Open Journal of Statistics, 2023, 13, 568-594
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Fitness of Four-Parameter Beta Distribution Function for Forecasting Gold Reserve and Its Production Lifespan in Ghana

Samuel Kwaku Obeng[1,2*], Christiana C. Nyarko[2], Lewis Brew[2], K. Sagary Nokoe[3]

[1]Research and Statistics Unit, Minerals Commission of Ghana, Accra, Ghana
[2]Department of Mathematics and Statistics, University of Mines and Technology, Tarkwa, Ghana
[3]Department 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) Fitness of Four-Parameter Beta Distribution Function for Forecasting Gold Reserve and Its Production Lifespan in Ghana. *Open Journal of Statistics*, **13**, 568-594.
https://doi.org/10.4236/ojs.2023.134028

Received: July 25, 2023
Accepted: August 20, 2023
Published: August 23, 2023

Abstract

Ghana, renowned for its abundant gold reserves, plays a significant role in the global mining industry. Effective management and accurate forecasting of these reserves are vital for sustainable resource utilization and economic planning. Forecasting gold reserves and estimating their production lifespan are complex tasks that require robust statistical models capable of capturing the underlying dynamics of gold deposit accumulation and extraction. To this end, the four-parameter Beta distribution function emerges as a promising candidate due to its flexibility and ability to handle non-negative data. This research aims to investigate the fitness and applicability of the four-parameter Beta distribution function for forecasting Ghana's gold reserves and estimating the production lifespan of this precious resource. The empirical paper relied mainly on quarterly secondary datasets on gold reserve between the years 2009 and 2022 secured from the Minerals Commission of Ghana, Accra. Several known statistical distributions including Beta, Weibull, Normal, Logistic and Gamma were explored with Maximum Likelihood Estimation (MLE) and evaluated using model selection criteria as AIC 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 four-parameter Beta distribution provided the best fit for gold reserve in Ghana. At a 99.9% confidence level and considering the current annual average gold production estimate of 3,700,031.248 to 4,302,647.888 ounces, the projected lifespan of gold production in Ghana extends to the year 1,953,765. This astounding estimate suggests that the country's gold reserves are expected to sustain production for an extended period, providing a critical resource for economic development and supporting the mining industry well into the distant future.

Keywords

Gold Reserve, Four-Parameter Beta Distribution Function, Goodness of Fit Statistics

1. Introduction

Ghana, renowned for its abundant gold reserves, plays a significant role in the global mining industry [1]. Effective management and accurate forecasting of these reserves are vital for sustainable resource utilization and economic planning [2]. Forecasting gold reserves and estimating their production lifespan are complex tasks that require robust statistical models capable of capturing the underlying dynamics of gold deposit accumulation and extraction [3]. To this end, the four-parameter Beta distribution function emerges as a promising candidate due to its flexibility and ability to handle non-negative data [4]. This research aims to investigate the fitness and applicability of the four-parameter Beta distribution function for forecasting Ghana's gold reserves and estimating the production lifespan of this precious resource.

The use of statistical distributions, such as the four-parameter Beta distribution function, has proven effective in various domains for modeling non-negative data [5] [6]. The four-parameter Beta distribution, being more flexible than other commonly used distributions, can better adapt to the irregularities and asymmetries often observed in gold reserve data [7].

Moreover, previous research in the field of resource forecasting and natural resource management has demonstrated the utility of the Beta distribution function for modeling finite reserves and predicting production lifespans [8] [9]. By extending this approach to Ghana's gold reserves, this study builds upon existing knowledge and applies it to a context of significant practical importance.

Additionally, Ghana's economy heavily relies on gold mining, and the industry faces increasing pressure to implement sustainable practices [10]. A reliable forecasting model using the four-parameter Beta distribution function can aid in making informed decisions about the timing and scale of gold extraction, ensuring long-term viability, and minimizing negative environmental and socio-economic impacts.

Overall, investigating the fitness of the four-parameter Beta distribution function for forecasting gold reserves and estimating their production lifespan in Ghana has the potential to contribute valuable insights to resource management practices in the country and serve as a template for similar studies with significant mining activities. Therefore, the main objective of the paper is to evaluate and investigate the fitness and applicability of the four-parameter Beta distribution function for forecasting Ghana's gold reserves and estimating the production lifespan of this precious resource through statistical modelling.

2. Materials and Methods

The study mainly employed quarterly secondary datasets on gold reserve between the years 2009 and 2022 secured from the Minerals Commission of Ghana, Accra. This quarterly data is mandatory and backed by law (*i.e.*, MINERALS AND MINING (HEALTH, SAFETY AND TECHNICAL REGULATIONS, 2012) (L.I. 2182), Regulation 28—Monthly and quarterly returns) for the mining companies to provide or submit to the Minerals Commission. Moreover, the mining companies use geological surveys to estimate the gold reserves. Geological surveys to estimate gold reserves involve a systematic and comprehensive study of the geological conditions in a particular region to determine the presence and quantity of gold deposits. The process typically includes several stages and methodologies. Here is an overview of the main processes involved in geological surveys to estimate gold reserves:

1) Desk Study and Literature Review: The first step is to conduct a thorough desk study and literature review. This involves gathering and reviewing all available geological, geochemical, and geophysical data, historical mining records, academic publications, and reports from previous exploration activities in the area. The goal is to gain an understanding of the geological history of the region and identify areas with the highest potential for gold deposits.

2) Geological Mapping: Geological mapping involves field surveys to create detailed maps of the geological formations in the area. Geologists study the rock types, structures, and mineral assemblages present to identify favorable geological environments for gold deposition, such as faults, folds, and shear zones.

3) Geochemical Sampling: Geochemical sampling involves collecting soil, rock, and stream sediment samples in strategic locations across the survey area. These samples are then analyzed in the laboratory to identify anomalous concentrations of gold and other associated minerals. Geochemical anomalies can indicate the presence of gold deposits nearby.

4) Geophysical Surveys: Geophysical methods are used to detect variations in the physical properties of rocks beneath the surface. Airborne or ground-based surveys may include magnetic, electromagnetic, and gravity measurements. These surveys can help identify potential gold-bearing structures or alterations associated with gold mineralization.

5) Drilling: Once promising targets are identified through the above methods, drilling is conducted to obtain core samples from specific locations. Diamond drilling is the most common method for exploration drilling in gold projects. The core samples provide valuable information about the geology, mineralization, and grade of the deposit.

6) Assaying: Core samples obtained from drilling are sent to laboratories for assaying. Assaying involves analyzing the samples to determine the gold content and other relevant elements. The results of the assays are crucial in estimating

the potential gold resources and reserves in the deposit.

7) Resource Estimation: Based on the data collected from geological mapping, geochemical sampling, geophysical surveys, and drilling, resource estimation is performed using various methods such as polygonal estimation, inverse distance weighting, and kriging. This estimation process divides the potential gold-bearing areas into categories like measured, indicated, and inferred resources based on the confidence level of the data.

8) Economic Evaluation: Once the resource estimation is complete, an economic evaluation is conducted to assess the viability of mining the gold deposit. Factors such as gold prices, production costs, mining methods, and infrastructure availability are considered to determine the economic feasibility of the project.

9) Reporting: The results of the geological survey, resource estimation, and economic evaluation are compiled into a technical report, compliant with international reporting standards like those set by the Committee for Mineral Reserves International Reporting Standards (CRIRSCO). This report is essential for attracting investors and securing necessary permits for further exploration and potential mining operations.

It is important to emphasize that geological surveys for gold estimation are complex and require expertise in various scientific disciplines. Additionally, the accuracy of the estimates can be influenced by the scale and intensity of exploration efforts, as well as the variability of gold deposits within the surveyed area. Regular updates and further exploration may be necessary to refine the estimates as more data becomes available.

Distributions such as Beta, Logistic, Weibull, Normal, Gamma were considered. Parameter Estimation used was Maximum Likelihood Estimation (MLE). The Model/Distribution Selection Criteria used were AIC, 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:

- Like 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, like 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 paper used a confidence level of 99.9% due to various factors, including the risk tolerance of stakeholders, the accuracy of data, and the importance of minimizing errors in reserve estimates. Selecting a specific confidence level, such as 99.9%, for gold reserve estimation is a decision that involves a trade-off between the level of confidence in the estimate and the associated uncertainty and cost. The following are the rationales for selecting a 99.9% confidence level for gold reserve estimation:

1) Risk Mitigation: Gold reserves are of significant economic and strategic importance for countries and mining companies. Choosing a high confidence level, like 99.9%, helps mitigate the risk of underestimating the actual reserves. Therefore, selecting a conservative estimate gives a higher probability of having reserves that can be economically viable and sustainable in the long term.

2) Long-Term Planning: Gold mining projects typically have long time horizons, and the decisions made based on reserve estimates have lasting impacts. A high confidence level provides more assurance for long-term planning and investment decisions. It reduces the chances of unforeseen production shortfalls and allows for more accurate financial projections.

3) Investor Confidence: Investors play a crucial role in funding gold mining projects. A higher confidence level in reserve estimates can boost investor confidence, making it more likely for investors to support the project. Greater certainty in reserve estimates can attract more investment, improving the project's chances of success.

4) Regulatory Requirements: Some countries or regulatory bodies may mandate a certain confidence level for reporting gold reserves. These regulations are designed to ensure transparency, comparability, and accountability in the mining industry.

5) Financial Reporting: Companies that publicly report their gold reserves must comply with accounting and reporting standards, such as those set by the CRIRSCO. These standards often require reporting reserves at specific confidence levels to provide consistency and comparability across different projects and companies.

6) Geological Complexity: The complexity of geological formations can introduce uncertainties in reserve estimation. In regions with complex geology, a higher confidence level may be chosen to account for potential geological risks and uncertainties.

7) Data Quality and Quantity: The availability and quality of data influence the confidence level selected for reserve estimation. In areas where data is limited or uncertain, a higher confidence level might be chosen to reflect the inherent uncertainty in the estimates.

8) Project Scale and Investment: Large-scale gold mining projects involve substantial investment, and the consequences of reserve estimation errors can be significant. A higher confidence level may be chosen to ensure the project's viability and protect investors' interests.

The paper covered quarterly gold reserves between 2009 and 2022. Several distribution functions were evaluated as initial step towards identifying the likely candidate. The initial distribution fitting for gold reserve using the XLSTAT software revealed that the Beta4 (*i.e.*, four-parameter Beta), Logistic, two-parameter

Weibull, Normal, two-parameter Gamma distributions perfectly fit the gold reserve data with their Kolmogorov-Smirnoff test p-value greater than 0.8. However, the best distribution among them is the four-parameter Beta distribution (K-S: 0.999999999999967) (Table 1).

3. Model Formulation and Parameter Estimation

3.1. Concepts of the Four-Parameter Beta Distribution

Beta distribution is a continues probability distribution which is used for experiments that involve two possible outcomes: success and failure, just like the Bernoulli and the Binomial distributions. The Beta distribution falls within the exponential family of distributions. Although, the Binomial and the Beta distributions both involve two possible outcomes (success or failure), Binomial distribution models the number of successes (say, x) while Beta distribution models the probability (say, p) of success. That is, while the probability is the parameter for the Binomial distribution, it is rather a random variable for the Beta distribution. In other words, the Beta distribution is the distribution on probabilities. The PDF of the first type of Beta distribution (that is Beta type 1) is given as

Table 1. Initial Distribution fitting for Gold Reserve using XLSTAT Software

Distribution	K-S-(p-value)
Beta4	0.999999999999967
Chi-square	<0.0001
Erlang	<0.0001
Exponential	<0.0001
Fisher-Tippett (1)	<0.0001
Fisher-Tippett (2)	0.0618
Gamma (1)	<0.0001
Gamma (2)	0.9109
GEV	0.8069
Gumbel	<0.0001
Log-normal	0.9109
Logistic	0.9180
Normal	0.9109
Normal (Standard)	<0.0001
Student	<0.0001
Weibull (1)	<0.0001
Weibull (2)	0.8325
Weibull (3)	0.8013

Source: Authors own estimation, 2023.

$$f(x;\alpha,\beta)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \text{ with } \alpha,\beta>0,$$

$$x\in[0,1] \text{ and } B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

where $E(X)=\frac{\alpha}{\alpha+\beta}$ and $VAR(X)=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.

However, the one being considered for this study is Beta with four parameters. The PDF of the four parametric Beta distribution is given as

$$f(x;\alpha,\beta,c,d)=\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}, \text{ with } \alpha,\beta>0,$$

$$x\in[c,d] \text{ and } c,d\in R$$

where $E(X)=c+\frac{(d-c)\alpha}{\alpha+\beta}$ and $VAR(X)=\frac{(d-c)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

The difference between the first Beta distribution (Beta2) and the Four parameter distribution (Beta4 or the generalized Beta distribution) is that, while the former takes values in the [0, 1] range, the later (that is the four-parameter distribution) takes values in the [c, d] range where the *c* (*i.e.* the min value) and *d* (*i.e.* the max value) can take any value. The Beta4 has an advantage over the Beta2 since the Beta4 is flexible and versatile and can take any value. That is, it provides description to many kinds of data.

3.2. Legitimacy and Parameter Derivations for the Four-Parameter Beta Distribution

For the Beta4 distribution to be legitimate, then;

$$\int_c^d f(x;\alpha,\beta,c,d)\,\mathrm{d}x=1$$

Therefore,

$$\int_c^d f(x;\alpha,\beta,c,d)\,\mathrm{d}x=\int_c^d\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}\mathrm{d}x \quad (1)$$

Let $y=\frac{x-c}{d-c}$ which implies, $1-y=\frac{d-x}{d-c}$, $x=y(d-c)+c$, where $\mathrm{d}x=(d-c)\mathrm{d}y$.

Also, when $x=c$, $y=0$ and when $x=d$, $y=1$.

Substituting the above into Equation (1), we have;

$$\int_c^d f(x;\alpha,\beta,c,d)\,\mathrm{d}x=\int_0^1\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y \quad (2)$$

But $B(\alpha,\beta)=\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\mathrm{d}y$ where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$\int_c^d f(x;\alpha,\beta,c,d)\,\mathrm{d}x=\frac{B(\alpha,\beta)}{B(\alpha,\beta)}$$

$$\therefore \int_c^d f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x = 1$$

Hence the proof. This shows that the Beta4 distribution is legitimate.

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

$$\mathrm{CDF} = \int_{-\infty}^{c} f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x + \int_c^d f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x + \int_d^{\infty} f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x$$

But $\int_c^d f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x = \int_0^1 \frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y$

$\therefore$ if $t = y$, then

$$\mathrm{CDF} = \int_0^y \frac{t^{\alpha-1}\left(1-t\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}t$$

Since $\int_{-\infty}^{c} f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x = \int_d^{\infty} f\left(x;\alpha,\beta,c,d\right)\mathrm{d}x = 0$.

This $\int_0^y t^{\alpha-1}\left(1-t\right)^{\beta-1}\mathrm{d}t$ then becomes the incomplete Beta distribution which is presented as $B\left(y;\alpha,\beta\right)$.

$$\therefore \mathrm{CDF} = I_y\left(\alpha,\beta\right) = \frac{B\left(y;\alpha,\beta\right)}{B\left(\alpha,\beta\right)}$$

This simplified regularized incomplete Beta function is the cumulative density function of the Beta distribution which is related to the cumulative density function of a random variable y that follows a Binomial distribution with probability of single success p and number of Bernoulli trails n.

$$F\left(k;n,p\right) = P\left(Y \le k\right) = I_{1-p}\left(n-k,k+1\right) = 1 - I_p\left(k+1,n-k\right)$$

where

$$I_0\left(\alpha,\beta\right) = 0$$

$$I_1\left(\alpha,\beta\right) = 1$$

$$I_y\left(\alpha,1\right) = y^a$$

$$I_y\left(1,\beta\right) = 1 - \left(1-y\right)^{\beta}$$

$$\Rightarrow I_y\left(\alpha,\beta\right) = 1 - I_{1-y}\left(\beta,\alpha\right)$$

$$I_y\left(\alpha+1,\beta\right) = I_y\left(\alpha,\beta\right) - \frac{y^{\alpha}\left(1-y\right)^{\beta}}{\alpha B\left(\alpha,\beta\right)}$$

$$I_y\left(\alpha,\beta+1\right) = I_y\left(\alpha,\beta\right) + \frac{y^{\alpha}\left(1-y\right)^{\beta}}{\beta B\left(\alpha,\beta\right)}$$

$$\int B\left(y;\alpha,\beta\right)\mathrm{d}y = yB\left(y;\alpha,\beta\right) - B\left(y;\alpha+1,\beta\right)$$

$$B\left(y;\alpha,\beta\right) = \left(-1\right)^{\alpha} B\left(\frac{y}{y-1};\alpha,1-\alpha-\beta\right)$$

This implies

$$\mathrm{CDF} = I_y(\alpha,\beta) = \frac{B\left(y;\alpha,\beta\right)}{B\left(\alpha,\beta\right)} = \frac{\left(-1\right)^{\alpha} B\left(\frac{y}{y-1};\alpha,1-\alpha-\beta\right)}{B\left(\alpha,\beta\right)} \tag{3}$$

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

$$E(X)=\int_c^d xf(x;\alpha,\beta,c,d)\,\mathrm{d}x=\int_c^d x\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}\mathrm{d}x \tag{4}$$

Employing the technique used for equation 2 where $x=y(d-c)+c$, we have;

$$E(X)=\int_0^1\big(y(d-c)+c\big)\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y$$

$$E(X)=\int_0^1 y(d-c)\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y+\int_0^1 c\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y$$

$$E(X)=\int_0^1(d-c)\frac{y^{\alpha+1-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y+\int_0^1 c\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y$$

But $B(\alpha,\beta)=\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,\mathrm{d}y=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\int_0^\infty y^{\alpha+1-1}(1-y)^{\beta-1}\,\mathrm{d}y=\dfrac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$.

This implies:

$$E(X)=(d-c)\frac{\dfrac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}+c\frac{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \tag{5}$$

$$E(X)=(d-c)\frac{\dfrac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}+c$$

$$E(X)=c+\frac{(d-c)\alpha}{\alpha+\beta} \tag{6}$$

As well, for the variance of the Beta4 distribution, we have;

$$VAR(X)=E(X^2)-[E(X)]^2$$

$$\therefore E(X^2)=\int_c^d x^2 f(x;\alpha,\beta,c,d)\,\mathrm{d}x=\int_c^d x^2\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}\mathrm{d}x \tag{7}$$

Employing the technique used for equation 2 where $x=y(d-c)+c$, we have;

$$E(X^2)=\int_0^1\big(y\{d-c\}+c\big)^2\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y \tag{8}$$

$$E(X^2)=\int_0^1\left\{y^2(d-c)^2+2c(d-c)y+c^2\right\}\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y$$

$$E\left(X^2\right)=\int_0^1 y^2\left(d-c\right)^2\frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y+\int_0^1 2c\left(d-c\right)y\frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y$$
$$+\int_0^1 c^2\frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y$$

But $\int_0^1 y\frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y=\int_0^1\frac{y^{\alpha+1-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y=\frac{\alpha}{\alpha+\beta}$ and $\int_0^1\frac{y^{\alpha-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y=1$

$$\therefore E\left(X^2\right)=\int_0^1\left(d-c\right)^2\frac{y^{\alpha+2-1}\left(1-y\right)^{\beta-1}}{B\left(\alpha,\beta\right)}\mathrm{d}y+2c\left(d-c\right)\frac{\alpha}{\alpha+\beta}+c^2$$

But $\int_0^1 y^{\alpha+2-1}\left(1-y\right)^{\beta-1}\mathrm{d}y=\frac{\Gamma\left(\alpha+2\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta+2\right)}=\frac{\alpha\left(\alpha+1\right)\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)\Gamma\left(\alpha+\beta\right)}$

$$\therefore E\left(X^2\right)=\left(d-c\right)^2\frac{\dfrac{\alpha\left(\alpha+1\right)\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)\Gamma\left(\alpha+\beta\right)}}{\dfrac{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta\right)}}+2c\left(d-c\right)\frac{\alpha}{\alpha+\beta}+c^2$$

$$\therefore E\left(X^2\right)=\left(d-c\right)^2\frac{\dfrac{\alpha\left(\alpha+1\right)\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)\Gamma\left(\alpha+\beta\right)}}{\dfrac{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta\right)}}+2c\left(d-c\right)\frac{\alpha}{\alpha+\beta}+c^2$$

$$\therefore E\left(X^2\right)=\left(d-c\right)^2\frac{\alpha\left(\alpha+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)}+2c\left(d-c\right)\frac{\alpha}{\alpha+\beta}+c^2 \tag{9}$$

But $VAR\left(X\right)=E\left(X^2\right)-\left[E\left(X\right)\right]^2$

$$\therefore VAR\left(X\right)=\left(d-c\right)^2\frac{\alpha\left(\alpha+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)}+2c\left(d-c\right)\frac{\alpha}{\alpha+\beta}+c^2$$
$$-\left[c+\frac{\left(d-c\right)\alpha}{\alpha+\beta}\right]^2$$

$$\therefore VAR\left(X\right)=\frac{\left(d-c\right)^2\alpha\left(\alpha+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)}+\frac{2c\left(d-c\right)\alpha}{\alpha+\beta}+c^2$$
$$-\left[c^2+\frac{\left(d-c\right)^2\alpha^2}{\left(\alpha+\beta\right)^2}+\frac{2c\left(d-c\right)\alpha}{\alpha+\beta}\right]$$

$$\therefore VAR\left(X\right)=\frac{\left(d-c\right)^2\alpha\left(\alpha+1\right)}{\left(\alpha+\beta\right)\left(\alpha+\beta+1\right)}-\frac{\left(d-c\right)^2\alpha^2}{\left(\alpha+\beta\right)^2}$$

$$\therefore VAR\left(X\right)=\frac{\left(\alpha+\beta\right)\left(d-c\right)^2\alpha\left(\alpha+1\right)-\left(\alpha+\beta+1\right)\left(d-c\right)^2\alpha^2}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)}$$

$$\therefore VAR\left(X\right)=\frac{\left(d-c\right)^2\left[\left(\alpha+\beta\right)\alpha\left(\alpha+1\right)-\left(\alpha+\beta+1\right)\alpha^2\right]}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)}$$

$$VAR(X)=\frac{(d-c)^2\left[\alpha^3+\alpha^2+\alpha^2\beta+\alpha\beta-\alpha^3-\alpha^2-\alpha^2\beta\right]}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\therefore VAR(X)=\frac{(d-c)^2\,\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\therefore VAR(X)=\frac{(d-c)^2\,\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{10}$$

Moreover, the Moment Generation Function (MGF) of the Beta4 distribution is given as;

$$E(X^n)=\int_c^d x^n f(x;\alpha,\beta,c,d)\,\mathrm{d}x=\int_c^d x^n\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}\mathrm{d}x \tag{11}$$

Employing the technique used for Equation (2) where $x=y(d-c)+c$, we have;

$$E(X^n)=\int_0^1\left[(y\{d-c\}+c)^n\right]\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y \tag{12}$$

$$\text{But }\ (a+b)^n=\sum_{r=0}^{n}\binom{n}{r}a^r b^{n-r},\ r=0,1,2,\cdots,n \tag{13}$$

$$\therefore E(X^n)=\int_0^1\sum_{r=0}^{n}\binom{n}{r}(y\{d-c\})^r c^{n-r}\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y \tag{14}$$

This implies;

$$E(X^n)=\int_0^1\sum_{r=0}^{n}\binom{n}{r}(d-c)^r c^{n-r}\frac{y^{\alpha+r-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}y$$

$$\text{But }\ \int_0^1 y^{\alpha+r-1}(1-y)^{\beta-1}\mathrm{d}y=\frac{\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha+\beta+r)}$$

$$\therefore E(X^n)=\sum_{r=0}^{n}\binom{n}{r}(d-c)^r c^{n-r}\frac{\dfrac{\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha+\beta+r)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$$

$$\therefore \mathrm{MGF}=E(X^n)=\sum_{r=0}^{n}\binom{n}{r}(d-c)^r c^{n-r}\frac{\Gamma(\alpha+r)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)} \tag{15}$$

The Maximum Likelihood Estimation (MLE) of the Beta4 distribution is also given as;

$$L(x_i;\alpha,\beta,c,d)=\prod_{i=1}^{n}f(x_i;\alpha,\beta,c,d) \tag{16}$$

$$L(x_i;\alpha,\beta,c,d)=\prod_{i=1}^{n}\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}} \tag{17}$$

Taking antilog of both sides, we have;

$$\ln L\left(x_i;\alpha,\beta,c,d\right)=\sum_{i=1}^{n}\ln\left[\frac{1}{B(\alpha,\beta)}\frac{(x-c)^{\alpha-1}(d-x)^{\beta-1}}{(d-c)^{\alpha+\beta-1}}\right]$$

$$\ln L\left(x_i;\alpha,\beta,c,d\right)=\sum_{i=1}^{n}\left\{\ln\left((x-c)^{\alpha-1}(d-x)^{\beta-1}\right)-\ln\left(B(\alpha,\beta)(d-c)^{\alpha+\beta-1}\right)\right\}$$

Therefore, the log likelihood function is given as;

$$\ln L\left(x_i;\alpha,\beta,c,d\right)=\sum_{i=1}^{n}\left\{\ln(x-c)^{\alpha-1}+\ln(d-x)^{\beta-1}-\ln B(\alpha,\beta)-\ln(d-c)^{\alpha+\beta-1}\right\}$$

$$\begin{aligned}\ln L\left(x_i;\alpha,\beta,c,d\right)=&-\ln B(\alpha,\beta)+\left[(\alpha-1)\ln(x-c)+(\beta-1)\ln(d-x)\right]\\&-(\alpha+\beta-1)\ln(d-c)\end{aligned}\tag{18}$$

Differentiating equation 18 with respect to α,β,c,d as in equations 19, 20, 21, 22 respectively,

$$\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}\alpha}=\mathrm{d}\frac{-\ln B(\alpha,\beta)+(\alpha-1)\ln(x-c)-(\alpha+\beta-1)\ln(d-c)}{\mathrm{d}\alpha}\tag{19}$$

$$\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}\beta}=\mathrm{d}\frac{-\ln B(\alpha,\beta)+(\beta-1)\ln(d-x)-(\alpha+\beta-1)\ln(d-c)}{\mathrm{d}\beta}\tag{20}$$

$$\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}c}=\mathrm{d}\frac{(\alpha-1)\ln(x-c)-(\alpha+\beta-1)\ln(d-c)}{\mathrm{d}c}\tag{21}$$

$$\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}d}=\mathrm{d}\frac{(\beta-1)\ln(d-x)-(\alpha+\beta-1)\ln(d-c)}{\mathrm{d}d}\tag{22}$$

Therefore, equating Equations (19)-(22) to zero and solving them simultaneously using numerical methods, the Maximum Likelihood Estimates of α,β,c,d are produced.

Also, the Cramer-Rao lower bound inequality attainsed for each of the estimated parameters for α,β,c,d are as follows in Equations (23)-(26) respectively;

$$\mathit{Variance}\left(\hat{\alpha}(x)\right)\geq\frac{1}{nI(\alpha)}\text{ or }\frac{1}{I(\alpha)}=\frac{1}{E\left[\left[\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}\alpha}\right]^2\right]}\tag{23}$$

$$\mathit{Variance}\left(\hat{\beta}(x)\right)\geq\frac{1}{nI(\beta)}\text{ or }\frac{1}{I(\beta)}=\frac{1}{E\left[\left[\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}\beta}\right]^2\right]}\tag{24}$$

$$\mathit{Variance}\left(\hat{c}(x)\right)\geq\frac{1}{nI(c)}\text{ or }\frac{1}{I(c)}=\frac{1}{E\left[\left[\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}c}\right]^2\right]}\tag{25}$$

$$\mathit{Variance}\left(\hat{d}(x)\right)\geq\frac{1}{nI(d)}\text{ or }\frac{1}{I(d)}=\frac{1}{E\left[\left[\frac{\mathrm{d}\ln L\left(x_i;\alpha,\beta,c,d\right)}{\mathrm{d}d}\right]^2\right]}\tag{26}$$

3.3. Forecasting Future Values with the Inverse CDF for the Beta4 Distribution

The inverse cumulative distribution function (CDF), also known as the quantile function or percent-point function, for the Beta4 distribution, is derived as follows;

$$\mathrm{CDF}=\int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}t = I_y(\alpha,\beta)=\frac{B(y;\alpha,\beta)}{B(\alpha,\beta)}=\frac{(-1)^{\alpha}B\left(\frac{y}{y-1};\alpha,1-\alpha-\beta\right)}{B(\alpha,\beta)}$$

If p is the result of the probability that a single observation from the Beta4 distribution with parameters α,β,c,d in the interval [0 *y*], then;

$$p=\int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}t$$

Integrating and making x the subject, we have;

$$y=[\mathrm{CDF}]^{-1}$$

This therefore means that the result of the value of y is an observation from the Beta4 distribution with parameters α,β,c,d that falls in the range [0 *y*] with probability *p*.

4. Results and Discussions

4.1. Evaluation of Statistical Distribution Functions in Modeling Gold Reserve

Figure 1 represents the scatterplot of the quarterly gold reserve between the first quarter of 2009 to the first quarter of 2022 in Ghana.

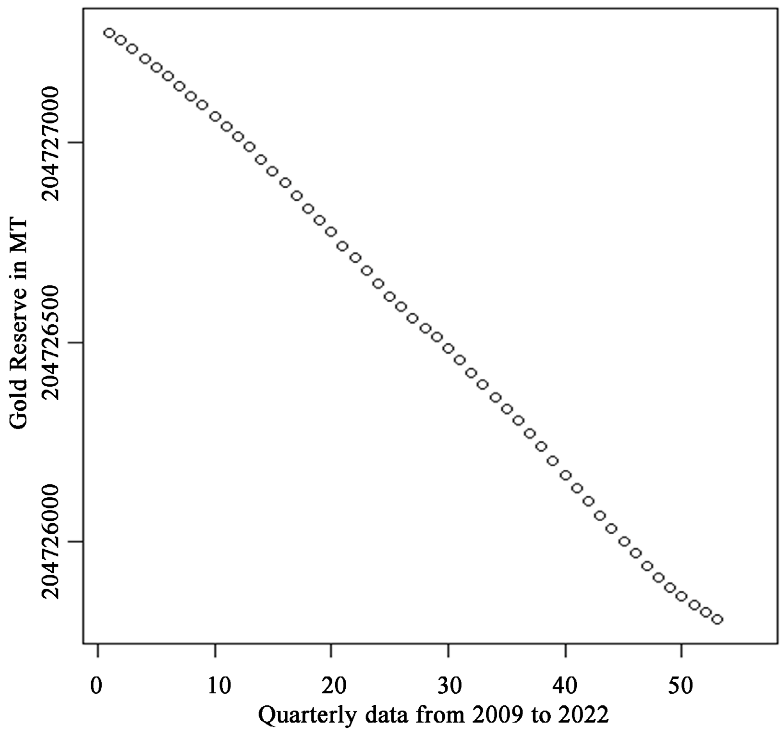
The minimum gold reserve (that is, 204,725,804.24 MT) was recorded in the first quarter of 2022 and the highest recorded in the first quarter of 2009. This is so because it is presumed that as more and more gold is being mined, the reserve depletes to a state of diminishing returns in terms of profitability [12] [13] [14] (Table 2).

The results in Table 3 indicate that the data is normally distributed [15]. This is presented graphically in Figures 2-4.

4.2. Application of the Four-Parameter Beta Distribution on Gold Reserve

The theoretical plot comparison of the major distributions identified is shown in Figure 4. It is clear from Figure 4 that the four-parameter Beta distribution fits the gold reserve data better.

Table 4 presents the estimated parameters of the fitted distributions to the quarterly gold reserve data. The p-values of all the parameters show that they are 99.9% significant to be part of the fitted models or distributions. Obviously, the fitted standard deviation (*i.e.*, 446.2630657) of the Beta4 distribution also showed that it fitted the gold reserve data better than the other distributions



Source of data: Minerals Commission of Ghana [11].

Figure 1. Scatterplot of quarterly gold reserve in metric tons (2009-2022).

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

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Reserve (MT)	53	204,725,804.2400	204,727,277.4300	204,726,555.9917	457.2600

Source of data: Minerals Commission of Ghana [11].

Table 3. Normality test on the gold production data.

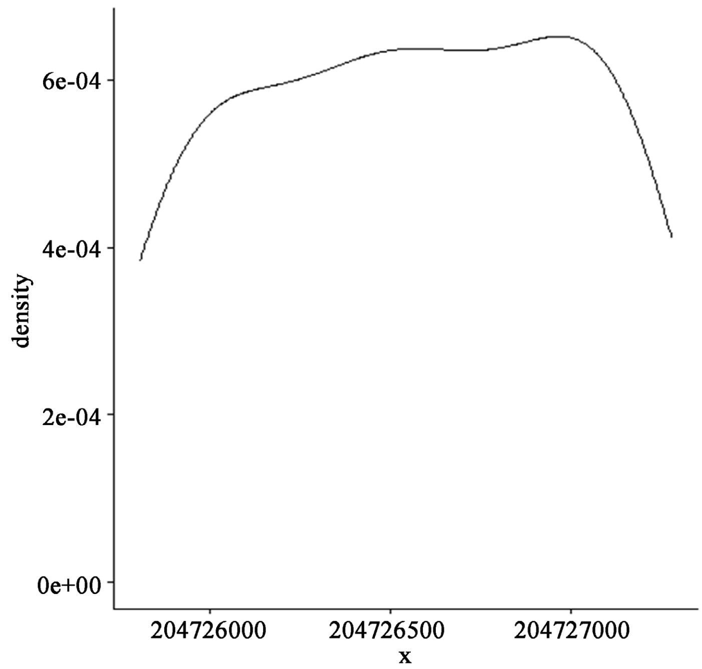
Variable	Lilliefors	Jarque-Bera
(Reserve_MT)	0.6301	0.1742

Source: Estimation from the gold production data.

since it is the least among them. Also, the average quarterly fitted gold reserve is 20,4726,550.71 ± 201.706 (±0.00010%) [*i.e.*, 204,726,349.004 - 204,726,752.416 metric tons]. Therefore, with the current annual average gold production estimate that lies between 3,700,031.248 and 4,302,647.888 ounces [16], it is estimated that Ghana may still produce gold even after 1,951,743 years to come (*i.e.*, the year 1,953,765) at 99.9% confidence level.

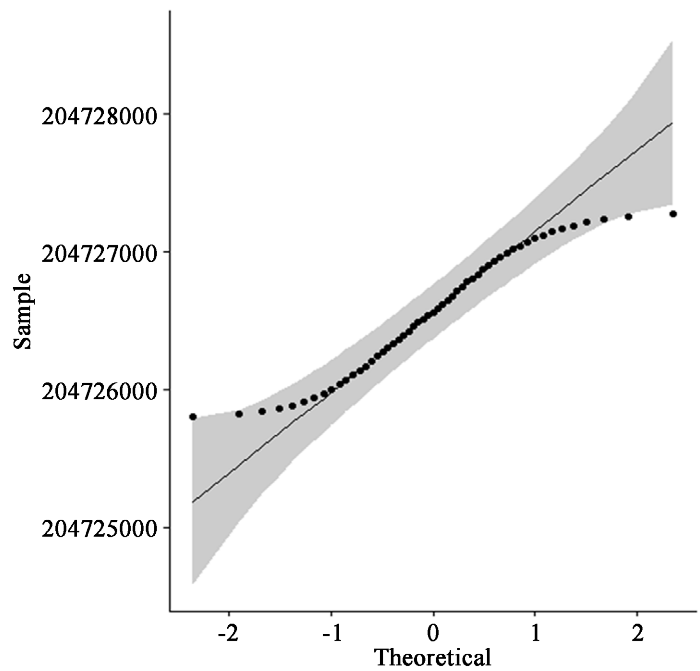
4.3. Application of the Reduced Modified Parameter Distribution on Gold Production

Table 5 represents the goodness of fit and the selection criteria for the gold reserve data. The results in **Table 5** show clearly that the four-parameter Beta



Source: Authors own estimation, 2023.

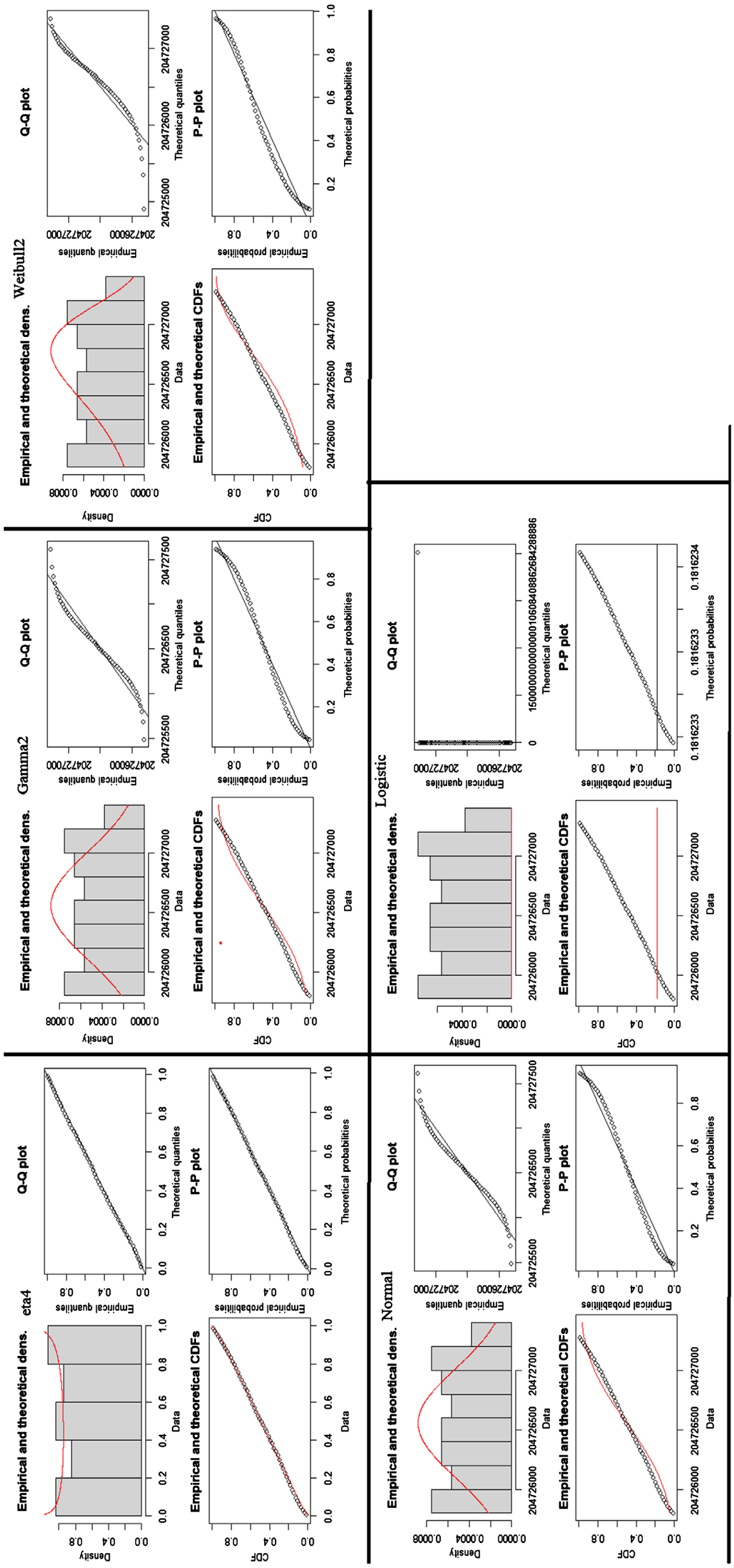
Figure 2. Normality test on the gold reserve data.



Source: Authors own estimation, 2023.

Figure 3. QQPLOT of gold reserve (2009-2022).

distribution (Beta4) performed better than the other distribution in terms of all the goodness of fit measures as well as with the selection criteria. The study revealed that the risk of rejecting Beta4 based on the Kolmogorov-Smirnoff test is 99.9999999999967%. Also, with the AIC and BIC values of 3.67433 and 7.614913



Source: Authors own estimation, 2023.

Figure 4. Theoretical plot of the distributions.

Table 4. Estimated parameters of the fitted distributions of the gold reserve.

Distribution	Parameters				Fitted Mean	Fitted Std. deviation
Weibull (2)	$k = 510608.6$	$\lambda = 204726779.9$			204,726,548.4	514.270394
Standard Error	54,933.6	58.3				
P-value	0.0000					
Gamma (2)	$k = 200{,}457{,}286{,}366.18$		$\beta = 0.001$		204,726,555.99	457. 457.26
Standard Error	0.0000		0.0003			
P-value	0.0000		0.0000			
Normal	$\mu = 204{,}726{,}555.99$		$\sigma = 457.26$		204,726,555.99	457. 457.26
Standard Error	62.21298		43.99183			
P-value	0.0000		0.0000			
Logistic	$\mu = 204{,}726{,}560.463923$		$s = 276.124914667133$		204,726,560.46	500.8351947
Standard Error	67.9038		30.52			
P-value	0.0000		0.0000			
Beta (4)	$\alpha = 0.929420$	$\beta = 0.9051$	$c = 204{,}725{,}789.4$	$d = 204{,}727{,}292.2$	204,726,550.71	446.2630657
Standard Error	0.1667796	0.1615	0.0000	0.0000		
P-value	0.0000	0.0000	0.0000	0.0000		

Source: Authors own estimation, 2023.

Table 5. Goodness of Fit and selection criteria for gold reserve.

Distribution	KS (P-value)	AD (P-value)	CM (P-value)	−2 log L	AIC	BIC
Weibull (2)	0.83252004	0.83288420	0.11722810	800.5284	804.5284	808.4690
Gamma (2)	0.9108644	0.74161380	0.09657750	798.6747	802.6747	806.6153
Normal	0.910865	0.74161321	0.09657741	798.6747	802.6747	806.6153
Logistic	0.918000	0.0313963	0.7889417	2493.653	2497.653	2501.593
Beta (4)	0.999999999999967	0.9999999	0.9999999	-0.3256704	3.67433	7.614913

Source: Authors own estimation, 2023.

respectively, Beta4 is adjudged as the best probability distribution in fitting the gold reserve quarterly data with the minimal error. This finding is therefore more accurate than those done by Kaba *et al.* [17] and Appiah *et al.* [18]. Findings of Kaba *et al.* [17], showed that, with a Beta (P-value of 0.75) distribution, the total average mining production fell within 210,414.86±3,301.59 in Bank Cubic Meters at 95% confidence level while that of Appiah *et al.* [18] 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 335866.94. Meanwhile, the proposed current model produced a Kolmogorov-Smirnov (K-S) of 0.999999999999967 which is the best as compared to those of Kaba *et al.* [17], Appiah *et al.* [18] and other research findings.

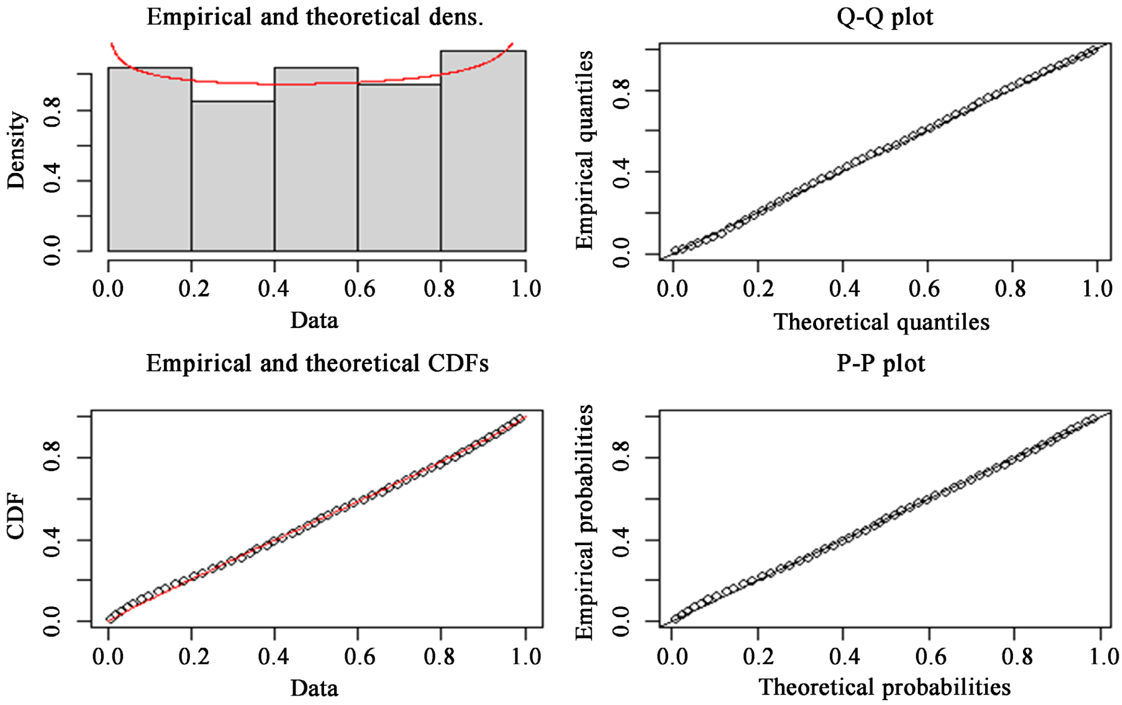
In **Table 6**, we can see that the actual gold reserve figures almost fall within the estimated intervals with the four-parameter Beta distribution.

Figure 5 represents the empirical plot of the four-parameter Beta distribution

Table 6. Comparison between the estimated intervals and the actual reserve values.

Quarter	Actual Reserve (MT)	Lower bound	Upper bound	Relative frequency	Density (Distribution)
1q-2015	204,726,615.1	204,726,380	204,726,610	0.1509	0.0007
2q-2015	204,726,587.6	204,726,380	204,726,610	0.1509	0.0007
3q-2015	204,726,559.1	204,726,380	204,726,610	0.1509	0.0007
4q-2015	204,726,533.9	204,726,380	204,726,610	0.1509	0.0007
1q-2016	204,726,512.4	204,726,610	204,726,840	0.1509	0.0007
2q-2016	204,726,484.6	204,726,610	204,726,840	0.1509	0.0007
3q-2016	204,726,455	204,726,610	204,726,840	0.1509	0.0007
4q-2016	204,726,421.3	204,726,610	204,726,840	0.1509	0.0007
1q-2017	204,726,392.5	204,726,610	204,726,840	0.1509	0.0007
2q-2017	204,726,362.7	204,726,610	204,726,840	0.1509	0.0007
3q-2017	204,726,331.4	204,726,610	204,726,840	0.1509	0.0007
4q-2017	204,726,302.8	204,726,610	204,726,840	0.1509	0.0007
1q-2018	204,726,272.7	204,726,840	204,727,070	0.1509	0.0007
2q-2018	204,726,238.7	204,726,840	204,727,070	0.1509	0.0007
3q-2018	204,726,203.2	204,726,840	204,727,070	0.1509	0.0007
4q-2018	204,726,167.2	204,726,840	204,727,070	0.1509	0.0007
1q-2019	204,726,132.7	204,726,840	204,727,070	0.1509	0.0007
2q-2019	204,726,101.4	204,726,840	204,727,070	0.1509	0.0007
3q-2019	204,726,065.5	204,726,840	204,727,070	0.1509	0.0007
4q-2019	204,726,032.7	204,726,840	204,727,070	0.1509	0.0007
1q-2020	204,726,000.5	204,727,070	204,727,300	0.1698	0.0007
2q-2020	204,725,972.5	204,727,070	204,727,300	0.1698	0.0007
3q-2020	204,725,937.8	204,727,070	204,727,300	0.1698	0.0007
4q-2020	204,725,911.5	204,727,070	204,727,300	0.1698	0.0007
1q-2021	204,725,884.4	204,727,070	204,727,300	0.1698	0.0007
2q-2021	204,725,862.9	204,727,070	204,727,300	0.1698	0.0007
3q-2021	204,725,842.7	204,727,070	204,727,300	0.1698	0.0007
4q-2021	204,725,824	204,727,070	204,727,300	0.1698	0.0007
1q-2022	204,725,804.2	204,727,070	204,727,300	0.1698	0.0007

Source: Authors own estimation, 2023.



Source: Authors own estimation, 2023.

Figure 5. Empirical plot of the RM3PWD on gold production data.

on the quarterly gold reserve data from the first quarter of 2009 to the first quarter of 2022. It is obvious from this plot that Beta4 model fitted the gold reserve data appropriately. It can therefore be concluded that Beta4 distribution is better fit to the quarterly gold reserve data.

4.4. Sensitivity Analysis

Performing a sensitivity analysis on the confidence level will help assess the robustness of the projected lifespan of gold production in Ghana. The goal is to understand how variations in the confidence level impact the projection and to identify the range of possible outcomes. Let us conduct the sensitivity analysis by varying the confidence level and observing its effects on the projected lifespan.

Given:

- Current annual average gold production estimate: 3,700,031.248 to 4,302,647.888 ounces;
- Projected lifespan of gold production at a 99.9% confidence level: 1,953,765 years.

Now, let us calculate the projected lifespan at different confidence levels:

1) 99.5% Confidence Level: At a 99.5% confidence level, the risk of rejecting Beta4 is $1 - 0.995 = 0.005$.

Projected Lifespan = $1/(1 - 0.005) * 1{,}953{,}765$ Projected Lifespan ≈ 1,955,757 years

2) 99.0% Confidence Level: At a 99.0% confidence level, the risk of rejecting Beta4 is $1 - 0.990 = 0.010$.

Projected Lifespan = 1/(1 − 0.010) * 1,953,765 Projected Lifespan ≈ 1,975,648 years

3) 95.0% Confidence Level: At a 95.0% confidence level, the risk of rejecting Beta4 is 1 − 0.950 = 0.050.

Projected Lifespan = 1/(1 − 0.050) * 1,953,765 Projected Lifespan ≈ 3,907,529 years

4) 90.0% Confidence Level: At a 90.0% confidence level, the risk of rejecting Beta4 is 1 − 0.900 = 0.100.

Projected Lifespan = 1/(1 − 0.100) * 1,953,765 Projected Lifespan ≈ 2,170,851 years

By performing the sensitivity analysis, we can observe how the projected lifespan of gold production changes with varying confidence levels. The results show that as the confidence level decreases, the projected lifespan tends to increase. This is because a lower confidence level allows for a wider range of possible outcomes, leading to a longer projected lifespan. Conversely, a higher confidence level results in a shorter projected lifespan, as it reflects a more conservative and narrow range of estimates.

It is essential to consider the implications of these variations in the confidence level while making decisions based on the projected lifespan of gold production. Stakeholders should carefully assess the level of risk they are willing to accept and how it may impact their long-term planning and investment decisions in the mining industry.

5. Conclusions

In conclusion, this study conducted a rigorous evaluation of various distribution functions to forecast gold reserves and estimate their production lifespan in Ghana between 2009 and 2022. The results from the distribution fitting analysis using the XLSTAT software indicated that several distributions, including the Beta4 (four-parameter Beta), Logistic, two-parameter Weibull, Normal, and two-parameter Gamma distributions, exhibited a good fit to the gold reserve data, with their Kolmogorov-Smirnoff test p-values greater than 0.8.

Among the tested distributions, the four-parameter Beta distribution emerged as the best-fitting model, with an exceptional Kolmogorov-Smirnoff test statistic of 0.999999999999967, indicating an excellent fit to the gold reserve data. This finding underscores the suitability of the four-parameter Beta distribution for accurately modeling non-negative data, such as gold reserves, and highlights its potential as a reliable forecasting tool in the context of natural resource management.

Based on the average quarterly fitted gold reserve of 204,726,550.71 ± 201.706 metric tons (or a range of 204,726,349.004 to 204,726,752.416 metric tons), the study estimates that Ghana may continue to produce gold for an astonishingly long duration. At a 99.9% confidence level and considering the current annual average gold production estimate of 3,700,031.248 to 4,302,647.888 ounces, the projected lifespan of gold production in Ghana extends to the year 1,953,765.

This astounding estimate suggests that the country's gold reserves are expected to sustain production for an extended period, providing a critical resource for economic development and supporting the mining industry well into the distant future.

This study's findings have substantial implications for Ghana's mining industry, policymakers, and stakeholders. The utilization of the four-parameter Beta distribution for gold reserve forecasting and production lifespan estimation provides valuable insights into sustainable resource management strategies. By understanding the vast potential of gold reserves and their long-lasting production capacity, decision-makers can formulate well-informed plans to harness this natural resource responsibly, promoting economic stability and environmental conservation in Ghana.

However, it is essential to acknowledge the inherent uncertainties in forecasting models and the potential impact of changing factors over time. Continuous monitoring and periodic reassessment of the forecasting model will be necessary to adapt to evolving conditions and ensure the accuracy of long-term projections.

In summary, the adoption of the four-parameter Beta distribution for forecasting gold reserves and production lifespan in Ghana offers a robust approach to resource management, positioning the country to leverage its abundant gold reserves sustainably and responsibly for generations to come.

Implications and significance for investors

1) Long-term Investment Opportunities: The study's findings, indicating an extensive production lifespan for Ghana's gold reserves, present significant long-term investment opportunities for investors in the mining sector. Investing in gold mining companies or exploration projects in Ghana can offer investors the potential for stable returns over an extended period.

2) Stability and Predictability: The utilization of the four-parameter Beta distribution to forecast gold reserves enhances the stability and predictability of investment decisions. With a reliable forecasting model, investors can have more confidence in their investment strategies, knowing that Ghana's gold production is expected to remain viable for a considerable period.

3) Risk Mitigation: The long production lifespan estimated for Ghana's gold reserves reduces the risk associated with short-term fluctuations in gold prices or mining operations. Investors can better navigate market volatility and economic uncertainties, knowing that the resource base is expected to be available for a prolonged duration.

4) Sustainable Mining Investments: Investors seeking to align their portfolios with sustainable and socially responsible practices can find opportunities in the Ghanaian gold mining sector. With a clear understanding of the long-term availability of gold reserves, investors can support mining companies that prioritize responsible mining practices and environmental stewardship.

5) Diversification Benefits: Investing in the gold mining sector in Ghana can provide diversification benefits for investors looking to spread their risks across

different asset classes and geographical regions. The long production lifespan offers stability and diversification potential to complement other investment holdings.

6) Informed Decision-making: The research findings equip investors with valuable information to make informed decisions regarding resource allocation and investment strategies. Understanding the estimated production lifespan of gold reserves in Ghana allows investors to tailor their investment time horizons and risk profiles accordingly.

7) Attracting Foreign Direct Investment (FDI): The evidence of a sustainable and extensive gold production lifespan in Ghana can attract foreign investors seeking stable and lucrative opportunities in the mining sector. Increased FDI can contribute to the development of the mining industry and the broader economy of the country.

8) Government Policies and Regulations: The long-term forecast of gold reserves can influence government policies and regulations related to the mining sector. Authorities may prioritize sustainable mining practices and implement policies that encourage responsible resource extraction, attracting further investments from socially conscious investors.

9) Infrastructure Development: With the assurance of prolonged gold production, investors can be more willing to finance infrastructure projects and logistics improvements necessary for efficient mining operations. This could lead to increased efficiency, reduced operational costs, and higher returns on investments.

10) Market Positioning: Companies operating in Ghana's gold mining sector can leverage the long production lifespan as a unique selling point to attract investors and secure funding for expansion projects. The longevity of the gold reserves enhances the attractiveness of the mining projects and can improve market positioning.

In conclusion, the research's implications and significance for investors are profound, providing valuable insights into the long-term investment potential in Ghana's gold mining sector. The estimated extensive production lifespan offers stability, diversification opportunities, and risk mitigation, making it an attractive destination for investors seeking sustainable and lucrative investments in the precious metals market.

Inherent uncertainties in forecasting for this model

1) Data Limitations: Forecasting models heavily depend on historical data to make predictions. In the case of gold reserves and production lifespan, the available historical data might be limited or subject to data gaps, errors, or inconsistencies. Such limitations can introduce uncertainty into the forecasting process.

2) Geological Variability: Geological factors play a crucial role in gold deposit formation, and these factors can be highly variable across different regions within Ghana. The model's ability to accurately capture the diverse geological conditions and their impact on gold reserves might be limited, leading to uncertainties in the forecasts.

3) Extraction Technology Advancements: Technological advancements in the mining industry can significantly influence gold extraction rates and the feasibility of accessing previously uneconomical reserves. Forecasting models may struggle to account for these future technological developments, leading to uncertainties in production lifespan estimates.

4) Price Fluctuations: Gold prices are subject to fluctuations influenced by global economic conditions, geopolitical events, and investor sentiment. Forecasting models might not fully capture these external market forces, leading to uncertainties in predicting future gold prices and their impact on reserve estimates.

5) Regulatory Changes: Changes in mining regulations and policies can impact the feasibility and profitability of mining operations in Ghana. Forecasting models might not account for potential shifts in the regulatory landscape, introducing uncertainties in estimating production lifespans.

6) Economic and Political Factors: The overall economic conditions and political stability of Ghana can affect mining activities and investments in the sector. Uncertainties surrounding economic and political developments can influence the long-term viability of gold production in the country.

7) Environmental Considerations: Increasing awareness of environmental sustainability might lead to stricter regulations and higher operational costs for mining companies. The model's inability to fully incorporate these future environmental considerations can introduce uncertainties into production lifespan estimates.

8) Exploration and Discovery: The discovery of new gold deposits or advancements in exploration technologies can impact reserve estimates. Forecasting models might not anticipate such discoveries, leading to uncertainties in forecasting gold reserves.

9) Market Demand: Future fluctuations in global gold demand and its correlation with production levels can introduce uncertainties in gold reserve forecasts. Unanticipated changes in demand might not be fully accounted for in the model.

10) Extraction Rate Variability: The extraction rate of gold reserves can be influenced by operational inefficiencies, labor issues, and unforeseen technical challenges. Variability in extraction rates might not be fully captured in the forecasting model, leading to uncertainties in production lifespan estimates.

In summary, forecasting gold reserves and production lifespan involves inherent uncertainties due to various factors, including data limitations, geological variability, technological advancements, market dynamics, regulatory changes, and environmental considerations. These uncertainties should be carefully considered when interpreting and utilizing the forecasting model's results, and periodic reassessment of the model is essential to account for evolving conditions and improve the accuracy of the forecasts.

6. Limitations of the Study

While the research on forecasting gold reserves and estimating their production lifespan in Ghana appears promising and valuable, it is important to recognize some of the limitations that may affect the study's findings and conclusions:

1) Data Limitations: The accuracy and reliability of any reserve estimation heavily depend on the quality and quantity of available data. Inadequate or incomplete data on geological information, drilling results, and historical production may introduce uncertainties in the reserve estimates. The use of insufficient or outdated data can impact the precision of the model's predictions.

2) Assumptions in the Model: Like any mathematical model, the four-parameter Beta distribution function used for reserve estimation relies on certain assumptions. These assumptions might include the stationary nature of the deposit's geology and gold grade distribution, which may not fully capture potential changes in geological conditions over time. The model's accuracy may be affected if these assumptions do not hold true for the specific gold deposit under consideration.

3) Geological Complexity: Gold deposits can exhibit complex and heterogeneous geological structures. The simplicity of the model used might not fully account for such complexities, leading to potential inaccuracies in reserve estimations. Variations in gold grades, mineralogy, and deposit geometry may not be adequately captured by the chosen modeling approach.

4) External Factors: Gold reserves and production can be influenced by external factors beyond the geological considerations, such as changes in gold prices, government policies, environmental regulations, and geopolitical events. The model might not account for the impacts of these external factors, leading to deviations between predicted and actual outcomes.

5) Uncertainty in Economic Conditions: The study might not fully consider the impact of fluctuating economic conditions on gold production. Economic factors, such as inflation rates, exchange rates, and market demand, can affect the feasibility and profitability of mining projects, thereby influencing production decisions and reserve estimates.

6) Technological Advancements: The study's estimation of production lifespan assumes current mining technologies and practices will continue throughout the lifespan of the mine. However, technological advancements in mining techniques could extend the mine's life or improve production efficiency, leading to potential deviations from the estimated lifespan.

7) Environmental and Social Considerations: The study might not extensively address environmental and social factors, such as environmental impacts, community relations, and potential conflicts. These factors can significantly affect mining operations and could lead to changes in mining plans and production rates.

8) Exploration Potential: The study's focus on existing reserves may not fully consider the potential for discovering new gold deposits through further exploration efforts. New discoveries could alter the reserve estimates and production lifespan projections.

To enhance the robustness of the study's findings, researchers and stakeholders in the mining industry should acknowledge and address these limitations. Regular updates to the data used in the model, sensitivity analysis for key assumptions, and considering the impact of external factors will contribute to more reliable and comprehensive reserve estimates and production lifespan predictions. Additionally, incorporating environmental, social, and economic factors into the analysis will offer a more holistic view of the mining project's feasibility and sustainability.

Acknowledgements

Special appreciation goes to Mr. Martin Kwaku Ayisi, CEO, Mr. Francis Frimpong, head of Research and Statistics, and Mr. Michael Kotoka, head of Monitoring and Evaluation department, all of Minerals Commission 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] Kofi, A.S., Loveridge, A.D.S., Obed, A., Oduro, N.B. and Yah, A.D.T.R. (2023) The Contributions of Ghana's Minerals Resources to the Global Economy: Using Bauxite as a Case Study.

[2] Aycaya-Paco, Y.B., Vilca-Mamani, L.D. and Torres-Cruz, F. (2023) Peru Mining: Analysis and Forecast of Mining Production in Peru Using Time Series and Data Science Techniques.

[3] Wilson, R., Mercier, P.H. and Navarra, A. (2022) Integrated Artificial Neural Network and Discrete Event Simulation Framework for Regional Development of Refractory Gold Systems. *Mining*, **2**, 123-154. https://doi.org/10.3390/mining2010008

[4] Perera, S.C. and Sethi, S.P. (2023) A Survey of Stochastic Inventory Models with Fixed Costs: Optimality of (s, S) and (s, S)-Type Policies—Discrete-Time Case. *Production and Operations Management*, **32**, 131-153. https://doi.org/10.1111/poms.13820

[5] Kim, S. and Lee, S. (2021) Maximum Likelihood Estimation of Probabilistic Non-Suppression Model for OECD NPP Electrical Fire Applying Non-Negative Continuous Distribution. *Fire Safety Journal*, **122**, Article ID: 103323. https://doi.org/10.1016/j.firesaf.2021.103323

[6] Shin, Y. and Park, J.S. (2023) Modeling Climate Extremes Using the Four-Parameter Kappa Distribution for r-Largest Order Statistics. *Weather and Climate Extremes*, **39**, Article ID: 100533. https://doi.org/10.1016/j.wace.2022.100533

[7] McDonald, J.B., Walton, D.B. and Chia, B. (2020) Distributional Assumptions and the Estimation of Contingent Valuation Models. *Computational Economics*, **56**, 431-460. https://doi.org/10.1007/s10614-019-09930-x

[8] Teseletso, L.S. and Adachi, T. (2022) Long-Term Sustainability of Copper and Iron Based on a System Dynamics Model. *Resources*, **11**, Article No. 37. https://doi.org/10.3390/resources11040037

[9] Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y. and Chen, H. (2021)

Artificial Intelligence in Sustainable Energy Industry: Status Quo, Challenges and Opportunities. *Journal of Cleaner Production*, **289**, Article ID: 125834. https://doi.org/10.1016/j.jclepro.2021.125834

[10] Amoah, P. and Eweje, G. (2023) Organisational Drivers and Sustainability Implementation in the Mining Industry: A Holistic Theoretical Framework. *Business Strategy and the Environment*. https://doi.org/10.1002/bse.3438

[11] Minerals Commission (2022) Gold Production and Its Price. Ghana Minerals Commission.

[12] Humphreys, D. (2013) Long-Run Availability of Mineral Commodities. *Mineral Economics*, **26**, 1-11. https://doi.org/10.1007/s13563-013-0033-5

[13] Birch, C. (2019) Optimisation of Mining Block Size for Narrow Tabular Gold Deposits. In: Widzyk-Capehart, E., Hekmat, A. and Singhal, R., Eds., *Proceedings of the 27th International Symposium on Mine Planning and Equipment Selection-MPES* 2018, Springer International Publishing, Berlin, 121-141. https://doi.org/10.1007/978-3-319-99220-4_10

[14] Mutele, L. and Carranza, E.J.M. (2023) Statistical Analysis of Gold Production in South Africa Using Arima Modelling Technique: Extrapolating Future Gold Production, Resources-Reserves Depletion, and Implication on South Africa's Gold Exploration.

[15] Khatun, N. (2021). Applications of Normality Test in Statistical Analysis. *Open Journal of Statistics*, **11**, 113-122. https://doi.org/10.4236/ojs.2021.111006

[16] 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* (*OJS*), **13**, 534-566. https://doi.org/10.4236/ojs.2023.134027

[17] 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

[18] 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