



Article

The Continuous Bernoulli Distribution: Mathematical Characterization, Fractile Regression, Computational Simulations, and Applications

Mustafa Ç. Korkmaz ¹, Víctor Leiva ^{2,*} and Carlos Martin-Barreiro ^{3,4}

¹ Department of Measurement and Evaluation, Artvin Çoruh University, Artvin 08100, Turkey; mcagatay@artvin.edu.tr

² School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile

³ Faculty of Natural Sciences and Mathematics, Escuela Superior Politécnica del Litoral ESPOL, Guayaquil 090902, Ecuador; cmmartin@espol.edu.ec

⁴ Faculty of Engineering, Universidad Espíritu Santo, Samborondón 0901952, Ecuador

* Correspondence: victor.leiva@pucv.cl or victorleivasanchez@gmail.com

Abstract: The continuous Bernoulli distribution is defined on the unit interval and has a unique property related to fractiles. A fractile is a position on a probability density function where the corresponding surface is a fixed proportion. This article presents the derivation of properties of the continuous Bernoulli distribution and formulates a fractile or quantile regression model for a unit response using the exponentiated continuous Bernoulli distribution. Monte Carlo simulation studies evaluate the performance of point and interval estimators for both the continuous Bernoulli distribution and the fractile regression model. Real-world datasets from science and education are analyzed to illustrate the modeling abilities of the continuous Bernoulli distribution and the exponentiated continuous Bernoulli quantile regression model.

Keywords: Bernoulli distribution; likelihood and Monte Carlo methods; point estimation; quantile function; R software; residual analysis

MSC: 60E05; 62J02



Citation: Korkmaz, M.Ç.; Leiva, V.; Martin-Barreiro, C. The Continuous Bernoulli Distribution: Mathematical Characterization, Fractile Regression, Computational Simulations, and Applications. *Fractal Fract.* **2023**, *7*, 386. <https://doi.org/10.3390/fractalfract7050386>

Academic Editor: Carlo Cattani

Received: 28 March 2023

Revised: 27 April 2023

Accepted: 29 April 2023

Published: 6 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The interest in new probability discrete and continuous distributions is increasing; see, for example, [1]. Recently, this interest has been present in distributions defined on the interval $[0, 1]$, which has raised great attention from researchers. These distributions have many applications in several fields of science; see, for example, [2–7]. Often, novel unit models are transported to unit intervals by transformations of well-known models [2], with studies on unit modeling increasing day by day [2–7].

Regression is a statistical approach that is widely employed [8] to explain the structure between a dependent (response) variable and independent (explanatory) variables or covariates. Such a structure can be linear or nonlinear, associating the response mean with the independent variables. When the dependent variable is stated on the interval $[0, 1]$, a regression structure employing the beta distribution [9] has been the most preferred method to explain the unit-dependent variable with the independent variables based on the conditional mean. Note that the beta model is formulated in relation to its mean. Then, its unit regression based on the mean was established in [9].

With similar ideas, alternative unit mean response regressions were introduced, among them are the beta rectangular [10], Birnbaum–Saunders [11], log-Bilal [12], log-Lindley [1], log-weighted exponential [13], and unit Lindley [14] models.

The dependent variables can be affected by situations related to atypical data or asymmetric distributions. When these situations exist, the mean is also affected strongly, and the inference from the model may give rise to possible inaccurate interpretations.

A fractile is a point on a probability density function (PDF) where the surface upon the function is a fixed proportion. Due to the inaccuracy mentioned, the quantile regression (QR) models are a robust alternative to mean regression models. The QR model was introduced in [15] and relates the response's conditional quantiles to independent variables' given values instead of explaining their conditional mean. One of the key benefits of using QR models over standard mean regressions is its lack of imposed distributional assumptions on the error term. This allows for greater flexibility in modeling different types of data [16].

To implement a QR model utilizing a probability distribution, it is necessary to parameterize it using its quantile function (QF). Fortunately, this parameterization can be utilized for any probability distribution with a closed-form QF, regardless of whether its mean has a closed form. This provides a high level of adaptability and enables us to model quantiles based on various distributions, as demonstrated in the following examples: the Kumaraswamy (KW) [17], L-logistic [18], log-extended exponential-geometric (LEEG) [19], log-symmetric [20], power Johnson SB, [21], unit-Birnbaum–Saunders [22], unit-Burr-XII, [2,23], unit-Chen [24], unit-Weibull [25], and Weibull–Marshall–Olkin [26] QR models. These types of models and structures have been largely applied to describe COVID-19 mortality [23,27].

The continuous Bernoulli (CB) model, as described in [4], is a probability distribution that exists on the interval $[0, 1]$ and has a single parameter related to the shape of the distribution. It belongs to the exponential family of distributions and offers a generalization of the uniform distribution defined on the unit interval. In their study, the authors of [4] proposed using the CB distribution to investigate the impact of Bernoulli variational autoencoders on intensity data that fall within the $[0, 1]$ range. The results of their research indicated that the CB distribution leads to significant improvements in various metrics and datasets, including the generation of sharper image samples, and suggested a wider range of potential performance-optimized variational autoencoders. Despite the promising findings of the CB distribution, as far as we know, its associated mathematical and statistical properties as well as its associated QR model have yet to be explored. Thus, the primary objectives of this article are as follows:

- (i) To derive some mathematical and statistical properties that have not been studied of the CB distribution stated on the interval $[0, 1]$, including estimation and inference.
- (ii) To formulate a QR model for a unit response employing the exponentiated CB (ECB) distribution. In this way, the applications of the CB distribution are extended. The formulated regression model is a robust alternative to the beta regression and other models proposed in the literature. Another motivation for the proposed QR model comes from a real-world dataset.

The present article is organized into seven main sections. Section 2 is dedicated to deriving mathematical properties of the continuous Bernoulli (CB) distribution, while Section 3 focuses on the various point estimation procedures for its parameters. A novel QR model is introduced in Section 4. Monte Carlo simulations are used to evaluate the performance of point and interval estimators for both the CB distribution and the associated QR model in Section 5. The results are illustrated utilizing a real-world dataset in Section 6. Finally, Section 7 sketches some concluding remarks.

2. The Continuous Bernoulli Distribution

This section provides some known and new functions of the CB distribution.

2.1. Cumulative Distribution and Probability Density Functions

In the fields of artificial intelligence, in the context of modeling pixel intensities, the CB distribution is a commonly encountered concept [28]. The PDF and cumulative distribution function (CDF) of the CB distribution are formulated as

$$F(x, \beta) = \begin{cases} x, & \beta = 1/2; \\ (1 - \beta) \left(\left(\frac{\beta}{1 - \beta} \right)^x - 1 \right) \frac{1}{(2\beta - 1)}, & \beta \neq 1/2; \end{cases} \quad (1)$$

and

$$f(x, \beta) = \begin{cases} 1, & \beta = 1/2; \\ (1 - \beta) \left(\frac{\beta}{1 - \beta} \right)^x \frac{1}{(2\beta - 1)} \log(\beta/(1 - \beta)), & \beta \neq 1/2; \end{cases}$$

respectively, with $x \in (0, 1)$, and $\beta \in (0, 1)$ being a parameter related to the shape of the distribution. Alternatively, the PDF of the CB distribution can be written as

$$f(x, \beta) = \begin{cases} 1, & \beta = 1/2; \\ \frac{2(1 - \beta) \tanh^{-1}(1 - 2\beta)}{(1 - 2\beta)} (\beta/(1 - \beta))^x, & \beta \neq 1/2; \end{cases}$$

where $\tanh^{-1}(w) = (1/2) \log((1 + w)/(1 - w))$, for $w \in (-1, 1)$, is the inverse hyperbolic tangent function. We denote the CB distribution of parameter β as $CB(\beta)$. Note that the CB distribution corresponds to the uniform distribution on the interval $[0, 1]$, that is, on $(0, 1)$, when $\beta = 1/2$. Observe that there is an interesting recursive relation between the CB PDF and its derivatives. For $\beta \neq 1/2$, this relation is stated as

$$\frac{d^m f(x, \beta)}{dx^m} = f(x, \beta) \log^{m+1} \left(\frac{\beta}{1 - \beta} \right), \quad m \in \{1, 2, \dots\}.$$

Hence, it can be concluded that the PDF with a concave CDF is decreasing for $\beta < 1/2$, and the PDF with a convex CDF is increasing for $\beta > 1/2$. Note that the term $\sum_{i=1}^n X_i$ is a sufficient statistic for estimating β based on a sample of size n , X_1, \dots, X_n namely, collected from the CB distribution.

2.2. Inverse Cumulative Distribution Function

From (1), we can obtain the CB inverse CDF, which is given by

$$F^{-1}(x, \beta) = \begin{cases} x, & \beta = 1/2; \\ \frac{\log \left(1 + \left(\frac{2\beta - 1}{1 - \beta} \right) x \right)}{\log \left(\frac{\beta}{1 - \beta} \right)}, & \beta \neq 1/2. \end{cases} \quad (2)$$

In Figure 1, we can see the plot of (2) for the values $\beta = 0.25$, $\beta = 0.5$, and $\beta = 0.75$. To simulate a random variable (RV) with CB distribution, we propose Algorithm 1. In simulation studies related to the CB distribution, generating data from this distribution is important. Then, we can use Algorithm 1 for this purpose, which takes advantage of (2) to generate a sample of size n .

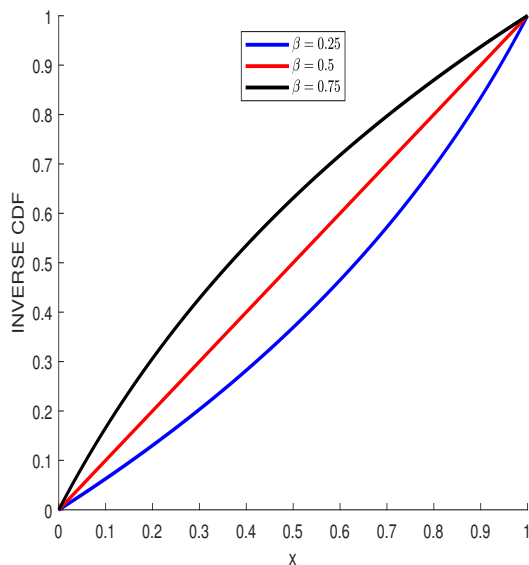


Figure 1. The CB inverse CDF for the indicated values of β .

Algorithm 1: Approach that simulates the CB distribution using the inverse-transform method

```

begin
  input :  $n, \beta$ 
  output:  $n\_vector, \bar{x}$  (sample mean),  $s^2$  (sample variance)
  1. Obtain a value  $r$  from  $U \sim \text{continuous uniform}(0, 1)$ ;
  2. Append, using (2),  $F^{-1}(r, \beta)$  to  $n\_vector$ ;
  3. Repeat steps 1–2 until length of  $n\_vector$  is equal to  $n$ ;
  4. Compute  $\bar{x}$  and  $s^2$  with the values  $n\_vector$ .
end
    
```

Computational experiments were conducted utilizing Algorithm 1. Table 1 displays the results generated with $\beta = 0.25$, while Table 2 provides the results generated with $\beta = 0.75$, using $n \in \{25, 50, 100, 200, 500, 1000\}$ and performing 1000 Monte Carlo simulations for each n .

The simulation results demonstrate the experimental convergence of Algorithm 1. The theoretical mean and variance values for the CB distribution with $\beta = 0.25$ are 0.4102 and 0.0785354, respectively, while the theoretical mean and variance values with $\beta = 0.75$ are 0.5898 and 0.0785354, respectively.

Table 1. Simulation results with $\beta = 0.25$.

n	Runtime (in seconds)	\bar{x}	s^2
25	0.0116	0.4106	0.0780877
50	0.0207	0.4110	0.0788680
100	0.0352	0.4101	0.0783473
200	0.0513	0.4098	0.0785829
500	0.1273	0.4104	0.0784173
1000	0.1951	0.4102	0.0784991

Table 2. Simulation results with $\beta = 0.75$.

n	Runtime (in seconds)	\bar{x}	s^2
25	0.0197	0.5895	0.0784745
50	0.0224	0.5897	0.0780702
100	0.0315	0.5885	0.0788916
200	0.0459	0.5887	0.0786282
500	0.0955	0.5894	0.0788145
1000	0.2136	0.5897	0.0784867

2.3. Hazard Rate Function

The hazard rate function (HRF) of the CB distribution is stated as

$$h(x, \beta) = \frac{(1 - \beta)^{1-x} \beta^x}{\beta - (1 - \beta)^{1-x} \beta^x} \log(\beta/(1 - \beta)), \quad x \in (0, 1), \quad \beta \neq 1/2.$$

The HRF is an increasing function in x because its derivative expressed as

$$\frac{dh(x, \beta)}{dx} = \frac{\beta(1 - \beta)}{(\beta - (1 - \beta)(\beta/(1 - \beta))^x)^2} (\beta/(1 - \beta))^x \log^2(\beta/(1 - \beta))$$

is positive for all β . Furthermore, note that $\lim_{x \rightarrow 1} h(x, \beta) = \infty$ and

$$\lim_{x \rightarrow 0} h(x, \beta) = \begin{cases} 1, & \beta = 1/2; \\ \frac{(1-\beta)}{2\beta-1} \log(\beta/(1 - \beta)), & \beta \neq 1/2. \end{cases}$$

Figure 2 displays the CB PDF, CDF, and HRF. From this figure, note the flexible shapes these functions have for the different parameter values, that is, for β .

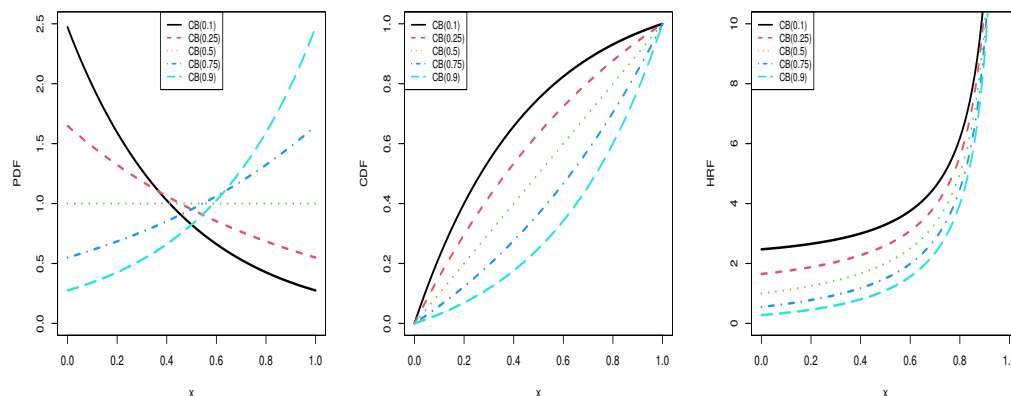


Figure 2. Plots of the CB PDF (left), CDF (center), and HRF (right) for the listed values of β .

2.4. Moments

Using the definition of the moment-generating function, for the case of the CB distribution, we obtain that

$$M_X(t) = E(e^{tX}) = \frac{(\beta(1 + e^t) - 1) \log(\beta/(1 - \beta))}{(2\beta - 1)(\log(\beta/(1 - \beta)) + t)}, \quad \beta \neq 1/2, t \in \mathfrak{R}.$$

Note that $M_X(t) = (e^t - 1)/t$ for $\beta = 1/2$. Thus, utilizing the relation given by $E(X^r) = d^r M_X(t)/dt^r|_{t=0}$, for $r \in \{1, 2, \dots\}$, the first four noncentral moments are, respectively, for $\beta \neq 1/2$, stated as

$$\begin{aligned}
 m_1 = E(X) &= \frac{\beta \log(\beta/(1-\beta)) + 1 - 2\beta}{(2\beta - 1) \log(\beta/(1-\beta))}, \\
 m_2 = E(X^2) &= \frac{\beta \log^2 \beta - 2\beta \log \beta \log(1-\beta) + \beta \log^2(1-\beta) - 2\beta \log \beta + 2\beta \log(1-\beta) + 4\beta - 2}{(2\beta - 1) \log^2(\beta/(1-\beta))}, \\
 m_3 = E(X^3) &= \frac{-\beta \log^3(1-\beta) + 3\beta(\log \beta - 1) \log^2(1-\beta) - 3\beta(\log^2 \beta + 2 - 2 \log \beta) \log(1-\beta)}{(2\beta - 1) \log^3(\beta/(1-\beta))} \\
 &\quad + \frac{\beta \log^3 \beta - 12\beta + 6\beta \log \beta - 3\beta \log^2 \beta + 6}{(2\beta - 1) \log^3(\beta/(1-\beta))}, \\
 m_4 = E(X^4) &= \frac{\beta \log^4 \beta - 4\beta \log^3 \beta \log(1-\beta) + 6\beta \log^2 \beta \log^2(1-\beta) - 4\beta \log^3(1-\beta) \log \beta}{(2\beta - 1) \log^4(\beta/(1-\beta))} \\
 &\quad + \frac{\beta \log^4(1-\beta) - 4\beta \log^3 \beta + 12\beta \log^2 \beta \log(1-\beta) - 12\beta \log^2(1-\beta) \log \beta + 4\beta \log^3(1-\beta) + 12\beta \log^2 \beta}{(2\beta - 1) \log^4(\beta/(1-\beta))} \\
 &\quad + \frac{-24\beta \log \beta \log(1-\beta) + 12\beta \log^2(1-\beta) - 24\beta \log \beta + 24\beta \log(1-\beta) + 48\beta - 24}{(2\beta - 1) \log^4(\beta/(1-\beta))}.
 \end{aligned}$$

Figure 3 shows some shapes of the CB moments. Note that, as β increases, the mean increases, whereas the skewness decreases. According to the values of β , the variance increases for $\beta < 1/2$ and decreases for $\beta > 1/2$, but a reverse situation is detected for the kurtosis.

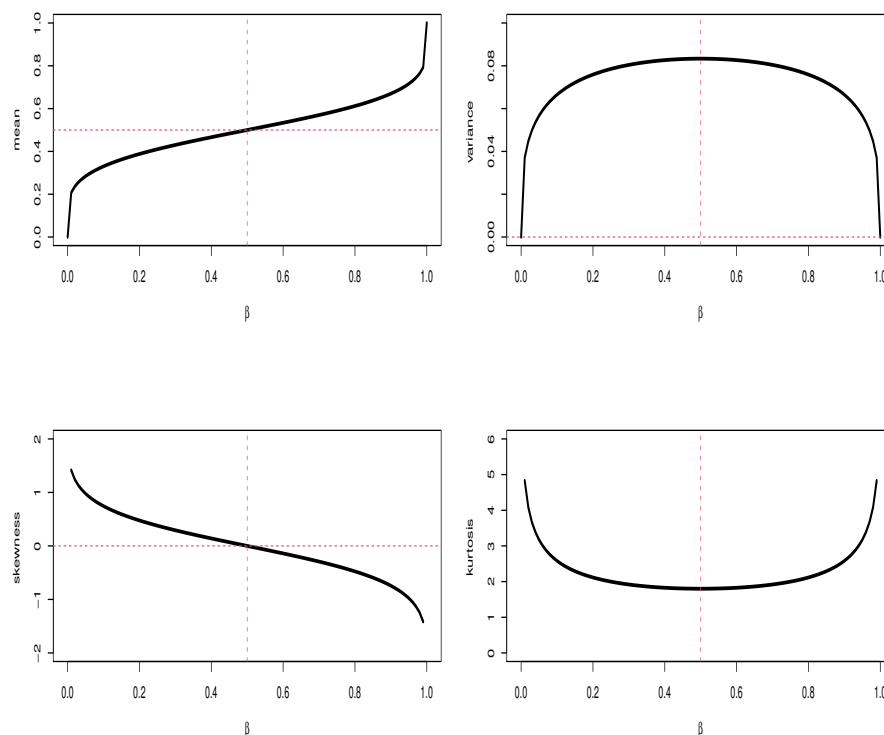


Figure 3. Plots of the mean, variance, skewness, and kurtosis of the CB distribution for the indicated values of β .

2.5. Quantile Function

The QF, defined by $F^{-1}(q, \beta) = x_q(\beta)$, for $q \in (0, 1)$, of the CB distribution can be obtained in a closed form. Let X be a CB distributed RV with CDF as defined in (1). Then, the CB QF is given by

$$x_q(\beta) = \frac{\log\left(1 + \frac{(2\beta - 1)q}{1 - \beta}\right)}{\log(\beta/(1 - \beta))}, \quad \beta \neq 1/2, q \in (0, 1). \tag{3}$$

Note that the CB QF is equal to q for $\beta = 1/2$, whereas the corresponding median is derived as

$$x_{0.5}(\beta) = (\log(\beta/(1 - \beta)))^{-1} \log\left(\frac{0.5}{1 - \beta}\right),$$

for $\beta \neq 1/2$. Notice that if the RV Q related to q in (3) has a uniform distribution on $(0, 1)$, which is denoted as $U(0, 1)$, then the RV associated with $x_Q(\beta)$ has the CB distribution.

2.6. Mean Residual Life Function

The mean residual life (MRL), also known as life expectancy, is a critical parameter in the field of reliability that characterizes the assumed distribution. Unlike the hazard rate function (HRF), which only considers the risk of immediate failure, the MRL summarizes the entire residual life distribution. The MRL function of a positive random variable at time x is defined as

$$\text{MRL}(x) = \frac{1}{1 - F(x)} \int_x^\infty (1 - F(y))dy,$$

where F is the lifetime model CDF. Therefore, the MRL function of the CB distribution is stated as

$$\text{MRL}(x) = \frac{\beta(1 - x) \log(\beta/(1 - \beta)) - \beta(1 + (\beta/(1 - \beta))^x) + (\beta/(1 - \beta))^x}{(\beta - (1 - \beta)(\beta/(1 - \beta))^x) \log(\beta/(1 - \beta))}, \beta \neq 1/2,$$

while $\text{MRL}(0)$ is obtained as

$$\text{MRL}(0) = \frac{\beta \log(\beta/(1 - \beta)) + 1 - 2\beta}{(2\beta - 1) \log(\beta/(1 - \beta))}, \beta \neq 1/2,$$

which is equal to the corresponding expected value, and then $\text{MRL}(1) = 0$. Observe that $\text{MRL}(x) = (1 - x)/2$, for $\beta = 1/2$.

We can comment about the shapes of the MRL function via its HRF. The following results hold for both discrete and continuous lifetimes [29]:

- If the HRF is (strictly) increasing, then the MRL is (strictly) decreasing.
- If the HRF is (strictly) decreasing, then the MRL is (strictly) increasing.
- The HRF is a constant function (exponential or geometric distribution) if and only if the MRL is a constant.

We can conclude that, since the HRF function of the CB distribution is strictly increasing, its MRL function is strictly decreasing. For detail about the MRL, see [30].

2.7. Incomplete Moments and Lorenz Curve

For $\beta \neq 1/2$ and $r \in \{1, 2, \dots\}$, the r th incomplete moment of the CB distribution is given by

$$\begin{aligned} I(x, r) &= \int_0^x z^r f(z, \beta) dz \\ &= \frac{x^r(1 - \beta)(r\Gamma(r, -x \log(\beta/(1 - \beta)))(-x \log(\beta/(1 - \beta)))^{-r} - \Gamma(r + 1)(-x \log(\beta/(1 - \beta)))^{-r} + (\beta/(1 - \beta))^x)}{2\beta - 1}, \end{aligned}$$

for $z < x$, where $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ are the gamma function and upper incomplete gamma function, respectively. For $\beta = 1/2$, the CB r th incomplete moment is equal to $x^{r+1}/(r + 1)$. Note that the equations $\lim_{x \rightarrow \infty} I(x, r) = E(X^r)$ and $F(y, \beta) = I(x, r = 0)$ hold.

As a result of the incomplete moments, the Lorenz curve of the CB distribution for $\beta \neq 1/2$ is expressed by

$$\begin{aligned} L(p) &= \frac{1}{E(X)} \int_0^{x_q(\beta)} x f(x, \beta) dx \\ &= \frac{I(x_q(\beta), 1)}{E(X)} = \frac{(\beta(2p-1) + 1 - p) \log\left(\frac{1+\beta(2p-1)-p}{1-\beta}\right) - 2p\beta + p}{1 - 2\beta + \beta \log(\beta/(1-\beta))}. \end{aligned}$$

Then, $L(p) = p^2$ is obtained for $\beta = 1/2$.

2.8. Stress Strength Reliability

The system with applied stress works smoothly if its strength is higher than the applied stress. The stress–strength parameter of a life distribution is an important indicator of reliability. This parameter is the working probability of the system and a reliability measure, which is defined as

$$R = P(X < Y) = \int_0^1 f_X(x) F_Y(x) dx,$$

where f_X is the PDF of the stress (X) and F_Y is the CDF of the strength (Y). Now, consider X and Y as two independent CB distributed RVs with parameters β_1 and β_2 , respectively. Then, the CB stress–strength parameter is given by

$$R = \frac{(2\beta_1(\beta_2-1)-\beta_2+1) \log\left(\frac{\beta_2}{1-\beta_2}\right) + (2\beta_2-1)\beta_1 \log\left(\frac{\beta_1}{1-\beta_1}\right)}{(2\beta_2-1)(2\beta_1-1) \log\left(\frac{\beta_1}{1-\beta_1} \frac{\beta_2}{1-\beta_2}\right)}.$$

Note that, when $\beta_1 = \beta_2 \neq 1/2$, then $R = 1/2$.

3. Point Estimation for the Parameter of the CB Distribution

This section provides some point estimation methods, such as maximum likelihood (ML), moment, percentile (PE), least square (LS), weighted square (WS), Anderson–Darling (AD), and Cramér–von Mises (CM) to determine the model parameter β . Note that, for the estimation processes of the above methods, we focus on $\beta \neq 1/2$.

3.1. Maximum Likelihood Method

Consider X_1, \dots, X_n as a sample of size n collected from the CB distribution, with x_1, \dots, x_n as its observed values. Thus, the associated log-likelihood function is established as

$$\ell(\beta) = n \log(1 - \beta) - n \log(2\beta - 1) + n \log(\log(\beta/(1 - \beta))) + \log(\beta/(1 - \beta)) \sum_{i=1}^n x_i. \quad (4)$$

Taking the first derivative of the expression stated in (4) with respect to β , we obtain

$$\frac{d\ell(\beta)}{d\beta} = -\frac{n}{\log(1 - \beta)} - \frac{2n}{\log(2\beta - 1)} + \frac{n}{\beta(1 - \beta) \log(\beta/(1 - \beta))} + \frac{1}{\beta(1 - \beta)} \sum_{i=1}^n x_i. \quad (5)$$

Since the equation $d\ell(\beta)/d\beta = 0$ obtained from (5) is a nonlinear function of β , the ML estimator $\hat{\beta}$ must be computed using numerical methods. For any smooth PDF, it is widely recognized that under regularity conditions, $\hat{\beta}$ has a normal distribution asymptotically with mean β and variance \mathcal{I}_β^{-1} , with $\mathcal{I}_\beta = -E(d^2\ell(\beta)/d\beta^2)$ being the Fisher information.

After performing several computations, the inverse of the Fisher information can be obtained as

$$\mathcal{I}_\beta^{-1} = \frac{(\beta(1 - \beta)(2\beta - 1) \log(\beta/(1 - \beta)))^2}{n(1 - \beta(1 - \beta)(4 + \log^2(\beta/(1 - \beta))))}. \quad (6)$$

Based on the previous results, the asymptotic $100(1 - \theta)\%$ confidence interval of the parameter β is given as

$$\hat{\beta} - z_{\theta/2} \frac{\beta(1-\beta)(2\beta-1) \log(\beta/(1-\beta))}{\sqrt{n(1-\beta(1-\beta)(4+\log^2(\beta/(1-\beta))))}} \leq \beta \leq \hat{\beta} + z_{\theta/2} \frac{\beta(1-\beta)(2\beta-1) \log(\beta/(1-\beta))}{\sqrt{n(1-\beta(1-\beta)(4+\log^2(\beta/(1-\beta))))}},$$

where $z_{\theta/2}$ refers to the $(\theta/2) \times 100$ th upper percentile of the standard normal distribution.

3.2. Moment Estimation

The parameter estimation for the CB distribution can be performed using the moment method, as its expected value can be expressed analytically. This method involves equating the first moment of the sample to the expected value of the CB distribution. By solving this equation, the moment estimate of the parameter β can be obtained as

$$\bar{x} - \frac{\beta \log(\beta/(1-\beta)) + 1 - 2\beta}{(2\beta - 1) \log(\beta/(1-\beta))} = 0, \tag{7}$$

where \bar{x} is the sample mean. Note that the equation stated in (7) must be solved employing a numerical method.

3.3. Percentile Estimation

The percentile method to estimate β in the CB distribution was originally proposed in [31]. Since the QF of the CB distribution has a closed form, this method can be applied straightforwardly. Consider $X_{(1)}, \dots, X_{(n)}$ as the order statistics of a sample of size n from the CB distribution and $x_{(1)}, \dots, x_{(n)}$ as their observed values. An estimate of the CDF can be obtained as $\hat{F}(x_{(i)}) = i/(n + 1)$, for $i \in \{1, \dots, n\}$. Then, an estimate of β can be reached by minimizing a certain function with respect to β . The expression for this function is given by

$$PE(\beta) = \sum_{i=1}^n \left(x_{(i)} - x_{q_i}(\beta) \right)^2, \tag{8}$$

and then the PE estimate of β is attained, where x_q is defined in (3). Therefore, the PE estimate of β is the solution on β evaluated at $\hat{\beta}$ of an equation formulated as

$$\begin{aligned} \frac{dPE(\beta)}{d\beta} &= \frac{2}{\beta(1-\beta) \log^2(\beta/(1-\beta))} \sum_{i=1}^n \left(x_{(i)} - \frac{\log\left(1 + \frac{i(2\beta-1)}{(n+1)(1-\beta)}\right)}{\log(\beta/(1-\beta))} \right) \\ &\times \frac{(\beta((2i-1-n)) + n - i + 1) \log\left(1 + \frac{i(2\beta-1)}{(n+1)(1-\beta)}\right) - i\beta \log(\beta/(1-\beta))}{(n+1)(1-\beta) + i(2\beta-1)} = 0. \end{aligned}$$

3.4. Other Methods of Estimation

Next, the LS, WS, AD, and CM methods of estimation for β are provided. According to these methods, the functions to be used for estimating β are expressed as

$$\begin{aligned} LS(\beta) &= \sum_{i=1}^n \left(F(x_{(i)}, \beta) - \frac{i}{n+1} \right)^2, \\ WS(\beta) &= \sum_{i=1}^n \frac{(n+2)(n+1)^2}{(n-i+1)i} \left(F(x_{(i)}, \beta) - \frac{i}{n+1} \right)^2, \\ LS(\beta) &= -n - \sum_{i=1}^n \frac{(2i-1)}{n} \left(\log\left(1 - F(x_{(n+1-i)}, \beta)\right) + \log(F(x_{(i)}, \beta)) \right), \\ CM(\beta) &= \frac{1}{12n} + \sum_{i=1}^n \left(F(x_{(i)}, \beta) - \frac{2i-1}{2n} \right)^2, \end{aligned}$$

In the CB distribution, a sample of size n is drawn from the population, with $x_{(1)}, \dots, x_{(n)}$ representing the ordered observed values of this sample. Therefore, the LS, WS, AD, and CM estimates of β are given by

$$\hat{\beta}_{\text{LS}} = \underset{\beta}{\operatorname{argmin}}\{\text{LS}(\beta)\}, \quad (9)$$

$$\hat{\beta}_{\text{WS}} = \underset{\beta}{\operatorname{argmin}}\{\text{WS}(\beta)\}, \quad (10)$$

$$\hat{\beta}_{\text{AD}} = \underset{\beta}{\operatorname{argmin}}\{\text{AD}(\beta)\}, \quad (11)$$

$$\hat{\beta}_{\text{CM}} = \underset{\beta}{\operatorname{argmin}}\{\text{CM}(\beta)\}, \quad (12)$$

respectively. It is important to emphasize that the equations presented in (4), (7), (8), (9), (10), (11) and (12) do not have explicit solutions and must be solved through numerical methods. However, these equations can be optimized directly using software such as R or MatLab.

4. Associated QR Model

This section states a QR model employing the exponentiated CB distribution.

4.1. The Exponentiated CB Distribution

Although the CB distribution has a closed-form mean function and QF, its PDF and CDF cannot be parameterized using these functions due to the dependence of the parameter β . Therefore, we propose an alternative QR model that is based on the ECB distribution, which is a generalized form of the CB distribution. The ECB CDF and PDF are expressed as

$$G(y, \alpha, \beta) = \begin{cases} y^\alpha, & \beta = 1/2; \\ \left(\frac{(1-\beta)((\beta/(1-\beta))^y - 1)}{(2\beta-1)} \right)^\alpha, & \beta \neq 1/2; \end{cases}$$

and

$$g(y, \alpha, \beta) = \begin{cases} \alpha y^{\alpha-1}, & \beta = 1/2; \\ \frac{\alpha(\beta/(1-\beta))^y \log(\beta/(1-\beta))}{(\beta/(1-\beta))^y - 1} \left(\frac{(1-\beta)((\beta/(1-\beta))^y - 1)}{(2\beta-1)} \right)^\alpha, & \beta \neq 1/2; \end{cases}$$

respectively, where $y \in (0, 1)$, $\alpha > 0$ is a parameter related to the shape of the distribution and $\beta \in (0, 1)$. The CDF of the ECB distribution is $G(y, \alpha, \beta) = (F(y, \beta))^\alpha$, and it can be named a Lehmann-type I CB distribution. For $\beta = 1/2$, the ECB distribution is the power function distribution.

The $\tau \times 100$ th quantile of the ECB distribution is presented by

$$Q(\tau, \alpha, \beta) = \begin{cases} \tau^{1/\alpha}, & \beta = 1/2; \\ \left(\log(\beta/(1-\beta)) \right)^{-1} \log \left(1 + \frac{(2\beta-1)\tau^{1/\alpha}}{1-\beta} \right), & \beta \neq 1/2; \end{cases}$$

where $\tau \in (0, 1)$.

By using the QF of the ECB distribution, we can now reparameterize it in terms of its quantiles. Observe that, as $\beta = 1/2$, the ECB distribution can be reparameterized as a particular case of the Kumaraswamy distribution based on the quantile form, as presented in [17]. Here, we consider $\beta \neq 1/2$ and provide the PDF form of the ECB distribution based on its QF.

Consider $\mu = Q(\tau, \alpha, \beta)$ and $\alpha = \log(\tau(\log((1 - \beta)((\beta/(1 - \beta))^\mu - 1)/(2\beta - 1)))^{-1})$. Hence, the reparameterized ECB CDF and PDF are introduced by

$$G(y, \beta, \mu, \tau) = \tau^{\log((1-\beta)((\beta/(1-\beta))^y-1)/(2\beta-1))} (\log((1-\beta)((\beta/(1-\beta))^\mu-1)/(2\beta-1)))^{-1} \tag{13}$$

and

$$g(y, \beta, \mu, \tau) = \frac{\log \tau (\beta/(1-\beta))^y \log(\beta/(1-\beta)) \tau^{\log((1-\beta)((\beta/(1-\beta))^y-1)/(2\beta-1))} (\log((1-\beta)((\beta/(1-\beta))^\mu-1)/(2\beta-1)))^{-1}}{\log((1-\beta)((\beta/(1-\beta))^\mu-1)/(2\beta-1)) ((\beta/(1-\beta))^y-1)}. \tag{14}$$

The reparameterized PDF stated in (14) is defined by the CDF presented in (13), where $\beta \in (0, 1)$ is a parameter related to the shape of the distribution, $\mu \in (0, 1)$ is the quantile parameter, and $\tau \in (0, 1)$ is a known quantile level. We denote the RV associated with the CDF given in (13) as $Y \sim \text{ECB}(\beta, \mu, \tau)$. Figure 4 illustrates the diverse shapes of the reparameterized PDF, which exhibit characteristics such as U-shape, unimodality, increasing, decreasing, and constancy.

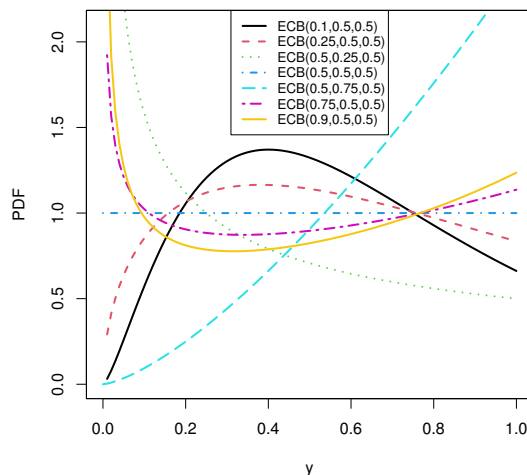


Figure 4. Plots of the PDF stated in (14) for the indicated values of parameters.

4.2. ECB Quantile Regression Model

Once the reparameterized ECB distribution is defined, a QR model can be introduced. Suppose we have n observations y_1, \dots, y_n from the reparameterized distribution with its PDF stated in (14). Then, the ECB QR model can be expressed as

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}. \tag{15}$$

Here, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$ is the vector of independent variable values, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ is the vector of unknown regression coefficients, and $g: (0, 1) \rightarrow \mathcal{R}$ is a strictly monotonic and twice differentiable link function used to relate independent variables with the conditional quantile of the dependent variable. Common link functions include the logit, probit, log-log, and Cauchy structures. Thus, μ_i is obtained by inverting the model given in (15) such that $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$, for $i \in \{1, \dots, n\}$. In this proposed regression model, we use the logit function for linking, stated as $g(\mu_i) = \log(\mu_i/(1 - \mu_i))$, for $i \in \{1, \dots, n\}$. If we take $\tau = 0.5$, the link function connects the median response to the independent variables in the ECB QR model.

4.3. ML Estimation of the Regression Coefficients

Assume that we have n independent RVs Y_1, \dots, Y_n following the ECB distribution, with β being a parameter related to the shape of the distribution, μ_i a quantile parameter,

and τ a known quantile level. The quantile parameter is obtained using the logistic function formulated as

$$\mu_i = \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}}, \tag{16}$$

where x_i is the vector of independent variable values for the i th observation, and β is the vector of unknown regression coefficients. The parameter vector $\theta = (\beta, \beta^\top)$ has $p + 2$ unknown elements. The log-likelihood function of the ECB regression model is expressed as

$$\begin{aligned} \ell(\theta) = & n \log(\log(\tau)) + n \log(\log(\beta/(1-\beta))) + \log(\beta/(1-\beta)) \sum_{i=1}^n y_i \\ & - \sum_{i=1}^n \log\left(\log\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right)\right) - \sum_{i=1}^n \log((\beta/(1-\beta))^{y_i} - 1) \\ & + \log(\tau) \sum_{i=1}^n \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{y_i}-1)}{(2\beta-1)}\right) \left(\log\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right)\right)^{-1}. \end{aligned} \tag{17}$$

Thus, the first derivatives according to the model coefficients established in (17) are given by

$$\begin{aligned} \frac{d\ell(\theta)}{d\beta_r} = & - \sum_{i=1}^n \frac{(\beta/(1-\beta))^{\mu_i} \mu_i (1-\mu_i) x_{ir} \log(\beta/(1-\beta))}{((\beta/(1-\beta))^{\mu_i}-1) \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right)}, \quad r \in \{0, 1, \dots, p\}, \\ & - \sum_{i=1}^n \frac{(\beta/(1-\beta))^{\mu_i} \mu_i (1-\mu_i) x_{ir} \log(\tau) \log(\beta/(1-\beta))}{\log^2\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right) ((\beta/(1-\beta))^{\mu_i}-1)} \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{y_i}-1)}{(2\beta-1)}\right), \\ \frac{d\ell(\theta)}{d\beta} = & \frac{n}{\beta(1-\beta) \log(\beta/(1-\beta))} + \frac{1}{\beta(1-\beta)} \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{y_i (\beta/(1-\beta))^{y_i}}{\beta(1-\beta) ((\beta/(1-\beta))^{y_i} - 1)} \\ & - \sum_{i=1}^n \frac{(\beta/(1-\beta))^{\mu_i} ((2\beta-1)\mu_i - \beta) + \beta}{\beta(1-\beta)(2\beta-1) ((\beta/(1-\beta))^{\mu_i} - 1) \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right)} \\ & + 2 \log(\tau) \sum_{i=1}^n \frac{\left((\beta(y_i - 1/2) - \frac{y_i}{2})(\beta/(1-\beta))^{y_i} + \frac{\beta}{2}\right)}{\beta(2\beta-1)(1-\beta) ((\beta/(1-\beta))^{y_i} - 1) \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right)} \\ & - 2 \log(\tau) \sum_{i=1}^n \frac{\left((\beta(\mu_i - 1/2) - \frac{\mu_i}{2})(\beta/(1-\beta))^{\mu_i} + \frac{\beta}{2}\right) \log\left(\frac{(1-\beta)((\beta/(1-\beta))^{y_i}-1)}{(2\beta-1)}\right)}{\beta(2\beta-1)(1-\beta) \log^2\left(\frac{(1-\beta)((\beta/(1-\beta))^{\mu_i}-1)}{(2\beta-1)}\right) ((\beta/(1-\beta))^{\mu_i} - 1)}. \end{aligned}$$

Once the derivatives are set to zero, the resulting equations should be solved using numerical methods since they are nonlinear functions. The formula presented in (17) can be directly maximized utilizing the R software to obtain these solutions. It is important to note that the asymptotic distribution of $(\hat{\theta} - \theta)$ is a multivariate normal $N_{p+2}(\mathbf{0}, \mathcal{I}^{-1}(\theta))$, where $\mathcal{I}(\theta)$ is the Fisher information matrix. However, for numerical applications, the $(p + 2) \times (p + 2)$ observed information matrix is often used instead of $\mathcal{I}(\theta)$ to make inference about the corresponding parameters.

4.4. Residual Analysis for Model Fitting

Validating a fitted regression model involves analyzing residuals. The randomized quantile (RQ) residual is a preferred method for this validation. The RQ residual, proposed in [32], is defined as

$$\hat{r}_i = \Phi^{-1}(G(y_i, \hat{\beta}, \hat{\mu}_i, \tau)), \quad i \in \{1, \dots, n\},$$

where $G(y, \beta, \mu, \tau)$ is the CDF of the CEB model and $\hat{\mu}_i$ is defined in (16). If the fitted regression is valid, the RQ quantile should be standard normal distributed.

5. Simulation Studies

In this section, we discuss simulation studies for point estimation of the CB parameter and estimation of the ECB QR parameters.

5.1. Computational Framework

The experiments were conducted on a computer with Windows 10 for 64 bits, 8 Giga-bytes of RAM, and an Intel Core i7-4510U 2-2.60 GHz processor. The optimization of the log-likelihood function can be performed directly by software such as R or Matlab.

In the case of using R, two libraries, named `ConstrOptim` and `maxLik`, can be employed to conduct the experimental results of the simulation studies and real data analysis when obtaining the point estimates of the corresponding parameters. The functions of these libraries numerically provide the parameter estimates and the associated estimated standard errors based on diverse algorithms. The Nelder–Mead optimization routine can be utilized to obtain these estimates. Such libraries may only be employed in specific contexts, which is their limitation. Our computations were carried out by the `maxLik` and `gofTest` functions of the R software.

For more details about R packages related to QR, the interested reader is referred to a package named `unitquantreg`, which was developed in the setting `stats::lm` package [33]. The `unitquantreg()` function of this package is flexible and allows us, by means of the `ns()` function, to utilize regression splines. This function is available in the `splines` package. Moreover, it is possible to employ the `gam()` function of an R package named `mgcv`.

The `unitquantreg` package can be secured from <https://github.com/AndrMenezes/unitquantreg> (accessed on 5 April 2023) and installed using

```
devtools::install_github('AndrMenezes/unitquantreg')
```

The `unitquantreg` package contains 15 distributions, and our new distribution can be added to this package.

5.2. Point Estimation

For the simulations of point estimation, we change n in $\{25, 50, 100, 200, 500, 1000\}$ with $M = 1000$ Monte Carlo replicates from the CB distribution, with $\beta \in \{0.25, 0.75\}$. The empirical mean, bias, and mean square error (MSE) of the estimators, given by $\text{Mean}(\hat{\beta}) = (1/M) \sum_{i=1}^M \hat{\beta}_i$, $\text{Bias}(\hat{\beta}) = (1/M) \sum_{i=1}^M (\beta - \hat{\beta}_i)$, and $\text{MSE}(\hat{\beta}) = (1/M) \sum_{i=1}^M (\beta - \hat{\beta}_i)^2$, were used to calculate statistical indicators. The results of the estimators are close to the true values of the parameters, and consistent, with MSE and bias decreasing as n increases. The moment estimate was used as the initial value for the ML estimates. The results are shown in Figure 5.

Simulation studies also were conducted to evaluate the behavior of 95% confidence intervals using the ML method. The coverage length (CL) and coverage probability (CP) were empirically calculated to assess this behavior. The CL and CP are defined by

$$\text{CL}(\hat{\beta}) = \frac{1}{M} \sum_{i=1}^M 3.92 s(\hat{\beta}_i), \quad (18)$$

$$\text{CP}(\hat{\beta}) = \frac{1}{M} \sum_{i=1}^M \mathbb{1}(\hat{\beta}_i - 1.96 s(\hat{\beta}_i) < \beta < \hat{\beta}_i + 1.96 s(\hat{\beta}_i)), \quad (19)$$

respectively. In the expression given in (18), $s(\hat{\beta}_i)$ is the estimated standard error of the ML estimator. In the formula stated in (19), $\mathbb{1}$ is the indicator function. The simulation results for the empirical CLs and CPs are presented in Figure 6. It is evident from the figure that the empirical CLs decrease as the sample size increases, while the empirical CPs remain close to 0.95.

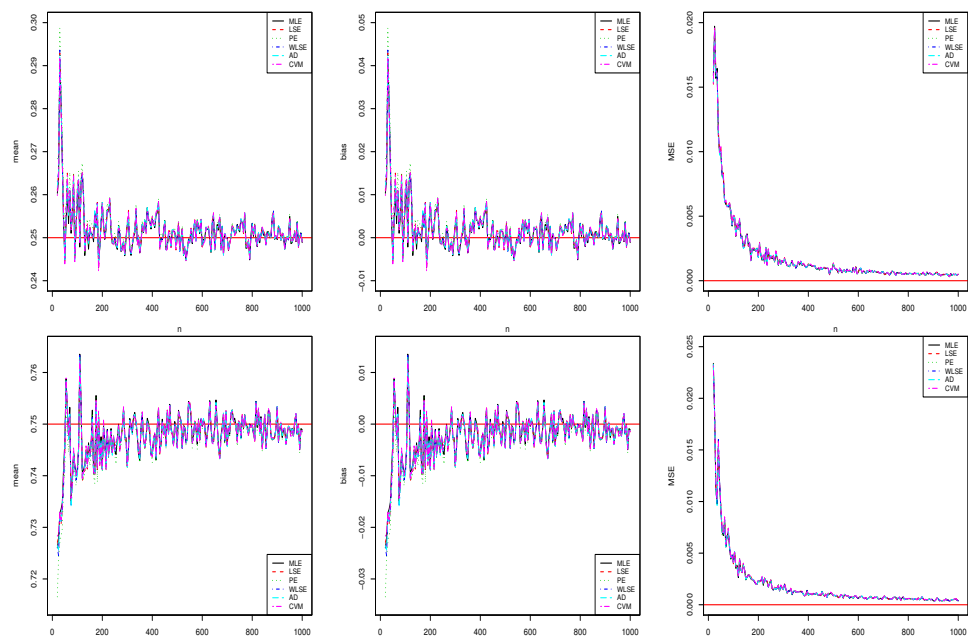


Figure 5. Plots of simulations for the point estimation with $\beta = 0.25$ (top) and $\beta = 0.75$ (bottom).

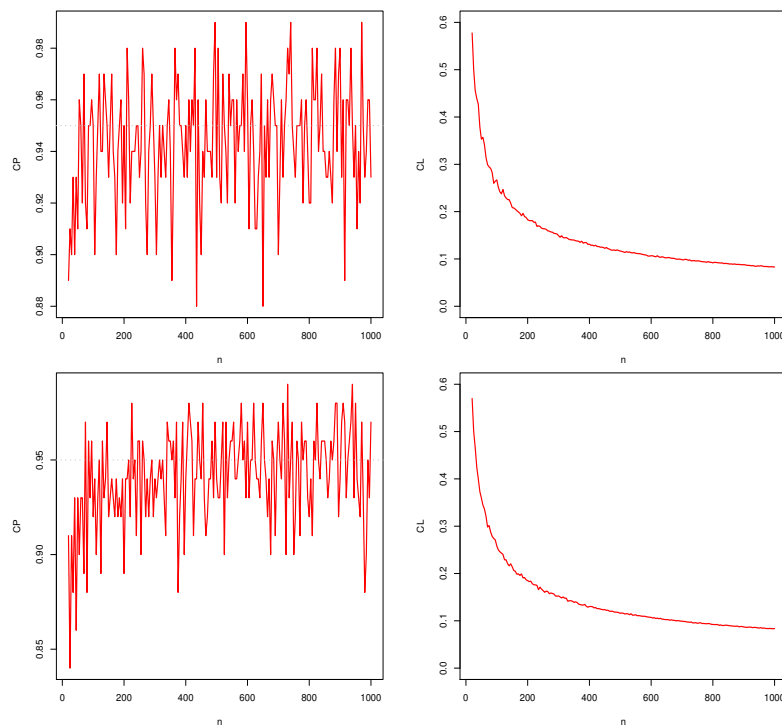


Figure 6. Plots of simulations for the CPs and CLs with $\beta = 0.25$ (top) and $\beta = 0.75$ (bottom).

5.3. Quantile Regression Model

Next, we conduct simulations to evaluate the ML estimators' performance for the ECB QR parameters. We measure their performance by examining the bias, MSE, CL, and CP across different sample sizes (n), known values of τ , true values of β , and various covariate values. To generate the values of the unit response, we use a structure stated as

$$y_i = \log \left(1 + \frac{(2\beta - 1)u_i^{1/\alpha_i}}{1 - \beta} \right) (\log(\beta/(1 - \beta)))^{-1}, \quad i \in \{1, \dots, n\},$$

where $u_i \sim U(0, 1)$ and $\alpha_i = \log(\tau(\log((1 - \beta)((\beta/(1 - \beta))^{H_i} - 1)/(2\beta - 1)))^{-1})$.

For this simulation study, we use $M = 1000$ Monte Carlo replicates and varying sample sizes from $n = 25$ to $n = 1000$, with step sizes of 25, 50, 100, 250, 500, and 1000. We also vary the values of τ and β , with τ taking on the values of 0.25, 0.50, and 0.75 and β taking on the values of 0.25, 0.5, and 0.75. The regression structure used is formulated as

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 z_{i1}, \quad i \in \{1, \dots, n\},$$

with $\beta_0 = -0.5, \beta_1 = 2$ and $z_{i1} \sim \text{Bernoulli}(0.5)$.

The simulation results for the ECB QR model are presented in Tables 3–8. It is observed that the empirical CLs exhibit a decreasing trend as the sample size increases, while the empirical CPs approach 0.95. The biases are negligible, and all MSEs approach zero.

Table 3. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.25$ and $\tau = 0.25$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	MSE($\hat{\beta}$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CL($\hat{\beta}$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)	CP($\hat{\beta}$)
25	0.0506	−0.0241	−0.0461	0.1578	0.2368	0.0524	1.3731	1.7408	0.8412	0.8880	0.9110	0.6430
50	0.0333	−0.0053	0.0057	0.0804	0.1169	0.0417	1.0516	1.3099	0.7733	0.9290	0.9490	0.7788
100	0.0120	−0.0009	−0.0064	0.0371	0.0593	0.0219	0.7197	0.9227	0.5588	0.9250	0.9430	0.8340
200	0.0078	−0.0008	−0.0013	0.0196	0.0298	0.0127	0.5310	0.6634	0.4251	0.9390	0.9360	0.8830
500	0.0004	0.0029	0.0001	0.0076	0.0122	0.0051	0.3351	0.4210	0.2758	0.9460	0.9510	0.9079
1000	−0.0010	−0.0001	0.0003	0.0035	0.0052	0.0027	0.2318	0.2967	0.1936	0.9500	0.9630	0.9240

Table 4. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.75$ and $\tau = 0.25$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	MSE($\hat{\beta}$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CL($\hat{\beta}$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)	CP($\hat{\beta}$)
25	0.0227	0.0017	−0.2301	0.2340	0.3634	0.1553	1.7324	2.1644	1.1733	0.9010	0.9190	0.6650
50	0.0296	−0.0311	−0.1406	0.1236	0.1746	0.0915	1.3305	1.5991	0.9841	0.9180	0.9360	0.7730
100	0.0171	−0.0141	−0.0833	0.0748	0.0979	0.0531	0.9991	1.1703	0.7680	0.9180	0.9320	0.8138
200	0.0027	−0.0024	−0.0442	0.0335	0.0463	0.0249	0.6950	0.8258	0.5717	0.9500	0.9440	0.8940
500	0.0046	−0.0025	−0.0206	0.0131	0.0184	0.0100	0.4437	0.5251	0.3749	0.9430	0.9500	0.9259
1000	0.0014	−0.0020	−0.0120	0.0069	0.0097	0.0050	0.3280	0.3790	0.2774	0.9520	0.9420	0.9390

Table 5. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.50$ and $\tau = 0.25$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)
25	0.0496	0.0033	0.0901	0.1766	1.1570	1.6256	0.9520	0.9480
50	0.0196	0.0110	0.0432	0.0875	0.7888	1.1547	0.9450	0.9500
100	0.0055	0.0012	0.0226	0.0465	0.5630	0.8153	0.9320	0.9410
200	0.0063	0.0032	0.0117	0.0214	0.4475	0.5820	0.9620	0.9540
500	0.0012	0.0001	0.0047	0.0088	0.2651	0.3638	0.9580	0.9470
1000	0.0022	0.0007	0.0025	0.0042	0.1919	0.2579	0.9420	0.9470

Table 6. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.50$ and $\tau = 0.75$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)
25	0.0466	−0.0147	0.1133	0.1821	1.2587	1.6439	0.9250	0.9450
50	0.0220	−0.0064	0.0443	0.0876	0.8032	1.1513	0.9470	0.9470
100	0.0146	0.0064	0.0204	0.0418	0.5626	0.8150	0.9520	0.9580
200	0.0085	−0.0130	0.0120	0.0212	0.4259	0.5761	0.9480	0.9510
500	0.0053	−0.0041	0.0045	0.0087	0.2624	0.3636	0.9460	0.9480
1000	−0.0002	0.0004	0.0024	0.0045	0.1901	0.2576	0.9430	0.9420

Table 7. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.25$ and $\tau = 0.50$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	MSE($\hat{\beta}$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CL($\hat{\beta}$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)	CP($\hat{\beta}$)
25	-0.0429	0.0319	-0.0166	0.2150	0.2637	0.0531	1.7293	1.8910	0.8087	0.9210	0.9260	0.6940
50	-0.0056	-0.0054	-0.0086	0.0798	0.1155	0.0305	1.1040	1.3311	0.6289	0.9380	0.9500	0.8060
100	-0.0070	0.0160	-0.0053	0.0431	0.0604	0.0147	0.8228	0.9506	0.4911	0.9550	0.9460	0.8700
200	-0.0007	-0.0009	-0.0023	0.0229	0.0310	0.0091	0.6091	0.6801	0.3655	0.9520	0.9490	0.8950
500	-0.0017	-0.0041	-0.0003	0.0096	0.0118	0.0038	0.3913	0.4336	0.2382	0.9560	0.9540	0.9300
1000	0.0008	-0.0013	0.0020	0.0049	0.0059	0.0019	0.2785	0.3076	0.1710	0.9590	0.9450	0.9420

Table 8. Empirical bias, MSE, 95% CP, and 95% CL for the ECB QR model with $\beta = 0.75$ and $\tau = 0.50$ for the indicated n and parameter with simulated data.

n	Bias($\hat{\beta}_0$)	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}$)	MSE($\hat{\beta}_0$)	MSE($\hat{\beta}_1$)	MSE($\hat{\beta}$)	CL($\hat{\beta}_0$)	CL($\hat{\beta}_1$)	CL($\hat{\beta}$)	CP($\hat{\beta}_0$)	CP($\hat{\beta}_1$)	CP($\hat{\beta}$)
25	-0.0501	0.0264	-0.1891	0.2898	0.3932	0.1172	1.9476	2.3130	1.0495	0.9120	0.9290	0.7404
50	-0.0283	0.0257	-0.1072	0.1507	0.2058	0.0638	1.4803	1.6950	0.8336	0.9260	0.9320	0.8029
100	-0.0165	0.0129	-0.0566	0.0955	0.1078	0.0322	1.1702	1.2454	0.6562	0.9300	0.9400	0.8805
200	-0.0105	0.0059	-0.0321	0.0540	0.0589	0.0169	0.8995	0.9262	0.4952	0.9440	0.9360	0.8958
500	-0.0007	-0.0005	-0.0150	0.0188	0.0209	0.0063	0.5222	0.5580	0.3107	0.9370	0.9470	0.9333
1000	-0.0013	0.0012	-0.0056	0.0090	0.0101	0.0028	0.3792	0.4032	0.2310	0.9370	0.9500	0.9306

6. Applications in Science and Education

This section provides two real data illustrations to show potential applications of the CB distribution and ECB QR model. All datasets were collected via: OECD.Stat with link <https://stats.oecd.org/> (accessed on 1 May 2023).

The OECD.Stat provides data on various development themes, including demography, education and training, health, labor, social protection and well-being, finance, as well as agriculture and fisheries, for both OECD and nonmember economies. Among these, we focus on two datasets. The first dataset includes the long-term unemployment rate of various countries, while the second dataset comprises the educational attainment proportion and better life index (BLI) measurements of different countries. We analyze the relationship between the educational attainment proportion and BLI using our QR model. All the datasets employed in this study are publicly available and can be secured directly from: <https://stats.oecd.org/index.aspx?DataSetCode=BLI2017> (accessed on 1 May 2023).

6.1. Application I

We analyze the proportion of long-term unemployment in different countries, defined as the ratio of the number of individuals who have been unemployed for one year or more to the total labor force (employed and unemployed individuals). We compare the performance of four one-parameter distributions, namely CB, log-Bilal [12], unit Lindley [14], and Topp–Leone (TL) [34], using indicators such as the estimated log-likelihood ($\hat{\ell}$), AIC (Akaike information criterion), and BIC (Bayesian information criterion), as well as Kolmogorov–Smirnov (KS), Cramer–von Mises (CM), and Anderson–Darling (AD) goodness-of-fit statistics.

We report the results of the data analysis utilizing ML estimation in Table 9, which shows that the CB distribution is the best model according to the lowest values of AIC, BIC, AD, CM, and KS statistics. Plots of fitted PDF and CDF for the unemployment data are provided in Figure 7.

Table 9. ML estimate and standard errors, in parentheses, as well as $\hat{\ell}$ and p-value of the goodness-of-fit test, in brackets, with unemployment data.

Distribution	$\hat{\beta}$	$\hat{\ell}$	AIC	BIC	AD	CM	KS
CB	0.261×10^{-13} (0.13×10^{-12})	92.8248	−183.6497	−182.0122	0.6018	0.1058	0.1332 [0.5097]
Log-Bilal	4.8700 (0.5694)	74.2699	−146.5399	−144.9023	5.7487	1.1128	0.3077 [0.0015]
TL	0.2996 (0.0486)	68.0997	−134.1995	−132.5619	35.2670	7.0160	0.7110 [< 0.001]
Unit-Lindley	29.7750 (4.8431)	92.2610	−182.5221	−180.8845	0.7448	0.1317	0.1532 [0.3344]

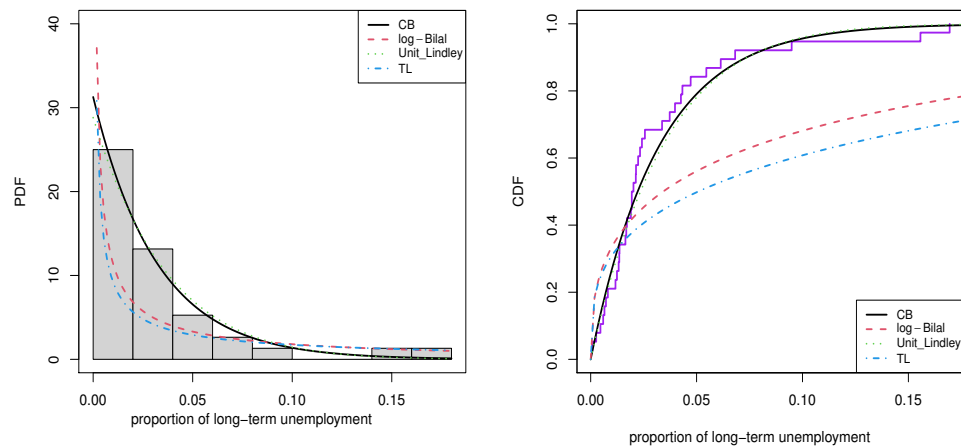


Figure 7. Plots of fitted PDF (left) and CDF (right) for the unemployment data.

6.2. Application II

Now, we focus on the modeling ability of the ECB QR and compare its results with those of the KW [17] and LEEG [19] QRs. The CDFs of the KW and LEEG distributions are, respectively, given by

$$G_{KW}(y; \beta, \mu, \tau) = 1 - (1 - y^\beta)^{\frac{\log(1-\tau)}{\log(1-\mu^\beta)}} \quad y \in (0, 1),$$

$$G_{LEEG}(y; \beta, \mu, \tau) = \frac{\tau y^\beta (\mu^\beta - 1)}{\tau (\mu^\beta - y^\beta) + \mu^\beta (y^\beta - 1)}, \quad y \in (0, 1),$$

where $\mu \in (0, 1)$ is the quantile parameter, $\beta > 0$ is a parameter related to the shape of the distribution, and τ is the known quantile level.

We investigate the relationship between the conditional quantile of educational attainment and the independent variables, including homicide rate (HR), dwellings without basic facilities (DWBF), and labor market insecurity (LMI) values of countries. The unit-dependent variable employed in this study is the proportion of educational attainment. This dataset was previously analyzed in [12] using the log-Bilal distribution for unit mean response regression modeling. The QR based on μ_i is given by

$$\mu_i = \frac{e^{\beta_0 + \beta_1 HR_i + \beta_2 DWBF_i + \beta_3 LMI_i}}{1 + e^{\beta_0 + \beta_1 HR_i + \beta_2 DWBF_i + \beta_3 LMI_i}}, \quad i \in \{1, \dots, 38\},$$

where μ_i is the quantile for all models.

We obtain the QR results for different quantile levels such as $\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$, separately. Tables 10–12 supply the results of the QR analysis for the different quantile levels mentioned. The findings from these tables indicate that when inferring about the ECB QR, all independent variables are statistically significant at any usual significance level based on all quantile levels.

The signs of β_0 and β_2 are positive, whereas the signs of β_1 and β_3 are negative. Hence, all independent variables have effects on the educational attainment of the countries. Since a direct positive correlation between the educational attainment and DWBF exists, as the educational attainment of countries increases, the DWBF increases as well. Moreover, there are negative relationships between educational attainment and the HR and LMI index, indicating that as the educational attainment of countries increases, the HR and LMI index decreases. Furthermore, the ECB QR model has minimal AIC/BIC with maximal log-likelihood values compared with other regression models at all quantile levels. Consequently, we conclude that the best model is the proposed regression structure.

Table 10. Results of fitted regressions with quantile $\tau = 0.25$ and model selection criteria for educational data.

Parameter	ECB			KW			LEEG		
	Estimate	SE	p-Value	Estimate	SE	p-Value	Estimate	SE	p-Value
β_0	1.9160	0.3172	<0.0001	1.3473	0.1730	<0.0001	1.4199	0.1599	<0.0001
β_1	-0.0914	0.0241	<0.0001	-0.0511	0.0130	<0.0001	-0.05588	0.0118	<0.0001
β_2	11.1800	2.3920	<0.0001	2.0563	1.1999	0.0866	2.3612	1.3068	0.0708
β_3	-21.2000	6.1390	<0.001	-7.1977	1.1596	<0.0001	-8.1841	1.7661	<0.0001
β	0.00058	0.0009	0.5367	7.0769	1.2104	<0.0001	9.1738	1.9582	<0.0001
$\hat{\ell}$		35.5794			33.5838			32.3569	
AIC		-61.1587			-57.1676			-54.7139	
BIC		-52.9708			-48.9797			-46.5259	

Table 11. Results of fitted regressions with quantile $\tau = 0.50$ and model selection criteria for educational data.

Parameter	ECB			KW			LEEG		
	Estimate	SE	p-Value	Estimate	SE	p-Value	Estimate	SE	p-Value
β_0	2.3830	1.7410	<0.0001	1.9595	0.1795	<0.0001	1.9907	0.1903	<0.0001
β_1	-0.0942	0.0207	<0.0001	-0.0639	0.0135	<0.0001	-0.0671	0.0134	<0.0001
β_2	10.9300	2.2890	<0.0001	2.3783	1.2035	0.0481	2.8145	1.4914	0.0591
β_3	-21.0700	1.8860	<0.0001	-8.9696	1.3384	<0.0001	-9.8119	2.0626	<0.0001
β	0.00097	0.0015	0.513	7.2166	1.2191	<0.0001	9.1091	1.9682	<0.0001
$\hat{\ell}$		34.6717			33.8642			32.2079	
AIC		-59.3433			-57.7286			-54.4158	
BIC		-51.1553			-49.5406			-46.2279	

Table 12. Results of fitted regressions with quantile $\tau = 0.75$ and model selection criteria for educational data.

Parameters	ECB			KW			LEEG		
	Estimate	SE	p-Value	Estimate	SE	p-Value	Estimate	SE	p-Value
β_0	3.0897	0.2262	<0.0001	2.5204	0.2044	<0.0001	2.7506	0.2707	< 0.0001
β_1	-0.0978	0.0229	<0.0001	-0.0772	0.0133	<0.0001	-0.0834	0.0157	<0.0001
β_2	9.7876	1.8659	<0.0001	2.7295	1.5489	0.0708	3.5070	1.8068	0.0523
β_3	-19.9055	1.4959	<0.0001	-10.0828	1.6659	<0.0001	-12.1760	2.6050	<0.0001
β	0.0022	0.0029	0.4530	8.2613	1.3396	<0.0001	9.0614	2.0043	<0.0001
$\hat{\ell}$		33.8923			33.6323			32.0757	
AIC		-57.7847			-57.2647			-54.1515	
BIC		-49.5967			-49.0768			-45.9635	

Figures 8–10 display the QQ (empirical quantile versus theoretical quantile) plot of the RQ residual for all QRs based on all quantile levels. These figures show that explaining the unit-dependent variable by independent variables is remarkable, and there is a good fit for the ECB QR model.

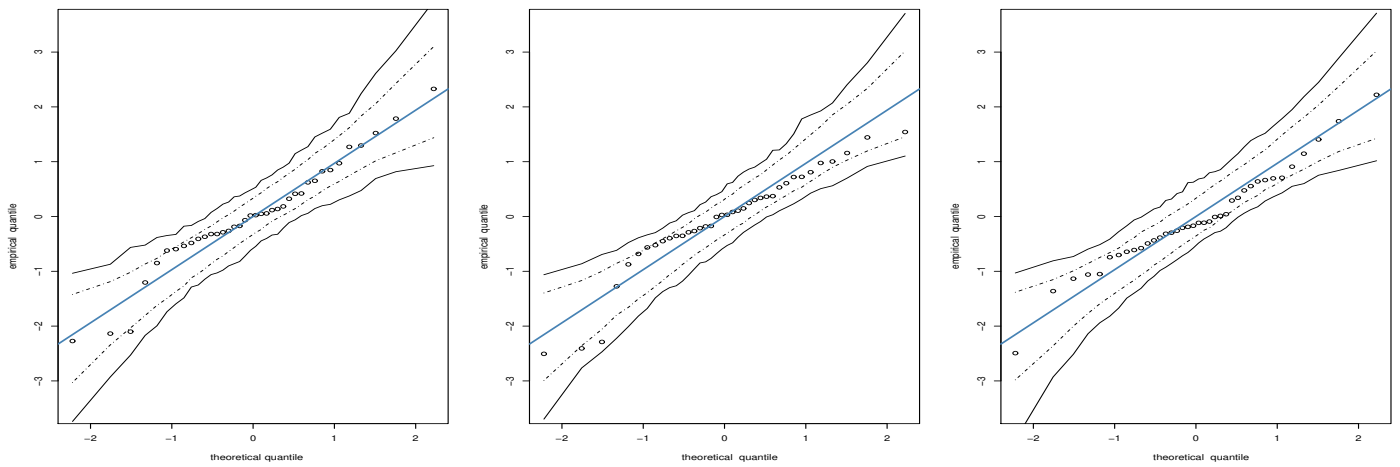


Figure 8. QQ plots of the RQ residual for KW (left), LEEG (center), and ECB (right) models based on $\tau = 0.25$ quantile level with educational data, where circles indicate the observed data.

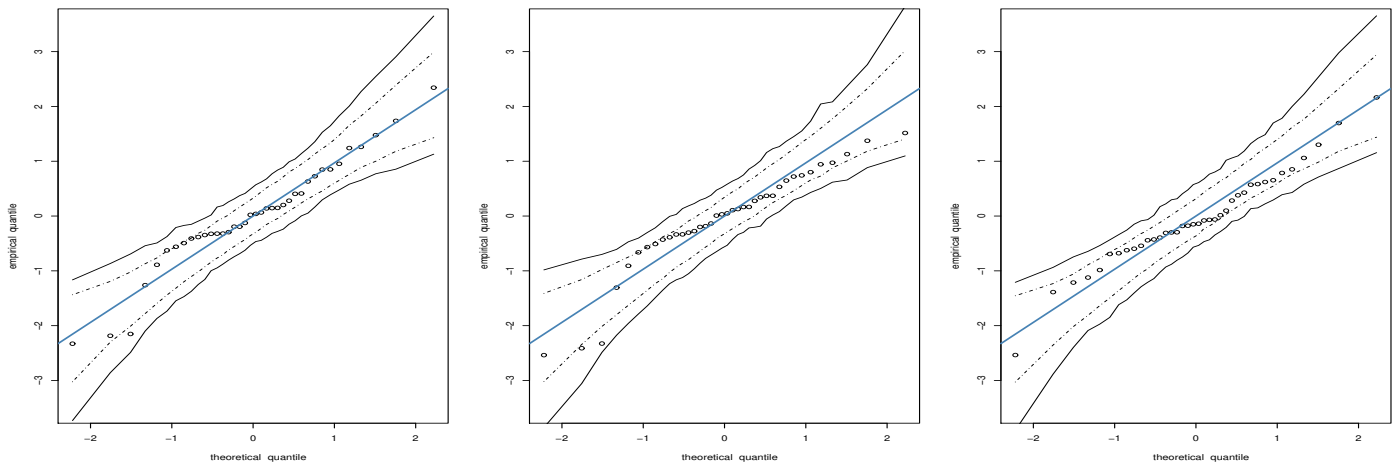


Figure 9. QQ plots of the RQ residual for KW (left), LEEG (center), and ECB (right) models based on $\tau = 0.50$ quantile level with educational data, where circles indicate the observed data.

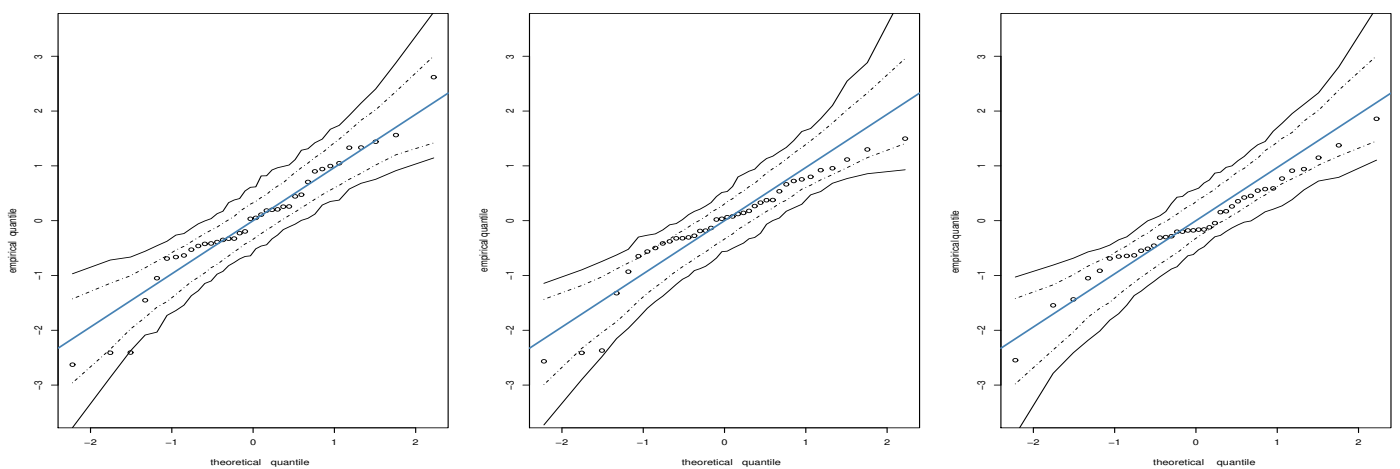


Figure 10. QQ plots of the RQ residual for KW (left), LEEG (center), and ECB (right) models based on $\tau = 0.75$ quantile level with educational data, where circles indicate the observed data.

7. Conclusions

This article presented new properties of the continuous Bernoulli distribution and introduced its quantile regression model with an exponentiated structure. Our quantile regression model incorporated independent variables that were associated with the quantile of the dependent variable using a link function related to a logit structure. To assess the performance of the estimators, we employed the Monte Carlo method under various settings. Our results indicated that the estimators work well in statistical terms, and additionally, the coverage probabilities approach the nominal value as the size of the sample increases.

The adequacy of the continuous Bernoulli distribution, exponentiated continuous Bernoulli quantile regression model, and other regression structures was compared by analyzing two real datasets based on the better life index of the countries. The comparison was carried out based on statistical indicators. Results indicated that the continuous Bernoulli distribution and its associated quantile regression model demonstrated a good coherence with the two sets based on real data, providing an alternative to other models presented in the literature on the topic.

Our quantile regression model yielded similar results to the unit mean regression models presented in [12,13]. Other related works [35–37] proposed the unit log-log, transmuted unit Rayleigh, and unit quantile regression models, respectively. These models analyzed the educational attainment values of OECD countries with a unit-dependent variable and reported a negative relationship between all responses and the independent variable of homicide rate, which is consistent with the findings of our study.

The examination we conducted showed that the continuous Bernoulli distribution can be very effective in analyzing data that are limited to the unit interval, whether or not independent variables are present. The use of both quantile and mean regression models can be easily carried out through the implementation of R codes by the authors of this article.

There are several avenues for future research that can build upon the present investigation. One possible direction is to incorporate multivariate, functional, temporal, and spatial structures, errors-in-variables, and partial least squares into the quantile regression framework, as well as studying their influence on diagnostics [38]. Other potential areas of investigation are the Tobit and Cobb–Douglas frameworks [39], which could be relevant to the topic of this study. Additionally, the analysis of censored observations and frailty models in the present context is another avenue for further research [40]. The authors are currently analyzing these issues and their findings are expected to be available in the future.

Author Contributions: Conceptualization, M.Ç.K., V.L. and C.M.-B.; data curation, M.Ç.K. and C.M.-B.; formal analysis, M.Ç.K., V.L. and C.M.-B.; investigation, M.Ç.K., V.L. and C.M.-B.; methodology, M.Ç.K., V.L. and C.M.-B.; writing—original draft, M.Ç.K. and C.M.-B.; writing—review and editing, V.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by FONDECYT grant number 1200525 (V.L.) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank two reviewers for their comments which helped to improve the presentation of this article.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Gómez-Déniz, E.; Sordo, M.A.; Calderín-Ojeda, E. The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insur. Math. Econ.* **2014**, *54*, 49–57. [[CrossRef](#)]
2. Korkmaz, M.Ç.; Chesneau, C. On the unit Burr-XII distribution with the quantile regression modeling and applications. *Comput. Appl. Math.* **2021**, *40*, 29. [[CrossRef](#)]
3. Kumaraswamy, P. A generalized probability density function for double-bounded random processes. *J. Hydrol.* **1980**, *46*, 79–88. [[CrossRef](#)]
4. Loaiza-Ganem, G.; Cunningham, J.P. The continuous Bernoulli: Fixing a pervasive error in variational autoencoders. *arXiv* **2019**, arXiv:1907.06845.
5. Smithson, M.; Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* **2006**, *11*, 54. [[CrossRef](#)]
6. Van Dorp, J.R.; Kotz, S. A novel extension of the triangular distribution and its parameter estimation. *J. R. Stat. Soc. D* **2002**, *51*, 63–79. [[CrossRef](#)]
7. Vasicek, O.A. *Probability of Loss on Loan Portfolio*; KMV Corporation: San Francisco, CA, USA, 1987.
8. Li, S.; Chen, J.; Li, B. Estimation and testing of random effects semiparametric regression model with separable space-time filters. *Fractal Fract.* **2022**, *6*, 735. [[CrossRef](#)]
9. Ferrari, S.; Cribari Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [[CrossRef](#)]
10. Bayes, C.L.; Bazán, J.L.; García, C. A new robust regression model for proportions. *Bayesian Anal.* **2012**, *7*, 841–866. [[CrossRef](#)]
11. Mazucheli, J.; Menezes, A.F.B.; Dey, S. The unit-Birnbaum-Saunders distribution with applications. *Chil. J. Stat.* **2018**, *9*, 47–57.
12. Altun, E.; El-Morshedy, M.; Eliwa, M. A new regression model for bounded dependent variable: An alternative to the beta and unit-Lindley regression models. *PLoS ONE* **2021**, *16*, e0245627. [[CrossRef](#)] [[PubMed](#)]
13. Altun, E. The log-weighted exponential regression model: Alternative to the beta regression model. *Commun. Stat. Theory Methods* **2021**, *50*, 2306–2321. [[CrossRef](#)]
14. Mazucheli, J.; Menezes, A.F.B.; Chakraborty, S. On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *J. Appl. Stat.* **2019**, *46*, 700–714. [[CrossRef](#)]
15. Koenker, R.; Bassett, G.J. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [[CrossRef](#)]
16. Bayes, C.L.; Bazán, J.L.; De Castro, M. A quantile parametric mixed regression model for bounded dependent variables. *Stat. Its Interface* **2017**, *10*, 483–493. [[CrossRef](#)]
17. Mitnik, P.A.; Baek, S. The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Stat. Pap.* **2013**, *54*, 177–192. [[CrossRef](#)]
18. Paz, R.F.; Balakrishnan, N.; Bazán, J.L. L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. *Braz. J. Probab. Stat.* **2019**, *33*, 455–479.
19. Jodrá, P.; Jiménez-Gamero, M.D. A quantile regression model for bounded responses based on the exponential-geometric distribution. *Revstat Stat. J.* **2020**, *4*, 415–436.
20. Saulo, H.; Dasilva, A.; Leiva, V.; Sanchez, L.; de la Fuente-Mella, H. Log-symmetric quantile regression models. *Stat. Neerl.* **2022**, *76*, 124–163. [[CrossRef](#)]
21. Cancho, V.G.; Bazán, J.L.; Dey, D.K. A new class of regression model for a bounded response with application in the study of the incidence rate of colorectal cancer. *Stat. Methods Med. Res.* **2020**, *29*, 2015–2033. [[CrossRef](#)]
22. Sanchez, L.; Leiva, V.; Galea, M. Saulo, H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Appl. Stoch. Model. Bus. Ind.* **2021**, *37*, 53–73. [[CrossRef](#)]
23. Ribeiro, T.F.; Cordeiro, G.M.; Peña-Ramírez, F.A.; Guerra, R.R. A new quantile regression for the COVID-19 mortality rates in the United States. *Comput. Appl. Math.* **2021**, *40*, 255. [[CrossRef](#)]
24. Korkmaz, M.Ç.; Emrah, A.; Chesneau, C.; Yousof, H.M. On the unit-Chen distribution with associated quantile regression and applications. *Math. Slovaca* **2022**, *72*, 765–786. [[CrossRef](#)]
25. Mazucheli, J.; Menezes, A.F.B.; Fernandes, L.B.; de Oliveira, R.P.; Ghitany, M.E. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on independent variables. *J. Appl. Stat.* **2020**, *47*, 954–974. [[CrossRef](#)] [[PubMed](#)]
26. Leiva, V.; Mazucheli, M.; Alves, B. A novel regression model for fractiles: Formulation, computational aspects, and applications to medical data. *Fractal Fract.* **2023**, *7*, 169. [[CrossRef](#)]
27. Shahin, A.I.; Almotairi, S. A deep learning BiLSTM encoding-decoding model for COVID-19 pandemic spread forecasting. *Fractal Fract.* **2021**, *5*, 175. [[CrossRef](#)]
28. Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. *arXiv* **2013**, arXiv:1312.6114.
29. Guiraud, P.; Leiva, V.; Fierro, R. A non-central version of the Birnbaum-Saunders distribution for reliability analysis. *IEEE Trans. Reliab.* **2009**, *58*, 152–160. [[CrossRef](#)]
30. Tang, L.; Lu, Y.; Chew, E. Mean residual life of lifetime distributions. *IEEE Trans. Reliab.* **1999**, *48*, 73–78. [[CrossRef](#)]
31. Kao, J.H. Computer methods for estimating Weibull parameters in reliability studies. *IRE Trans. Reliab. Qual. Control* **1958**, *PGRQC-13*, 15–22. [[CrossRef](#)]
32. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.

33. Mazucheli, M.; Alves, B.; Menezes, A.F.B.; Leiva, V. An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data. *Comput. Methods Programs Biomed.* **2022**, *221*, 106816. [[CrossRef](#)] [[PubMed](#)]
34. Topp, C.W.; Leone, F.C. A family of J-shaped frequency functions. *J. Am. Stat. Assoc.* **1955**, *50*, 209–219. [[CrossRef](#)]
35. Korkmaz, M.Ç.; Korkmaz, Z.S. The unit log-log distribution: A new unit distribution with alternative quantile regression modeling and educational measurements applications. *J. Appl. Stat.* **2023**, *50*, 889–908. [[CrossRef](#)] [[PubMed](#)]
36. Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. Transmuted unit Rayleigh quantile regression model: Alternative to beta and Kumaraswamy quantile regression models. *Univ. Politeh. Buchar. Sci. Bull. Ser. Appl. Math. Phys.* **2021**, *83*, 149–158.
37. Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. The unit folded normal distribution: A new unit probability distribution with the estimation procedures, quantile regression modeling and educational attainment applications. *J. Reliab. Stat. Stud.* **2022**, *15*, 261–298. [[CrossRef](#)]
38. Figueroa-Zuniga, J.; Niklitschek, S.; Leiva, V.; Liu, S. Modeling heavy-tailed bounded data by the trapezoidal beta distribution with applications. *Revstat Stat. J.* **2023**, *20*, 387–404.
39. de la Fuente, H.; Rojas Fuentes, J.L.; Leiva, V. Econometric modeling of productivity and technical efficiency in the Chilean manufacturing industry. *Comput. Ind. Eng.* **2020**, *139*, 105793. [[CrossRef](#)]
40. Leao, J.; Leiva, V.; Saulo, H.; Tomazella, V. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Stat. Med.* **2018**, *37*, 4421–4440. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.