International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 5
ISSN: 2088-5334

The Modified Structural Quasi Score Estimator for Poisson Regression Parameters with Covariate Measurement Error

Fevi Novkaniza [a,*], Khairil Anwar Notodiputro [b], I Wayan Mangku [c], Kusman Sadik [b]

[a] *Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Indonesia, Depok, West Java, 16424, Indonesia*
[b] *Department of Statistics, IPB University, Bogor, West Java, 16680, Indonesia*
[c] *Department of Mathematics, IPB University, Bogor, West Java, 16680, Indonesia*
*Corresponding author: *fevi.novkaniza@sci.ui.ac.id*

***Abstract*—This article proposed the Modified Structural Quasi Score (MSQS) estimators for Poisson regression parameters when a covariate is subject to measurement error. We study the situation when the true covariate in the Poisson regression model is unobserved, and the surrogate for this covariate is related to the true covariate by the additive measurement error model. We assumed that true covariate as a random variable with unknown density function distribution and its observable values as surrogates, which also has Poisson distribution. We applied the Empirical Bayes Deconvolution (EBD) method for estimating the true covariate density with a finite discrete support set. To estimate Poisson regression parameters, we construct an MSQS estimating equation based on proper functions for the mean and variance of the Poisson distributed surrogate. The MSQS estimator for the Poisson regression parameter is the root of the quasi-score function based on the quasi-likelihood method. We did some simulation scenarios for assessing the MSQS estimator by assuming the true covariate comes from Gamma distribution as a conjugate before Poison distribution. We compute the standard error of the mean, standard deviation, and bias of the MSQS estimator for various sample sizes to examine the estimator's appropriateness. The simulation showed that a combination of the finite discrete support set of surrogates based on the range values and smaller-scale parameter of Gamma distribution yields smaller values of bias estimator and the estimated standard deviation.**

***Keywords*— Covariate; measurement error; Poisson; quasi score; surrogate.**

Manuscript received 6 Jun. 2020; revised 28 Mar. 2021; accepted 16 Apr. 2021. Date of publication 31 Oct. 2022.

I. INTRODUCTION

The Poisson regression model is one of the generalized linear models used for analyzing count data where the response variable is a non-negative integer [1]-[5]. This model is relevant for the analysis of count data in social and natural sciences, such as infometric [6], transportation [7], insurance [8], predictors of length of stay among HIV patients [9] and other application. In the Poisson regression model, the response variable Y has a Poisson distribution with a rate parameter λ that depends on a covariate X: $\log \lambda = {}_0 + {}_1 X$ and the regression parameters ${}_0, {}_1)$ are needed to estimate. In this article, we restrict to the case of one unobserved covariate X measured by error, and we called it surrogate W. If covariate X is measured without error and Poisson regression parameters are produced by maximum likelihood (ML) estimation, it will be consistent and asymptotically efficient. However, if a covariate is measured with error, the ML estimator ignoring the covariate measurement error will lead to inconsistent estimators and bias. In this article, we consider the case event count as the response variable, and we work with unobserved covariate X measured with error. Because X it is an unobserved covariate, we use observable covariate W as a surrogate, and it modeled via additive measurement error model, $W = X + e$, where e is measurement error and independent of (Y, X) and identically distributed samples from a known density. Many methods have been proposed and mostly depend on the distribution of e is known [10]–[13]. However, the methods do not utilize the distribution of covariate X measured with error (functional measurement error) [14]-[18].

In contrast to the functional measurement error model, structural measurement error assumed X is a random variable and has distribution. If this distribution can be specified parametrically, it is possible to calculate them directly. We can get the conditional distribution of X given W as the posterior distribution if we get the prior distribution of X, denoted by $g(x)$. We used the Empirical Bayes

Deconvolution (EBD) method proposed by Efron [9] for estimating the density $g(x)$ of X using surrogates W. In EBD methods, $g(x)$ is an unknown prior density that has realizations $X_1, X_2, \dots, X_n$ is unobservable. Each X_i produces a surrogate W_i according to a known probability density such as Poisson distribution, $W_i \sim Poisson(X_i)$. We use deconvolution for estimating $g(x)$ from the surrogate $W_1, W_2, \dots, W_n$. Many literature studies on the deconvolution problem, as in Efron [9] which focus on the additive measurement error model. The Bayes deconvolution problem proposed by Efron [9] use likelihood approach with the prior density $g(x)$ belongs to an exponential family of densities on the space-X, denoted by T. The support set T is a finite discrete set, $T = (x_1, x_2, \dots, x_m)$, so the prior $g(x)$ is a n-vector $\boldsymbol{g} = (g_1, g_2, \dots, g_n)$ which specifying probability g_j on $x_j, j = 1, \dots, n$.

If the measurement error was ignored, it could lead to inconsistent estimators of the regression coefficients, and we need another method for estimating the regression coefficient. Many statistical methods are proposed for estimating a parameter of a regression model with covariate measurement error such as [19]–[29], especially for a Poisson regression model with covariate is subject to measurement error [19], [23], [24]. The quasi-likelihood method for Poisson regression models with covariate contains measurement error required posterior distribution of the unobserved covariate X given the observed covariate W. Structural Quasi Score (SQS) estimators for Poisson regression parameter was one of the estimators in Poisson regression where covariate contains measurement errors, but in [19] and [23], X and e have a normal distribution.

In this article, we also use the assumption of the non-differential measurement error, which means that the conditional distribution of Y given X is independent of W: $f_{Y|X,W} = f_{Y|X}$. This paper proposes a quasi-likelihood method for modeling count data by the Poisson regression model when one covariate is measured with error, and we called it a modified SQS estimator. We modified the quasi-score estimator proposed by Efron [14] for estimating the regression coefficient of the Poisson regression model. First, we estimated the density of unobserved covariate as the prior density of Poisson distributed surrogate by the EBD method. After getting the probability estimation for every discretization in support set of an unobserved covariate, we constructed the quasi-score function based on the mean and variance function of the Poisson surrogate.

This paper is organized as follows. In Section 2, we derive the quasi-likelihood approach for the Poisson regression model proposed by Efron [14] with measurement error in one covariate X. We compute mean and variance function from the observed sample W_i when the unobserved covariate X_i measured with errors by finding the roots of the quasi score estimating equation. Section 3 proposed a modified SQS estimator for the Poisson regression parameter by constructing a new quasi score function with mean and variance functions from the prior density estimate of X given W by the EBD method. Some simulation studies are carried out to assess the performance of the proposed estimator. The conclusion is given in Section 4.

II. Materials and Method

A. Poisson Regression with Covariate Measurement Error

In a Poisson regression model, the response Y has Poisson distribution with a parameter $\lambda = \lambda(X, \beta) = exp(\beta_0 + \beta_1 X)$ that depends on one covariate X measured with error as described in [10] and [14]. Let a sample observation of i.i.d pairs $(X_i, Y_i), i = 1, \dots, n$, the parameters β_0 and β_1 are estimated by the maximum likelihood method with the log-likelihood function of β:

$$l(\beta) = \sum_{i=1}^{n} [Y_i \ln \lambda(X_i, \beta) - \lambda(X_i, \beta)] \tag{1}$$

and derive to the estimating equation.

$$\sum_{i=1}^{n} \left[Y_i - exp\left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)\right](1, X_i)^t = 0$$

If covariate X_i is measured with error, we use a surrogate W_i which modeled in additive measurement model such that

$$W_i = X_i + e_i \tag{2}$$

where e_i is the measurement error, independent of (X_i, Y_i), and we assume normally distributed $e_i \sim N(0, \sigma_e^2)$ with known variance or can be estimated from independent replications of W_i. As in the measurement error model, we assume the structural measurement error model, so we treat covariate X as a random variable and distribute probability. Kukush *et al.* [23] assumes that $X_i \sim N(\mu_x, \sigma_x^2)$, and by conditioning on W, the SQS estimator is constructed which is consistent and it has derived from the quasi-likelihood method. The quasi-likelihood approach requires the specifications of the mean and variance function for the Poisson regression model. According to Thamerus [19], the first step to obtain a quasi-likelihood model in the observable variable is to set up the unobservable mean and variance function as the distribution of Y_i given X_i and define two conditional moments as the function:

$$E(Y_i|X_i) = \mu(X_i, \beta) = e^{\beta_0 + \beta_1 X_i} \tag{3}$$

$$Var(Y_i|X_i) = \sigma^2(X_i, \beta) = e^{\beta_0 + \beta_1 X_i} \tag{4}$$

where β is the vector of the Poisson regression parameters. Mean and variance function for the surrogate

$$E(Y_i|W_i) = m(W_i, \beta)$$

$$Var(Y_i|W_i) = V(W_i, \beta)$$

then a quasi-likelihood model in the surrogate defines as

$$m(W_i, \beta) = E(E(Y_i|X_i, W_i)|W_i) = E(\mu(X_i, \beta)|W_i) = E\left(e^{\beta_0 + \beta_1 X_i}|W_i\right) = e^{\beta_0}.E\left(e^{\beta_1 X_i}|W_i\right)$$

$$\begin{aligned} V(W_i, \beta) &= Var(E(Y_i|X_i, W_i)|W_i) + E(Var(Y_i|X_i, W_i)|W_i) \\ &= Var(\mu(X_i, \beta)|W_i) + E(\sigma^2(X_i, \beta)|W_i) \\ &= Var\left(e^{\beta_0 + \beta_1 X_i}|W_i\right) + E\left(e^{\beta_0 + \beta_1 X_i}|W_i\right) \end{aligned} \tag{5}$$

We know that

$$\begin{aligned} Var\left(e^{\beta_0 + \beta_1 X_i}|W_i\right) &= E\left[\left(e^{\beta_0 + \beta_1 X_i}\right)^2|W_i\right] - \left[E\left(e^{\beta_0 + \beta_1 X_i}|W_i\right)\right]^2 \\ &= E\left[e^{2\beta_0 + 2\beta_1 X_i}|W_i\right] - \left[e^{\beta_0}.E\left(e^{\beta_1 X_i}|W_i\right)\right]^2 \\ &= e^{2\beta_0}.E\left[e^{2\beta_1 X_i}|W_i\right] - \left[e^{\beta_0}.E\left(e^{\beta_1 X_i}|W_i\right)\right]^2 \end{aligned}$$

then

$$V(W_i, \beta) = e^{2\beta_0}.E\left(e^{2\beta_1 X_i}|W_i\right) - \left[e^{\beta_0}.E\left(e^{\beta_1 X_i}|W_i\right)\right]^2 + e^{\beta_0}.E\left(e^{\beta_1 X_i}|W_i\right) \tag{6}$$

From equation (6), we have to get the conditional distribution of X_i given W_i for compute expectations of the form $E(e^{kX_i}|W_i)$ and derive expressions or $m(W_i,\beta)$, and $V(W_i,\beta)$. Because we use a structural additive measurement model in equation (2), we will compute the conditional distribution of X_i given W_i as posterior distribution when the distribution of W_i is known using the empirical Bayes deconvolution method proposed by Efron [9]. Kukush *et al.* [14] give an unbiased estimating equation as the quasi score function:

$$s^n(\beta) = \sum_{i=1}^n s_i(\beta) = \sum_{i=1}^n \frac{Y_i - m(W_i,\beta)}{V(W_i,\beta)} . \frac{\partial m(W_i,\beta)}{\partial\beta} \quad (7)$$

The SQS estimator for the Poisson regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is defined as the root of the quasi score function $s(\beta) = 0$. For finding the root, we differentiate the mean function $m(W_i,\beta)$ with respect to β_0 and β_1 estimation is carried out by solving $s(\beta) = 0$ as a nonlinear equation system in β_0 and β_1 numerically.

B. Empirical Bayes Deconvolution Method

Let an unknown prior density $g(x)$ has an observed independent random sample of realizations $X_1, \ldots X_n$:

$$X_1, X_2, \ldots, X_n \sim g(x) \quad (8)$$

Each X_i independently produces an observed random variable W_i as a surrogate with known probability densities for W_i given X_i:

$$W_i|X_i \sim p(W_i|X_i) = Poisson(X_i), i = 1, \ldots, n \quad (9)$$

The EBD method can be used for estimating the prior density $g(x)$ using sample observation $W_1, \ldots W_n$. The EBD method is an estimation procedure $g(x)$ based on sample observations from $f(w)$ by using a likelihood approach to EBD problems with prior $g(x)$, which is modeled through exponential family density in space-X, denote by T. Support set T is assumed to be a finite discrete support set, and T is discretized as many points as W_i, and we consider it as a surrogate of unobserved X_i:

$$X \in T = \{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$$

The prior distribution $g(x)$ is an n-vector $\boldsymbol{g} = (g_1, \ldots, g_n)$ which specifies the probability g_j on x_j and modeled as an exponential family of densities on T,

$$g = g(\alpha) = exp\{Q\alpha - \phi(\alpha)\}\phi(\alpha) = log \sum_{j=1}^n exp\{Q_j^T\alpha\} \quad (10)$$

where $\alpha =$ p-dimensional vector and $Q =$ known $n \times p$ structure matrix. The-j component of $g(\alpha)$:

$$g_i(\alpha) = \Pr\{X = x_{(i)}\} = exp\{Q_j^T\alpha - \phi(\alpha)\}, j = 1, \ldots, n \quad (11)$$

Define $p_{ij} = p_i(W_i|X_i = x_{(j)})$ and denote P_i as n-vector $P_i = (p_{i1}, \ldots, p_{in})^T$, then the marginal probability for W_i:

$$f_i(\alpha) = \sum_{j=1}^n p_{ij} g_j(\alpha) = P_i^T g(\alpha)$$

For the maximum likelihood estimation, the loglikelihood function for the vector parameter $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ is:

$$l_i(\alpha) = log f_i(\alpha) = log P_i^T g(\alpha)$$

with p-dimensional first derivative vector and $p \times p$ -dimensional second derivative matrix

$$\dot{l}_i(\alpha) = \left(\ldots, \frac{\partial l_i(\alpha)}{\partial\alpha_h}, \ldots\right)^T, \quad \ddot{l}_i(\alpha) = \left(\ldots, \frac{\partial^2 l_i(\alpha)}{\partial\alpha_h \partial\alpha_k}, \ldots\right)$$

For W_i with n observation, the total loglikelihood $l(\alpha) = \sum_{i=1}^n l_i(\alpha)$ has first and the second derivative is:

$$\dot{l}(\alpha) = \sum_{i=1}^n \dot{l}_i(\alpha) = Q^T \sum_{i=1}^n B_i(\alpha) = Q^T B_+ \alpha$$

where

$$B_i(\alpha) = \{b_{i1}(\alpha), \ldots, b_{in}(\alpha)\}^T$$
$$b_{ij}(\alpha) = g_j(\alpha)\left\{\frac{p_{ij}}{f_i(\alpha)} - 1\right\}$$

and

$$-\ddot{l}_i(\alpha) = Q^T[B_i(\alpha)B_i(\alpha)^T + B_i(\alpha)g(\alpha)^T + g(\alpha)B_i(\alpha)^T - dig\{B_i(\alpha)\}]Q$$

According to Efron [9], maximum likelihood estimation $\hat{\alpha}$ for α satisfies:

$$Q^T B_+(\hat{\alpha}) = 0 \quad (12)$$

where $B_+(\hat{\alpha}) = \sum_{i=1}^n B_i(\hat{\alpha})$, so we get the prior distribution estimation of X_i, $\hat{g}_j(x_{(j)}), j = 1, \ldots, n$ for every discretization point in finite discrete support set, T. Based on the definition of marginal density, the marginal density of W_i is:

$$f_{W_i}(w_i) = \sum_{x_i} p(W_i|x_i) g(x_i) \quad (13)$$

C. The Modified Structural Quasi Score Estimator for Poisson Regression Parameters

The mean and variance functions of the surrogate model based on the quasi-likelihood approach in equation (5) and (6):

$$m(W_i,\beta) = e^{\beta_0}.E(e^{\beta_1 X_i}|W_i)$$
$$V(W_i,\beta) = e^{2\beta_0}.E(e^{2\beta_1 X_i}|W_i) - [e^{\beta_0}.E(e^{\beta_1 X_i}|W_i)]^2 + e^{\beta_0}.E(e^{\beta_1 X_i}|W_i)$$

have a term of the conditional distribution of X_i given W_i. We assume the conditional distribution of W_i given X_i, $f_{W_i|X_i}$ is the Poisson distributed with rate X_i and the prior density estimation $\hat{g}_i(x_i)$ for every discretization point in support set T are estimated by the EBD method. From equation (10), we get the marginal density $f_{W_i}(w_i)$, and we compute the joint probability distribution of X_i and W_i :

$$f_{X_i,W_i}(x_i,w_i) = f_{W_i|X_i}(w_i).\hat{g}_i(x_i) \quad (14)$$

The conditional probability of X_i given W_i is given as follows:

$$f_{X_i|W_i}(x_i) = \frac{f_{X_i,W_i}(x_i,w_i)}{f_{W_i}(w_i)} \quad (15)$$

Using equation (15), we can compute the mean function and variance function in equation (5) and (6) as a function of the Poisson regression parameter for $X_{(i)}, i = 1, \ldots, n$:

$$m(W_i,\beta) = e^{\beta_0}.\sum_{X_{(1)}}^{X_{(n)}} e^{(\beta_1 X_{(i)})} Pr(X_{(i)}|W_i)$$
$$m(W_i,\beta) = e^{\beta_0}[e^{(\beta_1 X_{(1)})} Pr(X_{(1)}|W_1) + e^{(\beta_1 X_{(2)})} Pr(X_{(2)}|W_2) + \cdots + e^{(\beta_1 X_{(n)})} Pr(X_{(n)}|W_n)]$$

Let $k_i = Pr(X_{(i)}|W_i), i = 1, \ldots, n$, then:

$$m(W_i,\beta) = k_1 e^{\beta_0 + \beta_1 X_{(1)}} + k_2 e^{\beta_0 + \beta_1 X_{(2)}} + \ldots + k_n e^{\beta_0 + \beta_1 X_{(n)}}$$

Variance function in equation (6) as follows:

$$V(W_i,\beta) = e^{2\beta_0}.E[e^{(2\beta_1 X_i^*)}|W_i] - [e^{\beta_0}.E[e^{(\beta_1 X_i^*)}|W_i]]^2 + e^{\beta_0}.E[e^{(\beta_1 X_i^*)}|W_i]$$

First, we computed $E\left[e^{(2\beta_1 X_{(i)})}|W_i\right]$ as follows:

$$E[e^{(2\beta_1 X_{(i)})}|W_i] = \sum_{X_{(1)}}^{X_{(n)}} e^{(2\beta_1 X_{(i)})} Pr(X_{(i)}|W_i)$$

$$= \left[e^{(2\beta_1 X_{(1)})} Pr\left(X_{(1)} \mid W_1\right) + e^{(2\beta_1 X_{(2)})} Pr(X_2^* \mid W_2) + \cdots + e^{(2\beta_1 X_{(n)})} Pr\left(X_{(n)} \mid W_n\right)\right]$$

then

$$V(W_i,\beta) = \left[k_1 e^{2\beta_0+2\beta_1 X_{(1)}} + k_2 e^{2\beta_0+2\beta_1 X_{(2)}} + \ldots + k_n e^{2\beta_0+2\beta_1 X_{(n)}}\right] - \left[k_1 e^{\beta_0+\beta_1 X_{(1)}} + k_2 e^{\beta_0+\beta_1 X_{(2)}} + \ldots + k_n e^{\beta_0+\beta_1 X_{(n)}}\right]^2 + \left[k_1 e^{\beta_0+\beta_1 X_{(1)}} + k_2 e^{\beta_0+\beta_1 X_{(2)}} + \ldots + k_n e^{\beta_0+\beta_1 X_{(n)}}\right]$$

For constructing the quasi-score function, we compute the partial derivative of mean and variance function with respect to $\boldsymbol{\beta}$. The estimated regression coefficient of the Poisson regression model is carried out by solving the quasi-score equation $s(\boldsymbol{\beta}) = \mathbf{0}$ as a nonlinear system in $\boldsymbol{\beta}$. We called it the modified SQS estimator for the parameter $\boldsymbol{\beta}$ as follows:

$$s^{(n)}(\beta_0) = \sum_{i=1}^{n} \frac{\partial m(W_i,\boldsymbol{\beta})}{\partial \beta_0} \frac{Y_i - m(W_i,\boldsymbol{\beta})}{V(W_i,\beta)}$$

$$s^{(n)}(\beta_1) = \sum_{i=1}^{n} \frac{\partial m(W_i,\boldsymbol{\beta})}{\partial \beta_1} \frac{Y_i - m(W_i,\boldsymbol{\beta})}{V(W_i,\beta)} = 0$$

III. Results and Discussion

We did some simulation scenarios to explain the estimation methods. The following steps were conducted:

Step 1: Generate covariate measured with error X_i and surrogate data $W_i, i = 1,..,n.$ We choose $n = 50, 100, 200, 300.$ The vector of covariate $X_i \sim Gamma(2,1)$, $X_i \sim Gamma(1,2)$ and surrogate data $W_i \sim Poisson(X_i)$

Step 2: Using surrogate data $W_1, \ldots W_n$, support set T for X_i is chosen based on the value of surrogate data W_i: $\mathrm{T} = \{x_{(1)}, x_{(2)}, .., x_{(n)}\}$. We choose 2 types of finite discrete support set $\mathrm{T} = [min(W_i), max(W_i)]$ and $\mathrm{T} = [Q_1(W_i), Q_3(W_i)]$ compute prior density estimation, $\hat{g}(x_i) = Pr\left(X_i = x_{(j)}\right), j = 1, .., n$ by EBD method using R package deconvolveR

Step 3: By setting the vector of the Poisson regression model parameter $\beta = (1, -1)$, we generate the response data Y_i from a Poisson distribution with parameter $\lambda_i = e^{1-x_{(i)}}, i = 1, .., 100$

Step 4: Compute the marginal density of W_i as in equation (13) and $f_{X_i,W_i}(x_i, w_i)$ in equation (14), then we compute conditional expectation $E\left(e^{\beta_1 X_i} \mid W_i\right)$ and $E\left(e^{2\beta_1 X_i} \mid W_i\right)$ based on the conditional distribution $f_{X_i|W_i}(x_i)$

Step 5: We compute mean and variance function as in equation (5) and (6) and construct quasi score function $s^{100}(\beta)$

Step 6: We find the root of $s^{100}(\beta) = 0$, as the modified SQS estimator for the Poisson Regression model parameter: $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1\right)$

Step 7: Step 1 to 5 are repeated 1000 times and compute the mean, standard error of the mean, standard deviation, and bias of modified structural quasi score estimator $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1\right)$

All the steps are described in a flowchart, as shown in Fig. 1. The simulation is done by R package [30] and result of simulation studies are shown in Fig 2-3 and Table 1.

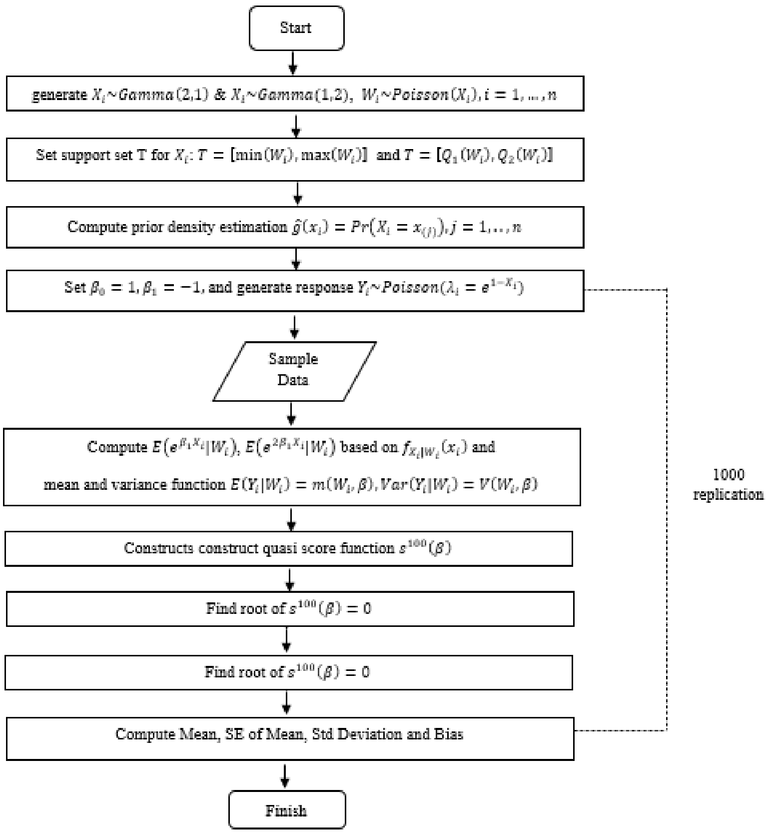


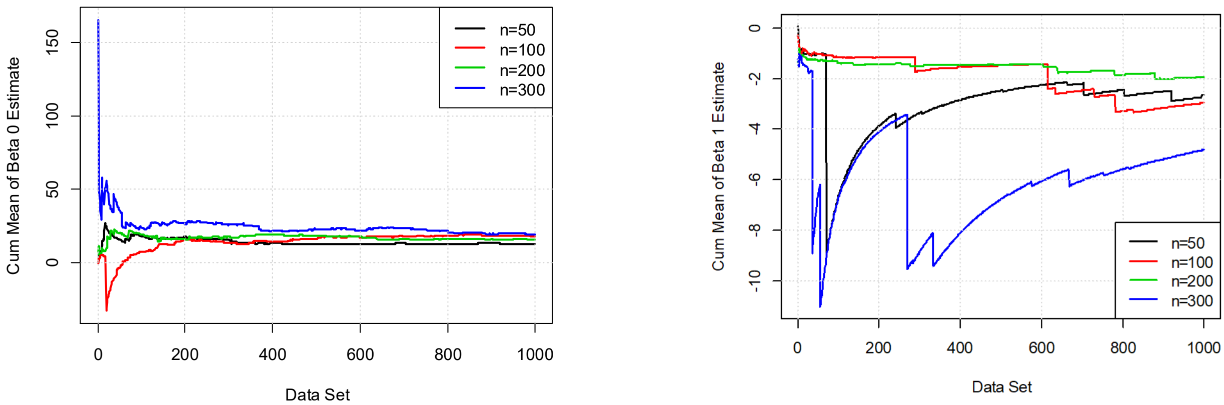
Fig. 1 Flowchart of simulation method

TABLE I

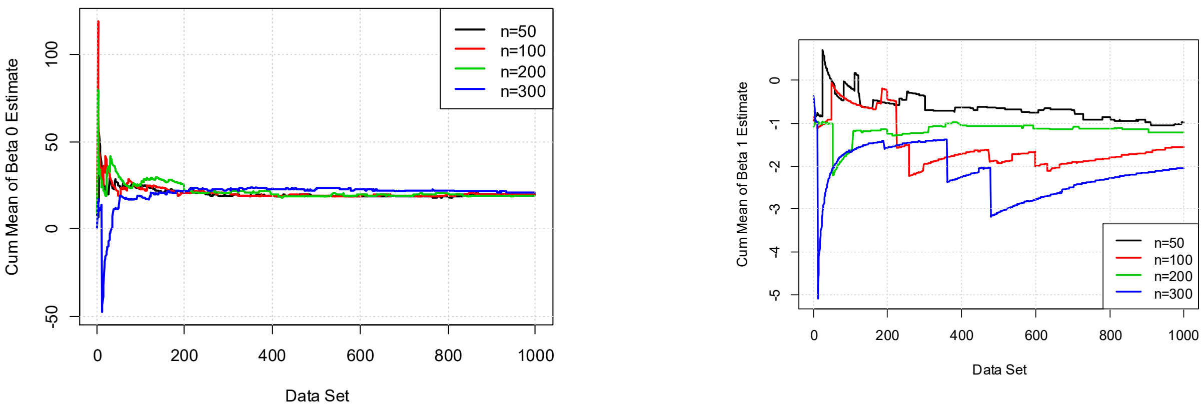
ESTIMATION RESULTS FOR POISSON REGRESSION PARAMETER BY MODIFIED STRUCTURAL QUASI SCORE ESTIMATOR $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$

Prior Distribution	T	n	$\hat{\beta}_0$				$\hat{\beta}_1$			
			Mean	SE Mean	StDev	Bias	Mean	SE Mean	StDev	Bias
Gamma (2,1)	$T=[Q_1(W_i),Q_2(W_i)]$	50	12.98	1.31	41.41	12.88	-2.65	0.79	25.16	-2.55
	$T=[\min(W_i),\max(W_i)]$	50	19.20	1.40	44.37	19.10	-0.99	0.24	7.64	-0.89
	$T=[Q_1(W_i),Q_2(W_i)]$	100	18.11	1.69	53.53	18.01	-2.97	0.88	27.95	-2.87
	$T=[\min(W_i),\max(W_i)]$	100	19.47	1.42	44.76	19.37	-1.55	0.43	13.59	-1.45
	$T=[Q_1(W_i),Q_2(W_i)]$	200	16.04	1.93	61.05	15.94	-1.95	0.27	8.60	-1.85
	$T=[\min(W_i),\max(W_i)]$	200	19.16	2.13	67.34	19.06	-1.21	0.15	4.63	-1.11
	$T=[Q_1(W_i),Q_2(W_i)]$	300	19.53	2.29	72.40	19.43	-4.82	1.82	57.51	-4.72
	$T=[\min(W_i),\max(W_i)]$	300	20.81	1.77	55.99	20.71	-2.04	0.66	20.87	-1.94

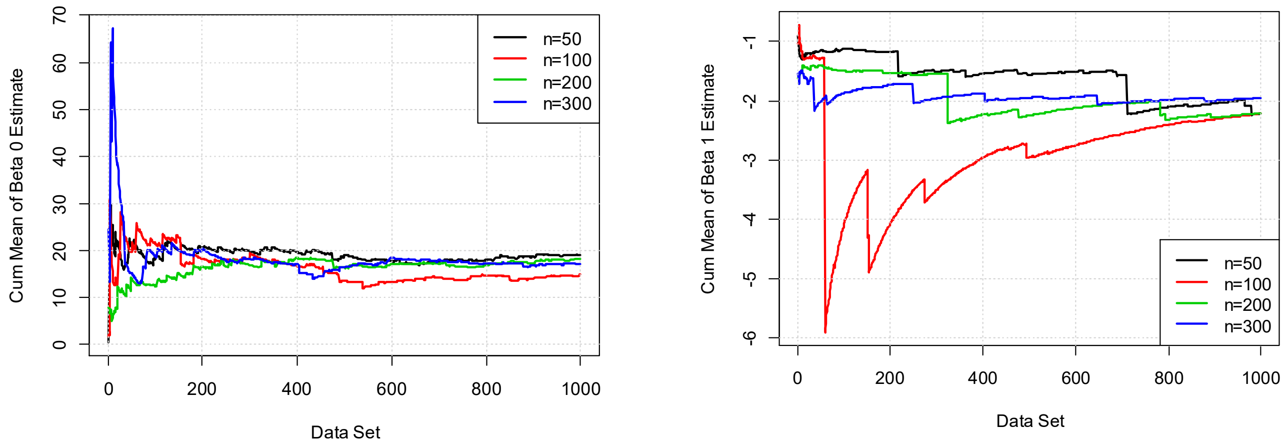
Prior Distribution	T	n	$\hat{\beta}_0$				$\hat{\beta}_1$			
			Mean	SE Mean	StDev	Bias	Mean	SE Mean	StDev	Bias
Gamma (1,2)	$T=[Q_1(W_i),Q_2(W_i)]$	50	19.02	1.72	54.37	18.92	-2.22	0.52	16.48	-2.12
	$T=[\min(W_i),\max(W_i)]$	50	21.93	1.53	48.52	21.83	-1.02	0.17	5.45	-0.92
	$T=[Q_1(W_i),Q_2(W_i)]$	100	14.94	2.22	70.10	14.84	-2.21	0.38	11.93	-2.11
	$T=[\min(W_i),\max(W_i)]$	100	22.86	1.56	49.29	22.76	-1.09	0.19	5.96	-0.99
	$T=[Q_1(W_i),Q_2(W_i)]$	200	18.38	1.68	53.11	18.28	-2.21	0.34	10.89	-2.11
	$T=[\min(W_i),\max(W_i)]$	200	22.04	1.88	59.31	21.94	-1.24	0.16	5.17	-1.14
	$T=[Q_1(W_i),Q_2(W_i)]$	300	17.15	1.75	55.40	17.05	-1.95	0.15	4.78	-1.85
	$T=[\min(W_i),\max(W_i)]$	300	25.11	1.55	48.96	25.01	-1.26	0.13	4.24	-1.16



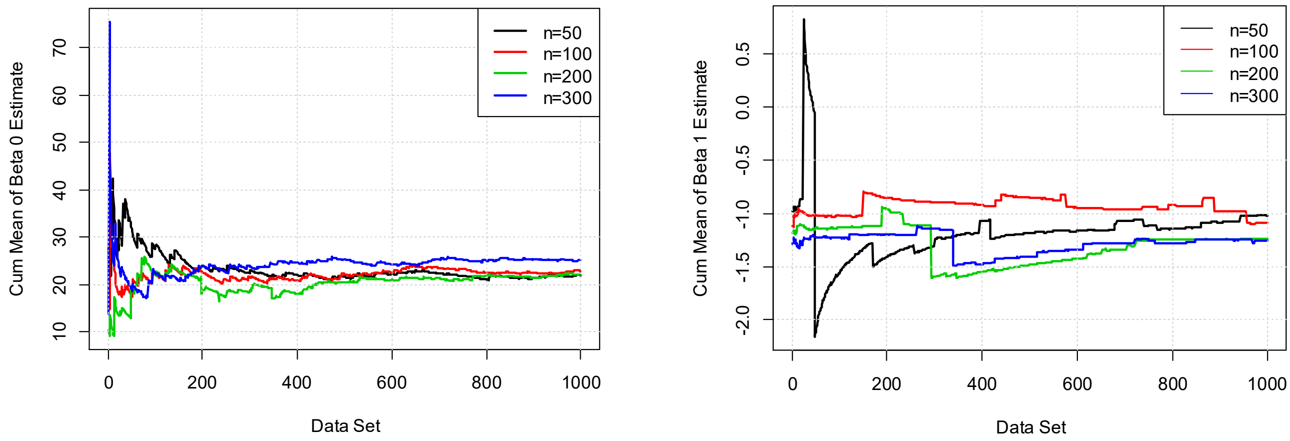
(a) Gamma(2,1), T=[$Q_1(W_i)$,$Q_3(W_i)$]



(b) Gamma (2,1), T=[Min(W_i),Max(W_i)]



(c) Gamma (1,2), T=[$Q_1(W_i)$,$Q_3(W_i)$]



(d) Gamma (1,2), T=[Min(W_i),Max(W_i)]

Fig. 2 (a)-(d) Cumulative Mean Plot of $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$ for $n = 50,100,200,300$

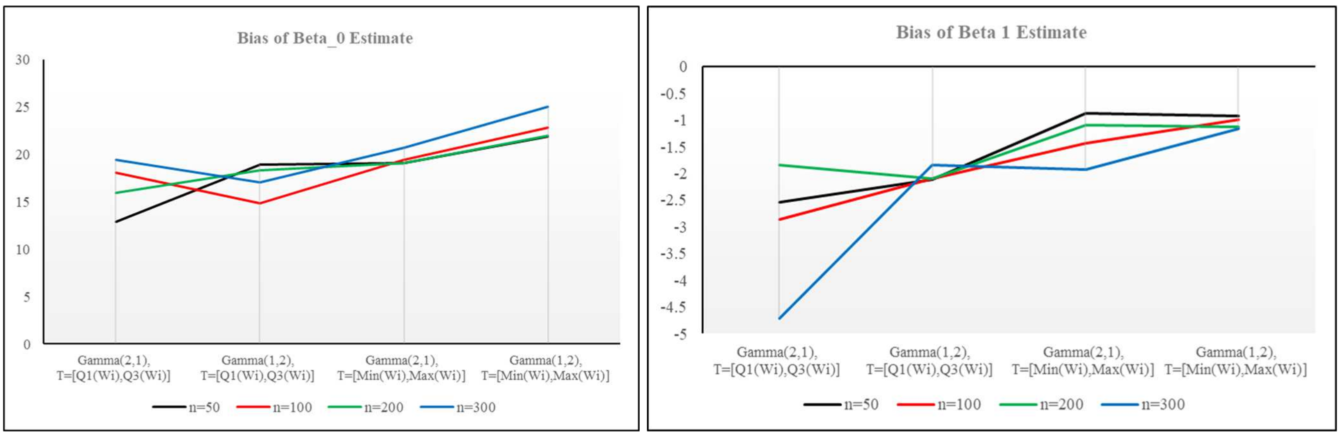


Fig. 3 Bias plot of modified SQS estimator $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$

IV. Conclusion

We considered the modified SQS estimator for the parameters of a Poisson regression model with measurement errors in a covariate. According to a known probability distribution, the true covariate measured with error comes from an unknown density function and has yielded observable values as a surrogate. We estimated the probability density function of true covariate as the prior density of Poisson distributed surrogate by the Empirical Bayes Deconvolution (EBD) method. After getting the probability estimation for every discretization in support set of the true covariate, we modified the mean and variance function in the quasi structural score estimating function proposed by Kukush *et al.* [23], and the modified SQS estimator is the root of the quasi structural score estimating function. For assessing the quality of the estimator, we did some simulation scenarios, and we compute the standard error of the mean, standard deviation, and the bias of the modified SQS estimator. As a result of the simulation, the bias of the modified SQS estimator has smaller values for the smaller-scale parameter value of Gamma distribution and becomes larger for the interquartile range discrete support set. In the next research, we will investigate the consistency of the modified SQS estimator and use more than one covariate contain measurement error in the Poisson regression model, and the estimation theory developed in this paper can be extended to the case where the distribution of the surrogate is not just a Poisson distribution but any discrete or continuous distribution.

ACKNOWLEDGMENT

The authors are grateful to the Indonesian Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP) of the Ministry of Finance of the Republic of Indonesia to fund this research through doctorate scholarship 20161141021499.

REFERENCES

[1] McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Boca Raton: Chapman Hall CRC.

[2] A. Agresti, Foundations of Linear and Generalized Linear Models Wiley Series in Probability and Statistics, 1st Edition.

[3] A. J. Dobson and A. G. Barrnett, An Introduction to Generalized Linear Models, Third Edition, Chapman & Hall/CRC Texts in Statistical Science.

[4] J. M. Hilbe, 2014. Modeling count data, Cambridge University Press.

[5] L. P. Fávero and P. Belfiore, "Regression Models for Count Data: Poisson and Negative Binomial". 2019.

[6] Z. Xie, "Predicting publication productivity for researchers: A piecewise Poisson model," *J. Informetr.*, vol. 14, no. 3, p. 101065, 2020, doi: 10.1016/j.joi.2020.101065.

[7] F. Pagliara and F. Mauriello, "Modelling the impact of High-Speed Rail on tourists with Geographically Weighted Poisson Regression," *Transp. Res. Part A Policy Pract.*, vol. 132, no. July 2018, pp. 780–790, 2020, doi: 10.1016/j.tra.2019.12.025.

[8] K. Chen, R. Huang, N. H. Chan, and C. Y. Yau, "Subgroup analysis of zero-inflated Poisson regression model with applications to insurance data," *Insur. Math. Econ.*, vol. 86, pp. 8–18, 2019, doi: 10.1016/j.insmatheco.2019.01.009.

[9] A. N. Shaaban, B. Peleteiro and M. R. O. Martins, "Statistical models for analyzing count data: predictors of length of stay among HIV patients in Portugal using a multilevel model, BMC Health Services Research (2021) 21:372 https://doi.org/10.1186/s12913-021-06389-1

[10] K. Adusumilli, D. Kurisu, T. Otsu, and Y.-J. Whang, "Inference on distribution functions under measurement error," *J. Econom.*, vol. 215, no. 1, pp. 131–164, 2020, doi: 10.1016/j.jeconom.2019.09.002.

[11] E. Kroc and B. D. Zumbo, "A transdisciplinary view of measurement error models and the variations of X=T+E," *J. Math. Psychol.*, vol. 98, p. 102372, 2020, doi: 10.1016/j.jmp.2020.102372.

[12] F. Elmagbri and R. M. Mnatsakanov, "Nonparametric density estimation based on the scaled Laplace transform inversion," *Trans. A. Razmadze Math. Inst.*, vol. 172, no. 3, pp. 440–447, 2018, doi: 10.1016/j.trmi.2018.09.003.

[13] M. B. Kamel and T. Sayed, "Cyclist-vehicle crash modeling with measurement error in traffic exposure," *Accid. Anal. Prev.*, vol. 144, no. January, p. 105612, Sep. 2020, doi: 10.1016/j.aap.2020.105612.

[14] B. Efron, "Empirical Bayes deconvolution estimates," *Biometrika*, vol. 103, no. 1, pp. 1–20, 2015, doi: 10.1093/biomet/asv068.

[15] A. A. Tsiatis, "Locally efficient semiparametric estimators for functional measurement error models", Biometrika 2004, 91, 4, pp. 835–848.

[16] P. Hall, "Semiparametric estimators of functional measurement error models with unknown error" *J. R. Statist. Soc.* B, 2007. 69, *Part* 3, *pp.* 429–446.

[17] M. Ghosh K. Sinha., Empirical Bayes estimation in finite population sampling under functional measurement error models", Journal of Statistical Planning and Inference, volume 137, Issue 9, 1 September 2007, p 2759-2773.

[18] S. Jadhav, S. Ma., Functional measurement error in functional regression, Canadian Journal of Statistics, volume 48, Issue 2 p. 238-258.

[19] M. Thamerus, "Modelling Count Data with Heteroscedastic Measurement Error in the Covariates," *Discuss. Pap. 58, SFB 386*, vol. 58, 1997.

[20] Rozliman N.A, Nur Ibrahim Adriana Irawati, and R. Mohammad Yunus, "Bayesian approach to errors-in-variables in regression models," *AIP Conf. Proc.*, vol. 1842, 2017, doi: 10.1063/1.4982856.

[21] M. Li, Y. Ma, and R. Li, "Semiparametric regression for measurement error model with heteroscedastic error," *J. Multivar. Anal.*, vol. 171, pp. 320–338, 2019, doi: 10.1016/j.jmva.2018.12.012.

[22] G. Y. Yi, "Statistical Analysis with Measurement Error or Misclassication: Strategy, Method and Application," 2017.

[23] A. Kukush, H. Schneeweis, and R. Wolf, "Three estimators for the poisson regression model with measurement errors," *Stat. Pap.*, vol. 45, no. 3, pp. 351–368, 2004, doi: 10.1007/BF02777577.

[24] L. Li, M. Palta, and J. Shao, "A measurement error model with a poisson distributed surrogate," *Stat. Med.*, vol. 23, no. 16, pp. 2527–2536, 2004, doi: 10.1002/sim.1838.

[25] Fuller W. 1987. Measurement error models, new york: John wiley. *Measurement Error Models,* 1987.

[26] Fuller WA. 1990. Prediction of true values for the measurement error model. *Statistical Analysis of Measurement Error Models and Applications* 112:41–57.

[27] Ghosh M, Sinha K, Kim D. 2006. Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models. *Scandinavian Journal of Statistics* 33(3):591–608.

[28] Buonaccorsi JP. 2010. *Measurement error: Models, Methods, and Applications*. CRC press.

[29] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.

[30] R Core Team R, "A language and environment for statistical computing". R Foundation for Statistical Computing. The R Project for Statistical Computing," Vienna, Austria, 2017.