Gulf Journal of Mathematics ISSN: 2309-4966
Vol 22, Issue 2 (2026) 1-20 https://doi.org/10.56947/gjom.v22i2.4171

POISSON REGRESSION WITH RIGHT-CENSORED COVARIATE.

YAO ISMAEL KOFFI[1*], KONAN JEAN GEOFFROY KOUAKOU[1] and OUAGNINA HILI[1]

Abstract. Right-censoring of covariates is common in count-data applications, yet standard Poisson regression typically assumes fully observed predictors. We consider a cross-sectional Poisson model with one covariate subject to random right-censoring and propose a likelihood-based estimator that incorporates both uncensored and censored observations. For censored cases, the likelihood contribution integrates over the conditional distribution beyond the censoring threshold. The resulting maximum-likelihood approach accounts for censoring-induced uncertainty without discarding data or using ad hoc substitutions. Simulation results and an empirical illustration highlight the advantages of the proposed method.

Keywords. Asymptotic properties, count data, Monte carlo simulation.
2020 Mathematics Subject Classification. Primary 62J12; Secondary 62F12

1. Introduction

Count data arise in many areas of the applied sciences, including epidemiology, insurance, and environmental research. Poisson regression remains a central tool for relating event counts to explanatory variables through a log-link mean structure [16, 1]. In practice, however, covariates entering the linear predictor are not always fully observed. A common complication is right-censoring of a covariate, where the covariate is exactly observed for some units but, for others, only known to exceed a unit-specific threshold. Such censoring may result from measurement limitations (e.g., instrument saturation), top-coding for confidentiality, or administrative reporting rules. Ignoring censoring may lead to biased parameter estimates and misleading uncertainty quantification, thereby undermining scientific conclusions [9].

In applied work, covariate censoring is often handled using simple ad hoc strategies. Restricting estimation to fully observed records excludes units with censored covariate values [10], reducing available information and potentially causing substantial efficiency losses when the censoring rate is high [6]. Another common practice replaces censored values [19] with fixed surrogates, such as the censoring

Date: Received: Jan 28, 2026; Revised: Feb 26, 2026; Accepted: March 11, 2026.
* Corresponding author

threshold or a fraction thereof, ignoring variability in the unobserved covariate and possibly distorting estimated effects [4]. Multiple imputation provides a more principled alternative by generating plausible values under an explicit model and combining estimates using Rubin-type rules [7, 15, 16]. Although performance may improve relative to substitution [5], additional modeling assumptions on the covariate distribution are required, and computational demands can be substantial, particularly with correlated predictors or nonlinear regression structures [13, 9]. Most methodological developments on censoring have been motivated by survival analysis, where censoring affects the event time rather than a predictor [8, 13, 4].

The present paper considers cross-sectional count data in which the response is fully observed but at least one covariate is subject to random right-censoring [8]. This setting differs fundamentally: censoring acts directly on a predictor in the mean function, and uncertainty in the linear predictor arises from incomplete covariate measurement. Consequently, procedures designed for censored event times [5] do not directly address likelihood contributions generated by censored covariates in generalized linear models [8].

A likelihood-based framework is introduced for Poisson regression with a randomly right-censored covariate [4]. All observations are retained by treating censored covariate values [10]as latent variables and integrating them out of the Poisson likelihood[16]. For uncensored units, the likelihood contribution corresponds to the standard Poisson term conditional on observed covariates[13]. For censored units, the contribution takes the form of an integral of the Poisson likelihood over the region beyond the censoring threshold [3], weighted by a model for the conditional distribution of the censored covariate [5] given the fully observed covariates [16]. The resulting objective function coherently reflects uncertainty due to censoring and avoids both deletion and deterministic replacement.

A numerical maximum-likelihood estimator is constructed using a Newton–Raphson scheme. Because the score function and observed information involve one-dimensional integrals, accurate evaluation can be achieved through standard quadrature or Riemann-sum approximations, yielding a practical algorithm for estimation and inference. The observed information matrix provides standard errors and Wald-type inference in the usual manner [12]. Large-sample properties of the estimator [20], including existence , consistency[22] , and asymptotic normality [21], are established under regularity conditions adapted to the covariate-censoring setting [18].

The remainder of the paper is organized as follows[18]. Section 2 introduces the Poisson regression [16] model with a right-censored [9] covariate and derives the likelihood. Section 3 presents the estimation procedure and implementation details. Section 4 examines finite-sample behavior through simulation experiments. Section 5 illustrates the method with an application to real data. Section 6 concludes with a discussion and possible extensions.

2. Model

Definition 2.1. A random variable Y_i is modeled by a Poisson distribution if

$$P(Y_i = y_i) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}. \tag{2.1}$$

with $\mathbb{E}(Y_i) = \lambda = \mathrm{Var}(Y_i)$.

The parameter λ_i is modeled by the log link, so the model can be rewritten as follows [2]:

$$\forall i = 1...n, \quad \begin{cases} Y_i \sim \mathcal{P}(\lambda) \\ \lambda = \exp(\boldsymbol{\beta}^\top \mathbf{X}_i) \end{cases} \tag{2.2}$$

where $\mathbf{X}_i = (1, X_{i2}, \ldots, X_{ip})^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and

$$\boldsymbol{\beta}^\top \mathbf{X}_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \;\; \text{with} \;\; X_{i1} = 1$$

The likelihood of the Poisson model regression is given by :

$$L_n(\boldsymbol{\beta}) = \prod_{i=1}^{n} \exp\left(y_i(\boldsymbol{\beta}^\top \mathbf{X}_i) - \exp(\boldsymbol{\beta}^\top \mathbf{X}_i)\right)\frac{1}{y_i!} \tag{2.3}$$

and the log-likelihood of the Poisson model is given by $\log L_n$:

$$\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left\{ y_i \beta^\top \mathbf{X}_i - \exp(\boldsymbol{\beta}^\top \mathbf{X}_i) - \log(y_i!) \right\}. \tag{2.4}$$

Assume that X_{ip} is randomly right-censored. Let $\tilde{\mathbf{X}}_i = (X_{i1}, \ldots, X_{i(p-1)})$ be the vector of fully observed covariates where $X_{i1} = 1$, $\tilde{\boldsymbol{\beta}} = (\beta_1 ...., \beta_{p-1})$, the coefficient of the uncensored covariates and C_i the censoring variable. The true value of X_{ip} is observed when $X_{ip} < c_i$, and is censored otherwise. Thus, the observed data are given by $\min(X_{ip}, c_i)$. The censoring indicator is denoted by δ_i, which takes the value 1 when the covariate is observed and 0 when it is censored.

The Poisson model with right censored covariate is given by:

$$P(Y_i = y_i \mid \mathbf{X}_i, \delta_i) = \begin{cases} \mathcal{P}(\lambda(\mathbf{X}_i)), & \text{if } \delta_i = 1, \\ P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip} > c_i), & \text{if } \delta_i = 0, \end{cases} \tag{2.5}$$

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{i(p-1)}, X_p) = (\tilde{\mathbf{X}}_i, X_{ip})$

The likelihood of the model is given by:

$$\begin{aligned} L_n(\Theta) &= \prod_{i=1}^{n} P(Y_i = y_i \mid \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\beta})^{\delta_i}\, P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip} > c_i, \tilde{\boldsymbol{\beta}}, \beta_p)^{1-\delta_i} \\ &= \prod_{i=1}^{n} P(Y_i = y_i \mid \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\beta})^{\delta_i} \Bigg[\frac{1}{S(c_i \mid \tilde{\mathbf{x}}_i)} \int_{c_i}^{\infty} P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p) \\ &\qquad\qquad \times f_{X_{ip}\mid\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})\, dx_{ip} \Bigg]^{1-\delta_i}. \end{aligned} \tag{2.6}$$

where $S(c_i \mid \tilde{\mathbf{x}}_i) = P(X_{ip} > c_i \mid \tilde{\mathbf{X}}_i)$ is the conditional survival function, and $f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})$ is the conditional density of the censored covariate given the observed covariates.

The log-likelihood of the Poisson model with a right-censored covariate is given by $\log L_n := \ell_n$.

$$\begin{aligned} \ell_n(\Theta) &= \sum_{i=1}^{n} \delta_i \log\Big\{ P(y_i \mid \mathbf{X}_i, \boldsymbol{\beta}) \Big\} \\ &\quad + \sum_{i=1}^{n} (1-\delta_i) \Bigg\{ -\log\Big(S(c_i \mid \tilde{\mathbf{x}}_i) \Big) + \log \int_{c_i}^{\infty} P\big(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p\big) \\ &\qquad\qquad \times f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})\, dx_{ip} \Bigg\} \\ &= \sum_{i=1}^{n} \delta_i \Big\{ y_i(\boldsymbol{\beta}^\top \mathbf{x}_i) - \exp(\boldsymbol{\beta}^\top \mathbf{x}_i) - \log(y_i!) \Big\} \\ &\quad + \sum_{i=1}^{n} (1-\delta_i) \log \Bigg\{ S(c_i \mid \tilde{\mathbf{x}}_i)^{-1} \int_{c_i}^{\infty} \exp\Big(y_i(\tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{x}}_i + \beta_p x_{ip}) \\ &\qquad\qquad - \exp(\tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{x}}_i + \beta_p x_{ip}) - \log(y_i!) \Big) \\ &\qquad\qquad \times f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})\, dx_{ip} \Bigg\}. \end{aligned} \tag{2.7}$$

where $\Theta = (\tilde{\boldsymbol{\beta}}^\top, \beta_p, \boldsymbol{\theta}^\top)^\top$. The integral appearing in the likelihood can be approximated by a Riemann sum in the estimation.

3. Estimation

Given that the conditional distribution of the censored covariate, given the observed covariates, is most often Gaussian (see [17]), it is assumed in this work that $X_{ip} \mid \tilde{\mathbf{X}}_i$ follows a conditional Normal distribution. Under this assumption, the conditional survival function admits the explicit form

$$S(c_i \mid \tilde{\mathbf{x}}_i) = 1 - \Phi\left(\frac{c_i - \tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{x}}_i}{\sigma}\right).$$

This framework can, however, be extended to accommodate asymmetric or heavy-tailed distributions when the data exhibit such characteristics.

Let us adopt the following notations:

$$g(\Theta) \quad = \quad \log \int_{c_i}^{\infty} P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p) \cdot f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})\, dx_{ip},$$

$$h(\mathbf{X}_i, \boldsymbol{\beta}) = P(y_i \mid \mathbf{X}_i, \boldsymbol{\beta}) \quad \text{and} \quad L(\tilde{\mathbf{X}}_i, \tilde{\boldsymbol{\beta}}, \sigma^2) = \log\left[1 - \Phi(\frac{c_i - \tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{x}}_i}{\sigma})\right].$$

So,

$$\begin{aligned}
\frac{\partial g(\Theta, x_{ip})}{\partial \Theta} &= \int_{c_i}^{+\infty} B(x_{ip}, \Theta) P(x_{ip} \mid y_i, \tilde{\mathbf{x}}_i, \Theta)\, dx_{ip}. \\
\frac{\partial}{\partial \beta_l} h(\mathbf{X}_i, \boldsymbol{\beta}) &= X_{il}\left(y_i + \exp(\boldsymbol{\beta}^\top \mathbf{X}_i)\right), \quad l = 1, \ldots, p, \\
\frac{\partial L(\tilde{\mathbf{X}}_i, \tilde{\boldsymbol{\beta}}, \sigma^2)}{\partial \tilde{\beta}_l} &= \frac{X_{il}}{\sigma} \cdot \frac{\phi(z_i)}{1 - \Phi(z_i)}, \\
\frac{\partial L(\tilde{\mathbf{X}}_i, \tilde{\boldsymbol{\beta}}, \sigma^2)}{\partial \sigma^2} &= -\frac{(c_i - \tilde{\boldsymbol{\beta}}^\top \tilde{X}_i)\phi(z_i)}{2(\sigma^2)^{3/2}(1 - \Phi(z_i))}.
\end{aligned}$$

where

$$z_i = \frac{c_i - \tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{x}}_i}{\sigma},$$

$$P(x_{ip} \mid y_i, \tilde{\mathbf{x}}_i, \Theta) = \frac{P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p) f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})}{\int_{c_i}^{+\infty} P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p) f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta}) dx_{ip}},$$

is the conditional density of the of covariate x_{ip}, given the observed data $(y_i, \tilde{\mathbf{x}}_i)$ and

$$B(\Theta, x_{ip}) = \frac{\partial}{\partial \Theta}\left\{\log P(Y_i = y_i \mid \tilde{\mathbf{X}}_i = \tilde{\mathbf{x}}_i, X_{ip}, \tilde{\boldsymbol{\beta}}, \beta_p) + \log f_{X_{ip}|\tilde{\mathbf{X}}_i}(x_{ip} \mid \tilde{\mathbf{x}}_i, \boldsymbol{\theta})\right\}$$

After some tedious but straightforward algebra, the second-order derivatives are given by:

$$
\begin{aligned}
\frac{\partial^2}{\partial\beta_l\partial\beta_m} h(\mathbf{X}_i,\boldsymbol{\beta}) &= X_{il}X_{im}\psi_i(\Theta), \quad m=1,\dots p \\
\frac{\partial^2 L(\tilde{\mathbf{X}}_i,\tilde{\boldsymbol{\beta}},\sigma^2)}{\partial\sigma^2\partial\tilde{\beta}_l} &= X_{il}s_i(\boldsymbol{\Theta}), \\
\frac{\partial^2 L(\tilde{\mathbf{X}}_i,\tilde{\boldsymbol{\beta}},\sigma^2)}{\partial\tilde{\beta}_m\partial\tilde{\beta}_l} &= X_{il}X_{im}t_i(\Theta), \\
\frac{\partial^2 L(\tilde{\mathbf{X}}_i,\tilde{\boldsymbol{\beta}},\sigma^2)}{\partial(\sigma^2)^2} &= \frac{3}{4(\sigma^2)^{5/2}} l_i(\Theta),
\end{aligned}
$$

for $i = 1\cdots n$, where

$$
\begin{aligned}
\psi_i(\Theta) &= \exp(\boldsymbol{\beta}^\top \mathbf{X}_i), \\
t_i(\Theta) &= \frac{1}{\sigma^2}\frac{\phi'(Z_i)\,(1-\Phi(Z_i)) + \phi^2(Z_i)}{(1-\Phi(Z_i))^2}, \\
l_i(\Theta) &= (c_i - \tilde{\boldsymbol{\beta}}^\top\tilde{\mathbf{X}}_i)\,\frac{\phi(Z_i)}{1-\Phi(Z_i)}, \quad \text{and} \\
s_i(\Theta) &= -\frac{\phi(Z_i)}{2(\sigma^2)^{3/2}(1-\Phi(Z_i))} \\
&\quad + \frac{(1-\Phi(Z_i))\,(c_i - \tilde{\boldsymbol{\beta}}^\top\tilde{\mathbf{X}}_i)\,\phi'(Z_i) - \phi^2(Z_i)\,(c_i - \tilde{\boldsymbol{\beta}}^\top\tilde{\mathbf{X}}_i)}{\sigma^4(1-\Phi(Z_i))^2}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\frac{\partial^2 g(\Theta)}{\partial\Theta\partial\Theta^\top} &= \int_{c_i}^{\infty}\left[\frac{\partial B(x_{ip},\Theta)}{\partial\Theta}\cdot P(x_{ip}\mid y_i,\tilde{\mathbf{x}}_i,\Theta)\right]dx_{ip} \\
&\quad + \int_{c_i}^{\infty}\left[B^2(x_{ip},\Theta)\cdot P(x_{ip}\mid y_i,\tilde{\mathbf{x}}_i,\Theta)\right]dx_{ip} \\
&\quad - \left(\int_{c_i}^{\infty} B(x_{ip},\Theta)\cdot P(x_{ip}\mid y_i,\tilde{\mathbf{x}}_i,\Theta)dx_{ip}\right)^2 := r_i(\Theta).
\end{aligned}
$$

The log-likelihood score function for the parameter vector Θ is defined as

$$
U(\Theta) = \frac{\partial \ell_n(\Theta)}{\partial\Theta}. \tag{3.1}
$$

The score vector does not admit a closed-form expression and is therefore evaluated numerically. Starting from an initial value $\Theta^{(0)}$, the maximum likelihood estimator is obtained by applying a Newton–Raphson iteration based on a first-order Taylor expansion of the score function. Let $\Theta^{(k)}$ denote the estimate at iteration k. The next iterate $\Theta^{(k+1)}$ is defined as the solution to the linearized score equation:

$$U(\Theta^{(k+1)}) \approx U(\Theta^{(k)}) + \frac{\partial U(\Theta^{(k)})}{\partial \Theta^\top}(\Theta^{(k+1)} - \Theta^{(k)}) = 0.$$

Solving for $\Theta^{(k+1)}$, we obtain the Newton-Raphson update:

$$\begin{aligned}\Theta^{(k+1)} &= \Theta^{(k)} - \left(\frac{\partial U(\Theta^{(k)})}{\partial \Theta^\top}\right)^{-1} U(\Theta^{(k)}) \\ &= \Theta^{(k)} + \left(I_{\text{obs}}(\Theta^{(k)})\right)^{-1} U(\Theta^{(k)}),\end{aligned} \tag{3.2}$$

for $k = 0, 1, 2, \ldots$, where $U(\Theta^{(k)})$ is the score function evaluated at $\Theta^{(k)}$, and

$$I_{\text{obs}}(\Theta) = -\frac{\partial^2 \ell_n(\Theta)}{\partial \Theta \partial \Theta^\top}$$

denotes the observed Fisher information. This definition ensures consistency with the Newton-Raphson update.

Since the score and observed-information contributions involve integrals without closed-form solutions, we approximate these integrals numerically, using Riemann-sum quadrature in our implementation.

Given an initial value $\Theta^{(0)}$, the Newton–Raphson update in (3.2) is iterated until a convergence criterion is satisfied (e.g., a relative change in Θ or in the log-likelihood). A convenient starting point is the maximum likelihood estimate obtained from the complete-case analysis, where censored covariate values are treated as missing and the model is fitted using only the uncensored observations.

After convergence, the resulting estimator is denoted by $\hat{\Theta}_n$. The large-sample covariance matrix of the MLE is given by the inverse of the observed Fisher information matrix, defined as the negative Hessian of the log-likelihood:

$$\text{Var}(\hat{\Theta}_n) = I^{-1}_{\text{obs}}(\Theta).$$

A consistent estimator of this covariance matrix is obtained by evaluating the observed information at the MLE:

$$\widehat{\text{Var}}(\hat{\Theta}_n) = I^{-1}_{\text{obs}}(\hat{\Theta}_n).$$

We define the following notations for clarity :

- $S_n(\Theta) = \dfrac{\partial \ell_n(\Theta)}{\partial \Theta}$: the score function,
- $H_n(\Theta) = -\dfrac{\partial^2 \ell_n(\Theta)}{\partial \Theta \partial \Theta^\top}$: the observed information matrix,
- $F_n(\Theta) = \mathbb{E}[H_n(\Theta)]$: the expected (Fisher) information,
- I_k: the identity matrix of order k.

Also, we assume that $H_n(\Theta)$ is positive definite to ensure convergence and uniqueness of the estimator.

Using the approach proposed by [12], we use the Cholesky square root matrix to normalize the ML estimator. The left Cholesky square root $\mathbf{M}^{1/2}$ of a positive definite matrix $\mathbf{M}$ is the unique lower triangular matrix with strictly positive diagonal elements such that

$$\mathbf{M}^{1/2}(\mathbf{M}^{1/2})^\top = \mathbf{M}$$

For convenience, we define

$$\mathbf{M}^{\top/2} := (\mathbf{M}^{1/2})^\top, \quad \mathbf{M}^{-1/2} := (\mathbf{M}^{1/2})^{-1}, \quad \text{and} \quad \mathbf{M}^{-\top/2} := (\mathbf{M}^{\top/2})^{-1}.$$

Throughout this paper, we consider only the spectral norm of square matrices, denoted by $\|\cdot\|$. The spectral norm of a real-valued matrix $\mathbf{M}$ is defined as

$$\|\mathbf{M}\| = \big(\text{largest eigenvalue of } \mathbf{M}^\top\mathbf{M}\big)^{1/2} = \sup_{\|u\|=1} \|\mathbf{Mu}\|_2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm (or $\mathcal{L}_2$) of the vectors. We omit the subscript 2 in $\|\cdot\|_2$ since the spectral norm is induced by the Euclidean norm, and the context always makes the meaning clear. The smallest (respectively, largest) eigenvalue of a square matrix $\mathbf{M}$ is denoted by $\lambda_{\min}(\mathbf{M})$ (respectively, $\lambda_{\max}(\mathbf{M})$).

We begin by stating some regularity conditions:

- **C1** Covariates are bounded; that is, there exist compact sets $X \subset \mathbb{R}^p$.
- **C2** The true parameter value $\Theta_0 = (\tilde{\beta}_0^\top, \beta_{p0}, \boldsymbol{\theta}_0^\top)^\top$ lies in the interior of some known compact and convex set
  $\mathbf{K} = C\times G\times J \subset \mathbb{R}^{p+q}$, $(C \subset \mathbb{R}^{p-1}, G \subset \mathbb{R}, J \subset \mathbb{R}^q$ are the parameter spaces).
- **C3** There exist a positive constant λ_1 such that $n/\lambda_{\min}(F_n(\Theta_0)) \leq \lambda_1$ for every $n = 1, 2, ...$
- **C4** The censoring random variables C_i, for $i = 1, 2, \ldots,$ are bounded by some constant $M_i < \infty$.

Conditions **C1**–**C3** are standard in generalized linear regression model, see [12] for more details. Condition **C4** is specifically needed in the presence of covariate censoring.

For each $n = 1, 2, \ldots$ and $\varepsilon > 0$, we define the neighborhood

$$K_n(\varepsilon) := \left\{\Theta \in \mathbf{K} : (\Theta - \Theta_0)^\top F_n(\Theta - \Theta_0) \leq \varepsilon^2\right\}.$$

Our first result establishes that the solution to the score equation exists, remains within the neighborhood $K_n(\varepsilon)$ of Θ_0 for sufficiently large n, and is consistent for Θ_0.

Theorem 3.1. *Suppose that regularity conditions **C1**-**C4** hold, it follow that*

(i) $\mathbb{P}\big(\mathcal{S}_n(\hat{\Theta}_n) = 0\big) \to 1$ *as* $n \to \infty$ *(asymptotic existence),*

(ii) $\hat{\Theta}_n \xrightarrow{\mathbb{P}} \Theta_0$ *as* $n \to \infty$ *(consistency).*

Proof. Due to the positive definiteness of $F_n(\Theta)$ and the convexity of the set $\mathbf{K}$, there is at most one zero of the score function. This zero corresponds to both a local and a global maximum of the likelihood function, if it exists. The maximum likelihood estimator (MLE) $\hat{\Theta}_n$ is the unique zero of the score function, when such a zero exists; otherwise, it is defined as an arbitrary constant in $\mathbf{K}$. Let $\partial K_n(\varepsilon) := \left\{\Theta \in \mathbf{K} : (\Theta - \Theta_0)^\top F_n(\Theta - \Theta_0) = \varepsilon^2\right\}$ the boundary of $K_n(\varepsilon)$. For any $n \in \mathbb{N}$, $\varepsilon > 0$ the event

$$\ell_n(\Theta) - \ell_n(\Theta_0) < 0 \ \text{ for all } \ \Theta \in \partial K_n(\varepsilon), \tag{3.3}$$

implies that there is a local maximum inside of $\mathcal{K}(\varepsilon)$ and this maximum is located at Θ_n. The positive definiteness of H_n and the convexity of $\mathbf{K}$ will ensure that this maximum is global and unique.

We show that for every $\eta > 0$, there exist $\varepsilon > 0$ such that

$$\mathbb{P}\big(\ell_n(\Theta) - \ell_n(\Theta_0) < 0 \text{ for all } \Theta \in \partial K_n(\varepsilon)\big) \geq 1 - \eta. \tag{3.4}$$

In fact, equivalently to (3.4), we show that for every $\eta > 0$, there exist $\varepsilon > 0$ and $n_1 \in \mathbb{N}$ such that

$$\mathbb{P}\big(\ell_n(\Theta) - \ell_n(\Theta_0) \geq 0 \text{ for all } \Theta \in \partial K_n(\varepsilon)\big) < \eta. \tag{3.5}$$

The Taylor expansion of the log-likelihood becomes

$$\begin{aligned} \ell_n(\Theta) - \ell_n(\Theta_0) &= (\Theta - \Theta_0)^\top \mathcal{S}_n(\Theta_0) - \frac{1}{2}(\Theta - \Theta_0)^\top \mathcal{H}_n(\tilde{\Theta})(\Theta - \Theta_0)^\top \quad (3.6)\\ &:= (\Theta - \Theta_0)^\top \mathcal{S}_n(\Theta_0) - R_n(\Theta). \quad (3.7)\end{aligned}$$

where $\tilde{\Theta} = a\Theta + (1-a)\Theta_0$ with $0 \leq a \leq 1$ lies between Θ and Θ_0. Let $0 < c < 1/2$. Then we have

$$\begin{aligned} &\mathbb{P}\left(\ell_n(\Theta) - \ell_n(\Theta_0) \geq 0 \text{ for all } \Theta \in \partial K_n(\varepsilon)\right)\\ &= \mathbb{P}\left((\Theta - \Theta_0)^\top S_n(\Theta_0) \geq R_n(\Theta),\ R_n(\Theta) > c\varepsilon^2, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right)\\ &\quad + \mathbb{P}\left((\Theta - \Theta_0)^\top S_n(\Theta_0) \geq R_n(\Theta),\ R_n(\Theta) \leq c\varepsilon^2, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right).\end{aligned}$$

Let events

$$\Gamma := \left\{(\Theta - \Theta_0)^\top S_n(\Theta_0) > c\varepsilon^2, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right\}$$

and

$$\Sigma := \left\{R_n(\Theta) \leq c\varepsilon^2, \text{ for all } \Theta \in \partial \mathcal{K}_n(\varepsilon)\right\},$$

Let $\Lambda_n(\Theta) := \frac{1}{\varepsilon} F_n^{1/2}(\Theta - \Theta_0)$. Then

$$\begin{aligned} \Gamma &= \left\{\Lambda_n(\Theta)^\top F_n^{-1/2} S_n(\Theta_0) > c\varepsilon, \text{ for all } \Theta \in \partial \mathcal{K}_n(\varepsilon)\right\}\\ &\subseteq \left\{\sup_{\Theta \in \partial K_n(\varepsilon)} \Lambda_n(\Theta)^\top F_n^{-1/2} S_n(\Theta_0) > c\varepsilon\right\}\\ &\subseteq \left\{\sup_{\|\Lambda_n(\Theta)\|=1} \Lambda_n(\Theta)^\top F_n^{-1/2} S_n(\Theta_0) > c\varepsilon\right\}\\ &\subseteq \left\{\left\|F_n^{-1/2} S_n(\Theta_0)\right\| > c\varepsilon\right\} \quad (3.8)\end{aligned}$$

where the transition from the second to the third line comes from the fact that $\Theta \in \partial K_n(\varepsilon)$ implies $\|\Lambda_n(\Theta)\| = 1$. It follows that $\mathbb{P}(\Gamma) \leq \mathbb{P}(\left\|F_n^{-1/2} S_n(\Theta_0)\right\| > c\varepsilon)$

By Lemma 5 from [22], $\mathbb{E}\|F_n^{-1/2} S_n(\Theta_0)\|^2 = Q$ and Chebyshev's inequality implies

$$\mathbb{P}(\Gamma) \leq \frac{Q}{\varepsilon^2 c^2} \tag{3.9}$$

Let $\varepsilon^2 = \dfrac{2Q}{c^2\eta}$ implies that $\mathbb{P}(\Gamma) \leq \frac{\eta}{2}$.Now

$$\begin{aligned}
\Sigma &= \left\{\frac{1}{2}(\Theta-\Theta_0)^\top H_n(\tilde{\Theta})(\Theta-\Theta_0) \leq c\varepsilon^2, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right\} \\
&= \left\{\frac{1}{2}\Lambda_n(\Theta)^\top F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\Lambda_n(\Theta) \leq c, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right\} \\
&\subseteq \left\{\frac{1}{2}\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right)\Lambda_n(\Theta)^\top\Lambda_n(\Theta) \leq c, \text{ for all } \Theta \in \partial K_n(\varepsilon)\right\} \\
&= \left\{\frac{1}{2}\lambda_{\min}\left(\mathcal{F}_n^{-1/2} \mathcal{H}_n(\tilde{\Theta}) \mathcal{F}_n^{-1/2}\right) \leq c, \text{ for all } \Theta \in \partial \mathcal{K}_n(\varepsilon)\right\} \qquad (3.10)
\end{aligned}$$

Thus, $\mathbb{P}(\Sigma) \leq \mathbb{P}\left(\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right) \leq 2c \text{ for all } \Theta \in \partial K_n(\varepsilon)\right)$.
By lemma (3.16) $F_n^{-1/2} H_n(\Theta) F_n^{-1/2} \xrightarrow{\mathbb{P}} I_k$ uniformly in $\Theta \in K_n(\varepsilon)$, as $n \to \infty$. Thus [18] $\lambda_{\min}\left(F_n^{-1/2} H_n(\Theta) F_n^{-1/2}\right) \xrightarrow{\mathbb{P}} 1$ uniformly in $\Theta \in K_n(\varepsilon)$, as $n \to \infty$.

If $\tilde{\Theta} = a\Theta + (1-a)\Theta_0$ for some $0 \leq a \leq 1$ and $\Theta \in \mathcal{K}_n(\varepsilon)$, then

$$\begin{aligned}
\left\|F_n^{1/2}(\tilde{\Theta}-\Theta_0)\right\| &= \left\|F_n^{1/2}\left(a\Theta+(1-a)\Theta_0-\Theta_0\right)\right\| \\
&= a\left\|F_n^{1/2}(\Theta-\Theta_0)\right\| \\
&\leq \left\|F_n^{1/2}(\Theta-\Theta_0)\right\| \\
&\leq \varepsilon, \qquad (3.11)
\end{aligned}$$

because for all $\Theta \in K_n(\varepsilon)$

$$\begin{aligned}
(\Theta_n-\Theta_0)^\top F_n(t\Theta_n-\Theta_0) \leq \varepsilon^2 &\iff (\Theta_n-\Theta_0)^\top F_n^{1/2}F_n^{1/2}(\Theta_n-\Theta_0) \leq \varepsilon^2 \\
&\iff \left[F_n^{1/2}(\Theta_n-\Theta_0)\right]^\top F_n^{1/2}(\Theta_n-\Theta_0) \leq \varepsilon^2 \\
&\iff \|F_n^{1/2}(\Theta_n-\Theta_0)\|^2 \leq \varepsilon^2 \\
&\iff \|F_n^{1/2}(\Theta_n-\Theta_0)\| \leq \varepsilon \qquad (3.12)
\end{aligned}$$

and thus $\tilde{\Theta} \in K_n(\varepsilon)$. From the above, it follows that $\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right) \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$, since

$$\left|\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right) - 1\right| \leq \sup_{\Theta \in K_n(\delta)}\left|\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right) - 1\right|. \tag{3.13}$$

Therefore, for $n \geq n_1$,

$$\mathbb{P}\left(\lambda_{\min}\left(F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2}\right) \leq 2c \text{ for all } \Theta \in \partial K_n(\varepsilon)\right) \leq \frac{\eta}{2}, \quad \text{since } 2c < 1.$$

This implies that $\mathbb{P}(\Sigma) \leq \eta/2$. Finally,

$$\mathbb{P}\left(\ell_n(\Theta) - \ell_n(\Theta_0) \geq 0 \text{ for all } \Theta \in \partial K_n(\varepsilon)\right) \leq \mathbb{P}(\Gamma) + \mathbb{P}(\Sigma) \leq \eta, \tag{3.14}$$

for all $\Theta \in \partial K_n(\varepsilon)$. Hence, this proves equation (3.5), and thus the existence of (3.14) a unique global maximum in $K_n(\varepsilon)$, which coincides with $\hat{\Theta}_n$.

Consistency of $\hat{\Theta}_n$.

$$\begin{aligned}
\lambda_{\min}(F_n)\|\hat{\Theta}_n - \Theta_0\|^2 &= (\hat{\Theta}_n - \Theta_0)^\top \lambda_{\min}(\mathcal{F}_n) \mathrm{I}_k (\hat{\Theta}_n - \Theta_0) \\
&\leq (\hat{\Theta}_n - \Theta_0)^\top F_n (\hat{\Theta}_n - \Theta_0) \\
&= (\hat{\Theta}_n - \Theta_0)^\top F_n^{1/2} F_n^{1/2} (\hat{\Theta}_n - \Theta_0) \\
&= \left[F_n^{1/2}(\hat{\Theta}_n - \Theta_0)\right]^\top F_n^{1/2}(\hat{\Theta}_n - \Theta_0) \\
&= \|F_n^{1/2}(\hat{\Theta}_n - \Theta_0)\|^2 \\
&\leq \varepsilon^2, \qquad (3.15)
\end{aligned}$$

with probability tending to 1 as $n \to \infty$, by (i). By condition **C3**, $\lambda_{\min}(F_n) \to \infty$ as $n \to \infty$. Therefore, $\|\hat{\Theta}_n - \Theta_0\| \to 0$ with probability tending to 1 as $n \to \infty$, which concludes the proof. □

Theorem 3.2. *Under the assumption* ***C1*** − ***C4***, $\hat{\Theta}_n$, *converges in distribution to the Gaussian vector with*

$$F_n^{\top/2}(\hat{\Theta}_n - \Theta) \xrightarrow{d} N(0, I_k).$$

Proof. The proof method we adopt here is similar to that of [22] in the presence of random right censoring in generalized linear models.
We first state the following lemma:

Lemma 3.3. *Assume conditions* ***C1***–***C4***. *Then,*

$$\forall \varepsilon > 0, \qquad \sup_{\Theta \in K_n(\varepsilon)} \left\| F_n^{-1/2} H_n(\Theta) F_n^{-1/2} - I_k \right\| \tag{3.16}$$

converges in probability to 0 *as* $n \to \infty$.

The proof of the lemma is provided in the Appendix and follows standard arguments for generalized linear models [20]. The asymptotic normality of the normalized score vector $F_n^{-1/2} S_n$, where $S_n = \mathcal{S}_n(\Theta_0)$, is first established. Let $v \in \mathbb{R}^k$ be an arbitrary vector. It suffices to show that the linear combination $v^\top F_n^{-1/2} S_n$ converges in distribution to the standard normal distribution $\mathcal{N}(0, 1)$. Without loss of generality, assume that $\|v\| = 1$.

The result follows from the Lindeberg–Feller central limit theorem applied to linear combinations, together with the Cramér–Wold device, in accordance with classical results in multivariate asymptotic theory [20, 22].

We now rely on the asymptotic normality of the normalized score vector mentioned above to show that it also extends to the normalized maximum likelihood estimators $\hat{\Theta}_n$.

Let

$$\gamma(t) = S_n\left(\hat{\Theta}_n + t(\Theta_0 - \hat{\Theta}_n)\right),$$

then:

$$S_n(\Theta_0) - S_n(\hat{\Theta}_n) = \gamma(1) - \gamma(0) = \int_0^1 \gamma'(t)\, dt,$$

with:

$$\gamma'(t) = \mathcal{S}_n'\left(\hat{\Theta}_n + t(\Theta_0 - \hat{\Theta}_n)\right) \cdot (\Theta_0 - \hat{\Theta}_n).$$

Assuming that $S_n(\hat{\Theta}_n) = 0$ and that $S_n = S_n(\Theta_0)$, obtain:

$$\begin{aligned} S_n &= \left[\int_0^1 S_n'\left(\hat{\Theta}_n + t(\Theta_0 - \hat{\Theta}_n)\right) dt\right] \cdot (\hat{\Theta}_n - \Theta_0) \\ &= \left[\int_0^1 H_n(\tilde{\Theta}) dt\right] \cdot (\hat{\Theta}_n - \Theta_0) \\ &= \left[F_n^{1/2} \int_0^1 F_n^{-1/2} H_n F_n^{-1/2}(\tilde{\Theta}) dt - I_k + I_k\right] \cdot F_n^{1/2}(\hat{\Theta}_n - \Theta_0) \end{aligned}$$

where $0 \leq t \leq 1$ and $\tilde{\Theta}_n = \Theta_0 + t(\hat{\Theta}_n - \Theta_0)$. By multiplying both sides by $F_n^{1/2}$, we obtain:

$$F_n^{-1/2} S_n = \left(\int_0^1 F_n^{-1/2} H_n(\tilde{\Theta}) F_n^{-1/2} dt - I_k\right) F_n^{1/2}(\hat{\Theta}_n - \Theta_0) + F_n^{1/2}(\hat{\Theta}_n - \Theta_0)$$

Assume that $\hat{\Theta}_n \in K_n(\varepsilon)$, then $\tilde{\Theta}_n$ also belongs to this neighborhood. Moreover,

$$\left\| F_n^{-1/2} H_n(\tilde{\Theta}_n) F_n^{-1/2} - I_k \right\| \leq \sup_{\Theta \in K_n(\varepsilon)} \left\| F_n^{-1/2} H_n(\Theta) F_n^{-1/2} - I_k \right\|.$$

And since

$$\int_0^1 \sup_{\Theta \in K_n(\varepsilon)} \left\| F_n^{-1/2} H_n(\Theta) F_n^{-1/2} - I_k \right\| dt = \sup_{\Theta \in K_n(\varepsilon)} \left\| F_n^{-1/2} H_n(\Theta) \mathcal{F}_n^{-1/2} - I_k \right\|,$$

the lemma implies that

$$\int_0^1 \sup_{\Theta \in K_n(\varepsilon)} \left\| F_n^{-1/2} H_n(\Theta) F_n^{-1/2} - I_k \right\| dt \to 0 \quad \text{in probability, as } n \to \infty.$$

Hence, we thus have:

$$\int_0^1 \left\| F_n^{-1/2} H_n(\tilde{\Theta}_n) F_n^{-1/2} - I_k \right\| dt \to 0 \quad \text{in probability, as} n \ \to \infty.$$

Finally, combining the asymptotic normality of the normalized score vector and Slutsky's theorem, we obtain the desired result :

$$F_n^{1/2}(\hat{\Theta}_n - \Theta_0) \xrightarrow{d} N(0, I_k) \quad \text{as } n \to \infty.$$

□

4. Simulations

In this section, simulations are conducted to evaluate and compare the performance of the standard analysis in the absence of censoring with that of the complete-case analysis under various sample sizes and degrees of censoring. The data are generated according to the Poisson model (2.2) with:

$$\log(\lambda(\boldsymbol{\beta})) = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} \tag{4.1}$$

The covariates $X_{i1}, \ldots, X_{i3}$ are observed, and the covariate X_{i4} is subject to right random censoring. $X_{i1} = 1$, X_{i2} and X_{i3} are independently drawn from the following distributions: $X_{i2} \sim \text{Weibull}(2, 1/4)$, $X_{i3} \sim \exp(3)$. The covariate X_{i4} is such that $X_{i4}|\tilde{\mathbf{X}}_i \sim N(\tilde{\boldsymbol{\beta}}^\top \tilde{\mathbf{X}}_i; \sigma^2)$ were $\tilde{\mathbf{X}}_i = (X_{i1}, \ldots, X_{i3})$, $\tilde{\boldsymbol{\beta}} = (0.01, -0.08, 0.05)$ and $\beta_4 = -0.1$. The censoring thresholds C_i are generated respectively by $C_i \sim U(-1, 11)$ and $C_i \sim U\left(-2, \frac{1}{4}\right)$ to obtain light censoring 25% and heavy censoring 60%, respectively.

Numerical optimization is performed using the `optim` function in `R`.

Two censoring levels are considered, with average rates of 25% and 60%. For each configuration of the simulation design parameters (`sample size` × `censoring proportion`), $N = 500$ datasets are generated, and both the maximum likelihood estimator (MLE) and the complete-case (CC) estimator are computed. Results for $n = 500$ are reported in Table Table 1, whereas those for moderate sample sizes are presented in the Appendix; the conclusions remain qualitatively unchanged across sample sizes.

The no-censoring scenario serves as a benchmark. Under moderate censoring (25%), the MLE exhibits low bias and controlled variability across all sample sizes. Although the CC estimator sometimes shows slightly smaller bias, it does so at the expense of higher variability, leading to larger standard deviations (SD) and mean squared errors (MSE).

When censoring is substantial (60%), the differences become more pronounced. The proposed estimator remains relatively stable, with moderate variability and acceptable empirical SD, mean standard error (SE), and MSE, particularly for large samples, indicating good convergence properties even under heavy censoring. In contrast, the CC estimator deteriorates markedly, with increased bias, substantial variability, and inflated standard errors.

The QQ-plots (Figure 1 and Figure 2) suggest approximate normality of the estimates for $N = 500$; similar patterns are observed in the other scenarios (not shown). Overall, these findings support the theoretical results established in Theorems 3.1 and 3.2.

Table 1. Simulation results for the Poisson model with a censored covariate for $n = 500$. SD: empirical standard deviation. SE: average standard error. MSE: Mean Squared Error

		$\hat{\beta}_n$			
Methods	%Censoring	$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$
	0%				
No censoring	BIAS	0.004391	0.001333	0.002754	0.000224
	SD	0.105277	0.372740	0.132822	0.008372
	SE	0.102203	0.364772	0.124784	0.008433
	MSE	0.015050	0.146155	0.028084	0.008862
	25%				
MLE	BIAS	0.099624	0.007314	0.007482	0.00838
	SD	0.123194	0.440247	0.148643	0.01633
	SE	0.102001	0.369509	0.126107	0.015384
	MSE	0.036887	0.203236	0.034289	0.00784
CC	BIAS	0.011831	0.020548	0.004680	0.000159
	SD	0.144039	0.532467	0.173190	0.015831
	SE	0.137007	0.493460	0.170233	0.016484
	MSE	0.024193	0.286808	0.040925	0.008973
	60%				
MLE	BIAS	0.818961	0.096797	0.091951	1.002784
	SD	0.14185	0.385685	0.13805	0.062472
	SE	0.303713	1.062205	0.351031	0.227461
	MSE	0.63168	0.155405	0.022437	1.15885
CC	BIAS	3.213227	1.021331	0.018086	0.419419
	SD	7.001316	8.754504	10.350940	9.684503
	SE	3593.890358	7536.435075	10631.582807	7832.003817
	MSE	58.958213	77.584639	106.918447	93.733187

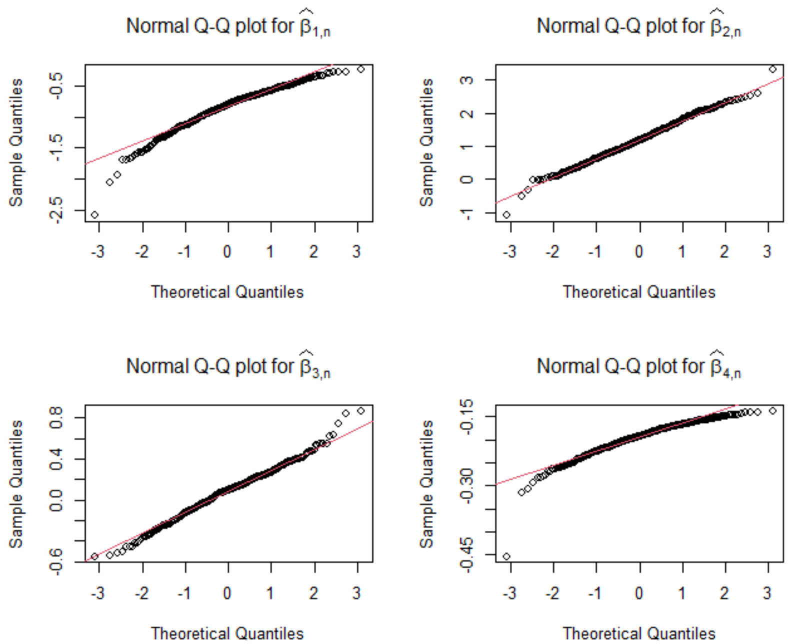

FIGURE 2. QQ-plot normalizes with 60% censoring covariate

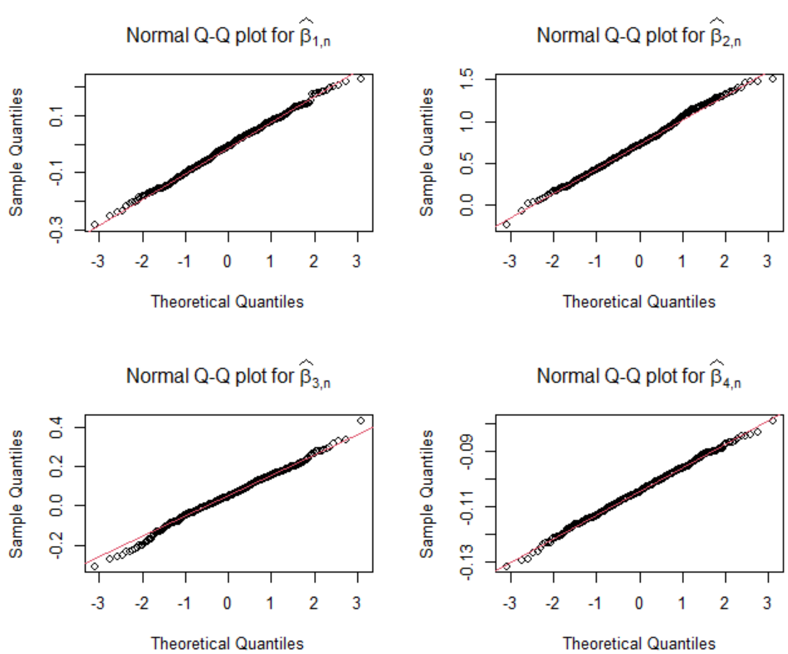

FIGURE 1. QQ-plot normalizes with 25% censoring covariate

5. An application in health economics

To demonstrate the empirical behavior of the proposed estimator, we analyze data from the National Medical Expenditure Survey (NMES) [11]. This dataset is frequently used to illustrate zero-inflated count models, but it also provides a standard benchmark for Poisson regression when the outcome is taken to be the number of physician office visits. We use the NMES 1988 extract distributed with the `R` package `AER` [14] under the name `NMES1988`. The data comprise $n = 4{,}406$ Medicare-covered individuals aged 66 years and older.

The response variable is the number of visits to a physician in an office setting, denoted by **ofp**. Covariates include demographic and socioeconomic characteristics—gender (**female**, coded 1 for female and 0 for male), marital status (**fstatus**, coded 1 if married and 0 otherwise), and years of schooling (**school**)—as well as insurance indicators for Medicaid coverage (**med**, 1 if yes and 0 otherwise) and supplemental private insurance (coded analogously). Health status is summarized by the number of reported chronic conditions (**chronic**) and self-assessed health. The latter is recorded on an ordinal scale (poor/average/excellent) and is represented in the analysis by two dummy variables: **health1** indicates poor health and **health2** indicates excellent health, with average health serving as the reference category. Age enters the model as **age**, scaled by a factor of 10 for numerical stability and interpretability.

In many applied settings, covariates may be incompletely observed because of survey design, data-processing rules, or confidentiality protections. Age is a common example. In self-administered questionnaires, respondents may decline to report an exact age beyond a personal threshold, and in administrative or public-use files, ages above a prespecified cutoff are often top-coded or collapsed into an open-ended category to reduce disclosure risk. These practices yield right-censored covariate measurements: for some individuals the exact age is observed, whereas for others only the information that age exceeds a cutoff is available. When the cutoff differs across individuals—for instance due to mode of administration or heterogeneous disclosure-control rules—the censoring mechanism can be viewed as random.

In the `NMES1988` data, **age** is fully observed. To assess the operating characteristics of our method under covariate censoring, we therefore construct a censored version of **age** by imposing artificial right censoring. This design is motivated by health-economics applications in which exact ages above certain thresholds are masked or coarsened for anonymity. We treat the censoring as non-informative in the sense that, conditional on the observed covariates, the censoring process is independent of the outcome and the unobserved portion of age. Accordingly, censoring thresholds are generated from a Uniform distribution and applied to **age**. We consider two censoring regimes, calibrated to yield average censoring proportions of approximately 24% and 53%, respectively. The fitted model is:

$$\log(\lambda(\beta)) = \beta_1 + \beta_2\,\mathbf{health1}_i + \beta_3\,\mathbf{health2}_i + \beta_4\,\mathbf{chronic}_i + \beta_5\,\mathbf{fstatus}_i + \beta_6\,\mathbf{school}_i \\ + \beta_7\,\mathbf{med}_i + \beta_8\,\mathbf{age}_i$$

Table 2 summarizes the maximum likelihood estimates and associated standard errors from Poisson regressions for the number of office-based physician visits (`ofp`) under three versions of the covariate `age`: fully observed, and artificially right-censored with average censoring proportions of 24% and 53%. Across the three settings, inference based on Wald statistics is broadly stable: most covariates remain statistically distinguishable from zero at conventional levels, indicating that the qualitative conclusions from hypothesis tests are not strongly affected by the imposed censoring.

In contrast, several point estimates and their standard errors exhibit noticeable variation as the censoring rate increases, underscoring that right-censoring of a continuous covariate can materially affect effect magnitudes even when significance is preserved. This pattern is evident for `age` itself and is also reflected in covariates such as `chronic`, `fstatus`, `school`, and `med`, whose estimated coefficients shift upward or downward with heavier censoring.

Notwithstanding these changes, the substantive interpretations align with prior empirical evidence. The coefficient on the number of chronic conditions (`chronic`) is positive, consistent with more frequent utilization among individuals reporting a higher chronic disease burden. The indicator for poor self-rated health (`health1`) is estimated to have an effect that differs in sign from `chronic`, suggesting that self-assessed health may capture dimensions of health status and care-seeking behavior not fully summarized by the chronic-condition count. Years of schooling (`school`) is positively associated with physician visits, potentially reflecting differences in health literacy or engagement with preventive care. Medicaid coverage (`med`) and marital status (`fstatus`) are also positively related to utilization, in line with earlier analyses of similar populations.

Overall, the results indicate that increasing censoring may have limited impact on the direction of effects and on test-based conclusions, but can meaningfully alter estimated effect sizes and uncertainty quantification. This reinforces the importance of explicitly accounting for covariate censoring when the goal is interpretation or comparison of covariate importance, particularly for continuous predictors such as `age`.

Variable	No censoring				Censoring = 24%				Censoring = 53%			
	Estimate	S.E.	Signif.	P-value	Estimate	S.E.	Signif.	P-value	Estimate	S.E.	Signif.	P-value
`Intercept`	0.2878	0.0175	***	$< 2e^{-16}$	0.3221	0.0173	***	$< 2e^{-16}$	0.3613	0.0260	***	$< 2e^{-16}$
`health1`	-0.3647	0.0303	***	$< 2e^{-16}$	-0.2980	0.0289	***	$< 2e^{-16}$	-0.3622	0.0444	***	$3.64e^{-16}$
`health2`	0.1715	0.0045	***	$< 2e^{-16}$	0.2021	0.0043	***	$< 2e^{-16}$	0.1667	0.0067	***	$< 2e^{-16}$
`chronic`	0.1241	0.0140	***	$< 2e^{-16}$	0.2659	0.0133	***	$< 2e^{-16}$	0.1686	0.0206	***	$2.38e^{-16}$
`fstatus`	0.0575	0.0136	***	$2.32e^{-05}$	0.1674	0.0130	***	$< 2e^{-16}$	0.0548	0.0197	**	0.00542
`school`	0.0435	0.0018	***	$< 2e^{-16}$	0.0760	0.0014	***	$< 2e^{-16}$	0.0543	0.0026	***	$< 2e^{-16}$
`med`	0.1440	0.0218	***	$3.97e^{-11}$	0.2876	0.0210	***	$< 2e^{-16}$	0.2146	0.0323	***	$3.20e^{-11}$
`age`	0.1139	0.0032	***	$< 2e^{-16}$	0.0392	0.0024	***	$< 2e^{-16}$	0.0927	0.0046	***	$< 2e^{-16}$

TABLE 2. Analysis of health care data. *Estimates ,S.E(standard errors) and significance codes:* `***` significant at the 0.1% level, `**` significant at the 1% level, `*` significant at the 5% level, `.` significant at the 10% level.

6. Conclusion

A likelihood-based estimation framework is developed for Poisson regression when a covariate is subject to random right-censoring. The central idea consists in retaining all observations by integrating the Poisson likelihood over the unobserved tail of the censored covariate, using a working model for its conditional distribution given the fully observed covariates. Under standard regularity conditions, the resulting maximum likelihood estimator is consistent and asymptotically normal, providing a principled basis for inference in the presence of partially observed covariate information. Simulation experiments indicate that the estimator substantially reduces the bias arising when censoring is ignored and yields stable finite-sample performance across a range of censoring levels, including moderate sample sizes. An empirical analysis based on healthcare data further illustrates the practical value of the approach: accounting for covariate censoring can materially improve model fit and sharpen interpretation of covariate effects.

Several extensions merit future investigation. The same likelihood construction may be adapted to other count-data models, including negative binomial and zero-inflated specifications. Relaxing the assumption of noninformative censoring—for example, by allowing dependence between the censoring mechanism and the outcome—would broaden applicability. Bayesian implementations could provide a coherent framework for propagating uncertainty regarding the censored covariate distribution and other nuisance parameters. Finally, because inference may be sensitive to misspecification of the conditional distribution assumed for the censored covariate, semiparametric or robust alternatives that reduce reliance on parametric assumptions constitute an important direction for further research.

References

1. Amani, K. M., Kouakou, K. J. G., and Hili, O. (2025). Marginalized zero-inflated Bell regression models for overdispersed count data. *Journal of Statistical Theory and Practice*, 19(2):22. Id/No 17. https://doi.org/10.1007/s42519-025-00433-7.
2. Ameraoui, A., Boukhetala, K., and Dupuy, J.-F. (2016). Bayesian estimation of the tail index of a heavy-tailed distribution under random censoring. *Computational Statistics & Data Analysis*, 104:148–168. https://doi.org/10.1016/j.csda.2016.06.009.
3. Atem, F., Sampene, E., and Greene, T. J. (2019). Improved conditional imputation for linear regression with a randomly censored predictor. *Statistical Methods in Medical Research*, 28(2):432–444. https://doi.org/10.1177/0962280217727033.
4. Atem, F. D., Qian, J., Maye, J. E., Johnson, K. A., and Betensky, R. A. (2017). Linear Regression with a Randomly Censored Covariate: Application to an Alzheimer's Study. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(2):313–328. https://doi.org/10.1111/rssc.12164.
5. Atem, F. D., Qian, J., Maye, J. E., Johnson, K. A., and Betensky, R. A. (2016). Multiple imputation of a randomly censored covariate improves logistic regression analysis. *Journal of Applied Statistics*, 43(15):2886–2896. https://doi.org/10.1080/02664763.2016.1155110.
6. Austin, P. C. and Brunner, L. J. (2003). Type I Error Inflation in the Presence of a Ceiling Effect. *The American Statistician*, 57(2):97–104. https://doi.org/10.1198/0003130031450.
7. Buuren Van, S. (2012). *Flexible Imputation of Missing Data.* https://doi.org/10.1201/b11826.

8. Christopher, Chu, H., Schneider, M. F., and Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, 26(23):4352–4374. https://doi.org/10.1002/sim.2836.
9. Dabana, H., Agbokou, K., and Gneyou, K. (2025). Local linear estimation of conditional probability density and mode under right censoring and left truncation: dependent data case. *Gulf Journal of Mathematics*, 20(1):338–359. https://doi.org/10.11648/j.gjom.20252001.28.
10. Das, D. and Das, P. (2025). On second order correctness of bootstrap in logistic regression. *Bernoulli*, 31(4). https://doi.org/10.3150/24-BEJ1827.
11. Deb, P. and Trivedi, P. K. (1997). Demand for Medical Care by the Elderly: A Finite Mixture Approach. *Journal of Applied Econometrics*, 12(3):313–336. https://doi.org/10.1002/(SICI)1099-1255(199705)12:3<313::AID-JAE440>3.0.CO;2-G.
12. Fahrmeir, L. and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *The Annals of Statistics*, 13(1):342–368.
13. Folefac D. Atem, A. (2017). Linear Regression with a Randomly Censored Covariate: Application to an Alzheimer's Study.
14. Kleiber, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer, New York. https://doi.org/10.1007/978-0-387-77318-6.
15. Kouakou, K. J. G., Amani, K. M., Bamba, L., Soro, B., and Hili, O. (2025). Statistical inference for the marginalized zero-inflated Bell regression model with missing covariates. *Chilean Journal of Statistics*, 16(1):36–59. https://doi.org/10.32372/ChJS.16-01-03.
16. Kouakou, K. J. G., Hili, O., and Dupuy, J. F. (2024). Estimation of zero-inflated bivariate Poisson regression with missing covariates. *Communications in Statistics - Theory and Methods*, 53(20):7216–7243. https://doi.org/10.1080/03610926.2024.XXXXX.
17. Lotspeich, S. C., Ashner, M. C., Vazquez, J. E., Richardson, B. D., Grosser, K. F., Bodek, B. E., and Garcia, T. P. (2024). Making Sense of Censored Covariates: Statistical Methods for Studies of Huntington's Disease. *Annual Review of Statistics and Its Application*, 11(1):255–277. https://doi.org/10.1146/annurev-statistics-040522-095944.
18. Maller, R. (2003). Asymptotics of regressions with stationary and nonstationary residuals. *Stochastic Processes and their Applications*, 105(1):33–67. https://doi.org/10.1016/S0304-4149(02)00263-6.
19. Mezhoud, M., Leulmi, R., and Leulmi, S. (2025). Asymptotic normality of adaptive recursive kernel estimator of the density in a right censored model. *Gulf Journal of Mathematics*, 13(1):77–91. https://doi.org/10.11648/j.gjom.20251301.19.
20. Nguyen, V. T. and Dupuy, J. F. (2021). Asymptotic results in censored zero-inflated Poisson regression. *Communications in Statistics - Theory and Methods*, 50(12):2759–2779. https://doi.org/10.1080/03610926.2019.1676442.
21. Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. https://doi.org/10.1007/978-1-4757-2545-2.
22. Xiao, Z. and Liu, L. (2006). Consistency and asymptotic normality of MLE of parameter vector in a randomly censored GLM with incomplete information. *Wuhan University Journal of Natural Sciences*, 11(2):333–338. https://doi.org/10.1007/BF02832116.

[1] Labaratoire de Statistique, de Probabilt´es et de Recherche Operationnelle, Institut Nationnal Polythectnique Felix Houphou¨et Boigny de Yamoussoukro
Email address: laspro@gmail.com