AIMS Mathematics, 11(10): 34975–35004.
DOI: 10.3934/math.20261362
Received: 29 June 2026
Revised: 13 September 2026
Accepted: 29 September 2026
Published: 26 October 2026

http://www.aimspress.com/journal/Math

Research article

A zero–one-inflated smooth unit-Laplace distribution for boundary-inflated proportional data

Fuxiang Liu[1,2] **and Tingting Li**[1,2,*]

[1] Three Gorges Mathematical Research Center, China Three Gorges University, Yichang, Hubei 443002, China

[2] Department of Mathematics, College of Mathematics and Physics, China Three Gorges University, Yichang, Hubei 443002, China

* **Correspondence:** Email: litingting@ctgu.edu.cn; Tel: +8615972809137.

Abstract: Bounded observations with exact endpoint values arise in many applications involving proportions, rates, and normalized indices. Such data require a model that is theoretically able to describe the continuous variation of observations in (0,1), while practically allowing positive probability masses at the exact boundary values 0 and 1. This paper proposes a zero–one-inflated smooth unit-Laplace distribution for boundary-inflated proportional data. The construction first transforms a Laplace random variable to a smooth one-parameter distribution on the open unit interval and then combines this continuous law with point masses at 0 and 1. The resulting model admits closed-form expressions for the distribution function, density function, quantile function, and likelihood. Consequently, in the nondegenerate case, all maximum likelihood estimators admit closed-form expressions and can be computed with linear complexity in the sample size. Several mathematical properties are derived, including endpoint behavior, density-shape characteristics, raw moments, the Lorenz curve, and the Gini coefficient. Large-sample properties are established through the Fisher information matrix. Numerical experiments are carried out to examine bias, root mean squared error, standard errors, and confidence-interval coverage. An application to school-level proportional observations from the Programme for International Student Assessment (PISA) 2022 database illustrates the usefulness of the proposed computational framework for modeling boundary inflation and distributional concentration.

Keywords: boundary-inflated data; smooth unit-Laplace distribution; zero–one-inflated model; computational inference; Monte Carlo simulation
Mathematics Subject Classification: 62E15, 62F12, 65C05, 65C20

1. Introduction

Response variables taking values in the unit interval are frequently encountered in empirical studies involving proportions, rates, indices, and bounded scores. Such data arise in medicine, economics, public health, environmental sciences, reliability analysis, and educational assessment. In large-scale educational assessment, for example, proportion-type indicators are often used to summarize student achievement, teacher qualifications, and school-level resource allocation [1–4]. Although these variables are naturally supported on [0, 1], t heir empirical distributions may exhibit strong asymmetry, near-boundary concentration, and positive probability masses at the endpoints 0 and 1.

The empirical features described above call for models that respect the bounded nature of fractional responses [5] and distinguish between exact endpoint observations and continuous observations in $(0, 1)$. Zero–one-inflated models address the former by assigning positive probability masses to 0 and 1 within a mixed discrete–continuous framework [6–8], whereas the distinction among these models lies primarily in the specification of the interior distribution. The beta family provides considerable flexibility in terms of skewness and modality, but its distribution and quantile functions require special-function evaluation or numerical inversion, and its maximum likelihood estimators are generally obtained through iterative procedures [9, 10]. This increases the computational burden of repeated model fitting and requires numerical convergence to be monitored. The Kumaraswamy distribution has explicit distribution and quantile functions, but its shape parameters still generally require numerical likelihood maximization [11]. In addition, continuous unit distributions such as the unit-gamma, unit-Lindley, unit-Weibull, unit-Gompertz, and unit-Burr III distributions provide further choices for the interior component [12–16]; depending on the particular family, however, they may involve special functions, multiple parameters, iterative estimation, or restrictions on density shape. Based on these continuous models, zero–one-inflated unit-gamma, inflated unit-Lindley, inflated Kumaraswamy, and inflated unit-Birnbaum–Saunders models have been developed to accommodate exact endpoint observations [17–20]. Nevertheless, introducing endpoint inflation does not eliminate the analytical and computational features inherited from the interior continuous distribution. Consequently, the present study aims to construct a complementary continuous model for the interior observations that can be embedded in a zero–one-inflated framework while retaining simple analytical expressions and a useful degree of adaptability to different interior distributional patterns, with closed-form estimators available for all model parameters in the nondegenerate case. By avoiding iterative numerical optimization, this structure makes likelihood-based inference more direct and computationally efficient.

To achieve the above objective, consider the Laplace distribution with location parameter μ and rate parameter $\theta > 0$ [21]. Let

$$Z \sim \text{Laplace}(\mu, \theta),\ \mu \in \mathbb{R},\ \theta > 0.$$

The Laplace distribution has the following simple structural property:

$$|Z - \mu| \sim \text{Exp}(\theta).$$

That is, its absolute centered deviation follows a one-parameter exponential distribution. This property provides a basis for constructing an analytically tractable unit-interval distribution and therefore motivates the choice of the Laplace distribution as the baseline distribution. A Laplace-logistic unit distribution has previously been constructed from the Laplace distribution [22], but its density must

be represented by two piecewise analytical expressions and is nondifferentiable at the joining point of the two branches. Accordingly, we consider constructing an alternative Laplace-based unit distribution with a unified smooth form. For this purpose, we introduce the following transformation:

$$X = \frac{1}{1 + \lambda(Z - \mu)^2}, \ \lambda > 0.$$

This transformation is based on the squared centered deviation and maps the two branches of the Laplace distribution that are symmetric about μ into a unified expression. At the same time, under this transformation,

$$\sqrt{\frac{1 - X}{X}} = \sqrt{\lambda}|Z - \mu| \sim \text{Exp}\left(\frac{\theta}{\sqrt{\lambda}}\right).$$

Therefore, the proposed transformation preserves the exponential-deviation structure of the Laplace baseline distribution while yielding a distribution with a unified smooth analytical form on $(0, 1)$.

The transformation introduced above induces a one-parameter distribution on $(0, 1)$, which we refer to as the smooth unit-Laplace (SUL) distribution. The exponential relationship preserved by this construction then allows us to derive elementary closed-form expressions for the cumulative distribution function, probability density function, and quantile function of the SUL distribution. Building on this continuous distribution, we further define the zero–one-inflated smooth unit-Laplace (ZOISUL) model by assigning positive probability masses to 0 and 1 and using the SUL distribution to describe the observations in $(0, 1)$.

The main contributions of this paper are summarized as follows:

- A new one-parameter SUL distribution is constructed, and its main distributional properties are derived, including endpoint behavior, density-shape characteristics, and raw moments.
- The SUL distribution is embedded into a zero–one-inflated mixture structure, leading to the ZOISUL model with interpretable endpoint and interior components.
- The Lorenz curve and the Gini coefficient are obtained to provide distributional summaries for concentration and inequality in bounded observations.
- Closed-form maximum likelihood estimators are derived, large-sample inference is established through the Fisher information matrix, and a summary-based likelihood computation scheme is formulated for efficient estimation and model comparison.
- The finite-sample and empirical performance of the proposed model is assessed through Monte Carlo simulations and an application to school questionnaire data from the Programme for International Student Assessment (PISA) 2022 released by the Organisation for Economic Co-operation and Development (OECD).

The remainder of this paper is organized as follows. Section 2 introduces the SUL distribution and the ZOISUL model. Section 3 develops likelihood-based inference and examines the finite-sample behavior of the proposed estimators. Section 4 presents the application to OECD PISA 2022 data and compares the proposed model with competing zero–one-inflated unit-interval models. Section 5 concludes with possible extensions.

2. Model construction and distributional properties

2.1. The smooth unit-Laplace distribution

Let Z follow a Laplace distribution with location parameter $\mu \in \mathbb{R}$ and rate parameter $\theta > 0$. Its probability density function is given by

$$g_Z(z;\mu,\theta) = \frac{\theta}{2}\exp\{-\theta|z-\mu|\},\ z \in \mathbb{R}.$$

As motivated in the Introduction, for $\lambda > 0$, define the smooth mapping

$$\varphi_\lambda : \mathbb{R} \to (0,1],\ \varphi_\lambda(z;\mu) = \frac{1}{1+\lambda(z-\mu)^2},$$

and let $X = \varphi_\lambda(Z;\mu)$. The mapping φ_λ is smooth on $\mathbb{R}$ and symmetric about μ. It attains its maximum value 1 at $z = \mu$ and is strictly decreasing as the absolute centered deviation $|z-\mu|$ increases. In particular,

$$\lim_{z\to-\infty}\varphi_\lambda(z;\mu) = 0,\ \varphi_\lambda(\mu;\mu) = 1,\ \lim_{z\to+\infty}\varphi_\lambda(z;\mu) = 0.$$

Thus, values close to the center μ are mapped near the upper endpoint of the unit interval, whereas the mapped values decrease toward 0 as $|z-\mu|$ increases. Although φ_λ attains the value 1 at $z = \mu$, the absolute continuity of Z implies that

$$\Pr(X = 1) = \Pr(Z = \mu) = 0,\ \Pr(0 < X < 1) = 1.$$

Therefore, the range of the transformation φ_λ is $(0,1]$, whereas the transformed random variable X is concentrated on the open unit interval $(0,1)$ with probability one. The parameter λ governs the rate at which the mapped values decrease as the centered deviation increases, thereby controlling the degree of compression from the original scale to the unit interval. For a fixed nonzero centered deviation, a larger value of λ maps the observation closer to the lower endpoint, whereas a smaller value maps it closer to the upper endpoint.

Using the preceding transformation and the properties of the Laplace distribution, define

$$\alpha = \frac{\theta}{\sqrt{\lambda}} > 0.$$

Then, for $0 < x < 1$, the cumulative distribution function is

$$F_X(x;\alpha) = \Pr(X \le x) = \Pr\left(|Z-\mu| \ge \sqrt{\frac{1-x}{\lambda x}}\right) = \exp\left\{-\alpha\sqrt{\frac{1-x}{x}}\right\}. \tag{2.1}$$

This result shows that θ and λ enter the distribution of X through α, yielding a parsimonious one-parameter family and facilitating the likelihood inference developed in Section 3. Moreover, because the transformation $\varphi_\lambda(z;\mu)$ is based on the squared deviation $(z-\mu)^2$, the nondifferentiability of the Laplace density at $z = \mu$ is not inherited by the induced unit distribution. Consequently, its cumulative distribution function and probability density function are both smooth on $(0,1)$. This leads to the following definition.

Definition 2.1 (Smooth unit-Laplace distribution)**.** A random variable X is said to follow a smooth unit-Laplace (SUL) distribution with shape parameter α, denoted by

$$X \sim \mathrm{SUL}(\alpha),\ \alpha > 0,$$

if its cumulative distribution function is given by (2.1).

Differentiating (2.1) with respect to x on $(0, 1)$ yields the Lebesgue density

$$f_X(x;\alpha) = \frac{\alpha}{2} x^{-3/2}(1-x)^{-1/2} \exp\left\{-\alpha\sqrt{\frac{1-x}{x}}\right\},\ 0 < x < 1. \tag{2.2}$$

Moreover, $f_X(x;\alpha) > 0$ for $0 < x < 1$. Because

$$\lim_{x\to 0^+} F_X(x;\alpha) = 0,\ \lim_{x\to 1^-} F_X(x;\alpha) = 1,$$

it follows that

$$\int_0^1 f_X(x;\alpha)\,\mathrm{d}x = F_X(1^-;\alpha) - F_X(0^+;\alpha) = 1.$$

Thus, (2.2) defines a proper probability density on the open unit interval $(0, 1)$.

Proposition 2.2. *Let $X \sim \mathrm{SUL}(\alpha)$, where $\alpha > 0$, and let $f_X(\cdot;\alpha)$ be given by* (2.2)*. Then,*

$$\lim_{x\to 0^+} f_X(x;\alpha) = 0^+,\ \lim_{x\to 1^-} f_X(x;\alpha) = +\infty.$$

Moreover, there exists a critical value

$$\alpha_0 = \frac{3}{\sqrt{3+2\sqrt{3}}} \approx 1.1800,$$

such that the following monotonicity properties hold:

(i) If $\alpha \geq \alpha_0$, then $f_X(x;\alpha)$ is increasing on $(0, 1)$;

(ii) If $0 < \alpha < \alpha_0$, then there exist $0 < x_1 < x_2 < 1$ such that $f_X(x;\alpha)$ is increasing on $(0, x_1]$, decreasing on $(x_1, x_2]$, and increasing on $(x_2, 1)$.

Proof. The boundary limits follow directly from (2.2). To establish monotonicity, set

$$t = \sqrt{\frac{1-x}{x}},\ t > 0.$$

A logarithmic differentiation gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\log f_X(x;\alpha) = \frac{t^2+1}{2}\,\phi_\alpha(t),\ \phi_\alpha(t) = \alpha\frac{t^2+1}{t} + \frac{1}{t^2} - 3$$

Clearly, the sign of $f_X'(x;\alpha)$ is determined by $\phi_\alpha(t)$. Moreover,

$$\phi_\alpha(t) \geq 0 \iff \alpha \geq h(t) := \frac{3t^2-1}{t(t^2+1)},$$

with the case $h(t) \leq 0$ being automatic. A direct calculation gives

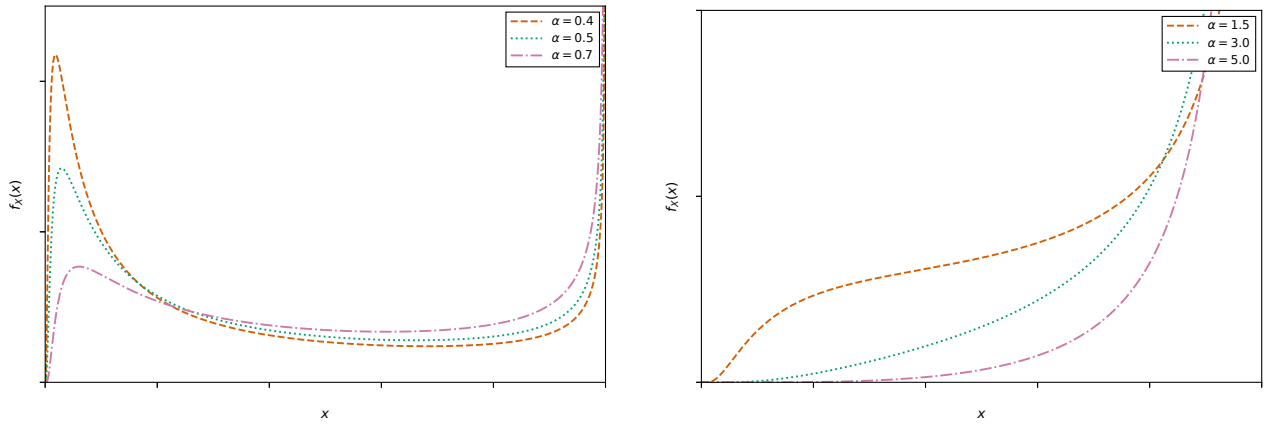
$$\sup_{t>0} h(t) = \frac{3}{\sqrt{3+2\sqrt{3}}} = \alpha_0.$$

Hence, if $\alpha \geq \alpha_0$, then $f_X(x;\alpha)$ is increasing on $(0,1)$. If $0 < \alpha < \alpha_0$, the equation $\phi_\alpha(t) = 0$ has two positive roots $t_1 < t_2$, and the sign changes from positive to negative and then to positive. Because $x = 1/(t^2+1)$, setting

$$x_1 = \frac{1}{t_2^2+1},\ x_2 = \frac{1}{t_1^2+1}$$

gives $0 < x_1 < x_2 < 1$, and the stated monotonicity follows. □

Proposition 2.2 characterizes the boundary behavior and monotonicity of the SUL density. To provide a more intuitive illustration of these properties, Figure 1 displays representative density curves of $f_X(x;\alpha)$ for selected values of α. The patterns shown agree with the theoretical results established in Proposition 2.2. In particular, the SUL family can generate both monotone and specific nonmonotone density shapes through the single shape parameter α. As α varies, the interior shape and the concentration of probability mass toward the boundaries change jointly. Thus, the one-parameter formulation provides a parsimonious description of these density patterns, and the corresponding distributional features are governed by the same shape parameter rather than being adjusted independently. Furthermore, the sharp increase in the SUL density curves shown in Figure 1 is mainly concentrated in a narrow region very close to 1. In practical applications, this local feature should be interpreted in conjunction with the model's fit over the remainder of the interval.



(a) (b)

Figure 1. Probability density functions of the SUL(α) distribution for selected parameter values: (a) $0 < \alpha < \alpha_0$; (b) $\alpha \geq \alpha_0$.

2.2. *The zero–one-inflated smooth unit-Laplace model*

Based on the SUL distribution defined in Section 2.1, we now define a discrete-continuous mixture distribution on the closed unit interval $[0,1]$ by assigning positive probability masses to both endpoints.

Definition 2.3 (Zero–one-inflated smooth unit-Laplace distribution)**.** A random variable Y on $[0, 1]$ is said to follow a zero–one-inflated smooth unit-Laplace (ZOISUL) distribution with parameters $\delta, \pi \in (0, 1)$ and $\alpha > 0$, denoted by

$$Y \sim \text{ZOISUL}(\delta, \pi, \alpha),$$

if

$$\Pr(Y = 0) = \delta(1 - \pi),\ \Pr(Y = 1) = \delta\pi,$$

and

$$\Pr(0 < Y < 1) = 1 - \delta,\ Y \mid (0 < Y < 1) \sim \text{SUL}(\alpha).$$

Under this parameterization, δ is the probability that an observation falls at either endpoint, π is the conditional probability that it equals 1 given that it is a boundary observation, and α characterizes the distributional shape of the nonboundary observations. Thus, the ZOISUL model retains the SUL distribution as the conditional model for interior observations while accommodating exact endpoint outcomes through an additional probability-allocation mechanism.

Proposition 2.4 (Parameter identifiability)**.** *Let $\mathcal{P} = (0, 1) \times (0, 1) \times (0, \infty)$. The ZOISUL family is identifiable on $\mathcal{P}$; that is, if*

$$Y_j \sim \text{ZOISUL}(\delta_j, \pi_j, \alpha_j),\ (\delta_j, \pi_j, \alpha_j) \in \mathcal{P},\ j = 1, 2,$$

and $Y_1 \overset{d}{=} Y_2$, then

$$(\delta_1, \pi_1, \alpha_1) = (\delta_2, \pi_2, \alpha_2).$$

Proof. Equality in distribution implies

$$\delta_1(1 - \pi_1) = \delta_2(1 - \pi_2),\ \delta_1\pi_1 = \delta_2\pi_2,$$

and hence, $\delta_1 = \delta_2$ and $\pi_1 = \pi_2$. Equality of the conditional distributions on $(0, 1)$ further gives

$$\exp\left\{-\alpha_1\sqrt{\frac{1-x}{x}}\right\} = \exp\left\{-\alpha_2\sqrt{\frac{1-x}{x}}\right\},\ 0 < x < 1.$$

Therefore, $\alpha_1 = \alpha_2$, which completes the proof. □

This result establishes a one-to-one correspondence between the parameter vector and the resulting ZOISUL distribution. Hence, the endpoint probability allocation characterized by (δ, π) and the conditional interior shape characterized by α constitute distinct and nonredundant features of the model.

Proposition 2.5 (Distribution function and generalized density)**.** *Let $Y \sim \text{ZOISUL}(\delta, \pi, \alpha)$. Then, its cumulative distribution function is*

$$F_Y(y; \delta, \pi, \alpha) = \begin{cases} 0, & y < 0, \\ \delta(1 - \pi), & y = 0, \\ \delta(1 - \pi) + (1 - \delta)\exp\left\{-\alpha\sqrt{\frac{1-y}{y}}\right\}, & 0 < y < 1, \\ 1, & y \geq 1. \end{cases} \tag{2.3}$$

Let

$$\nu = \delta_0 + \lambda_{(0,1)} + \delta_1,$$

where δ_0 *and* δ_1 *denote the Dirac measures at 0 and 1, respectively, and* $\lambda_{(0,1)}$ *denotes the Lebesgue measure on* $(0, 1)$*. Then the generalized density of* Y *with respect to* ν *is*

$$f_Y(y; \delta, \pi, \alpha) = \begin{cases} \delta(1-\pi), & y = 0, \\ (1-\delta)\dfrac{\alpha}{2} y^{-3/2}(1-y)^{-1/2} \exp\left\{-\alpha\sqrt{\dfrac{1-y}{y}}\right\}, & 0 < y < 1, \\ \delta\pi, & y = 1. \end{cases} \tag{2.4}$$

Proof. For $0 < y < 1$,

$$F_Y(y; \delta, \pi, \alpha) = \Pr(Y = 0) + \Pr(0 < Y \le y) = \delta(1-\pi) + (1-\delta)F_X(y; \alpha).$$

Substituting (2.1) gives (2.3). The endpoint jumps and (2.4) follow from the same mixture representation. □

Remark 2.6 (Boundary and degenerate cases)**.** The restrictions $\delta, \pi \in (0, 1)$ ensure the coexistence of the two endpoint masses and the interior continuous component. If boundary values are allowed, the model reduces to several special cases: $\delta = 0$ gives the SUL distribution, $\pi = 0$ or $\pi = 1$ allocates all endpoint mass to one endpoint, and $\delta = 1$ gives a Bernoulli-type distribution on $\{0, 1\}$.

The mixture structure above admits the following latent-variable representation.

Proposition 2.7 (Stochastic representation)**.** *Let* $B \sim \text{Bernoulli}(\delta)$*,* $C \sim \text{Bernoulli}(\pi)$*, and* $X \sim \text{SUL}(\alpha)$*, where* B, C*, and* X *are mutually independent. Define*

$$Y = BC + (1 - B)X. \tag{2.5}$$

Then,

$$Y \sim \text{ZOISUL}(\delta, \pi, \alpha).$$

Proof. If $B = 1$, then $Y = C$, and hence,

$$\Pr(Y = 0) = \Pr(B = 1, C = 0) = \delta(1-\pi),\ \Pr(Y = 1) = \Pr(B = 1, C = 1) = \delta\pi.$$

If $B = 0$, then $Y = X \sim \text{SUL}(\alpha)$, with weight $1 - \delta$. Therefore, the random variable defined by (2.5) has the endpoint masses and interior continuous distribution specified in Definition 2.3. □

This stochastic representation describes the hierarchical generative structure of the ZOISUL distribution. The variable B determines whether an observation belongs to the boundary component or the continuous interior component. When $B = 1$, the variable C selects between the two endpoints, 0 and 1. When $B = 0$, the observation is generated by the SUL random variable X. Therefore, exact endpoint observations arise only from the discrete inflation component, whereas all observations strictly within $(0, 1)$, including those arbitrarily close to an endpoint, arise from the continuous SUL component. This hierarchical structure also provides a basis for simulation and likelihood construction.

2.3. Moment properties

The following result gives the raw moments of the ZOISUL distribution.

Proposition 2.8 (Raw moments)**.** *Let $Y \sim \mathrm{ZOISUL}(\delta, \pi, \alpha)$. For any $r > 0$,*

$$\mathbb{E}(Y^r) = \delta\pi + (1-\delta)\mu_r(\alpha), \tag{2.6}$$

where $\mathbb{E}$ denotes the expectation operator, and $\mu_r(\alpha)$ denotes the rth raw moment of the SUL component, namely

$$\mu_r(\alpha) = \mathbb{E}(X^r),\ X \sim \mathrm{SUL}(\alpha).$$

Moreover,

$$\mu_r(\alpha) = \alpha \int_0^\infty \frac{e^{-\alpha t}}{(1+t^2)^r}\,\mathrm{d}t,\ r > 0,\ \alpha > 0. \tag{2.7}$$

Proof. By the mixture representation in Proposition 2.7, for any $r > 0$,

$$\mathbb{E}(Y^r) = \delta\pi + (1-\delta)\mathbb{E}(X^r),\ X \sim \mathrm{SUL}(\alpha),$$

which gives (2.6). Finally, substituting $t = \sqrt{(1-x)/x}$ in

$$\mu_r(\alpha) = \int_0^1 x^r f_X(x;\alpha)\,\mathrm{d}x$$

yields (2.7). □

For the SUL distribution, all raw moments admit the integral representation in (2.7), which generally lacks an elementary closed form. Nevertheless, the first two raw moments can be explicitly represented in terms of the sine integral and cosine integral functions [23, 24]. Define

$$\mathrm{Si}(x) = \int_0^x \frac{\sin t}{t}\,\mathrm{d}t,\ \mathrm{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}\,\mathrm{d}t,\ x > 0.$$

Then, the mean $\mu_1(\alpha)$ and second raw moment $\mu_2(\alpha)$ are

$$\mu_1(\alpha) = \alpha\left[\sin\alpha\ \mathrm{Ci}(\alpha) + \cos\alpha\left\{\frac{\pi}{2} - \mathrm{Si}(\alpha)\right\}\right],$$

and

$$\mu_2(\alpha) = \frac{\alpha}{2}\left[(\sin\alpha - \alpha\cos\alpha)\,\mathrm{Ci}(\alpha) + (\cos\alpha + \alpha\sin\alpha)\left\{\frac{\pi}{2} - \mathrm{Si}(\alpha)\right\}\right].$$

These special-function representations provide explicit expressions for the first two raw moments and facilitate the theoretical characterization of the mean and variance of the SUL distribution. For numerical evaluation, the sine and cosine integrals are computed using the `scipy.special.sici` function in Python's SciPy library. The resulting values are substituted into the expressions above to obtain $\mu_1(\alpha)$ and $\mu_2(\alpha)$.

Corollary 2.9 (Mean and variance)**.** *Let $Y \sim \text{ZOISUL}(\delta, \pi, \alpha)$. Then,*

$$\mathbb{E}(Y) = \delta\pi + (1-\delta)\mu_1(\alpha),$$

and

$$\mathbb{E}(Y^2) = \delta\pi + (1-\delta)\mu_2(\alpha).$$

Consequently, the variance of Y is

$$\text{Var}(Y) = \delta\pi + (1-\delta)\mu_2(\alpha) - \left[\delta\pi + (1-\delta)\mu_1(\alpha)\right]^2. \tag{2.8}$$

Proof. Taking $r = 1$ and $r = 2$ in Proposition 2.8 gives the first and second raw moments, respectively. Equation (2.8) then follows from $\text{Var}(Y) = \mathbb{E}(Y^2) - \{\mathbb{E}(Y)\}^2$. □

Remark 2.10. Equation (2.6) shows that the rth raw moment of the ZOISUL distribution can be decomposed into weighted contributions from the endpoint masses and the interior continuous component. This decomposition also facilitates parameter interpretation, initialization, likelihood-based inference, and simulation studies.

2.4. Quantile-based concentration and inequality measures

The Lorenz curve and the Gini coefficient are fundamental tools for measuring the concentration and inequality of non-negative random variables, with wide applications in income and wealth distributions, resource allocation, and reliability analysis [25–27]. For a random variable Y with mean $\mu = \mathbb{E}(Y) > 0$ and quantile function

$$Q_Y(p) = \inf\{y \in [0,1] : F_Y(y) \geq p\},$$

the Lorenz curve is defined by

$$L_Y(p) = \frac{1}{\mu}\int_0^p Q_Y(s)\,\mathrm{d}s,\ 0 \leq p \leq 1. \tag{2.9}$$

For the $\text{ZOISUL}(\delta, \pi, \alpha)$ distribution, the point masses at 0 and 1 lead to piecewise expressions for the quantile function, the Lorenz curve, and the associated Gini coefficient.

For notational simplicity, let

$$p_0 = \Pr(Y = 0) = \delta(1-\pi),\ p_1 = \Pr(Y = 1) = \delta\pi,$$

and denote

$$\mu = \mathbb{E}(Y) = p_1 + (1-\delta)\mu_1(\alpha),$$

where $\mu_1(\alpha) = \mathbb{E}(X)$ for $X \sim \text{SUL}(\alpha)$.

Lemma 2.11 (Quantile function)**.** *Let $Y \sim \text{ZOISUL}(\delta, \pi, \alpha)$. The quantile function of Y is given by*

$$Q_Y(p) = \begin{cases} 0, & 0 \leq p \leq p_0, \\ \dfrac{\alpha^2}{\alpha^2 + \left[\log\left(\dfrac{1-\delta}{p-p_0}\right)\right]^2}, & p_0 < p < 1-p_1, \\ 1, & 1-p_1 \leq p \leq 1. \end{cases} \tag{2.10}$$

Proof. From the piecewise form of F_Y, the jumps at 0 and 1 are p_0 and p_1, respectively. Hence, $Q_Y(p) = 0$ for $0 \le p \le p_0$, and $Q_Y(p) = 1$ for $1 - p_1 \le p \le 1$. For $p_0 < p < 1 - p_1$, the quantile lies in $(0, 1)$ and satisfies

$$p = p_0 + (1 - \delta)F_X(y; \alpha).$$

Substituting

$$F_X(y; \alpha) = \exp\left\{-\alpha\sqrt{\frac{1-y}{y}}\right\}$$

and solving for y gives the middle part of (2.10). □

To evaluate the integral in (2.9) over the continuous part, where $Q_Y(p) \in (0, 1)$, we require the lower incomplete first moment of the SUL(α) distribution.

Definition 2.12 (Lower incomplete first moment of the SUL distribution)**.** Let $X \sim \mathrm{SUL}(\alpha)$ with density $f_X(x; \alpha)$. For $0 < x < 1$, define

$$M_1(x; \alpha) = \mathbb{E}\{X\mathbf{1}_{X \le x}\} = \int_0^x t f_X(t; \alpha)\,\mathrm{d}t.$$

With the change of variable $u = \sqrt{(1-t)/t}$, we have

$$M_1(x; \alpha) = \alpha \int_{\sqrt{(1-x)/x}}^{\infty} \frac{e^{-\alpha u}}{1 + u^2}\,\mathrm{d}u.$$

In particular, $M_1(1; \alpha) = \mu_1(\alpha)$.

We now assemble the Lorenz curve by evaluating (2.9) piecewise.

Proposition 2.13 (Lorenz curve)**.** *Let $Y \sim \mathrm{ZOISUL}(\delta, \pi, \alpha)$ with mean $\mu = \mathbb{E}(Y) > 0$. Then,*

$$L_Y(p) = \begin{cases} 0, & 0 \le p \le p_0, \\ \dfrac{(1-\delta)M_1(x_p; \alpha)}{\mu}, & p_0 < p < 1 - p_1, \\ \dfrac{(1-\delta)\mu_1(\alpha) + p - (1 - p_1)}{\mu}, & 1 - p_1 \le p \le 1, \end{cases} \tag{2.11}$$

where

$$x_p = \frac{\alpha^2}{\alpha^2 + \left[\log\left(\dfrac{1-\delta}{p - p_0}\right)\right]^2}, \; p_0 < p < 1 - p_1.$$

Proof. For $0 \le p \le p_0$, $Q_Y(s) = 0$, and hence, $L_Y(p) = 0$. For $p_0 < p < 1 - p_1$, let $x_p = Q_Y(p)$. From

$$s = p_0 + (1 - \delta)F_X(y; \alpha),$$

we have $\mathrm{d}s = (1 - \delta)f_X(y; \alpha)\,\mathrm{d}y$. Therefore,

$$\int_0^p Q_Y(s)\,\mathrm{d}s = \int_{p_0}^p Q_Y(s)\,\mathrm{d}s = (1 - \delta)\int_0^{x_p} y f_X(y; \alpha)\,\mathrm{d}y = (1 - \delta)M_1(x_p; \alpha).$$

For $1 - p_1 \le p \le 1$, $Q_Y(s) = 1$ on $[1 - p_1, p]$, giving

$$\int_0^p Q_Y(s)\,\mathrm{d}s = (1-\delta)M_1(1;\alpha) + p - (1-p_1) = (1-\delta)\mu_1(\alpha) + p - (1-p_1).$$

Substitution into (2.9) yields (2.11). □

The Gini coefficient is defined through the Lorenz curve by

$$G_Y = 1 - 2\int_0^1 L_Y(p)\,\mathrm{d}p. \tag{2.12}$$

A direct integration of the piecewise Lorenz curve gives the following result.

Proposition 2.14 (Gini coefficient)**.** *Let $Y \sim \mathrm{ZOISUL}(\delta, \pi, \alpha)$ with mean $\mu = \mathbb{E}(Y) > 0$. Then, the Gini coefficient of Y is*

$$G_Y = 1 - \frac{2(1-\delta)^2}{\mu}\int_0^1 M_1(x;\alpha)f_X(x;\alpha)\,\mathrm{d}x - \frac{\{2(1-\delta)\mu_1(\alpha) + p_1\}p_1}{\mu}, \tag{2.13}$$

where $f_X(x;\alpha)$ is the density of the SUL distribution, and $M_1(x;\alpha)$ is the lower incomplete first moment defined in Definition 2.12.

Proof. From (2.11), $L_Y(p) = 0$ on $[0, p_0]$. For $p_0 < p < 1 - p_1$, let $x_p = Q_Y(p)$. Because

$$p = p_0 + (1-\delta)F_X(x_p;\alpha),$$

we have $\mathrm{d}p = (1-\delta)f_X(x;\alpha)\,\mathrm{d}x$, and therefore,

$$\int_{p_0}^{1-p_1} L_Y(p)\,\mathrm{d}p = \frac{(1-\delta)^2}{\mu}\int_0^1 M_1(x;\alpha)f_X(x;\alpha)\,\mathrm{d}x.$$

For $1 - p_1 \le p \le 1$, Eq (2.11) gives

$$\int_{1-p_1}^1 L_Y(p)\,\mathrm{d}p = \frac{(1-\delta)\mu_1(\alpha)p_1 + p_1^2/2}{\mu}.$$

Substituting these two expressions into (2.12) yields (2.13). □

To clarify the relationship between the Gini coefficient of the ZOISUL distribution and that of its continuous component, define the Gini coefficient of the SUL distribution by

$$G_X(\alpha) = 1 - \frac{2}{\mu_1(\alpha)}\int_0^1 M_1(x;\alpha)f_X(x;\alpha)\,\mathrm{d}x. \tag{2.14}$$

This leads to the following decomposition.

Corollary 2.15 (Decomposition of the Gini coefficient)**.** *Let $Y \sim \mathrm{ZOISUL}(\delta, \pi, \alpha)$. Then,*

$$G_Y = \frac{(1-\delta)^2\mu_1(\alpha)G_X(\alpha) + p_0(1-\delta)\mu_1(\alpha) + p_1(1-\delta)\{1-\mu_1(\alpha)\} + p_0p_1}{\mu}, \tag{2.15}$$

where

$$p_0 = \Pr(Y=0) = \delta(1-\pi),\ p_1 = \Pr(Y=1) = \delta\pi,\ \mu = p_1 + (1-\delta)\mu_1(\alpha).$$

Proof. From (2.14),

$$\int_0^1 M_1(x;\alpha) f_X(x;\alpha)\,\mathrm{d}x = \frac{\mu_1(\alpha)}{2}\{1 - G_X(\alpha)\}.$$

Substituting this expression into (2.13), together with $p_0 + p_1 = \delta$ and $1 - p_1 = 1 - \delta + p_0$, gives (2.15) after straightforward algebra. □

Remark 2.16. Equations (2.11) and (2.15) make the discrete-continuous mixture structure of the ZOISUL distribution explicit. The point mass at zero yields the initial flat segment of the Lorenz curve, the continuous SUL component determines the middle part through $M_1(x;\alpha)$, and the point mass at one gives the final linear segment. The decomposition in (2.15) separates the within-component contribution of the SUL distribution from the between-component effects induced by the endpoint masses. In particular, if $\delta = 0$, then $p_0 = p_1 = 0$, $\mu = \mu_1(\alpha)$, and hence, $G_Y = G_X(\alpha)$. As $\delta \to 1$ with fixed $\pi \in (0, 1)$, the distribution degenerates to a Bernoulli-type law on $\{0, 1\}$, and $G_Y \to 1 - \pi$.

3. Likelihood inference and simulation study

3.1. Closed-form maximum likelihood estimation

Let $Y_1, \cdots, Y_n$ be an independent random sample from the ZOISUL distribution. Define

$$n_0 = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i=0\}},\ n_1 = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i=1\}},\ n_2 = \sum_{i=1}^{n} \mathbf{1}_{\{0<Y_i<1\}},$$

where $\mathbf{1}_A$ denotes the indicator function of the event A. Clearly, $n = n_0 + n_1 + n_2$. Let

$$\mathcal{I} = \{i : 0 < Y_i < 1\}$$

denote the index set of observations falling in the open interval.

Up to an additive constant independent of the parameter vector (δ, π, α), the sample log-likelihood can be written as

$$\ell(\delta, \pi, \alpha) = \ell_1(\delta, \pi) + \ell_2(\alpha) + C,$$

where C is independent of (δ, π, α), and

$$\ell_1(\delta, \pi) = (n_0 + n_1)\log\delta + n_2 \log(1 - \delta) + n_0 \log(1 - \pi) + n_1 \log\pi, \tag{3.1}$$

$$\ell_2(\alpha) = n_2 \log\alpha - \alpha \sum_{i\in\mathcal{I}} \sqrt{\frac{1 - Y_i}{Y_i}}. \tag{3.2}$$

Thus, (δ, π) and α are separable in the log-likelihood and can therefore be estimated separately.

Proposition 3.1 (Maximum likelihood estimation)**.** *Let $Y_1, \cdots, Y_n$ be an independent random sample from* ZOISUL(δ, π, α)*. If $n_0 > 0$, $n_1 > 0$, and $n_2 > 0$, then the maximum likelihood estimators (MLEs) of δ, π, and α are*

$$\widehat{\delta} = \frac{n_0 + n_1}{n},\ \widehat{\pi} = \frac{n_1}{n_0 + n_1},\ \widehat{\alpha} = \frac{n_2}{\displaystyle\sum_{i\in\mathcal{I}} \sqrt{\frac{1 - Y_i}{Y_i}}}. \tag{3.3}$$

Moreover, these MLEs are unique.

Proof. Differentiating (3.1) and (3.2) with respect to δ, π, and α, respectively, gives

$$\frac{\partial \ell_1}{\partial \delta} = \frac{n_0 + n_1}{\delta} - \frac{n_2}{1-\delta}, \quad \frac{\partial \ell_1}{\partial \pi} = -\frac{n_0}{1-\pi} + \frac{n_1}{\pi},$$

and

$$\frac{\partial \ell_2}{\partial \alpha} = \frac{n_2}{\alpha} - \sum_{i \in \mathcal{I}} \sqrt{\frac{1-Y_i}{Y_i}}.$$

Solving the likelihood equations yields (3.3). The corresponding second derivatives are

$$\frac{\partial^2 \ell_1}{\partial \delta^2} = -\frac{n_0 + n_1}{\delta^2} - \frac{n_2}{(1-\delta)^2} < 0,$$

$$\frac{\partial^2 \ell_1}{\partial \pi^2} = -\frac{n_0}{(1-\pi)^2} - \frac{n_1}{\pi^2} < 0,$$

and

$$\frac{\partial^2 \ell_2}{\partial \alpha^2} = -\frac{n_2}{\alpha^2} < 0.$$

Because the log-likelihood is separable in δ, π, and α, the corresponding second derivatives are strictly negative throughout their respective parameter spaces, the log-likelihood is strictly concave. Hence, the stationary point in (3.3) is the unique global maximizer, establishing the uniqueness of the MLEs. □

Remark 3.2. If one of the three types of observations is absent, the corresponding estimator may lie on the boundary of the parameter space. For example, $n_1 = 0$ gives $\widehat{\pi} = 0$, $n_0 = 0$ gives $\widehat{\pi} = 1$, and $n_2 = 0$ makes the continuous-component parameter α unidentifiable from the sample. Thus, Eq (3.3) corresponds to the nondegenerate case in which all three components are observed.

The separable likelihood gives a direct computational procedure for fitting the proposed model. Algorithm 1 summarizes the estimation, standard-error calculation, and information-criterion computation used in the numerical experiments and empirical application.

Algorithm 1 Likelihood-based computation for the zero–one-inflated smooth unit-Laplace model.

Require: Observations $y_1, \ldots, y_n \in [0, 1]$.

1: Classify observations as $y_i = 0$, $y_i = 1$, or $0 < y_i < 1$.
2: Compute n_0, n_1, n_2 and the interior index set $\mathcal{I} = \{i : 0 < y_i < 1\}$.
3: **if** $n_0 > 0$, $n_1 > 0$, and $n_2 > 0$, **then**
4: Set $\widehat{\delta} = (n_0 + n_1)/n$ and $\widehat{\pi} = n_1/(n_0 + n_1)$.
5: Set $\widehat{\alpha} = n_2 \Big/ \sum_{i \in \mathcal{I}} \sqrt{(1-y_i)/y_i}$.
6: Evaluate $\ell(\widehat{\delta}, \widehat{\pi}, \widehat{\alpha})$ from (3.1)–(3.2).
7: Compute standard errors from the plug-in inverse Fisher information matrix.
8: Compute model-selection quantities, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).
9: **else**
10: Report the corresponding boundary estimate and indicate the nonidentifiable component, if any.
11: **end if**
12: **return** $\widehat{\delta}, \widehat{\pi}, \widehat{\alpha}$, standard errors, AIC, and BIC.

To justify the linear computational complexity of the proposed likelihood-based estimation procedure, we analyze the computational cost of Algorithm 1. Because all estimators in the non-degenerate case admit closed-form expressions, Algorithm 1 avoids nonlinear equation solving and iterative optimization. From the raw sample, the required summaries are n_0, n_1, n_2, and

$$S_n = \sum_{i \in \mathcal{I}} \sqrt{\frac{1 - Y_i}{Y_i}}, \ \mathcal{I} = \{i : 0 < Y_i < 1\}.$$

Thus, the computational cost can be written as

$$T(n, n_2) = c_1 n + c_2 n_2 + c_3 = \Theta(n), \ 0 \le n_2 \le n,$$

where c_1, c_2, c_3 are constants independent of n. The additional storage requirement is $O(1)$ apart from the input sample.

3.2. Fisher information matrix and asymptotic properties

Let $\eta = (\delta, \pi, \alpha)$. By the separable log-likelihood in Section 3.1, the unit-sample Fisher information matrix is

$$\boldsymbol{I}(\eta) = \operatorname{diag}\left(\frac{1}{\delta(1-\delta)}, \frac{\delta}{\pi(1-\pi)}, \frac{1-\delta}{\alpha^2}\right). \tag{3.4}$$

Proposition 3.3 (Large-sample properties of the maximum likelihood estimators)**.** *Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables from the* ZOISUL(δ, π, α) *distribution. Suppose that the true parameter vector $\boldsymbol{\eta} = (\delta, \pi, \alpha)^{\mathsf{T}} \in \mathcal{P} = (0,1) \times (0,1) \times (0, \infty)$. Then,*

$$\widehat{\boldsymbol{\eta}} \xrightarrow{\text{a.s.}} \boldsymbol{\eta}, \tag{3.5}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{\text{d}} \mathcal{N}_3\left(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\eta})^{-1}\right), \tag{3.6}$$

where

$$\boldsymbol{I}(\boldsymbol{\eta})^{-1} = \operatorname{diag}\left\{\delta(1-\delta), \frac{\pi(1-\pi)}{\delta}, \frac{\alpha^2}{1-\delta}\right\}.$$

Proof. Define

$$R_i = \begin{cases} \sqrt{\dfrac{1 - Y_i}{Y_i}}, & 0 < Y_i < 1, \\ 0, & Y_i \in \{0, 1\}, \end{cases} \quad \boldsymbol{U}_i = \left(\mathbf{1}_{\{Y_i \in \{0,1\}\}}, \mathbf{1}_{\{Y_i = 1\}}, R_i\right)^{\mathsf{T}}.$$

Because $R_i \mid \{0 < Y_i < 1\} \sim \text{Exp}(\alpha)$,

$$\text{E}(\boldsymbol{U}_i) = \boldsymbol{m} = \left(\delta, \delta\pi, \frac{1-\delta}{\alpha}\right)^{\mathsf{T}}, \ \text{E}\|\boldsymbol{U}_i\|^2 < \infty.$$

Let $\overline{\boldsymbol{U}}_n = n^{-1} \sum_{i=1}^n \boldsymbol{U}_i$. The closed-form maximum likelihood estimator satisfies

$$\widehat{\boldsymbol{\eta}} = g(\overline{\boldsymbol{U}}_n), \ g(a, d, r) = \left(a, \frac{d}{a}, \frac{1-a}{r}\right)^{\mathsf{T}}.$$

Because g is continuous at $\boldsymbol{m}$, the strong law of large numbers and the continuous mapping theorem [28] yield

$$\widehat{\boldsymbol{\eta}} \xrightarrow{\text{a.s.}} g(\boldsymbol{m}) = \boldsymbol{\eta}.$$

Furthermore, the multivariate central limit theorem and the delta method [28] give

$$\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{\text{d}} \mathcal{N}_3\left(\mathbf{0}, \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^{\mathsf{T}}\right),$$

where $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{U}_i)$, and $\boldsymbol{G}$ is the Jacobian matrix of g evaluated at $\boldsymbol{m}$. A direct calculation gives

$$\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^{\mathsf{T}} = \mathrm{diag}\left\{\delta(1-\delta), \frac{\pi(1-\pi)}{\delta}, \frac{\alpha^2}{1-\delta}\right\} = \boldsymbol{I}(\boldsymbol{\eta})^{-1}.$$

This proves (3.6). □

It follows from (3.6) that the asymptotic variances of the estimators are

$$\mathrm{Var}(\widehat{\delta}) \approx \frac{\delta(1-\delta)}{n}, \ \mathrm{Var}(\widehat{\pi}) \approx \frac{\pi(1-\pi)}{n\delta}, \ \mathrm{Var}(\widehat{\alpha}) \approx \frac{\alpha^2}{n(1-\delta)}.$$

Using the plug-in method, the unknown parameters in the asymptotic variances are replaced by their maximum likelihood estimates [28], yielding the following standard errors:

$$\mathrm{se}(\widehat{\delta}) = \sqrt{\frac{\widehat{\delta}(1-\widehat{\delta})}{n}}, \ \mathrm{se}(\widehat{\pi}) = \sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n\widehat{\delta}}}, \ \mathrm{se}(\widehat{\alpha}) = \frac{\widehat{\alpha}}{\sqrt{n(1-\widehat{\delta})}}.$$

Furthermore, based on the Wald large-sample normal approximation [29], approximate $100(1-\gamma)\%$ confidence intervals can be constructed. Let $z_{1-\gamma/2}$ denote the $1-\gamma/2$ quantile of the standard normal distribution. The intervals are

$$\widehat{\delta} \pm z_{1-\gamma/2}\sqrt{\frac{\widehat{\delta}(1-\widehat{\delta})}{n}}, \ \widehat{\pi} \pm z_{1-\gamma/2}\sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n\widehat{\delta}}}, \ \widehat{\alpha} \pm z_{1-\gamma/2}\frac{\widehat{\alpha}}{\sqrt{n(1-\widehat{\delta})}}.$$

Remark 3.4. Because δ and π are probability parameters, Wald intervals may exceed the interval $(0, 1)$ in small samples or when the estimates are close to the boundary. In applications, logit-transformed confidence intervals may be used to preserve the parameter-space constraints.

3.3. Simulation study

To investigate the finite-sample performance of the maximum likelihood estimators under different degrees of endpoint inflation, we consider

$$Y \sim \mathrm{ZOISUL}(\delta, \pi, \alpha).$$

The parameter values are chosen as

$$\delta \in \{0.05, 0.50, 0.95\}, \ \pi \in \{0.10, 0.50, 0.90\}, \ \alpha \in \{0.50, 1.20, 2.00\},$$

yielding $3^3 = 27$ parameter combinations. The sample sizes are

$$n \in \{500, 1000, 2000, 5000\},$$

and $R = 5000$ independent Monte Carlo replications are performed for each combination of parameter values and sample size.

For computational efficiency, the program uses the data-generating mechanism implied by the stochastic representation in Proposition 2.7 to simulate the corresponding sufficient statistics directly, without generating and storing every Y_i individually. In each replication, $\widehat{\delta}$, $\widehat{\pi}$, and $\widehat{\alpha}$ are computed using the closed-form expressions in Eq (3.3).

Owing to space limitations, the main text reports the following five representative parameter settings:

$$(\delta, \pi, \alpha) \in \{(0.50, 0.50, 1.20), (0.05, 0.10, 0.50), (0.05, 0.90, 2.00), (0.95, 0.10, 2.00), (0.95, 0.90, 0.50)\}.$$

These settings cover the central region of the parameter space, endpoint-inflation parameters close to the boundary, rare endpoint categories, and cases with relatively few observations from the continuous component. They therefore provide a broad assessment of the estimators under distinct data configurations. In particular, varying values of δ allow us to examine different degrees of boundary inflation, as δ directly determines the total probability mass assigned to the endpoints. Large values of δ generate substantial point masses at 0 and 1, whereas small values of δ reduce the contribution of the discrete component and emphasize the behavior of the continuous SUL component. Consequently, the simulation settings also illustrate the distinction between exact endpoint inflation and continuous concentration near the boundaries.

Following commonly used evaluation criteria for Monte Carlo simulation studies [30–32], estimation performance is assessed using the empirical mean, empirical standard deviation (SD), absolute bias (AB), absolute relative bias (ARB), root mean squared error (RMSE), and coverage probability (CP) of the 95% confidence interval. For any $\vartheta \in \{\delta, \pi, \alpha\}$, let $\widehat{\vartheta}_r$ denote the estimate obtained from the rth replication. The empirical mean and SD are defined as

$$\overline{\widehat{\vartheta}} = \frac{1}{R}\sum_{r=1}^{R}\widehat{\vartheta}_r, \ \mathrm{SD}(\widehat{\vartheta}) = \left[\frac{1}{R-1}\sum_{r=1}^{R}\left(\widehat{\vartheta}_r - \overline{\widehat{\vartheta}}\right)^2\right]^{1/2}.$$

The AB and ARB are given by

$$\mathrm{AB}(\widehat{\vartheta}) = \left|\overline{\widehat{\vartheta}} - \vartheta\right|, \ \mathrm{ARB}(\widehat{\vartheta}) = \frac{\left|\overline{\widehat{\vartheta}} - \vartheta\right|}{|\vartheta|} \times 100\%,$$

and the RMSE is

$$\mathrm{RMSE}(\widehat{\vartheta}) = \left[\frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\vartheta}_r - \vartheta\right)^2\right]^{1/2}.$$

Finally, the coverage probability of the 95% confidence interval is defined as

$$\mathrm{CP}_{95\%}(\widehat{\vartheta}) = \frac{1}{R}\sum_{r=1}^{R}\mathbf{1}\left\{\vartheta \in \mathrm{CI}_{r,95\%}(\widehat{\vartheta})\right\},$$

where $\mathrm{CI}_{r,95\%}(\widehat{\vartheta})$ denotes the Wald-type 95% confidence interval constructed in the rth replication using the asymptotic standard errors derived in Section 3.2.

Tables 1–5 present the simulation results for the five representative settings described above. All simulation results are reported to six decimal places. Overall, the empirical means of the three estimators are close to their corresponding true values, and their SDs and RMSEs decrease markedly as the sample size increases.

Under the central parameter setting, the ARBs of all three estimators are small, and their CPs remain close to the nominal level of 0.95. When the total endpoint mass is low and one endpoint category is particularly rare, $\widehat{\pi}$ exhibits greater variability, and its confidence interval shows some undercoverage at smaller sample sizes. Specifically, when $n = 500$, the CPs of $\widehat{\pi}$ under the settings $(0.05, 0.10, 0.50)$ and $(0.05, 0.90, 2.00)$ are 0.897200 and 0.896800, respectively. As the sample size increases, its SD and RMSE decrease steadily, whereas its CP improves to 0.942400 and 0.938600, respectively, at $n = 5000$. These results highlight the role of the effective number of endpoint observations in the estimation of π.

When $\delta = 0.95$, the expected proportion of observations from the continuous component is only 0.05. In these settings, $\widehat{\alpha}$ exhibits positive bias at smaller sample sizes. Under the parameter settings $(0.95, 0.10, 2.00)$ and $(0.95, 0.90, 0.50)$, its ARBs are 4.318362% and 4.270027%, respectively, when $n = 500$. The corresponding values decrease to 0.953827% and 1.001489% at $n = 2000$, and further decline to 0.353335% and 0.324299% at $n = 5000$. These reductions are accompanied by marked decreases in SD and RMSE, whereas the CPs remain close to 0.95. These findings demonstrate the improvement in the estimation of α as the effective number of continuous observations increases. Bias-correction procedures for extremely small effective sample sizes are beyond the scope of the present simulation study and remain a possible direction for future research.

Table 1. Monte Carlo summaries for MLEs under $(\delta, \pi, \alpha) = (0.50, 0.50, 1.20)$.

	$\widehat{\delta}$				$\widehat{\pi}$				$\widehat{\alpha}$			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Mean	0.500129	0.499966	0.500076	0.499925	0.500135	0.500273	0.500016	0.500121	1.205914	1.203270	1.201688	1.200785
SD	0.022616	0.015680	0.011053	0.007103	0.031769	0.022409	0.015763	0.009985	0.075891	0.054321	0.037453	0.024263
AB	0.000129	0.000034	0.000076	0.000075	0.000135	0.000273	0.000016	0.000121	0.005914	0.003270	0.001688	0.000785
ARB	0.025760	0.006720	0.015120	0.014976	0.027081	0.054574	0.003162	0.024165	0.492817	0.272520	0.140677	0.065380
RMSE	0.022615	0.015678	0.011052	0.007103	0.031766	0.022409	0.015761	0.009984	0.076114	0.054414	0.037488	0.024273
$\mathrm{CP}_{95\%}$	0.944200	0.949800	0.951000	0.950000	0.952600	0.950000	0.950000	0.953200	0.950600	0.946000	0.951600	0.945600

Table 2. Monte Carlo summaries for MLEs under $(\delta, \pi, \alpha) = (0.05, 0.10, 0.50)$.

	$\widehat{\delta}$				$\widehat{\pi}$				$\widehat{\alpha}$			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Mean	0.049994	0.050050	0.049886	0.050031	0.102893	0.100371	0.099546	0.100247	0.501338	0.500829	0.500279	0.500072
SD	0.009730	0.007062	0.004930	0.003091	0.061895	0.042782	0.029940	0.018794	0.023018	0.016469	0.011436	0.007207
AB	0.000006	0.000050	0.000114	0.000031	0.002893	0.000371	0.000454	0.000247	0.001338	0.000829	0.000279	0.000072
ARB	0.012000	0.100800	0.227800	0.062080	2.892641	0.371488	0.454181	0.246558	0.267533	0.165846	0.055721	0.014392
RMSE	0.009729	0.007061	0.004931	0.003091	0.061956	0.042779	0.029940	0.018793	0.023054	0.016488	0.011438	0.007206
$\mathrm{CP}_{95\%}$	0.930400	0.938000	0.940600	0.948800	0.897200	0.914600	0.930200	0.942400	0.951600	0.951600	0.950000	0.951800

Table 3. Monte Carlo summaries for MLEs under $(\delta, \pi, \alpha) = (0.05, 0.90, 2.00)$.

	$\widehat{\delta}$				$\widehat{\pi}$				$\widehat{\alpha}$			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Mean	0.049950	0.049884	0.050123	0.049954	0.899471	0.899828	0.900228	0.899820	2.003510	2.003453	2.001517	2.000435
SD	0.009682	0.006966	0.005026	0.003047	0.060409	0.043077	0.029808	0.019403	0.091242	0.065531	0.046022	0.028925
AB	0.000050	0.000116	0.000123	0.000046	0.000529	0.000172	0.000228	0.000180	0.003510	0.003453	0.001517	0.000435
ARB	0.100800	0.232000	0.245400	0.091920	0.058786	0.019141	0.025302	0.019970	0.175515	0.172665	0.075873	0.021748
RMSE	0.009682	0.006966	0.005027	0.003047	0.060405	0.043073	0.029806	0.019402	0.091300	0.065615	0.046043	0.028925
$CP_{95\%}$	0.936600	0.938800	0.943600	0.948000	0.896800	0.910200	0.930600	0.938600	0.951000	0.949600	0.951400	0.949800

Table 4. Monte Carlo summaries for MLEs under $(\delta, \pi, \alpha) = (0.95, 0.10, 2.00)$.

	$\widehat{\delta}$				$\widehat{\pi}$				$\widehat{\alpha}$			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Mean	0.949946	0.950140	0.949917	0.949976	0.100063	0.100337	0.100077	0.100002	2.086367	2.039487	2.019077	2.007067
SD	0.009697	0.006906	0.004816	0.003065	0.013799	0.009754	0.006858	0.004415	0.444669	0.297239	0.202935	0.125788
AB	0.000054	0.000140	0.000083	0.000024	0.000063	0.000337	0.000077	0.000002	0.086367	0.039487	0.019077	0.007067
ARB	0.005726	0.014758	0.008705	0.002564	0.062647	0.337199	0.076918	0.001784	4.318362	1.974370	0.953827	0.353335
RMSE	0.009696	0.006907	0.004816	0.003065	0.013798	0.009759	0.006857	0.004415	0.452936	0.299821	0.203809	0.125973
$CP_{95\%}$	0.934000	0.939200	0.952200	0.952000	0.946800	0.949200	0.949400	0.946400	0.949200	0.951400	0.952800	0.949200

Table 5. Monte Carlo summaries for MLEs under $(\delta, \pi, \alpha) = (0.95, 0.90, 0.50)$.

	$\widehat{\delta}$				$\widehat{\pi}$				$\widehat{\alpha}$			
n	500	1000	2000	5000	500	1000	2000	5000	500	1000	2000	5000
Mean	0.950217	0.950124	0.949893	0.949952	0.900178	0.900117	0.899841	0.899994	0.521350	0.510902	0.505007	0.501621
SD	0.009721	0.006837	0.004874	0.003086	0.013767	0.009720	0.006819	0.004387	0.112359	0.074488	0.051386	0.031764
AB	0.000217	0.000124	0.000107	0.000048	0.000178	0.000117	0.000159	0.000006	0.021350	0.010902	0.005007	0.001621
ARB	0.022863	0.013032	0.011253	0.005040	0.019759	0.013021	0.017642	0.000622	4.270027	2.180414	1.001489	0.324299
RMSE	0.009722	0.006837	0.004874	0.003086	0.013766	0.009720	0.006820	0.004387	0.114359	0.075274	0.051625	0.031802
$CP_{95\%}$	0.931600	0.945600	0.943800	0.943600	0.945200	0.948000	0.952200	0.950000	0.955000	0.948000	0.947600	0.949000

Overall, the simulation results indicate satisfactory finite-sample performance of the closed-form maximum likelihood estimators under the settings considered. The reductions in SD and RMSE, together with coverage probabilities generally close to the nominal level at larger sample sizes, provide numerical support for the accuracy of point estimation and the reliability of the associated confidence intervals as the sample size increases.

4. Application to boundary-inflated proportional data

This section illustrates the proposed computational modeling framework using a bounded empirical variable with endpoint observations. The example is taken from the PISA 2022 public database released by the Organisation for Economic Co-operation and Development (OECD) [33]. Specifically, we use the variable PROPAT7 from the school questionnaire file CY08MSP_SCH_QQQ (https://webfs.oecd.org/pisa2022/index.html), which measures the proportion of teachers with at least an International Standard Classification of Education (ISCED) Level 7 qualification. Because this variable is supported on [0, 1] and contains observations at both endpoints, it provides a suitable setting for assessing zero–one-inflated unit-interval models. In this context, the zero–one-inflated structure provides a practical interpretation of the two model components: The inflation component captures structural endpoint observations with proportions equal to zero or one, whereas the continuous component describes the

variation among observations strictly within $(0, 1)$. Thus, the two components jointly characterize both endpoint patterns and continuous heterogeneity in teacher qualification proportions. The purpose of the application is to evaluate model fit and interpretability for boundary-inflated bounded data rather than to develop a full educational policy model.

After excluding missing and invalid observations, $n = 18175$ valid school-level observations remain. Table 6 reports the main descriptive statistics of PROPAT7, including its quartiles, mean, range, and empirical endpoint proportions. The results show evident probability masses at both 0 and 1, with $\Pr(Y = 0) = 13.02\%$ and $\Pr(Y = 1) = 4.38\%$, indicating the necessity of using a zero–one-inflated model.

Table 6. Descriptive statistics and endpoint proportions for the PROPAT7 variable.

Min.	Q_1	Median	Mean	Q_3	Max.	$\Pr(Y = 0)$	$\Pr(Y = 1)$
0.0000	0.0588	0.1875	0.3249	0.5333	1.0000	13.02%	4.38%

To construct competing models, we consider five commonly used continuous distributions on the unit interval: the beta [9], unit-Lindley [13], unit-gamma [12], unit-Weibull [14], and Kumaraswamy [11] distributions. For $0 < x < 1$, the probability density functions of these five baseline continuous distributions are defined as follows:

$$\begin{aligned}
&\text{Beta: } g_1(x;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1},\\
&\text{unit-Lindley: } g_2(x;\theta) = \frac{\theta^2}{1+\theta}(1-x)^{-3}\exp\left\{-\frac{\theta x}{1-x}\right\},\\
&\text{unit-gamma: } g_3(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(-\log x)^{\alpha-1}x^{\beta-1},\\
&\text{unit-Weibull: } g_4(x;\alpha,\beta) = \alpha\beta x^{-1}(-\log x)^{\beta-1}\exp\left\{-\alpha(-\log x)^\beta\right\},\\
&\text{Kumaraswamy: } g_5(x;\alpha,\beta) = \alpha\beta x^{\alpha-1}(1-x^\alpha)^{\beta-1}.
\end{aligned}$$

For comparison, each baseline distribution is embedded in the same zero–one-inflated structure:

$$\Pr(Y = 0) = \delta(1-\pi),\ \Pr(Y = 1) = \delta\pi,$$

and

$$Y \mid (0 < Y < 1) \sim G(\cdot;\vartheta),$$

where $G(\cdot;\vartheta)$ denotes the corresponding continuous baseline distribution. For notational simplicity, the resulting inflated models are denoted by zero–one-inflated beta (ZOIB), zero–one-inflated unit-Lindley (ZOIUL), zero–one-inflated unit-gamma (ZOIUG), zero–one-inflated unit-Weibull (ZOIUW), and zero–one-inflated Kumaraswamy (ZOIK), respectively.

All models are fitted by maximum likelihood. For the ZOISUL and ZOIUL models, the parameter estimates and their standard errors are computed directly from the corresponding analytical expressions. For the remaining competing models, no closed-form maximum likelihood solution exists for the full parameter vector; hence, the parameter estimates are obtained by numerical maximization of the log-likelihood, and their standard errors are calculated from the inverse of the numerically evaluated

observed information matrix. The resulting parameter estimates and standard errors are reported in Table 7.

Table 7. Maximum likelihood estimates and standard errors for the fitted zero–one-inflated models.

	ZOISUL	ZOIB	ZOIUL	ZOIUG	ZOIUW	ZOIK
$\widehat{\mu}$	–	0.3724(0.0022)	–	–	–	–
$\widehat{\phi}$	–	1.8412(0.0182)	–	–	–	–
$\widehat{\alpha}$	0.4389(0.0036)	–	–	1.1928(0.0123)	0.5357(0.0057)	0.6974(0.0070)
$\widehat{\beta}$	–	–	–	0.7567(0.0096)	1.2203(0.0084)	1.1677(0.0124)
$\widehat{\theta}$	–	–	0.5220(0.0031)	–	–	–
$\widehat{\delta}$	0.1740(0.0028)	0.1740(0.0028)	0.1740(0.0028)	0.1740(0.0028)	0.1740(0.0028)	0.1740(0.0028)
$\widehat{\pi}$	0.2517(0.0077)	0.2517(0.0077)	0.2517(0.0077)	0.2517(0.0077)	0.2517(0.0077)	0.2517(0.0077)

To compare the ability of the candidate models to characterize endpoint inflation and the interior empirical distribution, Figure 2 presents the empirical histogram scaled by the total sample size, together with a nonparametric kernel density estimate (KDE) and the weighted fitted density curves of the six candidate models. Because the continuous observations are supported on $(0, 1)$, a reflection-based boundary correction is applied to the KDE to mitigate boundary effects [34, 35]. The KDE is further weighted by the proportion of continuous observations to ensure comparability with the empirical histogram and the fitted parametric densities. The empirical probability masses at 0 and 1 are represented by vertical lines.

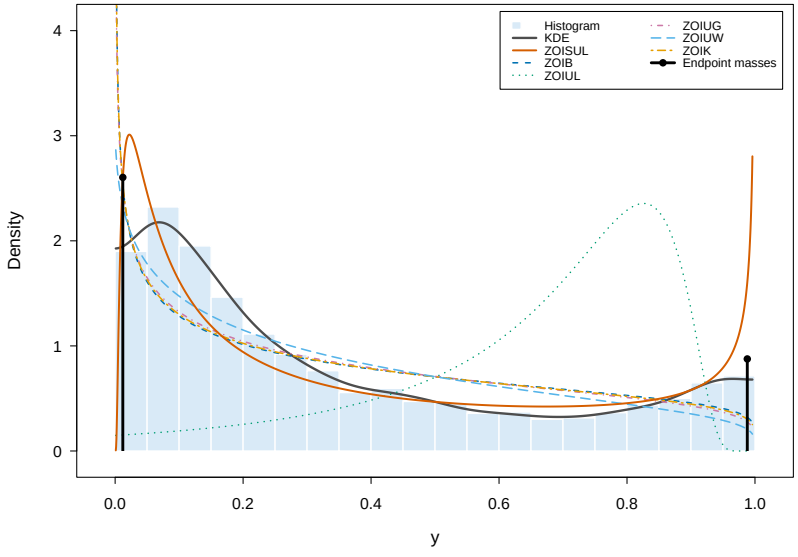


Figure 2. Fitted density curves of six zero–one-inflated unit-interval models for the empirical data.

Figure 2 shows that the candidate models differ in their characterization of the empirical distribution over the interior of the unit interval. Compared with the KDE curve, ZOISUL and ZOIUW show

relatively closer agreement with the main features of the empirical distribution, although their local fitting patterns differ. It should be noted that the sharp increase in the ZOISUL density as y approaches 1 reflects the inherent asymptotic behavior of its continuous SUL component near the upper boundary. This behavior is confined to a narrow region very close to the upper boundary and does not indicate corresponding empirical probability concentration throughout the upper range, nor does it materially affect the model's characterization of the empirical distribution over the main part of the interval. Accordingly, the model's agreement with the main features of the empirical distribution should be considered separately from its local behavior near the upper boundary.

Model fit is further evaluated using the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the mean squared error (MSE) between the empirical and fitted distribution functions, the Kolmogorov–Smirnov (KS) statistic, and its parametric-bootstrap-calibrated p-value. The MSE is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left\{F_n(y_i) - \widehat{F}(y_i)\right\}^2,$$

where F_n and $\widehat{F}$ denote the empirical and fitted distribution functions, respectively. Because the candidate models are discrete–continuous mixtures with endpoint probability masses, and their parameters are estimated from the data, the significance of the KS statistic is assessed using a parametric bootstrap. Specifically, bootstrap samples of the same size as the original sample are generated from each fitted model, the model parameters are re-estimated for each bootstrap sample, and the corresponding KS statistic is recalculated. A total of $B = 5000$ valid bootstrap replications are used to obtain the bootstrap-calibrated p-value and critical value. The results are reported in Table 8.

Table 8. Goodness-of-fit measures for the six candidate zero–one-inflated models.

	ZOISUL	ZOIB	ZOIUL	ZOIUG	ZOIUW	ZOIK
AIC	15012.9950	16580.1927	62162.6630	16458.1692	15967.6830	16532.3811
BIC	15036.4184	16611.4239	62186.0864	16489.4004	15998.9142	16563.6123
MSE	0.0017	0.0023	0.0664	0.0021	0.0014	0.0022
KS	0.0746	0.0926	0.4445	0.0869	0.0708	0.0895
p-value (KS)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

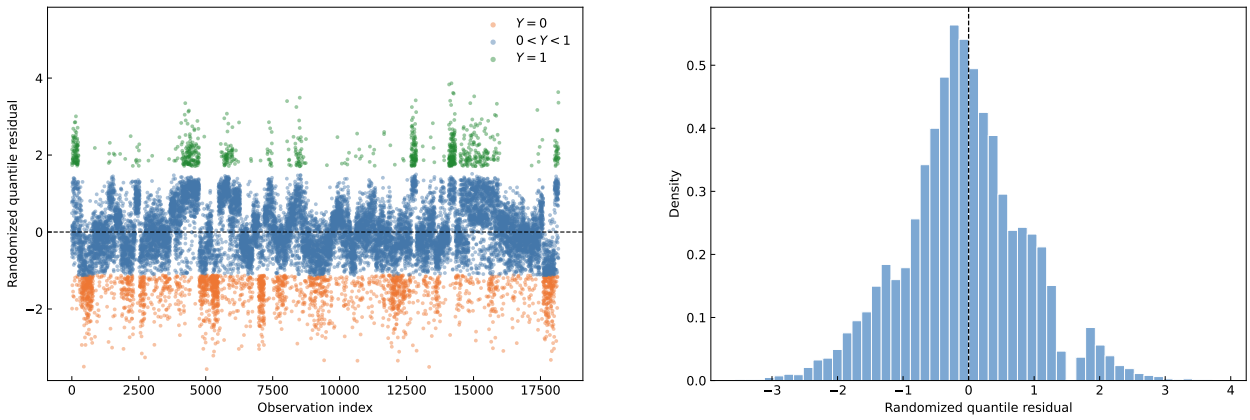
Table 8 shows that ZOISUL provides competitive relative fitting performance among the six candidate models, with substantially lower AIC and BIC values than the other models. Although ZOIUW yields slightly lower MSE and KS statistics, these criteria assess different aspects of model fit. AIC and BIC evaluate likelihood-based fit while accounting for model complexity, whereas MSE and KS measure discrepancies between the empirical and fitted cumulative distribution functions. Their rankings therefore need not coincide. Moreover, overall fit measures do not fully capture differences in fit across local intervals. Figure 2 shows that the ZOISUL density increases sharply within a narrow region very close to 1. The reported fit measures are calculated using the full sample and thus include the contributions of observations in this region. ZOISUL nevertheless achieves the lowest AIC and BIC, indicating that this local feature can coexist with its relative advantage under the information criteria.

The parametric-bootstrap-calibrated KS tests provide a complementary assessment of absolute goodness of fit. The test results are statistically significant for all candidate models, indicating

detectable discrepancies between the fitted and empirical distributions. In large samples, goodness-of-fit tests may be sensitive even to relatively small systematic departures, so their results should be interpreted alongside the magnitude of the discrepancies and other fit measures. These tests do not directly determine the relative ranking of candidate models. Regarding relative fit, ZOISUL achieves the lowest AIC and BIC among the models compared, indicating its advantage under these information criteria after accounting for model complexity.

Building on the preceding model comparisons, we further assessed the fitted ZOISUL model using randomized quantile residuals (RQRs). Figure 3(a) plots the residuals against the original record index, with different colors distinguishing observations at $Y = 0$, $0 < Y < 1$, and $Y = 1$. Figure 3(b) presents a density histogram of the residuals to illustrate their overall distribution. The residuals are constructed from the full discrete–continuous mixture distribution, with randomization over the cumulative probability intervals corresponding to the endpoint jumps. Consequently, the three observation categories occupy distinct residual intervals, producing the stratification in Figure 3(a); residuals within each category need not be symmetrically distributed around zero. Because the records are ordered by country and school, local segments of the scatterplot also display the residual patterns of the corresponding groups.

The histogram in Figure 3(b) is broadly unimodal and concentrated near zero, with a gap confined to a narrow interval on the positive side. This feature is consistent with sparse interior observations near the upper boundary: After transformation, a gap can arise between the residuals for interior observations and the randomized residuals for observations at $Y = 1$. Binning and randomization also affect its appearance. This local feature suggests a discrepancy between the model-assigned probability and the empirical distribution near the upper boundary, consistent with the KS tests calibrated by parametric bootstrap. These diagnostics provide additional insight into local fit, and the preceding AIC and BIC comparisons continue to support the relative advantage of ZOISUL among the candidate models.



(a) (b)

Figure 3. Randomized quantile residual diagnostics for the ZOISUL model fitted to the PROPAT7 data: (a) Residual scatterplot in the original record order; (b) Residual histogram.

Based on the overall fitting results, we further examine subgroup heterogeneity in the distribution of PROPAT7 with respect to school ownership and community location. Accordingly, the sample is

classified into four groups: public rural, private rural, public urban, and private urban schools. Table 9 summarizes the sample size and the empirical proportions of observations at 0 and 1 for each subgroup.

Table 9. Sample sizes and endpoint proportions by school group.

Group	Sample size	$\Pr(Y = 0)$	$\Pr(Y = 1)$
Public rural	9165	15.53%	5.68%
Private rural	1457	19.90%	3.71%
Public urban	4545	7.59%	3.67%
Private urban	2001	10.84%	1.80%

Table 9 reveals noticeable subgroup differences in both sample size and endpoint proportions. Public rural schools constitute the largest subgroup, indicating that they dominate the empirical sample. In terms of endpoint accumulation, private rural schools have the highest zero-inflation proportion, whereas public rural schools show a relatively high one-inflation proportion. These differences suggest heterogeneous distributional patterns in the proportion of teachers with at least an ISCED Level 7 qualification across school groups.

To further examine differences in the distribution of teachers with at least an ISCED Level 7 qualification across school groups, Figure 4 presents the empirical Lorenz curves and the corresponding Lorenz curves fitted by the ZOISUL model for the four groups. Lorenz curves provide an intuitive representation of distributional disparities across individuals or schools and have been applied in educational research to examine differences in learning outcomes and the allocation of educational resources [36]. Compared with summary measures of central tendency, Lorenz curves characterize disparities across the cumulative distribution, making them particularly useful for examining the concentration of teacher qualification proportions within different school groups. It should be noted that Lorenz curves summarize cumulative distributional differences and do not directly identify the specific sources or mechanisms underlying such disparities.

Figure 4 reveals clear differences in the Lorenz curves across the four school groups. The curve for private rural schools deviates the most from the line of equality, indicating the greatest concentration in the distribution of the proportion of teachers with at least an ISCED Level 7 qualification. In contrast, the curve for private urban schools lies closer to the line of equality, whereas the two public groups fall between these two cases. Meanwhile, the Lorenz curves fitted by the ZOISUL model are generally close to their empirical counterparts across all four groups, indicating that the model adequately captures the main distributional characteristics of the different school groups.

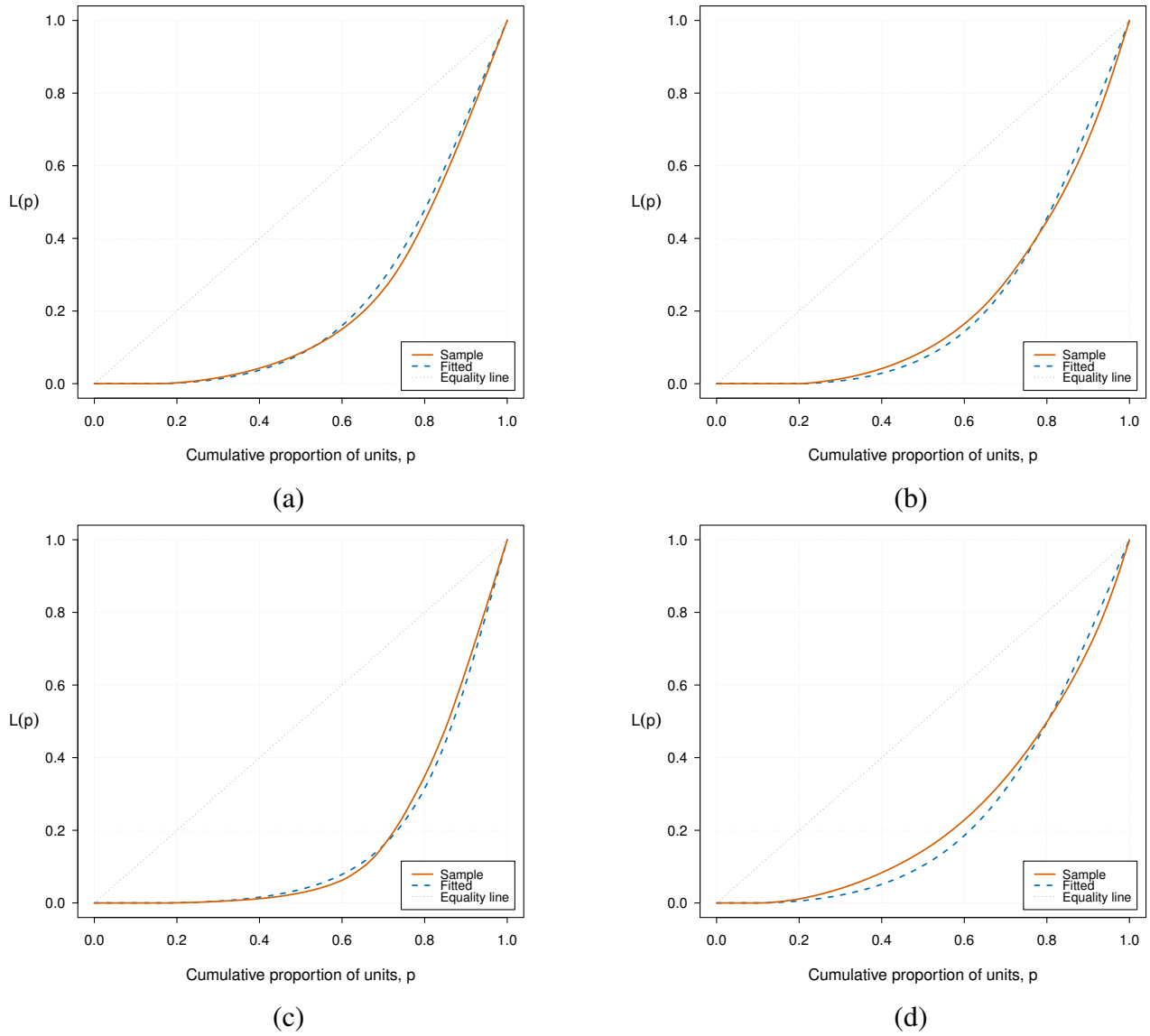


Figure 4. Sample and fitted Lorenz curves for the proportion of teachers with at least an ISCED Level 7 qualification across different school groups: (a) Public rural schools; (b) Private rural schools; (c) Public urban schools; (d) Private urban schools.

To further quantify the within-group distributional differences reflected by the Lorenz curves, we use the Gini coefficient to compare the four school groups. As a summary measure derived from the Lorenz curve, the Gini coefficient has been applied in educational research to characterize disparities in learning outcomes and educational resources across different populations and school contexts [37]. In this study, the Gini coefficient specifically measures the unevenness in the distribution of the proportion of teachers with at least an ISCED Level 7 qualification across schools within each group. A larger value indicates a more uneven distribution across schools and is not interpreted as a direct measure of educational inequality in a broader normative sense. It should also be noted that the Gini coefficient summarizes overall distributional dispersion and may not fully characterize inequality

patterns occurring in specific regions of the distribution. Therefore, in this study, the Gini coefficient is interpreted together with the Lorenz curves rather than as a standalone measure. To account for estimation uncertainty, Table 10 reports 95% bootstrap confidence intervals (95% CIs) for both the sample and theoretical Gini coefficients. The confidence intervals were obtained by resampling observations with replacement within each group and, for each bootstrap sample, recalculating the sample Gini coefficient, re-estimating the ZOISUL model, and deriving the corresponding theoretical Gini coefficient.

The results in Table 10 are broadly consistent with the distributional patterns shown by the Lorenz curves. Public urban schools have the highest sample Gini coefficient (0.6553), indicating the most uneven distribution of the proportion of teachers with at least an ISCED Level 7 qualification across schools within this group. In contrast, private urban schools have the lowest sample Gini coefficient (0.4967), indicating a relatively more even distribution, while the two rural groups fall between these two cases. The model-based Gini coefficients preserve this empirical ordering and are generally close to their sample counterparts.

Table 10. Estimated SUL shape parameters and bootstrap confidence intervals for Gini coefficients across school groups.

Group	$\widehat{\alpha}$	Sample Gini coefficient (95% CI)	Theoretical Gini coefficient (95% CI)
Public rural	0.4342	0.5627 (0.5566, 0.5687)	0.5469 (0.5409, 0.5525)
Private rural	0.4491	0.5641 (0.5490, 0.5792)	0.5681 (0.5546, 0.5814)
Public urban	0.4205	0.5313 (0.5244, 0.5382)	0.5170 (0.5102, 0.5237)
Private urban	0.4524	0.4967 (0.4845, 0.5086)	0.5223 (0.5120, 0.5323)

The bootstrap confidence intervals provide further information about the statistical relevance of the observed differences. For private rural schools, the confidence intervals overlap, and the absolute difference between the corresponding point estimates is small (0.0040), providing no clear evidence of statistically meaningful discrepancies. The wider intervals for private rural and private urban schools reflect the greater estimation uncertainty associated with these smaller subgroups. In contrast, the confidence intervals do not overlap for public rural, public urban, and private urban schools, suggesting possible discrepancies. Nevertheless, the differences for public rural and public urban schools are modest (0.0158 and 0.0143, respectively), whereas the private urban subgroup shows the largest discrepancy (0.0256), with the model-based coefficient exceeding the sample value. Because these intervals quantify the uncertainty of the two coefficients separately rather than that of their paired difference, their overlap is interpreted as evidence concerning agreement rather than as a formal test of equality.

Taken together, the Lorenz curves in Figure 4 and the Gini coefficients in Table 10 indicate that the ZOISUL model captures the broad ordering and overall levels of distributional inequality across the four school groups, although the degree of agreement varies across subgroups.

5. Conclusions

This paper developed a zero–one-inflated smooth unit-Laplace model for bounded data with endpoint observations. The model combines a smooth unit-interval transformation of a Laplace

random variable with point masses at 0 and 1, leading to a mixed discrete-continuous distribution with interpretable endpoint and interior components. Closed-form distributional quantities, the quantile function, the likelihood decomposition, and the maximum likelihood estimators were obtained. The Fisher information matrix and large-sample properties were also derived. Numerical simulations and the real-data application showed that the proposed model provides a computationally simple and interpretable framework for boundary-inflated bounded observations.

The model is useful when exact zeros and ones are structural features of the response distribution rather than data irregularities. By separating endpoint masses from the continuous component on $(0, 1)$, it can describe boundary inflation and interior distributional concentration within a unified likelihood framework.

Based on the structural characteristics and current scope of the proposed model, several extensions are worth further investigation. The continuous SUL component introduced in Section 2.1 involves only a single shape parameter α, with its density patterns characterized in Proposition 2.2 and illustrated in Figure 1. This specification ensures model parsimony and facilitates parameter estimation and is particularly suitable when the interior distribution can be adequately represented by these density patterns. However, for bounded data displaying pronounced multimodality, highly irregular interior shapes, or distributional features requiring separate control of central tendency and dispersion, more flexible continuous components may be preferable.

The upper-boundary behavior also deserves consideration when assessing model suitability. The continuous SUL density diverges as the upper boundary is approached for every $\alpha > 0$, so this feature cannot be removed by adjusting the shape parameter. The model may therefore be suitable when an increasing density near the upper boundary is compatible with the interior distribution. When the interior density instead approaches a finite limit or decreases toward the upper boundary, a continuous component with the corresponding boundary behavior may be more appropriate, particularly when inference concerns upper-tail probabilities. If the interior support is known to end strictly below the upper boundary, an appropriate support adjustment or truncation should also be considered. Future research may therefore explore additional shape parameters or alternative continuous constructions that allow more flexible control of both interior shape and boundary behavior while retaining the endpoint-inflation framework. In practical applications, model selection should take into account not only information criteria, but also the adequacy of the fitted density shape in regions relevant to the analysis, parameter interpretability, and the substantive purpose of the study.

Furthermore, the current framework is developed for univariate bounded responses. It can be extended to multivariate bounded data to characterize dependence structures and joint endpoint-inflation patterns. Another important direction is the development of regression extensions. Such extensions require appropriate link functions to preserve parameter constraints and careful interpretation of covariate effects, as covariates may simultaneously influence endpoint probabilities and the distributional behavior of observations within the interval. Developing flexible regression structures while maintaining model interpretability and valid statistical inference remains an important topic for future research.

Author contributions

Fuxiang Liu: Conceptualization, methodology, validation, supervision, project administration, writing–review and editing; Tingting Li: Conceptualization, methodology, software, formal analysis, data curation, visualization, writing–original draft. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

During the preparation of this work, the authors used ChatGPT for language polishing and formatting assistance. The authors reviewed, edited, and validated all AI-assisted outputs, and they take full responsibility for the content of the manuscript.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. OECD, *PISA 2022 results (Volume I): The state of learning and equity in education*, Paris: OECD Publishing, 2023. https://doi.org/10.1787/53f23881-en
2. OECD, *TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners*, Paris: OECD Publishing, 2019. https://doi.org/10.1787/1d0bc92a-en
3. S. G. Rivkin, E. A. Hanushek, J. F. Kain, Teachers, schools, and academic achievement, *Econometrica*, **73** (2005), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x
4. E. A. Hanushek, L. Woessmann, School resources and student achievement: A review of cross-country economic research, In: *Cognitive abilities and educational outcomes*, Cham: Springer, 2017, 149–171. https://doi.org/10.1007/978-3-319-43473-5_8
5. L. E. Papke, J. M. Wooldridge, Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *J. Appl. Econom.*, **11** (1996), 619–632.
6. R. Ospina, S. L. P. Ferrari, Inflated beta distributions, *Stat. Papers*, **51** (2010), 111–126. https://doi.org/10.1007/s00362-008-0125-4
7. R. Ospina, S. L. P. Ferrari, A general class of zero-or-one inflated beta regression models, *Comput. Stat. Data Anal.*, **56** (2012), 1609–1623. https://doi.org/10.1016/j.csda.2011.10.005
8. K. L. Mengersen, C. P. Robert, D. M. Titterington, *Mixtures: Estimation and applications*, John Wiley & Sons, Ltd, 2011. https://doi.org/10.1002/9781119995678
9. S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, *J. Appl. Stat.*, **31** (2004), 799–815. https://doi.org/10.1080/0266476042000214501
10. E. Gómez-Déniz, J. M. Sarabia, A family of generalised beta distributions: Properties and applications, *Ann. Data Sci.*, **5** (2018), 401–420. https://doi.org/10.1007/s40745-018-0143-6

11. M. C. Jones, Kumaraswamy's distribution: A beta-type distribution with some tractability advantages, *Stat. Methodol.*, **6** (2009), 70–81. https://doi.org/10.1016/j.stamet.2008.04.001
12. A. Grassia, On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions, *Aust. J. Stat.*, **19** (1977), 108–114. https://doi.org/10.1111/j.1467-842X.1977.tb01277.x
13. J. Mazucheli, A. F. B. Menezes, S. Chakraborty, On the one parameter unit-Lindley distribution and its associated regression model for proportion data, *J. Appl. Stat.*, **46** (2019), 700–714. https://doi.org/10.1080/02664763.2018.1511774
14. J. Mazucheli, A. F. B. Menezes, L. B. Fernandes, R. P. de Oliveira, M. E. Ghitany, The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, *J. Appl. Stat.*, **47** (2020), 954–974. https://doi.org/10.1080/02664763.2019.1657813
15. J. Mazucheli, A. F. Menezes, S. Dey, Unit-Gompertz distribution with applications, *Statistica*, **79** (2019), 25–43. https://doi.org/10.6092/issn.1973-2201/8497
16. K. Modi, V. Gill, Unit Burr-III distribution with application, *J. Stat. Manag. Syst.*, **23** (2020), 579–592. https://doi.org/10.1080/09720510.2019.1646503
17. C. R. da Silva, T. L. Pereira, L. M. A. Lima-Filho, P. J. Duarte-Neto, Zero-one-inflated unit gamma distribution and its application to basic sanitation data, *J. Stat. Theory Pract.*, **20** (2026), 53. https://doi.org/10.1007/s42519-026-00570-7
18. S. R. Bapat, R. Bhardwaj, On an inflated unit-Lindley distribution, *J. Stat. Res.*, **55** (2022), 299–311. https://doi.org/10.3329/jsr.v55i2.58807
19. F. Cribari-Neto, J. Santos, Inflated Kumaraswamy distributions, *An. Acad. Bras. Ciênc.*, **91** (2019), e20180955. https://doi.org/10.1590/0001-3765201920180955
20. G. Martínez-Flórez, R. Tovar-Falón, C. Barrera-Causil, Inflated unit-Birnbaum-Saunders distribution, *Mathematics*, **10** (2022), 667. https://doi.org/10.3390/math10040667
21. S. Kotz, T. J. Kozubowski, K. Podgórski, *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*, Birkhäuser Boston, 2001. https://doi.org/10.1007/978-1-4612-0173-1
22. V. S. Stojanović, T. J. Spasojević, M. Jovanović, Laplace-logistic unit distribution with application in dynamic and regression analysis, *Mathematics*, **12** (2024), 2282. https://doi.org/10.3390/math12142282
23. I. S. Gradshteyn, I. M. Ryzhik, *Table of integrals, series, and products*, Academic Press, 2014. https://doi.org/10.1016/C2010-0-64839-5
24. F. W. Olver, D. W. Lozier, R. Boisvert, C. W. Clark, *The NIST handbook of mathematical functions*, New York: Cambridge University Press, 2010.
25. B. C. Arnold, J. M. Sarabia, *Majorization and the Lorenz order with applications in applied mathematics and economics*, Cham: Springer, 2018. https://doi.org/10.1007/978-3-319-93773-1
26. M. Shaked, J. G. Shanthikumar, *Stochastic orders*, New York: Springer, 2007. https://doi.org/10.1007/978-0-387-34675-5

27. J. L. Gastwirth, The estimation of the Lorenz curve and Gini index, *Rev. Econ. Stat.*, **54** (1972), 306–316. https://doi.org/10.2307/1937992

28. G. Casella, R. Berger, *Statistical inference*, 2 Eds., New York: Chapman and Hall/CRC, 2024. https://doi.org/10.1201/9781003456285

29. A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Amer. Math. Soc.*, **54** (1943), 426–482. https://doi.org/10.1090/S0002-9947-1943-0012401-3

30. T. P. Morris, I. R. White, M. J. Crowther, Using simulation studies to evaluate statistical methods, *Stat. Med.*, **38** (2019), 2074–2102. https://doi.org/10.1002/sim.8086

31. J. Antonelli, M. Cefalu, N. Palmer, D. Agniel, Doubly robust matching estimators for high dimensional confounding adjustment, *Biometrics*, **74** (2018), 1171–1179. https://doi.org/10.1111/biom.12887

32. H. Ma, T. Nguyen, J. Jiang, Developing a practical measure: An asymmetric mean squared prediction error for small area estimation, *J. Amer. Stat. Assoc.*, **121** (2026), 1647–1660. https://doi.org/10.1080/01621459.2025.2555056

33. Organisation for Economic Co-operation and Development (OECD), *PISA 2022 database*, 2023. Available from: https://www.oecd.org/en/data/datasets/pisa-2022-database.html.

34. E. F. Schuster, Incorporating support constraints into nonparametric estimators of densities, *Commun. Stat. Theory Methods*, **14** (1985), 1123–1136. https://doi.org/10.1080/03610928508828965

35. B. W. Silverman, *Density estimation for statistics and data analysis*, New York: Routledge, 1998. https://doi.org/10.1201/9781315140919

36. D. Rodriguez-Segura, C. Campton, L. Crouchn, T. S. Slade, Looking beyond changes in averages in evaluating foundational learning: Some inequality measures, *Int. J. Educ. Dev.*, **87** (2021), 102411. https://doi.org/10.1016/j.ijedudev.2021.102411

37. J. Petersson, Using the Gini coefficient for assessing heterogeneity within classes and schools, *SN Soc. Sci.*, **3** (2023), 186. https://doi.org/10.1007/s43545-023-00772-x



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)