma} \right) > 0$, where μ , ξ and σ are the location, shape and scale parameters respectively. Moreover, the shape parameter ξ controls the three categories of the distributions. That is, when $\xi = 0$, we have the first type, which is the Gumbel light tailed distribution, while when $\xi > 0$, we have the Fréchet heavy tailed distribution and finally, when $\xi < 0$, we have the Weibull bounded distribution. This is shown graphically in **Figure 1**. Moreover, it should be noted that the location parameter μ is not the actual mean, it only represents the center of the distribution. Similarly, the scale parameter σ is not the standard deviation, it just governs the size of the deviations about the location parameter μ (**Figure 1**).

3.2. The 3-Parameter Distribution and Its Modification and Parameter Derivations of Gold Production

The Weibull 3-parameter distribution is an extension of the 2-parameter distribution where a third parameter known as location or threshold is used when the data points do not fall on the straight line but on a concave up or down curve. The probability density function (PDF) of the 3-parameter Weibull distribution

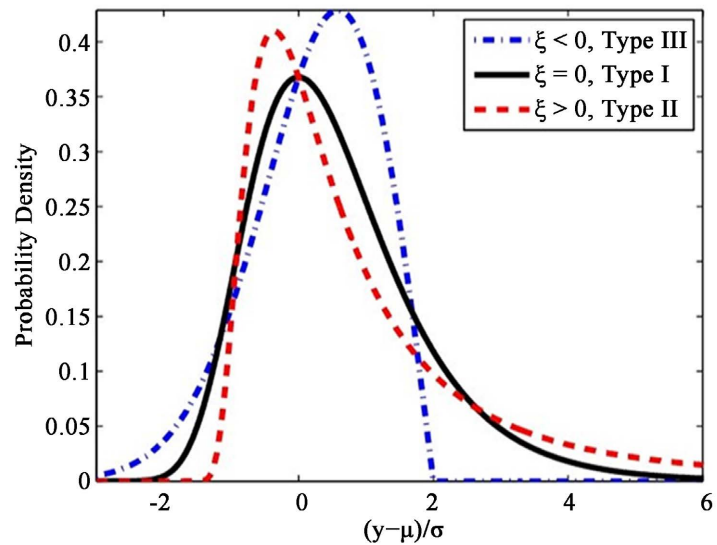


Figure 1. The three families of the generalized extreme value distribution.

is given by;

Weibull (3);

$$f(x; k, \lambda, Q) = \frac{k}{\lambda} \left(\frac{x-Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x-Q}{\lambda}\right)^k} \quad \text{for } x > Q, k > 0 \text{ and } \lambda > 0$$

where k is the shape parameter, λ is the scale parameter and Q is threshold parameter.

For the Weibull 3-parameter distribution to be legitimate, then;

$$\int_Q^\infty f(x; k, \lambda, Q) dx = 1$$

Therefore,

$$\int_Q^\infty f(x; k, \lambda, Q) dx = \int_Q^\infty \frac{k}{\lambda} \left(\frac{x-Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x-Q}{\lambda}\right)^k} dx \tag{1}$$

let $t = \left(\frac{x-Q}{\lambda}\right)^k$ which implies, $x-Q = \lambda t^{\frac{1}{k}}$ where $x = \lambda t^{\frac{1}{k}} + Q$ and

$$dx = \frac{1}{k} \lambda t^{\frac{1}{k}-1} dt.$$

Also, when $x = Q$, $t = 0$ and when $x = \infty$, $t = \infty$.

Substituting the above into equation 1, we have;

$$\begin{aligned} \int_Q^\infty f(x; k, \lambda, Q) dx &= \int_0^\infty \frac{k}{\lambda} \left(\frac{\lambda t^{\frac{1}{k}}}{\lambda} \right)^{k-1} e^{-t} \left(\frac{1}{k} \lambda t^{\frac{1}{k}-1} \right) dt \\ \int_Q^\infty f(x; k, \lambda, Q) dx &= \int_0^\infty e^{-t} dt \\ \int_Q^\infty f(x; k, \lambda, Q) dx &= \left[-e^{-t} \right]_0^\infty \\ \int_Q^\infty f(x; k, \lambda, Q) dx &= \left[-e^{-\infty} \right] - \left[-e^0 \right] \end{aligned} \tag{2}$$

$$\int_Q^\infty f(x; k, \lambda, Q) dx = [0] - [-1]$$

$$\therefore \int_Q^\infty f(x; k, \lambda, Q) dx = 1$$

Hence the proof. This shows that the Weibull 3-parameter distribution is legitimate.

Also, for the Cumulative Density Function (CDF), we have;

$$\int_{-\infty}^0 f(x; k, \lambda, Q) dx + \int_Q^x f(x; k, \lambda, Q) dx + \int_x^\infty f(x; k, \lambda, Q) dx = \int_0^t e^{-t} dt$$

Since $\int_{-\infty}^0 f(x; k, \lambda, Q) dx = \int_x^\infty f(x; k, \lambda, Q) dx = 0$ and

$$\int_Q^x f(x; k, \lambda, Q) dx = \int_0^t e^{-t} dt$$

$$\therefore \text{CDF} = \int_0^t e^{-t} dt$$

$$\text{CDF} = [-e^{-t}] - [-e^0]$$

$$\text{CDF} = 1 - e^{-t}$$

This implies

$$\text{CDF} = 1 - e^{-\left(\frac{x-Q}{\lambda}\right)^k} \quad (3)$$

Now, for the mean or expectation of the distribution, we have;

$$E(X) = \int_Q^\infty xf(x; k, \lambda, Q) dx = \int_Q^\infty x \frac{k}{\lambda} \left(\frac{x-Q}{\lambda}\right)^{k-1} e^{-\left(\frac{x-Q}{\lambda}\right)^k} dx \quad (4)$$

Employing the technique used for equation 2 where $x = \lambda t^{\frac{1}{k}} + Q$, we have;

$$E(X) = \int_0^\infty \left(\lambda t^{\frac{1}{k}} + Q\right) e^{-t} dt$$

$$E(X) = \int_0^\infty (Q) e^{-t} dt + \int_0^\infty \left(\lambda t^{\frac{1}{k}}\right) e^{-t} dt \quad (5)$$

But $\int_0^\infty e^{-t} dt = 1$

$$\therefore E(X) = Q + \int_0^\infty \left(\lambda t^{\frac{1}{k}}\right) e^{-t} dt$$

Also, let $\alpha - 1 = \frac{1}{k}$

$$\therefore E(X) = Q + \int_0^\infty (\lambda t^{\alpha-1}) e^{-t} dt$$

But $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$

$$\therefore E(X) = Q + \lambda \Gamma(\alpha)$$

$$\therefore E(X) = Q + \lambda \Gamma\left(\frac{1}{k} + 1\right) \quad (6)$$

As well, for the variance of the 3-parameter Weibull distribution, we have;

$$\text{VAR}(X) = E(X^2) - [E(X)]^2$$

$$\therefore E(X^2) = \int_Q^\infty x^2 f(x; k, \lambda, Q) dx = \int_Q^\infty x^2 \frac{k}{\lambda} \left(\frac{x-Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x-Q}{\lambda}\right)^k} dx \quad (7)$$

Employing the technique used for equation 4.2 where $x = \lambda t^{\frac{1}{k}} + Q$, we have;

$$E(X^2) = \int_0^\infty \left(\lambda t^{\frac{1}{k}} + Q \right)^2 e^{-t} dt$$

$$E(X^2) = \int_0^\infty \left[\left(\lambda t^{\frac{1}{k}} \right)^2 + Q^2 + 2Q\lambda t^{\frac{1}{k}} \right] e^{-t} dt \quad (8)$$

$$E(X^2) = \int_0^\infty [Q^2] e^{-t} dt + \int_0^\infty [2Q\lambda t^{\frac{1}{k}}] e^{-t} dt + \int_0^\infty \left[\left(\lambda t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt \quad (9)$$

Based on the techniques used for Equations (5) and (6),

$$\int_0^\infty [Q^2] e^{-t} dt + \int_0^\infty [2Q\lambda t^{\frac{1}{k}}] e^{-t} dt = Q^2 + 2Q\lambda \Gamma\left(\frac{1}{k} + 1\right)$$

$$\int_0^\infty \left[\left(\lambda t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = \lambda^2 \int_0^\infty t^{\frac{2}{k}} e^{-t} dt$$

Now, let $\alpha - 1 = \frac{2}{k}$

$$\therefore \int_0^\infty \left[\left(\lambda t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = \lambda^2 \int_0^\infty t^{\alpha-1} e^{-t} dt$$

$$\therefore \int_0^\infty \left[\left(\lambda t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = \lambda^2 \Gamma(\alpha) = \lambda^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$\therefore E(X^2) = Q^2 + 2Q\lambda \Gamma\left(\frac{1}{k} + 1\right) + \lambda^2 \Gamma\left(\frac{2}{k} + 1\right) \quad (10)$$

$$\therefore \text{VAR}(X) = Q^2 + 2Q\lambda \Gamma\left(\frac{1}{k} + 1\right) + \lambda^2 \Gamma\left(\frac{2}{k} + 1\right) - \left[Q + \lambda \Gamma\left(\frac{1}{k} + 1\right) \right]^2$$

$$\therefore \text{VAR}(X) = Q^2 + 2Q\lambda \Gamma\left(\frac{1}{k} + 1\right) + \lambda^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$- \left[Q^2 + \lambda^2 \Gamma^2\left(\frac{1}{k} + 1\right) + 2Q\lambda \Gamma\left(\frac{1}{k} + 1\right) \right]$$

$$\therefore \text{VAR}(X) = \lambda^2 \left[\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right) \right] \quad (11)$$

Moreover, the Moment Generation Function (MGF) of the 3-parameter Weibull distribution is given as;

$$E(X^n) = \int_Q^\infty x^n f(x; k, \lambda, Q) dx = \int_Q^\infty x^n \frac{k}{\lambda} \left(\frac{x-Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x-Q}{\lambda}\right)^k} dx \quad (12)$$

Employing the technique used for equation 2 where $x = \lambda t^{\frac{1}{k}} + Q$, we have;

$$E(X^n) = \int_0^\infty \left(\lambda t^{\frac{1}{k}} + Q \right)^n e^{-t} dt \quad (13)$$

$$\text{but } (a+b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r}, \quad r = 0, 1, 2, \dots, n \quad (14)$$

$$\therefore E(X^n) = \int_0^\infty \sum_{r=0}^n \binom{n}{r} \left(\lambda t^{\frac{1}{k}} \right)^r Q^{n-r} e^{-t} dt \quad (15)$$

This implies;

$$E(X^n) = \int_0^\infty \left[Q^n + n\lambda t^{\frac{1}{k}} Q^{n-1} + \frac{n(n-1)\lambda^2 t^{\frac{2}{k}} Q^{n-2}}{2!} + \frac{n(n-1)(n-2)\lambda^3 t^{\frac{3}{k}} Q^{n-3}}{3!} + \dots + \lambda^n t^{\frac{n}{k}} \right] e^{-t} dt \quad (16)$$

Integrating equation 15 with reference to equations 8, 9 and 10, we have;

$$E(X^n) = Q^n + n\lambda Q^{n-1} \Gamma\left(\frac{1}{k} + 1\right) + \frac{n(n-1)\lambda^2 Q^{n-2} \Gamma\left(\frac{2}{k} + 1\right)}{2!} + \frac{n(n-1)(n-2)\lambda^3 Q^{n-3} \Gamma\left(\frac{3}{k} + 1\right)}{3!} + \dots + \lambda^n \Gamma\left(\frac{n}{k} + 1\right) \quad (17)$$

With reference to Equation (15), Equation (17) or the MGF becomes;

$$E(X^n) = \sum_{r=0}^n \binom{n}{r} \lambda^r Q^{n-r} \Gamma\left(\frac{r}{k} + 1\right), \quad r = 0, 1, 2, \dots, n$$

The Maximum Likelihood Estimation (MLE) of the 3-parameter distribution is also given as;

$$L(x_i; k, \lambda, Q) = \prod_{i=1}^n f(x_i; k, \lambda, Q) \quad (18)$$

$$L(x_i; k, \lambda, Q) = \prod_{i=1}^n \frac{k}{\lambda} \left(\frac{x_i - Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x_i - Q}{\lambda}\right)^k} \quad (19)$$

Taking antilog of both sides, we have;

$$\ln L(x_i; k, \lambda, Q) = \sum_{i=1}^n \ln \left[\frac{k}{\lambda} \left(\frac{x_i - Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x_i - Q}{\lambda}\right)^k} \right]$$

$$\ln L(x_i; k, \lambda, Q) = \sum_{i=1}^n \left\{ \ln k - \ln \lambda + (k-1) \left[\ln(x_i - Q) - \ln \lambda \right] - \left(\frac{x_i - Q}{\lambda} \right)^k \right\}$$

Therefore, the log likelihood function is given as;

$$\ln L(x_i; k, \lambda, Q) = \sum_{i=1}^n \left\{ \ln k - k \ln \lambda + (k-1) [\ln(x_i - Q)] - \left(\frac{x_i - Q}{\lambda} \right)^k \right\} \quad (20)$$

Differentiating Equation (20) with respect to k, λ, Q as in Equations (21)-(23) respectively,

$$\frac{d \ln L(x_i; k, \lambda, Q)}{dk} = \sum_{i=1}^n \left\{ \frac{1}{k} - \ln \lambda + \ln(x_i - Q) - \left(\frac{x_i - Q}{\lambda} \right)^k \ln \left(\frac{x_i - Q}{\lambda} \right) \right\} \quad (21)$$

$$\frac{d \ln L(x_i; k, \lambda, Q)}{d\lambda} = \sum_{i=1}^n \left\{ \frac{k(x_i - Q) \left(\frac{x_i - Q}{\lambda} \right)^{k-1}}{\lambda^2} \right\} \quad (22)$$

$$\frac{d \ln L(x_i; k, \lambda, Q)}{dQ} = \sum_{i=1}^n \left\{ -\frac{(k-1)}{x_i - Q} + \frac{k}{\lambda} \left(\frac{x_i - Q}{\lambda} \right)^{k-1} \right\} \quad (23)$$

Therefore, equating Equations (21)-(23) to zero and solving them simultaneously using numerical methods, the Maximum Likelihood Estimates of k, λ, Q are produced.

Also, the Cramer-Rao lower bound inequality attained for each of the estimated parameters for k, λ, Q are as follows in Equations (24)-(26) respectively.

$$\text{Variance}(\hat{k}(x)) \geq \frac{1}{nI(k)} \quad \text{or} \quad \frac{1}{I(k)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{dk} \right]^2 \right]} \quad (24)$$

$$\text{Variance}(\hat{\lambda}(x)) \geq \frac{1}{nI(\lambda)} \quad \text{or} \quad \frac{1}{I(\lambda)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{d\lambda} \right]^2 \right]} \quad (25)$$

$$\text{Variance}(\hat{Q}(x)) \geq \frac{1}{nI(Q)} \quad \text{or} \quad \frac{1}{I(Q)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{dQ} \right]^2 \right]} \quad (26)$$

3.3. Modification of the Weibull 3-Parameter Distribution

This section modified the 3-parameter Weibull distribution by deducting the threshold parameter also from the scale parameter. This produced a better precision than the former. This was arrived at through a simulation.

Given the Weibull (3); $f(x; k, \lambda, Q) = \frac{k}{\lambda} \left(\frac{x - Q}{\lambda} \right)^{k-1} e^{-\left(\frac{x - Q}{\lambda} \right)^k}$ for $x > Q, k > 0$

and $\lambda > 0$ let $\frac{x - Q}{\lambda} = \frac{x - Q}{\lambda - Q}$ and also let $\frac{k}{\lambda} = \frac{k}{\lambda - Q}$.

This implies that;

$$f(x; k, \lambda, Q) = \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x - Q}{\lambda - Q} \right)^k} \quad \text{for } x > Q, k > 0 \text{ and } \lambda > 0 \quad (27)$$

where k is the shape parameter, λ is the scale parameter and Q is the threshold parameter.

3.3.1. Legitimacy of the Modified 3-Parameter Weibull Distribution

For the modified 3-parameter Weibull distribution (M3PWD) to be legitimate, then;

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = 1$$

Therefore,

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \int_0^{\infty} \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x - Q}{\lambda - Q} \right)^k} dx \quad (28)$$

Let $t = \left(\frac{x - Q}{\lambda - Q} \right)^k$ which implies, $x = (\lambda - Q)t^{\frac{1}{k}} + Q$ where

$$dx = \frac{1}{k} (\lambda - Q) t^{\frac{1}{k}-1} dt.$$

Also, when $x = 0$, $t = 0$ and when $x = \infty$, $t = \infty$.

Substituting the above into Equation (28), we have;

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \int_0^{\infty} \frac{k}{\lambda - Q} \left(t^{\frac{1}{k}-1} \right) e^{-t} \left[\frac{1}{k} (\lambda - Q) t^{\frac{1}{k}-1} \right] dt$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \int_0^{\infty} \frac{k}{\lambda - Q} \frac{1}{k} (\lambda - Q) t^{\frac{1}{k}-1} t^{\frac{1}{k}-1} e^{-t} dt$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \int_0^{\infty} t^0 e^{-t} dt$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \int_0^{\infty} e^{-t} dt \quad (29)$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \left[-e^{-t} \right]_0^{\infty}$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = \left[-e^{-\infty} \right] - \left[-e^0 \right]$$

$$\int_0^{\infty} f(x; k, \lambda, Q) dx = [0] - [-1]$$

$$\therefore \int_0^{\infty} f(x; k, \lambda, Q) dx = 1$$

Hence the proof. This shows that the modified 3-parameter Weibull distribution is legitimate.

3.3.2. Cumulative Density Function (CDF) of the Modified 3-Parameter Weibull Distribution

For the Cumulative Density Function (CDF) with reference to Equations (28) and (29), we have;

$$\int_{-\infty}^0 f(x; k, \lambda, Q) dx + \int_0^x f(x; k, \lambda, Q) dx + \int_x^{\infty} f(x; k, \lambda, Q) dx = \int_0^t e^{-t} dt$$

Since $\int_{-\infty}^0 f(x; k, \lambda, Q) dx = \int_x^{\infty} f(x; k, \lambda, Q) dx = 0$ and

$$\int_0^x f(x; k, \lambda, Q) dx = \int_0^t e^{-t} dt$$

$$\begin{aligned}\therefore \text{CDF} &= \int_0^t e^{-t} dt \\ \text{CDF} &= [-e^{-t}] - [-e^0] \\ \text{CDF} &= 1 - e^{-t}\end{aligned}$$

This implies

$$\text{CDF} = 1 - e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k} \quad (30)$$

Mean and Variance of the Modified 3-parameter Weibull distribution

For the mean or expectation of the modified Weibull 3-parameter distribution, we have;

$$E(X) = \int_0^\infty xf(x; k, \lambda, Q) dx = \int_0^\infty x \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q}\right)^{k-1} e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k} dx \quad (31)$$

Employing the techniques used for Equations (28) and (29) where

$x = (\lambda - Q)t^{\frac{1}{k}} + Q$, we have;

$$\begin{aligned}E(X) &= \int_0^\infty \left((\lambda - Q)t^{\frac{1}{k}} + Q \right) e^{-t} dt \\ E(X) &= (\lambda - Q) \int_0^\infty t^{\frac{1}{k}} e^{-t} dt + Q \int_0^\infty e^{-t} dt\end{aligned} \quad (32)$$

Let $\alpha - 1 = \frac{1}{k}$, which implies; $\alpha = \frac{1}{k} + 1$

$$\therefore E(X) = (\lambda - Q) \int_0^\infty (t^{\alpha-1}) e^{-t} dt + Q$$

But $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$

$$\begin{aligned}\therefore E(X) &= (\lambda - Q)\Gamma(\alpha) + Q \\ \therefore E(X) &= (\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) + Q\end{aligned} \quad (33)$$

As well, for the variance of the modified 3-parameter Weibull distribution, we have;

$$\begin{aligned}\text{VAR}(X) &= E(X^2) - [E(X)]^2 \\ \therefore E(X^2) &= \int_0^\infty x^2 f(x; k, \lambda, Q) dx = \int_0^\infty x^2 \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q}\right)^{k-1} e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k} dx\end{aligned} \quad (34)$$

Employing the techniques used for Equations (28) and 29 where

$x = (\lambda - Q)t^{\frac{1}{k}} + Q$, we have;

$$\begin{aligned}E(X^2) &= \int_0^\infty \left((\lambda - Q)t^{\frac{1}{k}} + Q \right)^2 e^{-t} dt \\ E(X^2) &= \int_0^\infty \left[\left((\lambda - Q)t^{\frac{1}{k}} \right)^2 + Q^2 + 2Q(\lambda - Q)t^{\frac{1}{k}} \right] e^{-t} dt\end{aligned} \quad (35)$$

$$E(X^2) = \int_0^\infty [Q^2] e^{-t} dt + \int_0^\infty \left[2Q(\lambda - Q)t^{\frac{1}{k}} \right] e^{-t} dt + \int_0^\infty \left[\left((\lambda - Q)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt$$

Based on the techniques used for Equations (28) and (29),

$$\int_0^\infty [Q^2] e^{-t} dt + \int_0^\infty \left[2Q(\lambda - Q)t^{\frac{1}{k}} \right] e^{-t} dt = Q^2 + 2Q(\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right)$$

$$\int_0^\infty \left[\left((\lambda - Q)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - Q)^2 \int_0^\infty t^{\frac{2}{k}} e^{-t} dt$$

Now let $\alpha - 1 = \frac{2}{k}$

$$\therefore \int_0^\infty \left[\left((\lambda - Q)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - Q)^2 \int_0^\infty t^{\alpha-1} e^{-t} dt$$

$$\therefore \int_0^\infty \left[\left((\lambda - Q)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - Q)^2 \Gamma(\alpha) = (\lambda - Q)^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$\therefore E(X^2) = Q^2 + 2Q(\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - Q)^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$\therefore VAR(X) = Q^2 + 2Q(\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - Q)^2 \Gamma\left(\frac{2}{k} + 1\right) - \left[Q + (\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) \right]^2$$

$$\therefore VAR(X) = Q^2 + 2Q(\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - Q)^2 \Gamma\left(\frac{2}{k} + 1\right) - \left[Q^2 + (\lambda - Q)^2 \Gamma^2\left(\frac{1}{k} + 1\right) + 2Q(\lambda - Q)\Gamma\left(\frac{1}{k} + 1\right) \right]$$

$$\therefore VAR(X) = (\lambda - Q)^2 \left[\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right) \right] \tag{36}$$

From Equation (36), we can see that the variance of the modified distribution is less than that of the original 3-parameter distribution since the scalar of $(\lambda - Q)^2$ might be less than λ^2 .

Moment Generation Function (MGF) of the Modified 3-parameter Weibull distribution

The Moment Generation Function (MGF) of the Modified 3-parameter Weibull distribution is given as;

$$E(X^n) = \int_0^\infty x^n f(x; k, \lambda, Q) dx = \int_0^\infty x^n \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k} dx \tag{37}$$

Employing the techniques used for Equation (28) and (29) where

$x = (\lambda - Q)t^{\frac{1}{k}} + Q$, we have;

$$E(X^n) = \int_0^{\infty} \left((\lambda - Q)t^{\frac{1}{k}} + Q \right)^n e^{-t} dt \quad (38)$$

$$\text{But } (a+b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r}, \quad r=0,1,2,\dots,n \quad (39)$$

$$\therefore E(X^n) = \int_0^{\infty} \sum_{r=0}^n \binom{n}{r} \left((\lambda - Q)t^{\frac{1}{k}} \right)^r Q^{n-r} e^{-t} dt \quad (40)$$

This implies;

$$E(X^n) = \int_0^{\infty} \left[Q^n + n(\lambda - Q)t^{\frac{1}{k}}Q^{n-1} + \frac{n(n-1)(\lambda - Q)^2 t^{\frac{2}{k}}Q^{n-2}}{2!} + \frac{n(n-1)(n-2)(\lambda - Q)^3 t^{\frac{3}{k}}Q^{n-3}}{3!} + \dots + (\lambda - Q)^n t^{\frac{n}{k}} \right] e^{-t} dt \quad (41)$$

Integrating (41) with reference to Equations (34)-(36), we have;

$$E(X^n) = Q^n + n(\lambda - Q)Q^{n-1}\Gamma\left(\frac{1}{k} + 1\right) + \frac{n(n-1)(\lambda - Q)^2 Q^{n-2}\Gamma\left(\frac{2}{k} + 1\right)}{2!} + \frac{n(n-1)(n-2)(\lambda - Q)^3 Q^{n-3}\Gamma\left(\frac{3}{k} + 1\right)}{3!} + \dots + (\lambda - Q)^n \Gamma\left(\frac{n}{k} + 1\right) \quad (42)$$

With reference to Equations (39)-(41), the MGF becomes;

$$\text{MGF} = E(X^n) = \sum_{r=0}^n \binom{n}{r} (\lambda - Q)^r Q^{n-r} \Gamma\left(\frac{r}{k} + 1\right), \quad r=0,1,2,\dots,n$$

3.4. The Maximum Likelihood Estimation (MLE) of the Modified 3-Parameter Distribution

The Maximum Likelihood Estimation (MLE) of the Modified 3-parameter distribution is also given as;

$$L(x_i; k, \lambda, Q) = \prod_{i=1}^n f(x_i; k, \lambda, Q) \quad (43)$$

$$L(x_i; k, \lambda, Q) = \prod_{i=1}^n \frac{k}{\lambda - Q} \left(\frac{x_i - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x_i - Q}{\lambda - Q}\right)^k} \quad (44)$$

Taking antilog of both sides, we have;

$$\ln L(x_i; k, \lambda, Q) = \sum_{i=1}^n \ln \left[\frac{k}{\lambda - Q} \left(\frac{x_i - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x_i - Q}{\lambda - Q}\right)^k} \right]$$

$$\ln L(x_i; k, \lambda, Q)$$

$$= \sum_{i=1}^n \left\{ \ln k - \ln(\lambda - Q) + (k-1) [\ln(x_i - Q) - \ln(\lambda - Q)] - \left(\frac{x_i - Q}{\lambda - Q} \right)^k \right\}$$

Therefore, the log likelihood function is given as;

$$\ln L(x_i; k, \lambda, Q) = \sum_{i=1}^n \left\{ \ln k - k \ln(\lambda - Q) + (k-1) \ln(x_i - Q) - \left(\frac{x_i - Q}{\lambda - Q} \right)^k \right\} \quad (45)$$

Differentiating Equation (45) with respect to k, λ, Q we obtain the respective Equations (46)-(48);

$$\frac{d \ln L(x_i; k, \lambda, Q)}{dk} = \sum_{i=1}^n \left\{ \frac{1}{k} - \ln(\lambda - Q) + \ln(x_i - Q) - \left(\frac{x_i - Q}{\lambda - Q} \right)^k \ln \left(\frac{x_i - Q}{\lambda - Q} \right) \right\} \quad (46)$$

$$\frac{d \ln L(x_i; k, \lambda, Q)}{d\lambda} = \sum_{i=1}^n \left\{ \frac{k}{\lambda - Q} - \frac{k(x_i - Q) \left(\frac{x_i - Q}{\lambda - Q} \right)^{k-1}}{(\lambda - Q)^2} \right\} \quad (47)$$

$$\frac{d \ln L(x_i; k, \lambda, Q)}{dQ} = \sum_{i=1}^n \left\{ \frac{k-1}{x_i - Q} + \frac{k}{\lambda - Q} \left(\frac{x_i - Q}{\lambda - Q} \right)^{k-1} \right\} \quad (48)$$

Therefore, equating Equations (46)-(48) to zero and solving them simultaneously using numerical methods, the Maximum Likelihood Estimates of k, λ, Q are produced.

3.4.1. Cramer-Rao Lower Bound Inequality for the M3PWD

Also, the Cramer-Rao lower bound inequality attained for each of the estimated parameters for k, λ, Q are as follows in Equations (49)-(51) respectively.

$$\text{Variance}(\hat{k}(x)) \geq \frac{1}{nI(k)} \text{ or } \frac{1}{I(k)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{dk} \right]^2 \right]} \quad (49)$$

$$\text{Variance}(\hat{\lambda}(x)) \geq \frac{1}{nI(\lambda)} \text{ or } \frac{1}{I(\lambda)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{d\lambda} \right]^2 \right]} \quad (50)$$

$$\text{Variance}(\hat{Q}(x)) \geq \frac{1}{nI(Q)} \text{ or } \frac{1}{I(Q)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda, Q)}{dQ} \right]^2 \right]} \quad (51)$$

3.4.2. Reduction of the Modified 3-Parameter Weibull Distribution

This section reduced the modified 3-parameter Weibull distribution by equating the shape parameter (k) to threshold parameter. That is, when the shape is same as the threshold. Where k is the shape parameter, λ is the scale parameter and Q is threshold parameter.

Given the Modified Weibull (3);

$$f(x; k, \lambda, Q) = \frac{k}{\lambda - Q} \left(\frac{x - Q}{\lambda - Q} \right)^{k-1} e^{-\left(\frac{x - Q}{\lambda - Q} \right)^k} \text{ for } x > Q, k > 0 \text{ and } \lambda > 0$$

If $Q = k$, it reduces to

$$f(x; k, \lambda) = \frac{k}{\lambda - k} \left(\frac{x - k}{\lambda - k} \right)^{k-1} e^{-\left(\frac{x-k}{\lambda-k}\right)^k} \quad \text{for } x > 0, k > 0 \text{ and } \lambda > 0 \quad (52)$$

where k is the shape parameter and λ is the scale parameter. In this reduced parameter distribution, there is no threshold parameter.

Legitimacy of the reduced Modified 3-parameter Weibull distribution

For the reduced Modified 3-parameter Weibull distribution (RM3PWD) to be legitimate, then;

$$\int_0^{\infty} f(x; k, \lambda) dx = 1$$

Therefore,

$$\int_0^{\infty} f(x; k, \lambda) dx = \int_0^{\infty} \frac{k}{\lambda - k} \left(\frac{x - k}{\lambda - k} \right)^{k-1} e^{-\left(\frac{x-k}{\lambda-k}\right)^k} dx \quad (53)$$

Let $t = \left(\frac{x-k}{\lambda-k}\right)^k$ which implies, $x = (\lambda - k)t^{\frac{1}{k}} + k$ where
 $dx = \frac{1}{k}(\lambda - k)t^{\frac{1}{k}-1} dt$.

Also, when $x = 0$, $t = 0$ and when $x = \infty$, $t = \infty$.

Substituting the above into Equation (53), we have;

$$\begin{aligned} \int_0^{\infty} f(x; k, \lambda) dx &= \int_0^{\infty} \frac{k}{\lambda - k} \left(t^{\frac{1}{k}-1} \right) e^{-t} \left[\frac{1}{k}(\lambda - k)t^{\frac{1}{k}-1} \right] dt \\ \int_0^{\infty} f(x; k, \lambda) dx &= \int_0^{\infty} \frac{k}{\lambda - k} \frac{1}{k} (\lambda - k) t^{\frac{1}{k}-1} t^{\frac{1}{k}-1} e^{-t} dt \\ \int_0^{\infty} f(x; k, \lambda) dx &= \int_0^{\infty} t^0 e^{-t} dt \\ \int_0^{\infty} f(x; k, \lambda) dx &= \int_0^{\infty} e^{-t} dt \\ \int_0^{\infty} f(x; k, \lambda) dx &= [-e^{-t}]_0^{\infty} \\ \int_0^{\infty} f(x; k, \lambda) dx &= [-e^{-\infty}] - [-e^0] \\ \int_0^{\infty} f(x; k, \lambda) dx &= [0] - [-1] \\ \therefore \int_0^{\infty} f(x; k, \lambda) dx &= 1 \end{aligned} \quad (54)$$

Hence the proof. This shows that the reduced modified 3-parameter Weibull distribution is legitimate.

Cumulative Density Function (CDF) of the reduced modified 3-parameter Weibull distribution

For the Cumulative Density Function (CDF) with reference to Equations (53) and (54), we have;

$$\int_{-\infty}^0 f(x; k, \lambda) dx + \int_0^x f(x; k, \lambda) dx + \int_x^{\infty} f(x; k, \lambda) dx = \int_0^t e^{-t} dt$$

Since $\int_{-\infty}^0 f(x; k, \lambda) dx = \int_x^{\infty} f(x; k, \lambda) dx = 0$ and $\int_0^x f(x; k, \lambda) dx = \int_0^t e^{-t} dt$

$$\therefore \text{CDF} = \int_0^t e^{-t} dt$$

$$\text{CDF} = [-e^{-t}] - [-e^0]$$

$$\text{CDF} = 1 - e^{-t}$$

This implies

$$\text{CDF} = 1 - e^{-\left(\frac{x-k}{\lambda-k}\right)^k} \quad (55)$$

Mean and Variance of the reduced Modified 3-parameter distribution.

For the mean or expectation of the reduced parameter distribution, we have;

$$E(X) = \int_0^{\infty} xf(x; k, \lambda) dx = \int_0^{\infty} x \frac{k}{\lambda-k} \left(\frac{x-k}{\lambda-k}\right)^{k-1} e^{-\left(\frac{x-k}{\lambda-k}\right)^k} dx$$

Employing the techniques used for Equations (53) and (54) where

$x = (\lambda - k)t^{\frac{1}{k}} + k$, we have;

$$E(X) = \int_0^{\infty} \left((\lambda - k)t^{\frac{1}{k}} + k \right) e^{-t} dt$$

$$E(X) = (\lambda - k) \int_0^{\infty} t^{\frac{1}{k}} e^{-t} dt + \int_0^{\infty} k e^{-t} dt \quad (56)$$

$$E(X) = (\lambda - k) \int_0^{\infty} t^{\frac{1}{k}} e^{-t} dt + k$$

Let $\alpha - 1 = \frac{1}{k}$, which implies; $\alpha = \frac{1}{k} + 1$

$$\therefore E(X) = (\lambda - k) \int_0^{\infty} (t^{\alpha-1}) e^{-t} dt + k$$

But $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$

$$\therefore E(X) = (\lambda - k) \Gamma(\alpha) + k$$

$$\therefore E(X) = (\lambda - k) \Gamma\left(\frac{1}{k} + 1\right) + k \quad (57)$$

As well, for the variance of the reduced Modified 3-parameter Weibull distribution, we have;

$$\text{VAR}(X) = E(X^2) - [E(X)]^2$$

$$\therefore E(X^2) = \int_0^{\infty} x^2 f(x; k, \lambda, Q) dx = \int_0^{\infty} x^2 \frac{k}{\lambda-k} \left(\frac{x-k}{\lambda-k}\right)^{k-1} e^{-\left(\frac{x-k}{\lambda-k}\right)^k} dx \quad (58)$$

Employing the techniques used for Equation (28) and (29) where

$x = (\lambda - k)t^{\frac{1}{k}} + k$, we have;

$$E(X^2) = \int_0^{\infty} \left((\lambda - k)t^{\frac{1}{k}} + k \right)^2 e^{-t} dt$$

$$E(X^2) = \int_0^\infty \left[\left((\lambda - k)t^{\frac{1}{k}} \right)^2 + k^2 + 2k(\lambda - k)t^{\frac{1}{k}} \right] e^{-t} dt \quad (59)$$

$$E(X^2) = \int_0^\infty [k^2] e^{-t} dt + \int_0^\infty [2k(\lambda - k)t^{\frac{1}{k}}] e^{-t} dt + \int_0^\infty \left[\left((\lambda - k)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt$$

With reference to Equations (28) and (29),

$$\int_0^\infty [k^2] e^{-t} dt + \int_0^\infty [2k(\lambda - k)t^{\frac{1}{k}}] e^{-t} dt = k^2 + 2k(\lambda - k)\Gamma\left(\frac{1}{k} + 1\right)$$

$$\int_0^\infty \left[\left((\lambda - k)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - k)^2 \int_0^\infty t^{\frac{2}{k}} e^{-t} dt$$

$$\text{Let } \alpha - 1 = \frac{2}{k}$$

$$\therefore \int_0^\infty \left[\left((\lambda - k)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - k)^2 \int_0^\infty t^{\alpha-1} e^{-t} dt$$

$$\therefore \int_0^\infty \left[\left((\lambda - k)t^{\frac{1}{k}} \right)^2 \right] e^{-t} dt = (\lambda - k)^2 \Gamma(\alpha) = (\lambda - k)^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$\therefore E(X^2) = k^2 + 2k(\lambda - k)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - k)^2 \Gamma\left(\frac{2}{k} + 1\right) \quad (60)$$

$$\therefore \text{VAR}(X) = k^2 + 2k(\lambda - k)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - k)^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$- \left[k + (\lambda - k)\Gamma\left(\frac{1}{k} + 1\right) \right]^2$$

$$\therefore \text{VAR}(X) = k^2 + 2k(\lambda - k)\Gamma\left(\frac{1}{k} + 1\right) + (\lambda - k)^2 \Gamma\left(\frac{2}{k} + 1\right)$$

$$- \left[k^2 + (\lambda - k)^2 \Gamma^2\left(\frac{1}{k} + 1\right) + 2k(\lambda - k)\Gamma\left(\frac{1}{k} + 1\right) \right]$$

$$\therefore \text{VAR}(X) = (\lambda - k)^2 \left[\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right) \right] \quad (61)$$

From Equation (61), we can see that the variance of the reduced distribution is less than that of the three-parameter distribution since the scalar of $(\lambda - k)^2$ might be less than λ^2 even with reference to the two-parameter Weibull distribution.

Moment Generation Function (MGF) of the reduced Modified 3-parameter Weibull distribution

The Moment Generation Function (MGF) of the reduced Modified 3-parameter Weibull distribution is given as;

$$E(X^n) = \int_0^\infty x^n f(x; k, \lambda) dx = \int_0^\infty x^n \frac{k}{\lambda - k} \left(\frac{x - k}{\lambda - k} \right)^{k-1} e^{-\left(\frac{x-k}{\lambda-k}\right)^k} dx \quad (62)$$

Employing the techniques used for Equation (53) and (54) where $x = (\lambda - k)t^{\frac{1}{k}} + k$, we have;

$$E(X^n) = \int_0^\infty \left[(\lambda - k)t^{\frac{1}{k}} + k \right]^n e^{-t} dt \quad (63)$$

$$\text{But } (a + b)^n = \sum_{r=0}^n \binom{n}{r} a^r b^{n-r}, \quad r = 0, 1, 2, \dots, n$$

$$\therefore E(X^n) = \int_0^\infty \sum_{r=0}^n \binom{n}{r} \left[(\lambda - k)t^{\frac{1}{k}} \right]^r k^{n-r} e^{-t} dt \quad (64)$$

This implies;

$$E(X^n) = \int_0^\infty \left[k^n + n(\lambda - k)t^{\frac{1}{k}}k^{n-1} + \frac{n(n-1)(\lambda - k)^2 t^{\frac{2}{k}}k^{n-2}}{2!} + \frac{n(n-1)(n-2)(\lambda - k)^3 t^{\frac{3}{k}}k^{n-3}}{3!} + \dots + (\lambda - k)^n t^{\frac{n}{k}} \right] e^{-t} dt \quad (65)$$

Integrating Equation (65) with reference to Equations (59)-(61), we have;

$$E(X^n) = k^n + n(\lambda - k)k^{n-1}\Gamma\left(\frac{1}{k} + 1\right) + \frac{n(n-1)(\lambda - k)^2 k^{n-2}\Gamma\left(\frac{2}{k} + 1\right)}{2!} + \frac{n(n-1)(n-2)(\lambda - k)^3 k^{n-3}\Gamma\left(\frac{3}{k} + 1\right)}{3!} + \dots + (\lambda - k)^n \Gamma\left(\frac{n}{k} + 1\right) \quad (66)$$

3.4.3. The Maximum Likelihood Estimation (MLE) of the Reduced Modified 3-Parameter Weibull Distribution

The Maximum Likelihood Estimation (MLE) of the reduced Modified 3-parameter Weibull distribution is also given as;

$$L(x_i; k, \lambda) = \prod_{i=1}^n f(x_i; k, \lambda) \quad (67)$$

$$L(x_i; k, \lambda) = \prod_{i=1}^n \frac{k}{\lambda - k} \left(\frac{x_i - k}{\lambda - k} \right)^{k-1} e^{-\left(\frac{x_i - k}{\lambda - k}\right)^k} \quad (68)$$

Taking antilog of both sides, we have;

$$\ln L(x_i; k, \lambda) = \sum_{i=1}^n \ln \left[\frac{k}{\lambda - k} \left(\frac{x_i - k}{\lambda - k} \right)^{k-1} e^{-\left(\frac{x_i - k}{\lambda - k}\right)^k} \right]$$

$$\ln L(x_i; k, \lambda) = \sum_{i=1}^n \left\{ \ln k - \ln(\lambda - k) + (k-1) [\ln(x_i - k) - \ln(\lambda - k)] - \left(\frac{x_i - k}{\lambda - k} \right)^k \right\}$$

Therefore, the log likelihood function is given as;

$$\ln L(x_i; k, \lambda) = \sum_{i=1}^n \left\{ \ln k - k \ln(\lambda - k) + (k-1) \ln(x_i - k) - \left(\frac{x_i - k}{\lambda - k} \right)^k \right\} \quad (69)$$

Differentiating Equation (69) with respect to k, λ we obtain the respective Equations (70) and (71);

$$\frac{d \ln L(x_i; k, \lambda)}{dk} = \sum_{i=1}^n \left\{ \frac{1}{k} - \ln(\lambda - k) + \frac{k}{\lambda - k} + \ln(x_i - k) - \left(\frac{x_i - k}{\lambda - k} \right)^k \left[\frac{k}{\lambda - k} + \ln \left(\frac{x_i - k}{\lambda - k} \right) \right] \right\} \quad (70)$$

$$\frac{d \ln L(x_i; k, \lambda)}{d\lambda} = \sum_{i=1}^n \left\{ \frac{k}{\lambda - k} - k(x_i - k) \left(\frac{x_i - k}{\lambda - k} \right)^{k-1} (\lambda - k)^{-2} \right\} \quad (71)$$

Therefore, equating Equations (70) and (71) to zero and solving them simultaneously using numerical methods, the Maximum Likelihood Estimates of k, λ are produced.

Cramer-Rao lower bound inequality

Also, the Cramer-Rao lower bound inequality attained for each of the estimated parameters for k, λ are as follows in Equations ((72), (73)) respectively;

$$\text{Variance}(\hat{k}(x)) \geq \frac{1}{nI(k)} \text{ or } \frac{1}{I(k)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda)}{dk} \right]^2 \right]} \quad (72)$$

$$\text{Variance}(\hat{\lambda}(x)) \geq \frac{1}{nI(\lambda)} \text{ or } \frac{1}{I(\lambda)} = \frac{1}{E \left[\left[\frac{d \ln L(x_i; k, \lambda)}{d\lambda} \right]^2 \right]} \quad (73)$$

3.5. Forecasting Future Values with the Inverse CDF for the Weibull Distributions

The inverse cumulative distribution function (CDF), also known as the quantile function or percent-point function, is a mathematical function that allows for determining the value at which a given probability occurs in a probability distribution [44].

Therefore, for the original 3-parameter Weibull distribution, the CDF is given as (see Equation (3));

$$\text{CDF} = 1 - e^{-\left(\frac{x-Q}{\lambda} \right)^k}$$

If p is the result of the probability that a single observation from the original 3-parameter Weibull distribution with parameters k, λ, Q in the interval $[0, x]$, then;

$$p = 1 - e^{-\left(\frac{x-Q}{\lambda} \right)^k}$$

Taking natural log of both sides and making x the subject, we have;

$$x = Q + \lambda [-\ln(1-p)]^{\frac{1}{k}} \quad (74)$$

This therefore means that the result of the value of x is an observation from the original 3-parameter Weibull distribution with parameters k, λ, Q that falls in the range $[0, x]$ with probability p .

Also, for the Modified 3-parameter Weibull distribution, the CDF is given as (see Equation (30));

$$\text{CDF} = 1 - e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k}$$

If p is the result of the probability that a single observation from the Modified 3-parameter Weibull distribution with parameters k, λ, Q in the interval $[0, x]$, then;

$$p = 1 - e^{-\left(\frac{x-Q}{\lambda-Q}\right)^k}$$

Taking natural log of both sides and making x the subject, we have;

$$x = Q + (\lambda - Q) [-\ln(1-p)]^{\frac{1}{k}} \quad (75)$$

This therefore means that the result of the value of x is an observation from the Modified 3-parameter Weibull distribution with parameters k, λ, Q that falls in the range $[0, x]$ with probability p .

Similarly, for the Reduced Modified 3-parameter Weibull distribution, the CDF is given as (see Equation (54));

$$\text{CDF} = 1 - e^{-\left(\frac{x-k}{\lambda-k}\right)^k}$$

If p is the result of the probability that a single observation from the Reduced Modified 3-parameter Weibull distribution with parameters k, λ in the interval $[0, x]$, then;

$$p = 1 - e^{-\left(\frac{x-k}{\lambda-k}\right)^k}$$

Taking natural log of both sides and making x the subject, we have;

$$x = k + (\lambda - k) [-\ln(1-p)]^{\frac{1}{k}} \quad (76)$$

This therefore means that the result of the value of x is an observation from the Reduced Modified 3-parameter Weibull distribution with parameters k, λ that falls in the range $[0, x]$ with probability p .

4. Results and Discussions

4.1. Evaluation of Statistical Distribution Functions in Modeling Gold Production

Figure 2 represents the scatter plot of the quarterly gold produced between 2009 to 2021 in Ghana. In the first quarter of 2009, gold produced was 727,907 ounces. This figure went up to 774,443 ounces in the second quarter of same

year (*i.e.*, 2009) and continued this upwards move to 1,124,809.3 ounces in the first quarter of 2013 but fell drastically to 757,990 ounces in the fourth quarter of 2015. It then began to pick up in the first quarter of 2016 to 978,893 ounces and continued this right movement until in the third quarter of the year 2020 when it started falling from 1,224,868 ounces in the second quarter of 2020 to 927,781 ounces in the third quarter of 2020.

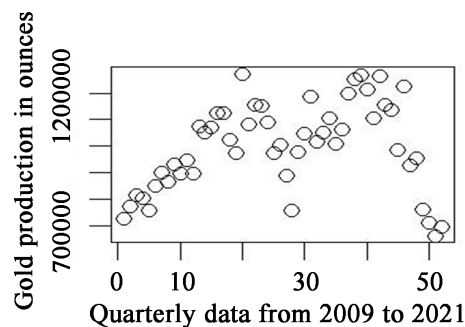
The lowest gold production was recorded in the third quarter of 2021 (*i.e.*, 661,277 ounces) with the highest being recorded in the fourth quarter of 2013 (*i.e.*, 1,273,786.88 ounces). Although COVID-19 was severe in 2020 in Ghana, that year recorded higher gold production than the year 2021.

From the look at the scatterplot, the data cannot be fitted linearly. This depicts a non-linear family of distributions [45] (**Table 2 & Table 3**).

The results indicate that the data is normally distributed [46]. This is presented graphically in **Figure 3** and **Figure 4**.

Now, to fit a distribution to the gold production data, we must check whether it is independent and identically distributed. That is, in probability theory, the random variable must be independent and identically distributed [47].

Table 4 represents the independent and identically distributed (iid) test. From the test, the data is iid from lag 4 since their p-values are greater than 0.05 [48]. This depicts the nature of the data since it is a quarterly data. This justifies the fitting of the distribution function to the gold production data [47].



Source of data: Minerals Commission of Ghana [12].

Figure 2. Scatterplot of the Quarterly gold production in ounces between 2009 and 2021.

Table 2. Descriptive statistics of quarterly gold production (2009-2021).

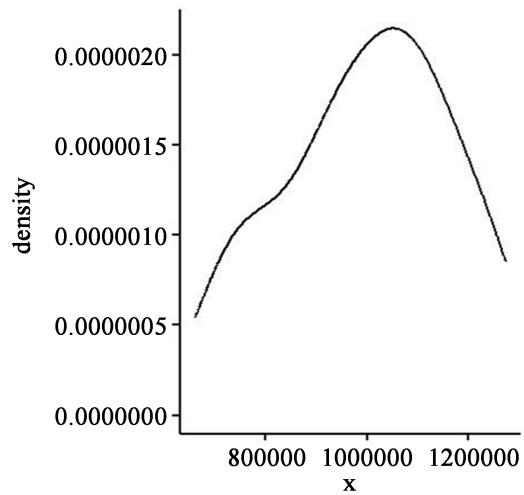
Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Gold (Oz)	52	661,277.000	1,273,786.880	999,333.944	165,732.913

Source of data: Minerals Commission of Ghana [12].

Table 3. Normality Test on the gold production data.

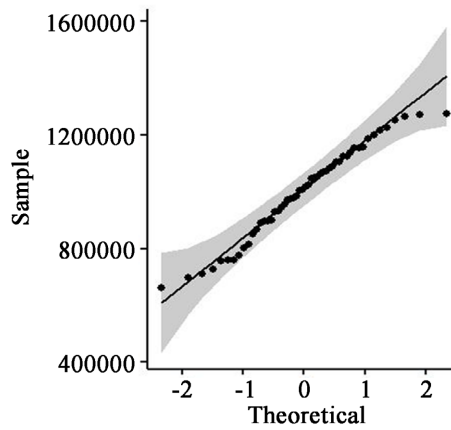
Variable	Shapiro-Wilk	Anderson-Darling	Lilliefors	Jarque-Bera
(Gold_Oz)	0.2094	0.4581	0.6686	0.3643

Source: Estimation from the gold production data.



Source: Authors own estimation, 2023.

Figure 3. Normality Test on the gold production data.



Source: Authors own estimation, 2023.

Figure 4. QQPLOT of gold production (2009-2021).

Table 4. IID Test on the gold production data.

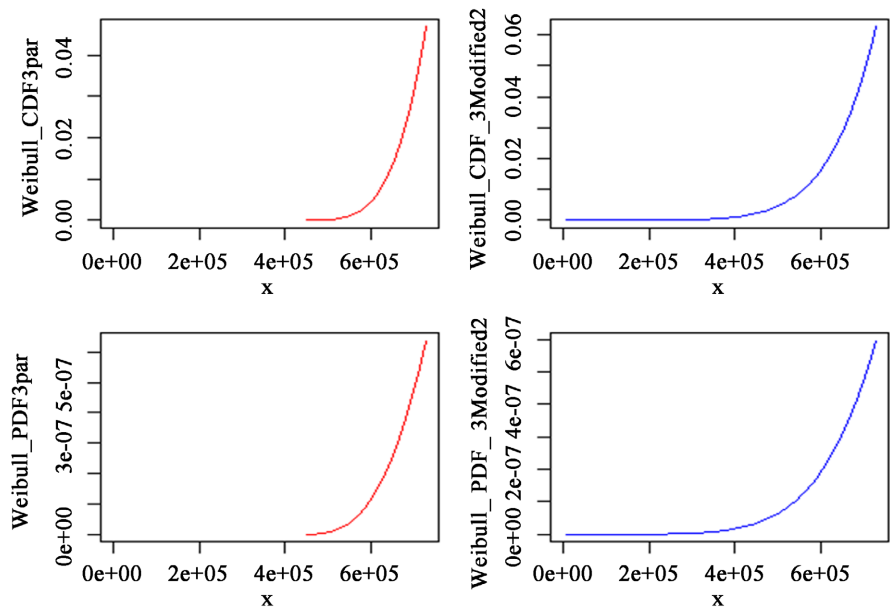
Lag	J [x, x]	P-value
1	33.351	0.000
2	19.840	0.000
3	11.447	0.003
4	5.169	0.075
5	1.336	0.513
6	0.622	0.733
7	0.398	0.820
8	0.439	0.803
9	0.776	0.678
10	3.498	0.174

Source: Authors own estimation, 2023.

4.2. Application of the Reduced Modified Parameter Distribution on Gold Production

The theoretical plot comparison between the new and the old distribution is shown in **Figure 5**. It is clear from **Figure 5** that the new plot is closer to the y-axis than the original 3-parameter Weibull distribution. This shows that the two are significantly different but have similar shapes.

Table 5 presents the estimated parameters of the fitted distributions to the



Source: Authors own estimation, 2023.

Figure 5. Theoretical plot of the new (in blue) and old (in red) distributions.

Table 5. Estimated parameters of the fitted distributions of the gold production.

Distribution	K	Λ	Q	Fitted Mean	Fitted Std. deviation
Weibull (3)	3.8810	609,471.5201	449,280.3234	1,000,763.4749	158,998.6203
Standard Error	0.3878	15,570.4049	26,572.7288		
Z value	9.335	249,660,000,000	741,200,000,000		
P-value	0.0000	0.0000	0.0000		
M3PWD	7.1371	1,068,303.5082	7.1371	1,000,334.8918	165,077.3099
Standard Error	0.7848	21,887.4978	0.7848		
Z value	9.5894	126,180,000,000	9.5894		
P-value	0.0000	0.0000	0.0000		
RM3PWD	7.1371	1,068,303.5082		1,000,334.8918	165,077.3099
Standard Error	0.7848	21,887.4978			
Z value	9.5894	126,180,000,000			
P-value	0.0000	0.0000			

Source: Authors own estimation, 2023.

quarterly gold production data. The p-values of all the parameters show that they are 99.9% significant to be part of the fitted models or distributions. Also, the average quarterly fitted gold production is $1000334.8918 \pm 75,327.080$ ($\pm 7.5\%$) [*i.e.*, 925,007.812 - 1,075,661.972]. This indicates that the average annually fitted gold production lies between 3700031.248 and 4302647.888 ounces at 99.9% confidence level. Therefore, the predicted gold production for the year 2022 is 3.7million ounces at 99.9% confidence level. Therefore, the predicted gold production for the year 2022 is 3.7million ounces at 99.9% confidence level.

Table 6 represents the goodness of fit and the selection criteria for the gold production data. The results in **Table 6** show clearly that the modified and the reduced distributions performed better than the original three-parameter Weibull distribution in terms of all the goodness of fit measures as well as with the selection criteria. The paper revealed that the risk of rejecting the new model based on the Kolmogorov-Smirnoff test is 99.3%. Also, with the AIC and BIC values of 1399.003 and 1402.905 respectively, the new model is adjudged as the best probability distribution in fitting the gold production quarterly data with the minimal error. This finding is therefore more accurate than those done by Kaba *et al.* [22] and Appiah *et al.* [23]. Findings of Kaba *et al.* [22], showed that, with a Beta (P-value of 0.75) distribution, the total average mining production fell within $210,414.86 \pm 3301.59$ in Bank Cubic Meters at 95% confidence level while that of Appiah *et al.* [23] revealed that Gompertz stochastic model was identified to give the best approximation of gold production trends in Ghana with R-Square of 0.9402 with RMSE of 335,866.94. Meanwhile, the proposed current model produced a Kolmogorov-Smirnov (K-S) of 0.993 which is the best as compared to those of Kaba *et al.* [22], Appiah *et al.* [23] and other research findings.

In **Table 7**, we can see that the actual gold production figures almost fall within the estimated intervals with the new reduced modified Weibull distribution.

Figure 6 represents the empirical plot of the modified distribution on the quarterly gold production data between 2009 and 2021. It is obvious from this plot that the new model fitted the data appropriately. It can therefore be concluded that this current model deduced from the three-parameter Weibull distribution is better fit to the quarterly gold production data.

5. Conclusions

Based on the analysis conducted, it is concluded that the reduced modified

Table 6. Goodness of fit and selection criteria for gold production.

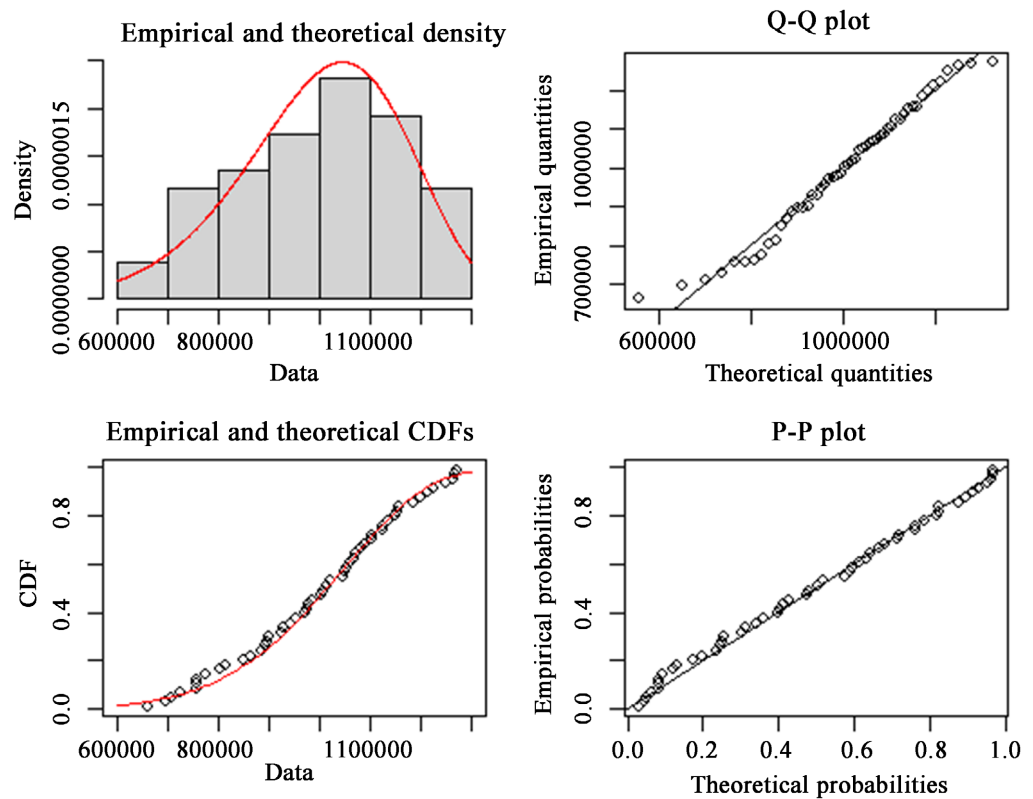
Distribution	KS (P-value)	AD (P-value)	CM (P-value)	-2 log L	AIC	AICc	BIC
Weibull (3)	0.9438	0.915	0.9431	1394.737	1401.000	1401.000	1407.000
M3PWD	0.9896	0.9772	0.9931	1395.003	1401.000	1401.000	1407.000
RM3PWD	0.9930	0.9772	0.9931	1395.003	1399.003	1399.247	1402.905

Source: Authors own estimation, 2023.

Table 7. Comparison between the estimated intervals and the actual production values.

Quarter	Actual Gold (oz)	Lower bound	Upper bound	Relative frequency	Density (Distribution)
1q-2009	727,907	600,000	668,000	0.0192	0.0183
2q-2009	774,443	668,000	736,000	0.0577	0.0332
3q-2009	813,962	668,000	736,000	0.0577	0.0332
4q-2009	803,520	668,000	736,000	0.0577	0.0332
1q-2010	756,932	736,000	804,000	0.0962	0.0556
2q-2010	849,423	736,000	804,000	0.0962	0.0556
3q-2010	901,003	736,000	804,000	0.0962	0.0556
4q-2010	867,069	736,000	804,000	0.0962	0.0556
1q-2011	931,342	736,000	804,000	0.0962	0.0556
2q-2011	897,387	804,000	872,000	0.0577	0.0860
3q-2011	945,740	804,000	872,000	0.0577	0.0860
4q-2011	895,396	804,000	872,000	0.0577	0.0860

Source: Authors own estimation, 2023.



Source: Source: Authors own estimation, 2023.

Figure 6. Empirical plot of the RM3PWD on gold production data.

3-parameter Weibull distribution provides a perfect fit for modeling gold production in Ghana. The statistical analysis and goodness-of-fit tests support the sui-

tability of this distribution for describing the behavior of gold production in the mining industry. This finding indicates that the reduced modified 3-parameter Weibull distribution can serve as an effective tool for understanding and predicting gold production trends in Ghana. While the reduced modified 3-parameter Weibull distribution provides a perfect fit for gold production, it is important to recognize that other external factors can influence production levels. Factors such as changes in mining techniques, geological conditions, environmental regulations, or geopolitical factors may impact gold production. It is recommended that future modeling activities incorporate these additional factors to enhance its predictive capabilities and improve operational efficiency.

Implications and significance for investors

The finding that the reduced modified 3-parameter Weibull distribution provides the best fit for gold production in Ghana carries significant implications for forecasting gold production and holds great significance for investors. The identified distribution can be leveraged to set production goals and make informed investment decisions in the following ways:

1) **Accurate Production Forecasts:** By utilizing the reduced modified 3-parameter Weibull distribution, analysts can generate more precise and reliable forecasts of gold production in Ghana. The distribution's ability to capture the shape, scale, and location parameters of the data allows for improved estimation of future production levels. This accuracy enhances decision-making by providing stakeholders with a clearer understanding of the expected production volumes.

2) **Risk Assessment and Mitigation:** The identified distribution enables a comprehensive assessment of risk associated with gold production in Ghana. By analyzing the distribution's properties, such as the shape parameter, analysts can quantify the likelihood of extreme events, such as production disruptions or significant fluctuations. This assessment aids in the development of risk management strategies, contingency plans, and operational adjustments to mitigate the potential negative impacts on production and investments.

3) **Setting Realistic Production Goals:** The identified distribution assists in setting realistic and achievable production goals in Ghana's gold mining operations. By analyzing the distribution's parameters and tail behavior, stakeholders can establish production targets that account for both the average production levels and the potential for extreme outcomes. This approach allows for more robust planning and resource allocation, improving operational efficiency and optimizing overall production processes.

4) **Informed Investment Decisions:** The choice of the reduced modified 3-parameter Weibull distribution for gold production forecasting provides valuable insights for investors. The distribution's characteristics help assess the expected range of production outcomes and associated risks. Investors can utilize this information to make informed decisions regarding funding, project evaluations, and portfolio diversification strategies. Understanding the probabilistic

nature of gold production through the identified distribution enhances the ability to evaluate the potential returns and risks associated with gold mining investments in Ghana.

In summary, the finding that the reduced modified 3-parameter Weibull distribution offers the best fit for gold production in Ghana has significant implications for forecasting and investment decisions. The distribution enables accurate production forecasts, facilitates risk assessment and mitigation, assists in setting realistic production goals, and provides valuable insights for investors. By leveraging the identified distribution, stakeholders can make more informed decisions and optimize their strategies in the dynamic gold mining industry in Ghana.

Details of the external factors

Indeed, external factors can have a significant impact on gold production levels and, consequently, influence the chosen statistical distribution's accuracy and reliability in forecasting. Several examples of these external factors and their potential implications are as follows:

1) Changes in Mining Techniques: Advancements in mining technologies and techniques can lead to changes in production processes, extraction rates, and efficiencies. For instance, the introduction of new equipment or innovative extraction methods may alter the shape and scale parameters of the production distribution. This can impact the accuracy of the forecasting model if the chosen statistical distribution does not adequately capture these changes, potentially resulting in biased forecasts.

2) Geological Conditions: Variations in geological conditions, such as changes in ore grades, mineral composition, or deposit characteristics, can influence gold production. Different geological conditions may require adjustments to the distribution parameters, affecting the model's accuracy. For example, if the distribution assumes a constant average production rate but the geological conditions exhibit a declining trend in ore grades, the forecasts may overestimate future production levels.

3) Environmental Regulations: Stringent environmental regulations and compliance requirements can affect gold production by imposing limitations on mining activities. These regulations may lead to production disruptions, reduced output, or changes in the mining processes to comply with environmental standards. The impact of such regulations may not be adequately captured by the chosen statistical distribution, potentially leading to forecast inaccuracies.

4) Geopolitical Factors: Geopolitical events and factors, such as changes in government policies, political instability, trade disputes, or legal frameworks, can significantly influence gold production. For instance, the introduction of new mining regulations or the imposition of export restrictions can impact the distribution's parameters and result in deviations from historical production patterns. Failure to account for these geopolitical factors may compromise the reliability of the forecasting model.

The implications of these external factors for the accuracy and reliability of the forecasting model are substantial. Ignoring or underestimating their influence can lead to biased forecasts and inadequate risk assessments. It is crucial to regularly update and refine the chosen statistical distribution to incorporate new information and adapt to changing external conditions. Continuous monitoring of external factors, robust sensitivity analysis, and ongoing calibration of the model based on actual production data and external indicators are essential to improve the accuracy and reliability of the forecasting model in the presence of these influential factors.

In brief, external factors such as changes in mining techniques, geological conditions, environmental regulations, and geopolitical factors can significantly impact gold production and the chosen statistical distribution's accuracy and reliability in forecasting. It is vital to consider and incorporate these factors into the forecasting model to enhance its ability to provide accurate and reliable forecasts in a dynamic and evolving mining industry.

Acknowledgements

Special appreciation goes to Mr. Martin Kwaku Ayisi, CEO of the Minerals Commission and Mr. Francis Frimpong, the Deputy Manager of Research and Statistics for their assistance and suggestions in this research paper.

Conflicts of Interest

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

References

- [1] Srivastava, M., Rao, A., Parihar, J.S., Chavriya, S. and Singh, S. (2023) What Do the AI Methods Tell Us about Predicting Price Volatility of Key Natural Resources: Evidence from Hyperparameter Tuning. *Resources Policy*, **80**, Article ID: 103249. <https://doi.org/10.1016/j.resourpol.2022.103249>
- [2] De-Arteaga, M., Feuerriegel, S. and Saar-Tsechansky, M. (2022) Algorithmic Fairness in Business Analytics: Directions for Research and Practice. *Production and Operations Management*, **31**, 3749-3770. <https://doi.org/10.1111/poms.13839>
- [3] Chimunhu, P., Topal, E., Ajak, A.D. and Asad, W. (2022) A Review of Machine Learning Applications for Underground Mine Planning and Scheduling. *Resources Policy*, **77**, Article ID: 102693. <https://doi.org/10.1016/j.resourpol.2022.102693>
- [4] Panagiotelis, A., Smith-Miles, K. and Duffield, S. (2012) Forecasting Gold Production: A Time Series Modelling Approach. *European Journal of Operational Research*, **221**, 186-194.
- [5] World Gold Council (n.d.) Gold Demand Trends. <https://www.gold.org/goldhub/research/gold-demand-trends>
- [6] Matrokhina, K.V., Trofimets, V.Y., Mazakov, E.B., Makhovikov, A.B. and Khaykin, M.M. (2023) Development of Methodology for Scenario Analysis of Investment Projects of Enterprises of the Mineral Resource Complex. *Journal of Mining Institute*, **259**, 112-124. <https://doi.org/10.31897/PMI.2023.3>

- [7] Wijayasekera, S.C., Hussain, S.A., Paudel, A., Paudel, B., Steen, J., Sadiq, R. and Hewage, K. (2022) Data Analytics and Artificial Intelligence in the Complex Environment of Megaprojects: Implications for Practitioners and Project Organizing Theory. *Project Management Journal*, **53**, 485-500. <https://doi.org/10.1177/87569728221114002>
- [8] Pereira, V., Tuffour, J., Patnaik, S., Temouri, Y., Malik, A. and Singh, S.K. (2021) The Quest for CSR: Mapping Responsible and Irresponsible Practices in an Intra-Organizational Context in Ghana's Gold Mining Industry. *Journal of Business Research*, **135**, 268-281. <https://doi.org/10.1016/j.jbusres.2021.06.024>
- [9] Statista (2021) Gold Price and Production Trend. Statista.
- [10] Schwartz, F.W., Lee, S. and Darrah, T.H. (2021) A Review of the Scope of Artisanal and Small-Scale Mining Worldwide, Poverty, and the Associated Health Impacts. *GeoHealth*, **5**, e2020GH000325. <https://doi.org/10.1029/2020GH000325>
- [11] World Gold Council (2022) World Gold Price and Production Trends by Country. World Gold Council.
- [12] Minerals Commission (2022) Gold Production and Its Price. Ghana Minerals Commission.
- [13] Statista (2022) Gold Price and Production Trend. Statista.
- [14] Sasu, D.D. (2021) Number of Chinese Migrants in African Countries 2000-2019. *Statistica*, **25**. <https://www.statista.com/statistics/1259725/stock-of-chinese-migrants-in-africa-by-region/>
- [15] Osumanu, I.K. (2020) Small-Scale Mining and Livelihood Dynamics in North-Eastern Ghana: Sustaining Rural Livelihoods in a Changing Environment. *Progress in Development Studies*, **20**, 208-222. <https://doi.org/10.1177/1464993420934223>
- [16] Ofosu, G., Dittmann, A., Sarpong, D. and Botchie, D. (2020) Socio-Economic and Environmental Implications of Artisanal and Small-Scale Mining (ASM) on Agriculture and Livelihoods. *Environmental Science and Policy*, **106**, 210-220. <https://doi.org/10.1016/j.envsci.2020.02.005>
- [17] Kumah, R. (2022) Determinants and Implications of Chinese Involvement in Informal Artisanal and Small-Scale Gold Mining in Ghana: Perspectives from the Grassroots. <https://www.proquest.com/openview/6d37688ee4c21f067b87782e211a856c/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [18] Livieris, I.E., Pintelas, E. and Pintelas, P. (2020) A CNN-LSTM Model for Gold Price Time-Series Forecasting. *Neural Computing and Applications*, **32**, 17351-17360. <https://doi.org/10.1007/s00521-020-04867-x>
- [19] Hess, J.P. (2021) A Multi-Level Analysis of Sustainability Practices in Ghana: Examining the Timber, Cocoa, and Gold Mining Industries. *International Journal of Organizational Analysis*, **30**, 760-777. <https://doi.org/10.1108/IJOA-01-2020-2011>
- [20] Matroushi, S.M. (2011) Hybrid Computational Intelligence Systems Based on Statistical and Neural Networks Methods for Time Series Forecasting: The Case of Gold Price. Lincoln University, Lincoln.
- [21] Yazdani-Chamzini, A., Yakhchali, S.H., Volungevičienė, D. and Zavadskas, E.K. (2012) Forecasting Gold Price Changes by Using Adaptive Network Fuzzy Inference System. *Journal of Business Economics and Management*, **13**, 994-1010. <https://doi.org/10.3846/16111699.2012.683808>

- [22] Kaba F.A., Temeng V.A. and Eshun, P.A. (2016) Application of Discrete Event Simulation in Mine Production Forecast. *Ghana Mining Journal*, **16**, 40-48. <https://doi.org/10.4314/gmj.v16i1.5>
- [23] Appiah, S.T., Buabeng, A. and Odoi, B. (2018) Comparative Study of Mathematical Models for Ghana's Gold Production. *Ghana Mining Journal*, **18**, 78-83. <https://doi.org/10.4314/gm.v18i1.10>
- [24] Hafezi, R. and Akhavan, A. (2018) Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network. *AUT Journal of Modeling and Simulation*, **50**, 71-82.
- [25] Chai, J., Zhao, C., Hu, Y. and Zhang, Z.G. (2021) Structural Analysis and Forecast of Gold Price Returns. *Journal of Management Science and Engineering*, **6**, 135-145. <https://doi.org/10.1016/j.jmse.2021.02.011>
- [26] Adhikari, R. and Agrawal, R.K. (2013) An Introductory Study on Time Series Modeling and Forecasting. ArXiv Preprint ArXiv: 1302.6613.
- [27] Sarkar, S., Zhu, X., Melnykov, V. and Ingrassia, S. (2020) On Parsimonious Models for Modeling Matrix Data. *Computational Statistics and Data Analysis*, **142**, Article ID: 106822. <https://doi.org/10.1016/j.csda.2019.106822>
- [28] Simsek, S., Kursuncu, U., Kibis, E., AnisAbdellatif, M. and Dag, A. (2020) A Hybrid Data Mining Approach for Identifying the Temporal Effects of Variables Associated with Breast Cancer Survival. *Expert Systems with Applications*, **139**, Article ID: 112863. <https://doi.org/10.1016/j.eswa.2019.112863>
- [29] Ribeiro, M.T., Singh, S. and Guestrin, C. (2018) Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 1527-1535. <https://doi.org/10.1609/aaai.v32i1.11491>
- [30] Salim, R. (2013) A Review on Conditional Extreme Value Analysis. Hal Science. <https://hal.science/hal-00770546/>
- [31] Stupfler, G. (2019) On a Relationship sssbetween Randomly and Non-Randomly Thresholded Empirical Average Excesses for Heavy Tails. *Extremes*, **22**, 749-769. <https://doi.org/10.1007/s10687-019-00351-5>
- [32] Allouche, M., El Methni, J. and Girard, S. (2022) A Refined Weissman Estimator for Extreme Quantiles. *Extremes*, 1-28. <https://doi.org/10.1007/s10687-022-00452-8>
- [33] Lèbre, É., Owen, J.R., Stringer, M., Kemp, D. and Valenta, R.K. (2021) Global Scan of Disruptions to the Mine Life Cycle: Price, Ownership, and Local Impact. *Environmental Science and Technology*, **55**, 4324-4331. <https://doi.org/10.1021/acs.est.0c08546>
- [34] Clarkson, D., Eastoe, E. and Leeson, A. (2023) The Importance of Context in Extreme Value Analysis with Application to Extreme Temperatures in the USA and Greenland. *Journal of the Royal Statistical Society Series C: Applied Statistics*, Article ID: Qlad020. <https://doi.org/10.1093/jrsssc/qlad020>
- [35] Liu, Q., Huang, X. and Zhou, H. (2022) The Flexible Gumbel Distribution: A New Model for Inference about the Mode. ArXiv: 2212.01832.
- [36] Bozóki, S. and Pataricza, A. (2022) Extreme Value Analysis for Time-Variable Mixed Workload. *Periodica Polytechnica Electrical Engineering and Computer Science*, **66**, 1-11. <https://doi.org/10.3311/PPee.17671>
- [37] Yu, X., Zhao, Z., Zhang, X., Zhang, Q., Liu, Y., Sun, C. and Chen, X. (2021) Deep-Learning-Based Open Set Fault Diagnosis by Extreme Value Theory. *IEEE Transactions on Industrial Informatics*, **18**, 185-196. <https://doi.org/10.1109/TII.2021.3070324>

- [38] Asante, E., Owusu, G. and Afriyie, E. (2018) Application of the Weibull Distribution Function for Characterizing and Modeling Gold Grades in the Spent Ore Stockpile at Bogoso Gold Limited, Ghana. *Journal of Sustainable Mining*, **17**, 131-137.
- [39] Abraham, B. and Ledolter, J. (2012) *Statistical Methods for Forecasting*. John Wiley & Sons, Hoboken.
- [40] Sikhimbayeva, D., Zulkharnay, A., Zhakupov, A., Yessilov, A. and Kuttybay, M. (2021) Analysis of Factors Affecting to the Development of Sub-Production Industry of the Republic of Kazakhstan. *Montenegrin Journal of Economics*, **17**, 41-57. <https://doi.org/10.14254/1800-5845/2021.17-3.4>
- [41] Han, J. (2021) Prediction of Precipitation Rate Based on Stationary Extreme Value Theory. *American Journal of Applied Mathematics*, **9**, 186-191. <https://doi.org/10.11648/j.ajam.20210905.13>
- [42] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. <https://doi.org/10.1007/978-1-4612-5449-2>
- [43] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (2012) *Extremes and Related Properties of Random Sequences and Processes*. Springer Science & Business Media, Berlin.
- [44] Harar, P., Elbrächter, D., Dörfler, M. and Johnson, K.D. (2022) Redistributor: Transforming Empirical Data Distributions. ArXiv: 2210.14219.
- [45] Krausch, N., Barz, T., Sawatzki, A., Gruber, M., Kamel, S., Neubauer, P. and Cruz Bournazou, M.N. (2019) Monte Carlo Simulations for the Analysis of Non-Linear Parameter Confidence Intervals in Optimal Experimental Design. *Frontiers in Bioengineering and Biotechnology*, **7**, Article 122. <https://doi.org/10.3389/fbioe.2019.00122>
- [46] Khatun, N. (2021) Applications of Normality Test in Statistical Analysis. *Open Journal of Statistics*, **11**, 113-122. <https://doi.org/10.4236/ojs.2021.1111006>
- [47] Majumdar, S.N., Pal, A. and Schehr, G. (2020) Extreme Value Statistics of Correlated Random Variables: A Pedagogical Review. *Physics Reports*, **840**, 1-32. <https://doi.org/10.1016/j.physrep.2019.10.005>
- [48] Choi, J.W., Nandram, B. and Choi, B. (2022) Combining Correlated P-Values from Primary Data Analyses. *International Journal of Statistics and Probability*, **11**, 12-27. <https://doi.org/10.5539/ijsp.v11n6p12>