Scientific Research Publishing

Open Journal of Social Sciences, 2024, 12, 396-416
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Parametric and Statistical Analysis on Corruption Perception Index (CPI) of Countries of the World Based on World Governance Index (WGI) Data of Transparency International (TI)

A. K. M. Raquibul Bashar[1], Chris P. Tsokos[2]

[1]Department of Mathematics & Computer Science, Augustana College, Rock Island, IL, USA
[2]Department of Mathematics & Statistics, University of South Florida, Tampa, FL, USA
Email: akmraquibulbashar@augustana.edu, ctsokos@usf.edu

How to cite this paper: Bashar, A. K. M. R., & Tsokos, C. P. (2024). Parametric and Statistical Analysis on Corruption Perception Index (CPI) of Countries of the World Based on World Governance Index (WGI) Data of Transparency International (TI). *Open Journal of Social Sciences, 12,* 396-416.
https://doi.org/10.4236/jss.2024.129023

Received: July 8, 2024
Accepted: September 21, 2024
Published: September 24, 2024

Abstract

In this study, we utilized data from Transparency International (TI) on 175 countries and from the World Bank (WB) on over 200 countries to develop the Worldwide Governance Indicators (WGI), which cover six dimensions of governance: 1) Voice and Accountability, 2) Political Stability, 3) Government Effectiveness, 4) Regulatory Quality, 5) Rule of Law, and 6) Control of Corruption. We identified and verified the statistical significance of the probability distribution function (PDF) for the Corruption Perception Index (CPI) scores. This PDF allows us to probabilistically characterize CPI score behavior and determine the likelihood of different score ranges, making it a valuable tool for unsupervised clustering and supervised classification that can be implemented in the domain of Artificial Intelligence of studying issues in social science.

Keywords

CPI (Corruption Perception Index), WGI (World Governance Indicator), PDF (Probability Distribution Function), CDF (Cumulative Distribution Function), GOF (Goodness-of-Fit)

1. Introduction

Corruption (Transparaency International, 1993) is a form of dishonest or unethical conduct by a person entrusted with a position of authority, often to acquire personal benefit (Caiden, 2019). Corruption may include many activities including bribery and embezzlement, though it may also involve practices that are legal in many countries. Government, or political, corruption occurs when an officeholder or other governmental employee acts in an official capacity for personal gain. Stephen D. Morris, a professor of politics, writes that political corruption is the illegitimate use of public power to benefit a private interest (Morris, 1991, 2001). Economist, Ian Senior defines corruption as an action to 1) secretly provide 2) a good or a service to a third party 3) so that he or she can influence certain actions which 4) benefit the corrupt, a third party, or both 5) in which the corrupt agent has authority (Senior, 1998). Daniel Kaufmann, from the World Bank, extends the concept to include "legal corruption" in which power is abused within the confines of the law—as those with power often have the ability to make laws for their protection (Kaufmann et al., 1999). From these expert opinions, and statements, it was a case where we wanted to identify the probability distribution function of these score indices, so that one will be able to calculate probabilities of these scores at various cut-off points and these probabilities can be incorporated as part of unsupervised clustering and classification based on the some third party data source to validate the clustering process (Bashar & Tsokos, 2017, 2019b). The results of this study are the extension of previous study conducted on Democracy Index Scores (DIS) of different countries of the world by Economists Intelligence Units (Bashar & Tsokos, 2019a).

2. Previous Studies & Rationale for This Topic

The rapid advancement of AI technologies is transforming industries in all sectors. Small businesses, a vital component of the economy, are increasingly integrating AI into their operations to improve efficiency (Farahani, 2024). Recognizing this impact is critical as it influences economic stability, innovation, and societal progress. The existing literature and data on CPI lay a crucial groundwork for furthering the use of AI to assess and rank countries worldwide. Currently, there is limited evidence from extensive research on how AI affects sociological issues such as corruption. For instance, a study examined the impact of corruption on FDI stock (Barassi & Zhou, 2012). Another study conducted a non-parametric analysis on CPI (McAdam & Rummel, 2004). Additionally, a semi-parametric analysis investigated the influence of corruption on entrepreneurial activity (Polemis, 2019). Furthermore, a survey of economic analysis of corruption omitted parametric analysis and probability distribution (Aidt, 2003; Halkos & Tzeremes, 2010). This research seeks to address this gap by evaluating the stochastic characteristics of this data. By doing so, governments, non-governmental organizations, small businesses, and corporations can utilize this mathematical framework to assess the corruption score of any country and allocate funding to various organizations of the respective countries.

3. Data Acquisition & Methodology

3.1. Data Source

The Corruption Perception Index (CPI) (Clark, 2018; Index, 2018) was created in 1995 to gauge public sector corruption perceptions in different countries worldwide.

Transparency International (TI) developed a method that involved selecting reliable data sources, standardizing the data, aggregating the re-scaled data, and reporting a measure of uncertainty. Using this approach, TI gathered data from various global sources for 175 countries, categorizing the world into six regions. Data from the World Bank (WB) was also utilized as a credible source for the corruption index (Kaufmann et al., 2010).

3.2. Methodology Involved

In this study of parametric analysis of Corruption Perception Index (CPI) scores, we have established the official mathematical setup and formulation of the probability distribution function (PDF) of the data. Then we have tested the statistical significance for the found PDFs using Kolmogorov-Smirnov test (Massey Jr., 1951), Anderson-Darling (Anderson & Darling, 1954), and χ^2 goodness-of-fit test (Cochran, 1952) suggested by Prof. Christopher P. Tsokos (Tsokos, 1972) when anyone identifies the PDF of any data-set. The purpose of following this statistical parametric analysis technique is to provide the readers an academic pathway of practicing the process of determination of probability distribution function for special or use case data relevant to the subject matter. In this study, we will be utilizing a data-set gathered by Transparency International, which includes information from 175 countries worldwide. Our goal is to analyze the CPI scores computed by Transparency International and then determine the probability distribution function (PDF) to describe the overall distribution of the CPI scores across the 175 countries. We have identified both the theoretical probability distributions as well as estimated the maximum likelihood estimates of the parameters relevant to each of the identified p.d.f. In this study, we have presented these processes in the following sections in setup where we will present the theoretical p.d.f.s, analytical structures of these p.d.f.s, and then we will present the estimates of the parameters in a table.

4. Estimating the Probability Distribution Function (p.d.f.) of Overall Corruption Perception Index (CPI) Scores

After acquiring the data about the CPI scores, we have inspected the overall shape of the distribution by the means of histogram with a smooth kernel curve imposed over the shape of the data as it is shown in the following **Figure 1**. In order to find the best theoretical p.d.f, three statistical tests, namely, Kolomogrov-Smirnov (Massey Jr., 1951), Anderson-Darling (Anderson & Darling, 1954) and Chi-square (Chernoff & Lehmann, 1954) were conducted. All of these goodness-of-fit tests agreed to the hypothesis of theoretical fit of the identified p.d.f.

4.1. Theoretical Probability Distribution Function (PDF) & Various Properties

We have found that, the best fitted distribution is the **"Mixture of 4-Gaussian PDF"** data as the initial shape of the histogram showed in **Figure 1**. The statistical significance test results given in Table 1 validate our finding as well. Thus, we proceeded with the fitted **Mixture of 4-Gaussian p.d.f.** of the CPI scores of 175 countries of the world.

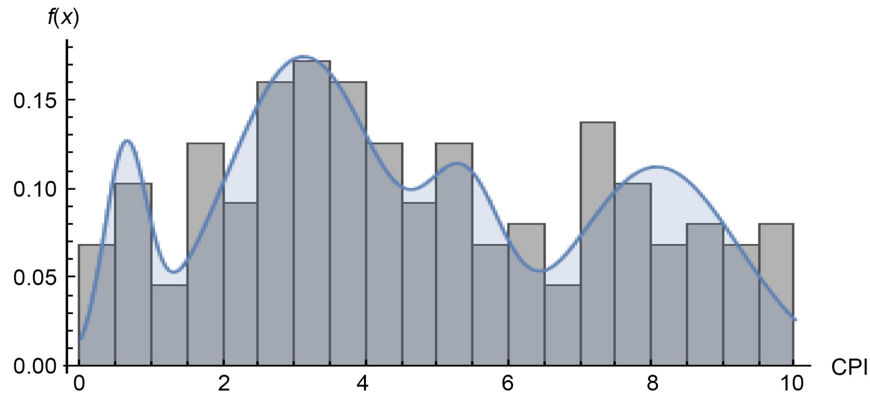


Figure 1. Overall distribution fitting of CPI (Corruption Perception Index).

Table 1. Goodness-of-Fit summary for overall CPI's.

	α	*p*-vaule	H_0: Data follow the identified Probability Distribution Function (PDF)
Kolmogorov-Smirnov	0.05	0.999	Do Not Reject
Anderson-Darling	0.05	0.994	Do Not Reject
Chi-Squared	0.05	0.977	Do Not Reject

Upon conducting three goodness-of-fit tests, it was determined that the probability distribution of CPI scores conforms to a 4-Mixed Gaussian probability density function (PDF) as shown in Figure 1. The Gaussian mixture model (Reynolds, 2009) is characterized by two sets of values: the mixture component weights and the component means and variances/covariance. In the case of CPI scores, a Gaussian mixture model with 4 components is utilized, with each component having a mean (μ_k) and standard deviation (σ_k) for the univariate case. The analytical structure is represented by Equation (1) for the CPI probability density function.

4.1.1. Theoretical PDF

$$f(x)=\sum_{i=1}^{k}\phi_i N\left(x\mid\mu_i,\sigma_i^2\right),$$

with, (1)

$$N\left(x\mid\mu_i,\sigma_i\right)=\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{\left(x-\mu_i\right)^2}{\sigma_i^2}\right),\quad -\infty\le X\le\infty$$

Here, $\sum_{i=1}^{k}\phi_i=1$.

4.1.2. Various Properties of This Theoretical PDF

The r^{th} moment can be found by

$$\mu_r = \int_{-\infty}^{\infty} x^r f(x)\delta x \tag{2}$$

From Equation (2) anyone can calculate the first four significant moments of the p.d.f.

The cumulative distribution function of the CPI scores is given in Equation (3) below:

$$F_x(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\delta t$$
$$= \frac{w_1 \operatorname{erfc}\left(\frac{\mu_1 - x}{\sqrt{2}\sigma_1}\right)}{2(w_1 + w_2 + w_3 + w_4)} + \frac{w_2 \operatorname{erfc}\left(\frac{\mu_2 - x}{\sqrt{2}\sigma_2}\right)}{2(w_1 + w_2 + w_3 + w_4)} + \frac{w_3 \operatorname{erfc}\left(\frac{\mu_3 - x}{\sqrt{2}\sigma_3}\right)}{2(w_1 + w_2 + w_3 + w_4)} + \frac{w_4 \operatorname{erfc}\left(\frac{\mu_4 - x}{\sqrt{2}\sigma_4}\right)}{2(w_1 + w_2 + w_3 + w_4)} \tag{3}$$

Furthermore, the moment generating function of 1 is given by

$$E\left(e^{tX}\right) = M_x(t) = \int e^{tX} f(x)\delta x$$
$$= \frac{w_1 e^{\frac{1}{2}\sigma_1^2 t^2 + \mu_1 t} + w_2 e^{\frac{1}{2}\sigma_2^2 t^2 + \mu_2 t} + w_3 e^{\frac{1}{2}\sigma_3^2 t^2 + \mu_3 t} + w_4 e^{\frac{1}{2}\sigma_4^2 t^2 + \mu_4 t}}{w_1 + w_2 + w_3 + w_4} \tag{4}$$

In the following section, we have presented the estimated m.l.e.s of the identified p.d.f either in a table or in an analytical form of the relevant function.

4.2. Maximum Likelihood Estimates of the Parameters

From Table 2, we see that the mean of the overall CPI scors is 4.73 and median is 4.313 and this data is slightly right skewed with a value of 0.249.

Table 2. Descriptive statistics of overall Corruption Perception Index (CPI).

Descriptive Statistics of CPI				
Mean	Median	Standard Deviation	Skewness	Kurtosis
4.739	4.313	2.679	0.249	1.98

The mean and the variance are 4.75 and 7.34, respectively, with standard deviation of 2.71. For our data, the approximate maximum likelihood estimates (MLEs) of the parameters (μ_i , σ_i , and ϕ_i , here, $i = 1,2,3,4$) of 1 are given in the following Table 3:

Table 3. MLEs of CPI scores of 175 countries of the world.

MLEs of CPI scores							
$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\mu}_3$	$\hat{\sigma}_3$	$\hat{\mu}_4$	$\hat{\sigma}_4$
9.18	0.454	5.6	0.67	2.93	1.46	7.54	0.442

Thus, the estimated analytical form of the subject PDF is given by

$$f(x)=\begin{cases}0.113e^{-0.385(x-8.063)^2}+0.085e^{-1.80(x-5.382)^2} & \\ +0.175e^{-0.412(x-3.118)^2}+0.114e^{-5.634(x-0.622)^2}, & 0\le X\le 10\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

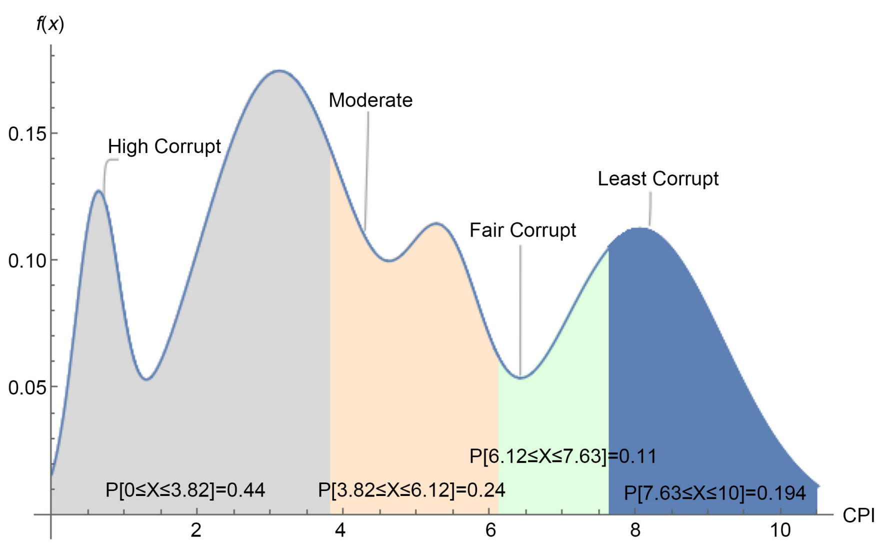
In a similar method, we have presented the moment generating function (6) of the identified p.d.f of CPI scores with its estimated parameter values from the data.

$$\begin{aligned}M_X t &= 0.085e^{0.044t^2+0.622t}+0.112e^{0.138t^2+5.38t}\\ &+0.482e^{0.61t^2+3.12t}+0.32e^{0.65t^2+8.063t}\end{aligned} \tag{6}$$

4.3. Real Life Relevance of This Study

The graph of the p.d.f. give in Equation (5) is given below in Figure 2.

Suppose a country is chosen randomly from the 175 countries. In this scenario, one can determine the probability of the country falling into one of four corruption categories. Figure 2 displays the Cumulative Distribution Function (c.d.f) of CPI scores. This graph is valuable for calculating the likelihood of a country having a CPI score below a certain threshold. For instance, the probability of a country having a CPI score less than 3.82 is approximately 44%. Furthermore, if we are interested in the probability of a country having a CPI score less than or equal to 7.63, Figure 2 shows that this probability is around 79%. Equation (6) provides the Moment Generating Function (MGF) that can be used to calculate higher order moments and, consequently, the mean and variance of the Mixed Gaussian PDF. When a country is randomly selected from the 175 countries, the expected CPI score is 4.86. The calculated variance is $\sigma_{cpi}^2 = 7.34$ and the standard deviation is $stdev_{cpi} = 2.71$. These estimates closely match the basic statistics given in Table 2, indicating the quality of the fit of the Mixed 4-Gaussian PDF.



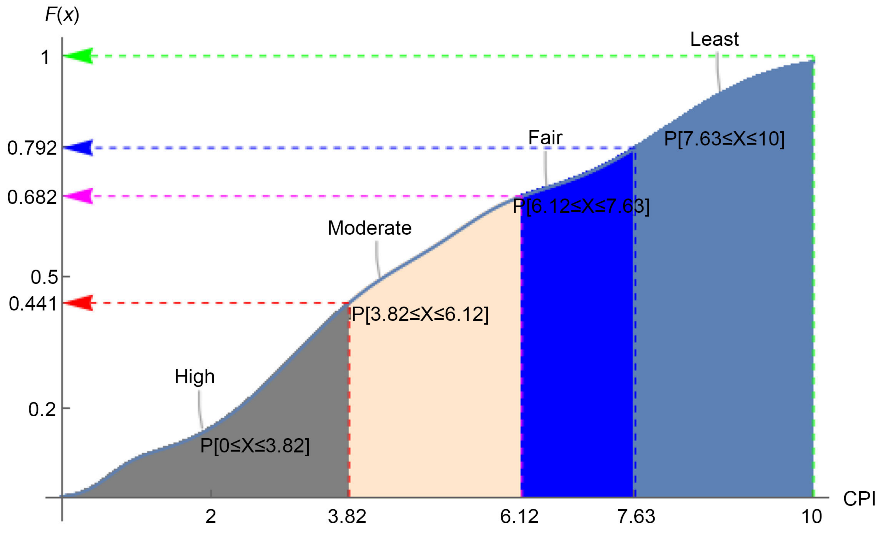


Figure 2. P.D.F. & C.D.F. of CPI scores.

In the upcoming sections, we will delve into the procedure for determining the Probability Density Function (PDF) for each of the four categories (represented by the 4 peaks of the histogram of CPI scores shown in **Figure 1**) of the 175 countries in the world.

5. Estimating the Probability Distribution Function (p.d.f.) of *Least Corrupted* Countries of the World

The upcoming sections will outline the same methodological approach to explain the process of deducing the probability density function (p.d.f) of 4 categories of CPI scores. The probability distribution defines the probabilistic behavior for countries classified as Least Corrupted with scores ranging from 7.63 to 10, as described in Equation (9). To begin, we conducted a detailed analysis of the descriptive statistics for Least Corrupted countries. The histogram for these countries is displayed in **Figure 3**. Based on this histogram, it is evident that we need to utilize a mixed probability density function for fitting purposes.

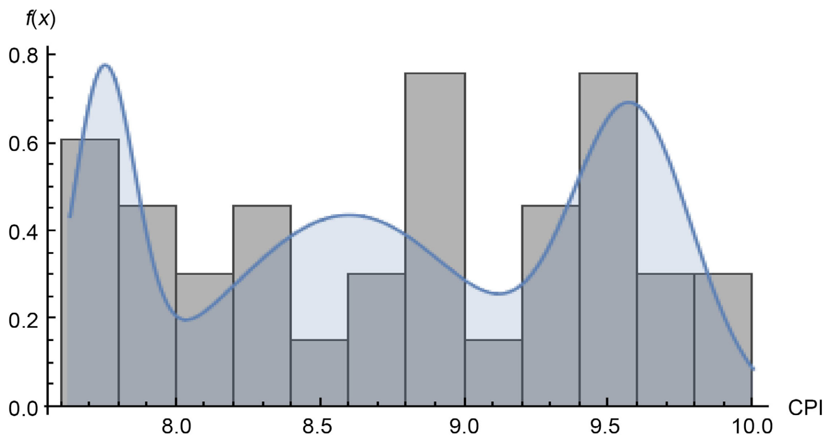


Figure 3. PDF fitted to histogram of least corrupted countries of the world.

5.1. Theoretical p.d.f of Least Corrupted Countries (7.63 ≤ CPI ≤ 10)

After performing three goodness-of-fit tests on the current data related to the Least corrupted countries, it has been determined that the data adheres to a distribution called "Mixed of 3-Gaussian PDF." This distribution has been confirmed by three different methods of goodness-of-fit, as outlined in Table 4.

Table 4. Goodness-of-fit summary for least corrupted scores.

	α	*p*-value	H_0: Data follow the identified Probability Distribution Function (PDF)
Kolmogorov-Smirnov	0.05	0.9492	Do Not Reject
Anderson-Darling	0.05	0.9720	Do Not Reject
Chi-Squared	0.05	0.7121	Do not Reject

Thus, the fitted theoretical PDF of the CPI scores of "Least Corrupted Countries" is given by

$$f(x)=\begin{cases}\dfrac{w_1 e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi}\sigma_1(w_1+w_2+w_3)}+\dfrac{w_2 e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi}\sigma_2(w_1+w_2+w_3)}+\dfrac{w_3 e^{-\frac{(x-\mu_3)^2}{2\sigma_3^2}}}{\sqrt{2\pi}\sigma_3(w_1+w_2+w_3)}, & x>0\\ 0, & \text{otherwise}\end{cases} \quad (7)$$

The cumulative distribution function of the CPI scores of Least Corrupted countries of the world is given by

$$F(x)=P(X\le x)=\frac{w_1\,\mathrm{erfc}\left(\frac{\mu_1-x}{\sqrt{2}\sigma_1}\right)}{2(w_1+w_2+w_3)}+\frac{w_2\,\mathrm{erfc}\left(\frac{\mu_2-x}{\sqrt{2}\sigma_2}\right)}{2(w_1+w_2+w_3)}+\frac{w_3\,\mathrm{erfc}\left(\frac{\mu_3-x}{\sqrt{2}\sigma_3}\right)}{2(w_1+w_2+w_3)} \quad (8)$$

5.2. M.L.E.s of the CPI Scores of the Least Corrupted Countries

The estimated m.l.e.s of the parameters that drive the estimated 3-Mixed Gaussian PDF are given in Table 5 below:

Table 5. MLEs of least corrupted countries of the world.

MLEs of Least Corrupted Countries					
$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\mu}_3$	$\hat{\sigma}_3$
8.1	0.32	8.84	0.033	9.51	0.211

With the weights $\hat{w}_1$, $\hat{w}_2$, and $\hat{w}_3$ having values 0.43, 0.18, and 0.39 respectively where $\sum_{k=1}^{3} w_i = 1$. So, the analytical structure of the estimated PDF is given by

$$f(x)=\begin{cases}0.66e^{-11.63(-9.58+x)^2}+0.44e^{-2.87(-8.59+x)^2}+0.72e^{-42.05(-7.75+x)^2}, & 7.63\le X\le 10\\ 0, & \text{otherwise}\end{cases} \quad (9)$$

5.3. Real-Life Relevance of the Identified p.d.f.

The graph of the PDF of Equation (9) is given by **Figure 4** below:

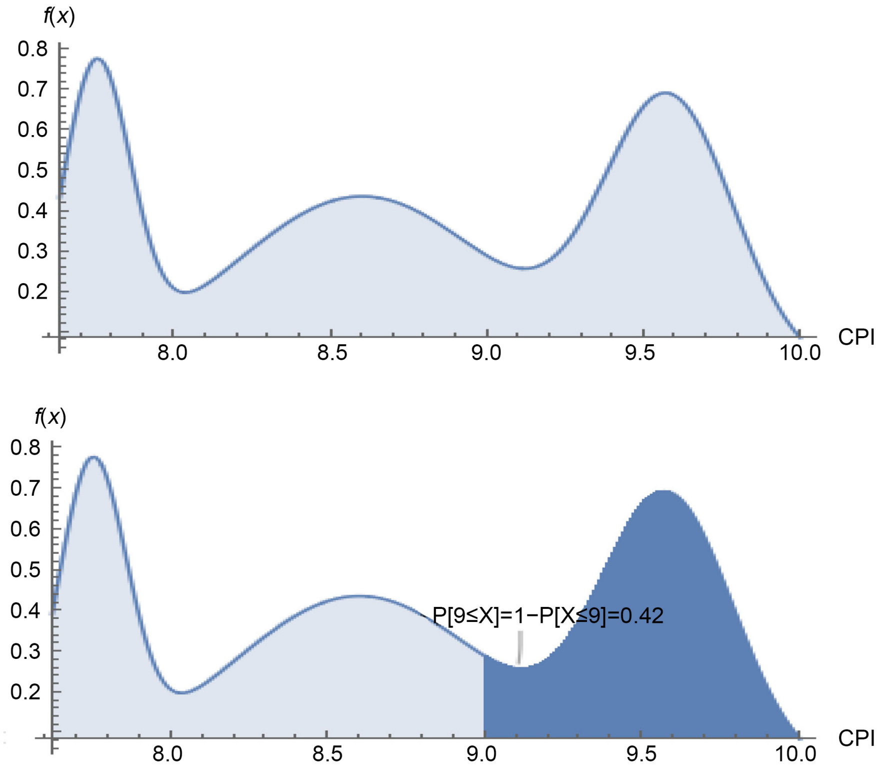
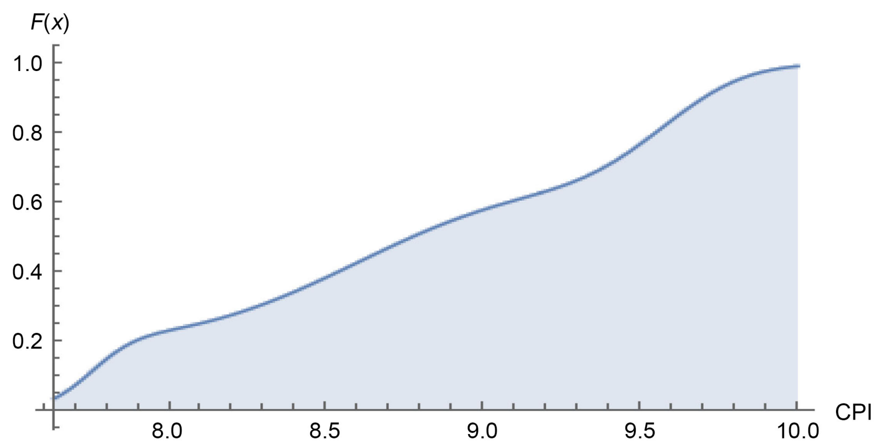


Figure 4. Plotting PDF of least corrupted countries of the world.

The information provided in the graph referenced in **Figure 4** can be very helpful for calculating the probability of a randomly selected country from this subset of the population having a score greater than 9. In the graph, the probability of a country having a score more than 9 (i.e. $P(X \geq 9) = 1 - P(X \leq 9)$) is approximately 0.42, as indicated in the same figure.

Figure 5 can be utilized to find the same probability for interested readers and scholars.



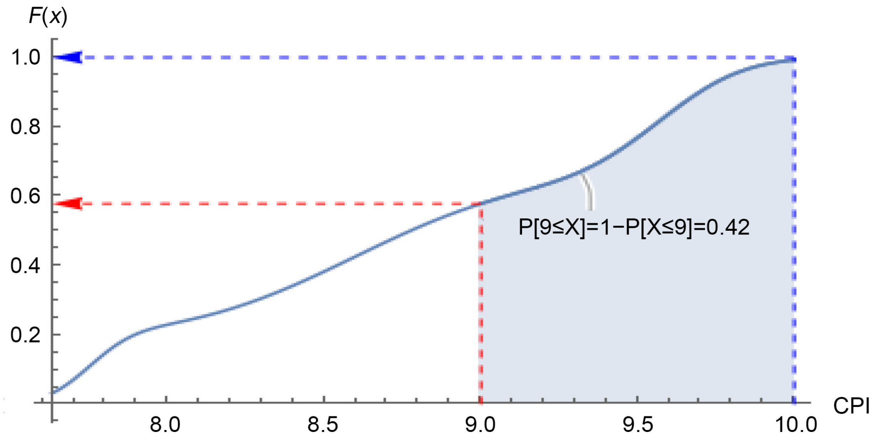


Figure 5. CDF of least corrupt countries of the world.

6. Estimating the Probability Distribution Function (p.d.f.) of Fairly Corrupted Countries of the World

As shown in **Figure 2**, the identified cut-off points for the entire CPI scores data, ranging from 6.12 to 7.63. Next, we will proceed to determine the probability distribution that represents the probabilistic behavior of countries with moderate corruption. We will follow the same steps used in the earlier sections.

6.1. Theoretical p.d.f. of Fairly Corrupted Countries (6.12 ≤ CPI ≤ 7.63)

This specific portion of the data is somewhat skewed to the left with a value of 0.785, and it has an average of 7.038. The graph displaying the distribution of Fairly Corrupted Countries is shown below in **Figure 6**. Based on the graph, it seems like we will need to use a left-skewed probability density function to model this data.

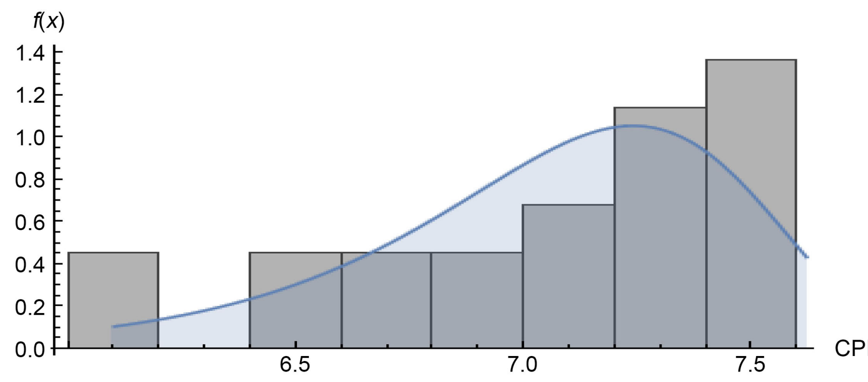


Figure 6. PDF fitted to histogram to the fairly corrupted countries.

Based on the three goodness-of-fit tests applied to the current data of moderately corrupt countries, it has been determined that the data exhibits probabilistic behavior characterized by the Gumbel PDF. The selection of this model is supported by the results of the three goodness-of-fit methods presented in **Table 6**.

These results confirm that the most suitable probability density function for the data related to moderately corrupt countries is the Gumbel distribution.

Table 6. Goodness-of-fit summary for fairly corrupted scores.

	α	*p*-value	H_0: Data follow the identified Probability Distribution Function (PDF)
Kolmogorov-Smirnov	0.05	0.877	Do Not Reject
Anderson-Darling	0.05	0.814	Do Not Reject
Chi-Squared	0.05	0.956	Do not Reject

Thus, the fitted theoretical PDF of the subject data is given by

$$f(x)=\begin{cases}\dfrac{e^{\frac{x-\alpha}{\beta}-e^{\frac{x-\alpha}{\beta}}}}{\beta}, & -\infty\le x\le\infty\\ 0, & \text{otherwise}\end{cases} \quad (10)$$

The c.d.f. of the CPI of Fairly corrupt countries of the world is given by

$$F(x)=P(X\le x)=1-e^{-e^{\frac{x-\alpha}{\beta}}}, \quad (11)$$

6.2. M.L.E.s of the CPI Scores of the Fairly Corrupted Countries

The following Table 7 shows the estimated m.l.e.s of the parameters presented in Equation (10). Based on the estimated values of the table, (this is shown in these studies (Bashar & Tsokos, 2019a, 2019b, 2017; Bashar, 2019)) we have presented the p.d.f. in Equation (12). These values of the m.l.e.s of the parameters of the identified p.d.f. of subset of the CPI scores categorized as *Fairly Corrupted* countries are given in Table 7 below.

Table 7. MLEs of fairly corrupted countries of the world.

MLEs of Fairly Corrupted Countries CPI scores	
$\hat{\alpha}$	$\hat{\beta}$
7.237	0.319

So, the analytical form of the PDF of Fairly corrupted countries of the world is given as follows:

$$f(x)=\begin{cases}2.87e^{2.87(x-7.239)-e^{2.87(x-7.239)}}, & 6.1\le X\le 7.62\\ 0, & \text{otherwise}\end{cases} \quad (12)$$

6.3. Real-Life Relevance of the Identified p.d.f.

The graph of the PDF of Equation (12) is given in Figure 7 below.

The mean and standard deviation of the least corrupted data subset are 7.038 and 0.199, respectively. This means that if a country is chosen at random from this subset, we anticipate its CPI score to be around 7.038. Moreover, the likelihood of a country having a CPI score between 6.5 and 7 is 0.283, as indicated in **Figure 7**. Its graph of the cumulative distribution function is given in **Figure 7** below:

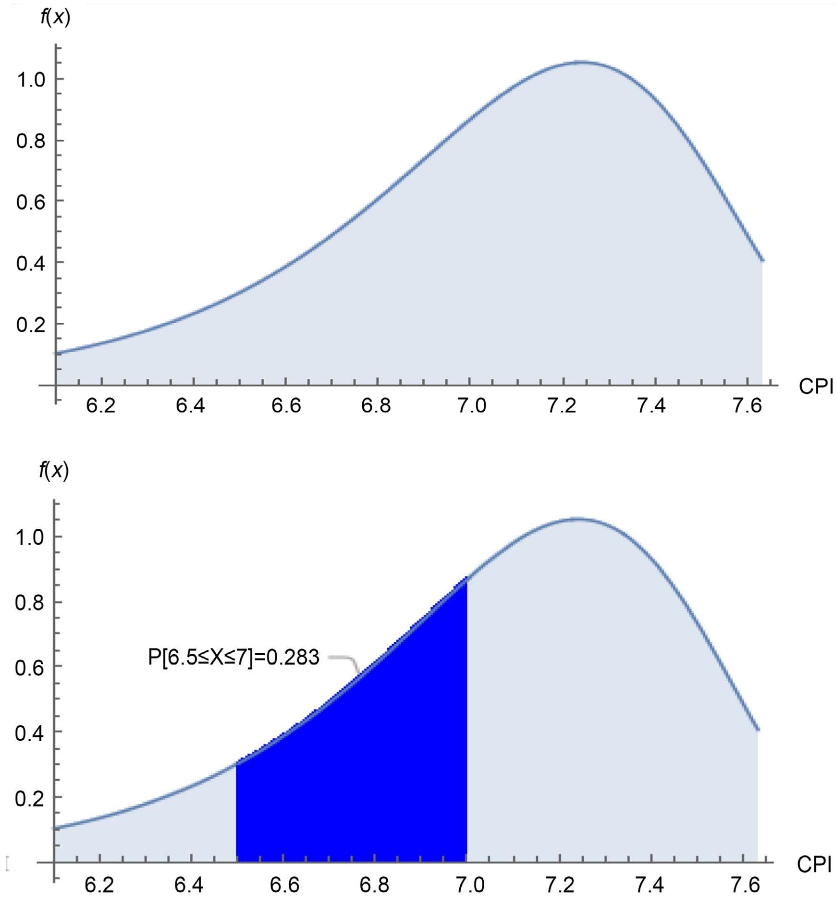
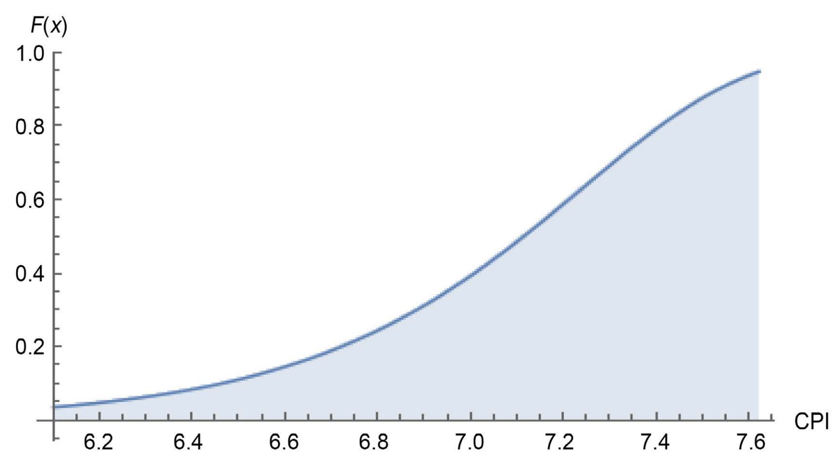


Figure 7. Plotting PDF of fairly corrupted countries of the world.

You can use **Figure 8** to calculate the probability of a randomly selected country from this population subset scoring between 6.8 and 7.4. This can be computed as the difference between the probabilities of the score being less than or equal to 7.4 and 6.8, which amounts to 0.55.



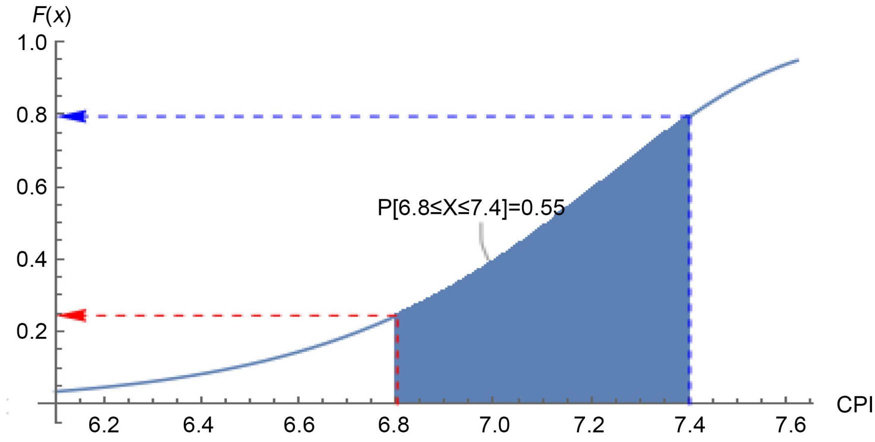


Figure 8. CDF of fairly corrupt countries of the world.

7. Estimating the Probability Distribution Function (p.d.f.) of Moderately Corrupted Countries of the World

In this part of our research, we will begin by identifying the probability distribution that represents the likelihood of the CPI data for Moderately Corrupted nations. To achieve this, we have applied the same methodology used to determine the probability density function (PDF) of CPI scores for the two preceding categories. To do this, we initially examined the fundamental descriptive statistics of "Moderately Corrupted" countries.

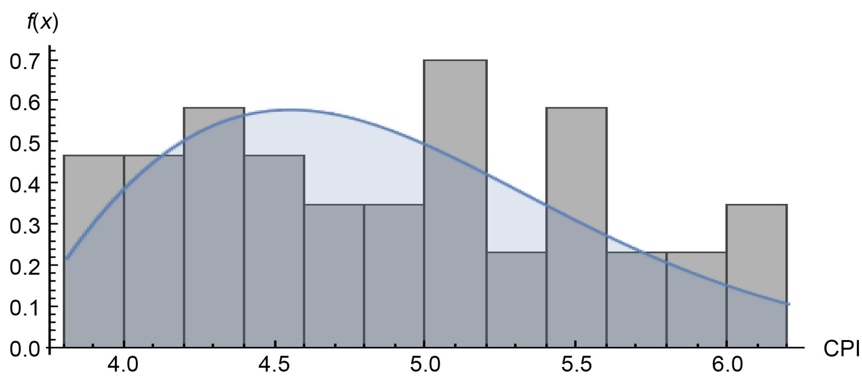


Figure 9. Histogram of moderately corrupted countries of the world.

In the provided **Table 8**, it can be observed that the average for this specific portion of the data is 4.89. The histogram displaying Moderately Corrupted Countries can be found in **Figure 9**. Based on this histogram, it seems necessary to employ a probability density function that is skewed to the right.

Table 8. Descriptive statistics of moderately corrupted countries of the world

Moderately Corrupted Countries CPI scores				
Mean	Median	Standard Deviation	Skewness	Kurtosis
4.89	4.93	0.68	0.132	1.92

7.1. Theoretical p.d.f. of Moderately Corrupted Countries ($3.82 \le \text{CPI} \le 6.12$)

The analysis of the subset of the overall CPI scores' categorized as *Moderate corrupted countries* data using three goodness-of-fit tests indicates that the data exhibits probabilistic behavior described by the Weibull probability density function (PDF). The selection of the Weibull PDF is justified by the results of the three goodness-of-fit tests presented in Table 9. These results confirm that the Weibull 3-P probability density function is the best fit for the data from Fairly Corrupted countries.

Table 9. Goodness-of-Fit summary for moderately corrupted countries of the world.

	α	*p*-value	H_0: Data follow the identified Probability Distribution Function (PDF)
Kolmogorov-Smirnov	0.05	0.881	Do Not Reject
Anderson-Darling	0.05	0.8572	Do Not Reject
Chi-Squared	0.05	0.985	Do not Reject

Thus, the fitted theoretical PDF of the subset of the CPI scores data is given by-

$$f(x)=\begin{cases}\dfrac{\alpha e^{-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}}\left(\dfrac{x-\gamma}{\beta}\right)^{\alpha-1}}{\beta}, & x>\gamma\\ 0, & \text{otherwise}\end{cases} \tag{13}$$

The cumulative distribution function of the Moderately Corrupted Countries of the world is given by

$$F(x)=P(X\le x)=1-e^{-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}} \tag{14}$$

7.2. M.L.E.s of the CPI Scores of the Moderately Corrupted Countries

The MLEs of this PDF of Equation (13) are given in the following Table 10.

Table 10. MLEs of moderately corrupted countries of the world.

MLEs of the Moderately Corrupted Countries		
$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
1.938	1.41	3.64

The analytical form of this p.d.f. with estimated parameters is given as follows:

$$f(x)=\begin{cases}0.989e^{-0.544(x-3.64)^{1.82}}(x-3.64)^{0.819}, & x>3.64\\ 0, & \text{Otherwise}\end{cases} \tag{15}$$

The cumulative distribution function of the Moderately corrupted country CPI scores with the estimated MLEs of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ is given below:

$$F(x)=\begin{cases}P(X \le x)=1-e^{-0.544(x-3.65)^{1.82}}, & 3.65 < X \\ 1, & X > 7.62\end{cases} \tag{16}$$

7.3. Real-Life Relevance of the Identified p.d.f.

The mean and standard deviation of the data subset for moderately corrupted countries are 4.89 and 0.499 respectively. This implies that if a country is chosen randomly from this group, the anticipated CPI score would be around 4.89. Moreover, the likelihood of a country having a CPI score between 4.5 and 5.5 is 0.478, illustrated in Figure 10. The graph representing the Probability Density Function (PDF) described in Equation (15) can be observed in Figure 10 below.

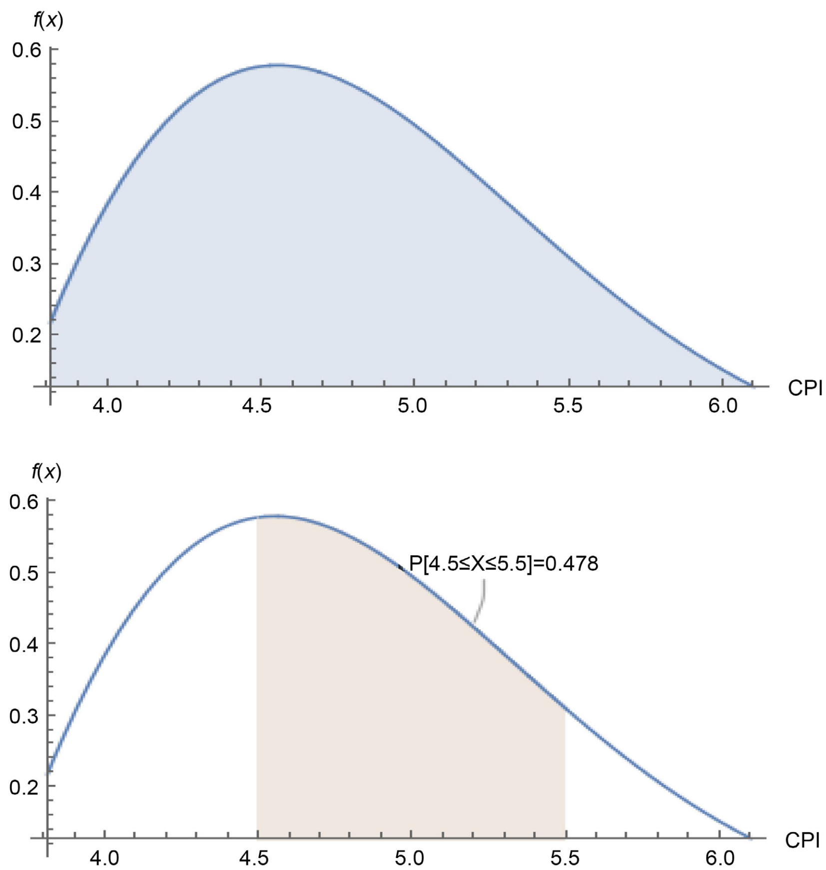


Figure 10. Plotting PDF of moderately corrupted countries of the world.

The method of plotting shown in Figure 11 is especially helpful for calculating the probability of randomly selecting a country from this specific population subset and determining the likelihood of that country scoring between 5 and 5.5 (i.e. $P(5 \le X \le 5.5)$), which is 0.20 according to the subset data.

8. Estimating the Probability Distribution Function (p.d.f.) of Highly Corrupted Countries of the World

In this part of our analysis on CPI scores, we will move forward with identifying the probability distribution that represents the likelihood of the CPI data specifically for countries categorized as Highly Corrupted. To achieve this, we have applied the same procedures used in determining the Probability Density Function (PDF) of CPI scores for the three previous categories. We initiated this process by examining the fundamental descriptive statistics of the Highly Corrupted countries.

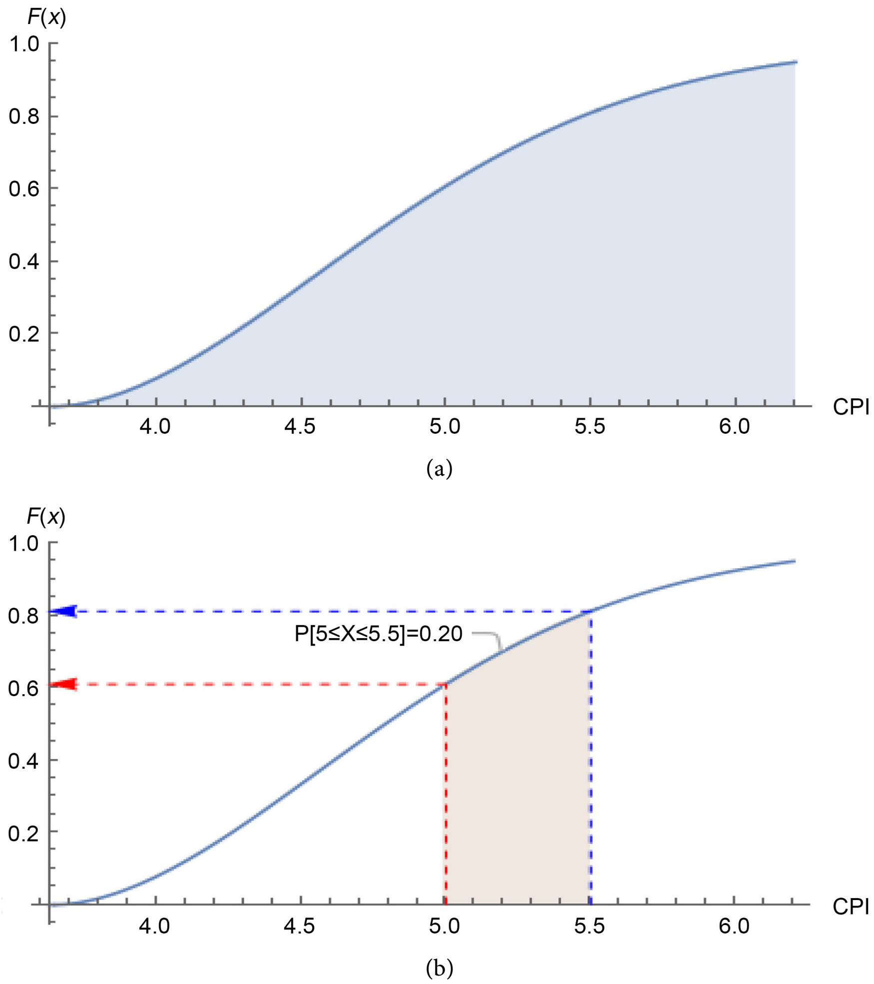


Figure 11. Plotting CDF of moderately corrupted countries of the world. (a) CDF of moderate corruption country; (b) CDF of moderate for $P(5.0 \le X \le 5.5)$.

Table 11. Descriptive statistics of highly corrupted countries of the world.

Highly Corrupted Countries CPI Scores				
Mean	Median	Standard Deviation	Skewness	Kurtosis
2.269	2.578	1.061	−0.427	1.965

From Table 11 above, we can see that the distribution of the CPI scores of Highly Corrupted countries are some kind of left skewed as it is shown in Figure 12.

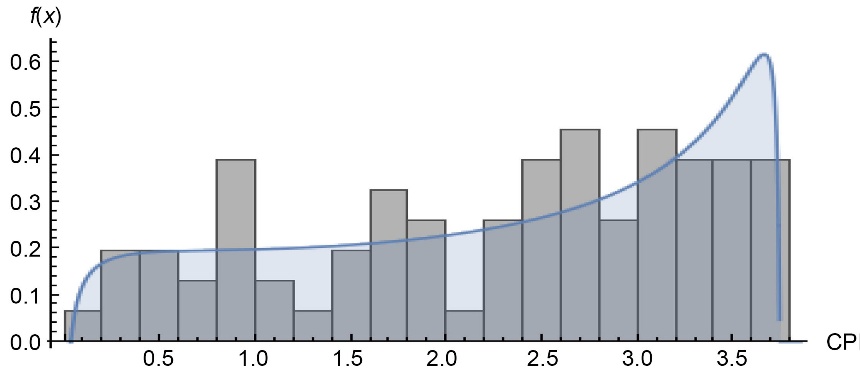


Figure 12. Histogram of highly corrupted countries of the world.

8.1. Theoretical p.d.f. of *Highly Corrupted* Countries ($0 \le CPI \le 3.82$)

When applying three goodness-of-fit tests to identify the most suitable candidate for representing the highly corrupted data subset, it was determined that the Johnson SB (4p) (Parresol, 2003) probability density function accurately portrayed the probabilistic characteristics of the highly corrupted data subset. This fact was further confirmed by the goodness-of-fit tests, as shown in Table 12 below:

Table 12. Goodness-of-Fit summary for highly corrupted countries of the world.

	α	*p*-value	H_0: Data follow the identified Probability Distribution Function (PDF)
Kolmogorov-Smirnov	0.05	0.992	Do Not Reject
Anderson-Darling	0.05	0.1342	Do Not Reject
Chi-Squared	0.05	0.747	Do not Reject

The analytical structure of the **Johnson SB (4p)** is given by Equation (17) below:

$$f(x)=\begin{cases}\dfrac{\delta\sigma e^{-\frac{1}{2}\left(\gamma+\delta\log\left(\frac{x-\mu}{\mu+\sigma-x}\right)\right)^2}}{\sqrt{2\pi}(x-\mu)(\mu+\sigma-x)}, & \mu < x < \mu+\sigma \\ 0, & \text{otherwise}\end{cases} \tag{17}$$

8.2. M.L.E.s of the CPI Scores of the Highly Corrupted Countries

The estimated MLEs (maximum likelihood estimates) for PDF in Equation (17) are given in the following Table 13:

Table 13. MLEs of highly corrupted countries PDF.

MLEs of Highly Corrupted Countries			
γ	δ	μ	σ
-0.364	0.552	0.0288	3.712

The fitted PDF with estimated parameters is given by Equation (18) below:

$$f(x)=\begin{cases}\dfrac{0.82e^{-\frac{1}{2}\left(0.55\log\left(\frac{x-0.029}{3.741-x}\right)-0.36\right)^2}}{(x-0.029)(3.741-x)}, & \text{if } 0<X<3.65\\ 0, & \text{Otherwise}\end{cases} \quad (18)$$

The cumulative distribution function is given by

$$F(x)=P(X\le x)$$

$$=\begin{cases}\dfrac{1}{2}\text{erfc}\left(-\dfrac{0.552\log\left(\dfrac{x-0.029}{3.74-x}\right)-0.36}{\sqrt{2}}\right), & 0.03<x<1.88\\ \dfrac{1}{2}\left(\text{erf}\left(\dfrac{0.552\log\left(\dfrac{x-0.029}{3.74-x}\right)-0.36}{\sqrt{2}}\right)+1\right), & 1.88\le x<3.74\\ 1, & x\ge 3.74\end{cases} \quad (19)$$

8.3. Real-Life Relevance of the Identified p.d.f.

The graph of the PDF of Equation (18) is given in Figure 13.

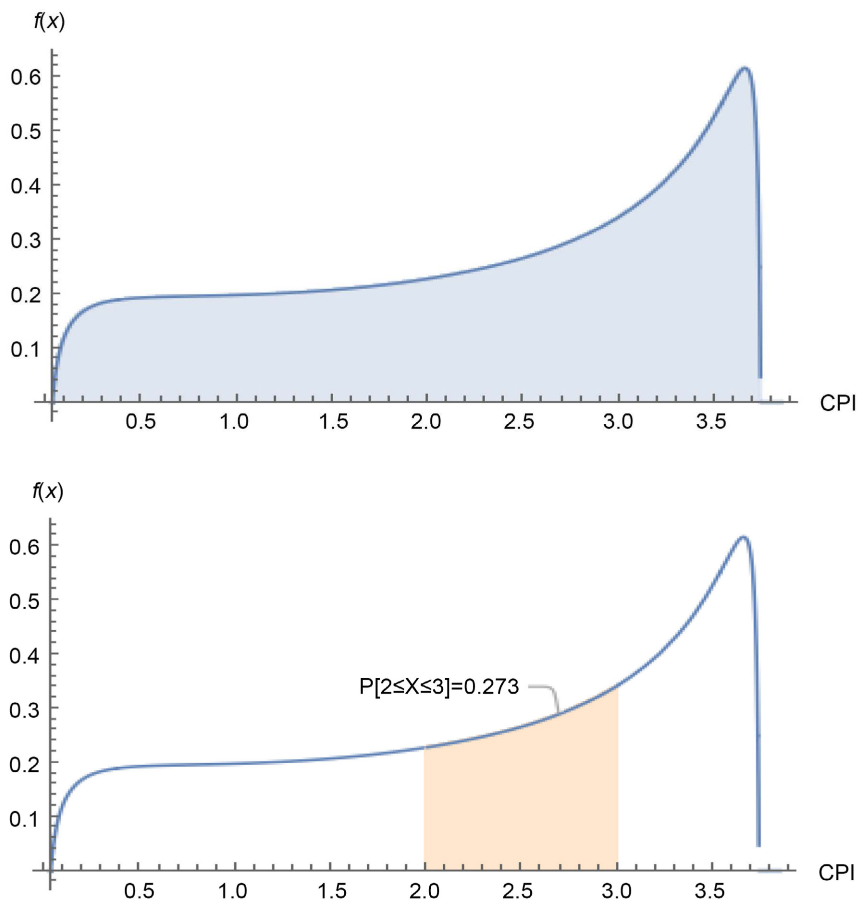


Figure 13. Plotting PDF of highly corrupted countries of the world.

The mean and variance of the subset of country data with high levels of corruption are 2.273 and 1.1648, respectively. This means that if a country is randomly chosen from this group, we anticipate its CPI score to be around 2.3. Moreover, the likelihood of a country having a CPI score between 2 and 3 is 0.273, as illustrated in **Figure 14**. The CDF plot of the Equation (19) is given below:

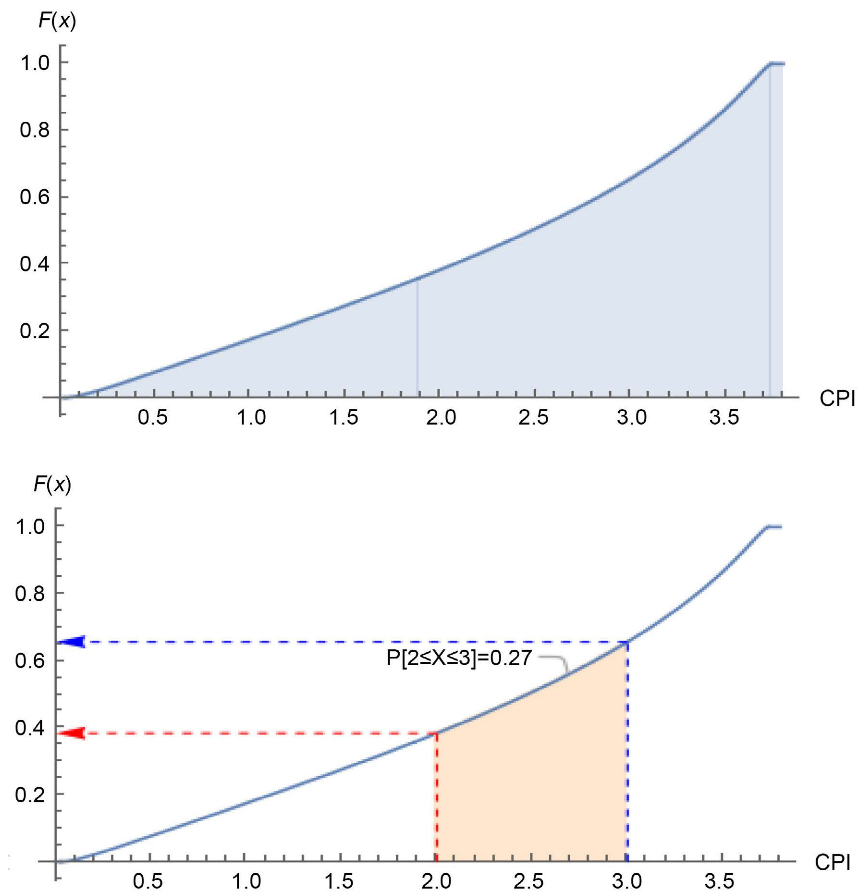


Figure 14. Plotting CDF of highly corrupted countries of the world.

The information presented in **Figure 14** can be valuable for determining the likelihood of a randomly selected country from a specific population subset having a score between 2 and 2.5. The probability of this occurrence, represented as $P(2 \le X \le 3)$, is 0.27.

9. Conclusion & Suggestions for Future Work

In our recent research, we analyzed the Corruption Perception Index scores provided by Transparency International and the World Bank to determine the probability distribution function (p.d.f.) of the data. Our analysis involved categorizing 175 countries into four groups based on their level of corruption: Least Corrupted, Fairly Corrupted, Moderately Corrupted, and Highly Corrupted. We then identified the p.d.f. and c.d.f. of the CPI scores for each group, finding that the data could be characterized as a Mixture of 4-Gaussian PDF for all 175 countries, a

Mixture of 3-Gaussian p.d.f. for the Least Corrupted countries, a Gumbel PDF for the Fairly Corrupted countries, a Weibull 3-P p.d.f. for the Moderately Corrupted countries, and a Johnson (SB) 4-p p.d.f. for the Highly Corrupted countries. Our analysis allows for the determination of expected values and confidence limits for the CPI scores, and the results can be further utilized in Bayesian, machine learning, and artificial intelligence analyses to gain additional insights into the CPI scores.

Acknowledgements

Sincere thanks to the members of JAMP for their professional performance, and special thanks to managing editor Hellen XU for a rare attitude of high quality.

Conflicts of Interest

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

References

Aidt, T. S. (2003). Economic Analysis of Corruption: A Survey. *The Economic Journal, 113,* F632-F652. https://doi.org/10.1046/j.0013-0133.2003.00171.x

Anderson, T. W., & Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association, 49,* 765-769. https://doi.org/10.1080/01621459.1954.10501232

Barassi, M. R., & Zhou, Y. (2012). The Effect of Corruption on FDI: A Parametric and Non-Parametric Analysis. *European Journal of Political Economy, 28,* 302-312. https://doi.org/10.1016/j.ejpoleco.2012.01.001

Bashar, A. K. M. R. (2019). *Probabilistic Modeling of Democracy, Corruption, Hemophilia A and Prediabetes Data.* Ph.D. Thesis, University of South Florida.

Bashar, A. K. M. R., & Tsokos, C. P. (2017). Parametric Analysis of Factor 8 (F8) Hemophilia A. *International Journal of Mathematical Sciences in Medicine (IJMSM), 1,* 1-10.

Bashar, A. K. M. R., & Tsokos, C. P. (2019a). Statistical Classification of Democracy Index Scores of Countries of the World. *Scholars Journal of Arts, Humanities and Social Sciences.*

Bashar, A. K. M. R., & Tsokos, C. P. (2019b). Statistical Parametric Analysis on Democracy Data. *Open Access Library Journal, 6,* 1-18. https://doi.org/10.4236/oalib.1105828

Caiden, G. E. (2019). Dealing with Administrative Corruption. In T. Cooper (Ed.), *Handbook of Administrative Ethics* (pp. 429-455). Routledge.

Chernoff, H., & Lehmann, E. L. (1954). The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit. *The Annals of Mathematical Statistics, 25,* 579-586. https://doi.org/10.1214/aoms/1177728726

Clark, A. K. (2018). Measuring Corruption: Transparency International's "Corruption Perceptions Index". *Public Policy and Governance, 29,* 3-22. https://doi.org/10.1108/s2053-769720170000029001

Cochran, W. G. (1952). The χ^2 Test of Goodness of Fit. *The Annals of Mathematical Statistics, 23,* 315-345. https://doi.org/10.1214/aoms/1177729380

Farahani, M. S. (2024). Applications of Artificial Intelligence in Social Science Issues: A Case Study on Predicting Population Change. *Journal of the Knowledge Economy, 15,* 3266-3296. https://doi.org/10.1007/s13132-023-01270-4

Halkos, G. E., & Tzeremes, N. G. (2010). Corruption and Economic Efficiency: Panel Data Evidence. *Global Economic Review, 39,* 441-454. https://doi.org/10.1080/1226508x.2010.533854

Index, C. P. (2018). *Corruption Perception Index.* Transparancy International.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The Worldwide Governance Indicators: Methodology and Analytical Issues (World Bank Policy Research Working Paper No. 5430).* The World Bank.

Kaufmann, D., Kraay, A., & Zoido, P. (1999). *Governance Matters.*

Massey Jr., F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association, 46,* 68-78. https://doi.org/10.1080/01621459.1951.10500769

McAdam, P., & Rummel, O. (2004). Corruption: A Non-Parametric Analysis. *Journal of Economic Studies, 31,* 509-523. https://doi.org/10.1108/01443580410569253

Morris, S. (2001). Political Correctness. *Journal of Political Economy, 109,* 231-265. https://doi.org/10.1086/319554

Morris, S. D. (1991). *Corruption & Politics in Contemporary Mexico.* University of Alabama Press.

Parresol, B. R. (2003). *Recovering Parameters of Johnson's SB Distribution, Volume 31.* US Department of Agriculture, Forest Service, Southern Research Station.

Polemis, M. (2019). Is the Effect of Corruption on Entrepreneurial Activity Nonmonotonic? A Semi-Parametric Panel Data Analysis. *Economics Bulletin, 39,* 2976-2989.

Reynolds, D. (2009). Gaussian Mixture Models. In S. Z. Li, & A. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 659-663). Springer. https://doi.org/10.1007/978-0-387-73003-5_196

Senior, I. (1998). An Economic View of Corruption. *Journal of Interdisciplinary Economics, 9,* 145-161. https://doi.org/10.1177/02601079x9800900203

Transparaency International (1993). *What Is Corruption? The Global Coalition against Corruption.*

Tsokos, C. P. (1972). *Probability Distributions: An Introduction to Probability Theory with Applications.* Duxbury Press.