Communications for Statistical Applications and Methods
2021, Vol. 28, No. 6, 627–641

https://doi.org/10.29220/CSAM.2021.28.6.627
Print ISSN 2287-7843 / Online ISSN 2383-4757

Generalized nonlinear percentile regression using asymmetric maximum likelihood estimation

Juhee Lee[a], Young Min Kim[1, a]

[a]Department of Statistics, Kyungpook National University, Korea

Abstract

An asymmetric least squares estimation method has been employed to estimate linear models for percentile regression. An asymmetric maximum likelihood estimation (AMLE) has been developed for the estimation of Poisson percentile linear models. In this study, we propose generalized nonlinear percentile regression using the AMLE, and the use of the parametric bootstrap method to obtain confidence intervals for the estimates of parameters of interest and smoothing functions of estimates. We consider three conditional distributions of response variables given covariates such as normal, exponential, and Poisson for three mean functions with one linear and two nonlinear models in the simulation studies. The proposed method provides reasonable estimates and confidence interval estimates of parameters, and comparable Monte Carlo asymptotic performance along with the sample size and quantiles. We illustrate applications of the proposed method using real-life data from chemical and radiation epidemiological studies.

Keywords: asymmetric maximum likelihood estimation, nonlinear regression, percentile, quantile

1. Introduction

Quantile regression (Koenker and Bassett Jr., 1978) is one of the most common statistical techniques used to conduct statistical inference for conditional quantile functions. The method can be used to construct models to estimate the percentile of conditional distribution and provide robustness to outliers. Thus, quantile regression has been applied in many fields of study, such as economics, clinical studies, and epidemiology. (Eide and Showalter, 1998; Zietz *et al.*, 2008; Austin *et al.*, 2005; Beyerlein, 2014).

Like classical regression methods based on minimizing sums of squared residuals which can estimate conditional mean models, quantile regression uses the method to minimize sums of weighted absolute residuals for estimating conditional quantile models. Since the quantile regression utilizes the absolute loss function, the quantile estimators can often be difficult to obtain the optimal solutions, i,e., it is not continuously differentiable (Newey and Powell, 1987). Thus Newey and Powell (1987) proposed asymmetric least square (ALS) estimation, which uses the square loss function called *expectile* regression to distinguish it from the general quantile regression. Efron (1991, 1992) extended the ALS idea to maximum likelihood estimation, which is called the asymmetric maximum likelihood (AML) method and can be applied to members of exponential families, such as normal, binomial, exponential, and Poisson distributions.

For quantile regression methods to apply to models with a nonlinear relationship between the covariates and the response, Koenker *et al.* (1994) used smoothing splines methods for the nonlinear

[1] Corresponding author: Department of Statistics, Kyungpook National University, 80, Daehak-ro, Buk-gu, Daegu 41566, Korea. E-mail: kymmyself@knu.ac.kr

Published 30 November 2021 / journal homepage: http://csam.or.kr

quantile model, and He *et al.* (1998) investigated the bivariate quantile smoothing spline and proposed penalized bivariate quantile B-splines methods. Karlsson (2007) applied quantile regression to nonlinear longitudinal data. Geraci and Bottai (2007) presented a likelihood-based approach for the estimation of nonlinear quantile regression using asymmetric Laplace density. Wang (2012) considered the Bayesian nonlinear quantile regression model using asymmetric Laplace distribution. Geraci (2019) developed the nonlinear conditional quantile method to use when data are clustered within two-level nested designs. All of the above mentioned methods were developed based on quantile regression.

In this study, we propose an extension of the method discussed in Efron (1992), the application of asymmetric maximum likelihood estimation (AMLE) to generalized nonlinear models, and use the parametric bootstrap method to obtain confidence intervals for the estimated parameters and their smoothing functions.

The paper is organized as follows: In Section 2, we introduce the AMLE method for the generalized nonlinear percentile model and the parametric bootstrap method to compute 95% confidence intervals of each percentile estimate. In Section 3, we conduct simulation studies considering three distributions of response in the exponential family to evaluate the statistical and computational performance of the proposed method. In Section 4, we use real-life data such as chlorine, and Life Span Study (LSS) cohort data to carry out the studies. In Section 5, we summarize and discuss our results.

2. Generalized nonlinear percentile regression using asymmetric maximum likelihood estimation

The generalized nonlinear percentile regression uses the asymmetric maximum likelihood estimation (AMLE) method to estimate the generalized nonlinear models. In the next two subsections, we first describe the generalized nonlinear models and then provide the AMLE method to estimate the parameter coefficients for the generalized nonlinear models. In Section 2.2, we present the algorithms of AMLE and the parametric bootstrap methods to compute the variance estimates of the parameter coefficients of interest. In addition, we rewrite the expectile regression as the percentile regression to clarify the meaning of the method in this manuscript.

2.1. Generalized nonlinear models

Suppose that data consists of $(\mathbf{x}, y)$, where $\mathbf{x}$ is a p-dimensional covariates vector, and y is a response. We assume that y belongs to an exponential family defined as,

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}, \tag{2.1}$$

where the parameter θ is the natural parameter, and ϕ is the dispersion parameter. We can rewrite (2.1) as,

$$f(y;\eta) = \exp\{y\eta - \psi(\eta)\}. \tag{2.2}$$

where $\eta = \theta/a(\phi)$ and $\psi(\eta) = -b(\theta)/a(\phi) + c(y,\phi)$. In generalized linear regression, we assumed that

$$E(y_t) = g^{-1}\left(\mathbf{x}_t^T\boldsymbol{\beta}\right) \tag{2.3}$$

where $(\mathbf{x}_t, y_t)$ is the t^{th} observed vector, and $g(\cdot)$ is the link function that connects the response, y_t, and the linear predictor function, $\mathbf{x}_t^T\boldsymbol{\beta}$. In this model, we estimate a p-dimensional parameter vector

Table 1: Canonical link functions for the distributions, which are members of exponential families. Note that in the Gamma distribution, κ is a scale parameter

Distribution	Distribution mean	Canonical link function
Normal(μ, σ^2)	μ	μ
Poisson(λ)	λ	$\log\lambda$
Bernoulli(p)	p	$\log\{p/(1-p)\}$
Gamma(α, κ)	$\alpha\kappa$	$-(\alpha\kappa)^{-1}$

$\boldsymbol{\beta}$ from the models. The link function g, which transforms the mean $E(y_t)$ to the natural parameter in (2.3), is called the canonical link, and it is preferred because it leads to models with desirable statistical properties (McCullagh and Nelder, 1989).

Table 1 shows the link functions corresponding to the conditional distribution of the response variable given the covariates. Generally, we take a link function as an identity, logarithm, logit and negative inverse link function for normal, Poisson, Bernoulli or binomial, and Gamma distribution, respectively. Compared with other link functions, the canonical link function in the Gamma distribution does not cover the whole space of the real number because the predictor function might be negative even though the mean of the Gamma distribution must be a positive value. So, in this case, the non-canonical link function like a logarithm link function is as an alternative.

$$E(y_t) = g^{-1}(h(\mathbf{x}_t, \boldsymbol{\beta})), \qquad \text{for} \quad t = 1, \ldots, n, \tag{2.4}$$

We extend generalized linear models to generalized nonlinear models using a nonlinear predictor function $h(\cdot)$ in the equation (2.4). In usual, the function is prespecified in data fields because the predictor function $h(\cdot)$ describes the relationship between the covariates and the response. For example, the excess relative risk models (4.2) is popular-used in the radiation epidemiology. The model with an identity $h(\cdot)$ is equal to generalized linear model.

2.2. Asymmetric maximum likelihood method

The AML method uses deviance to estimate the parameters of generalized nonlinear percentile models. In general, the deviance is defined as

$$\begin{aligned} D(\mu, \mu^*) &= 2\mathrm{E}_\eta\left[\log\frac{f(y;\eta)}{f(y;\eta^*)}\right] \\ &= 2\left[(\eta-\eta^*)\mu - \{\psi(\eta)-\psi(\eta^*)\}\right], \end{aligned} \tag{2.5}$$

where the two different parameters η and η^* are from (2.2) and μ, respectively, and μ^* are population mean functions. It represents the distance of two distributions. In this study, we are interested in the generalized percentile model using the nonlinear function $h(\cdot)$. The asymmetric version of the deviance function depending on a positive constant w, called a weight, is

$$D_w(\mu, \mu^*) = \begin{cases} D(\mu, \mu^*), & \text{if } \mu \le \mu^*, \\ wD(\mu, \mu^*), & \text{if } \mu > \mu^*. \end{cases} \tag{2.6}$$

From this deviance function, we compute the weight of the part of $\mu > \mu^*$ that corresponds to $\tau \in (0, 1)$ which is called a percentile rank. With the dataset $\{(\mathbf{x}_t, y_t) : t = 1, 2, \ldots, n\}$ in $(p+1)$-dimensional space, we can consider the surface $\mathcal{L}^{(\tau)} = \{(\mathbf{x}, y) : y = \mu(\mathbf{x}, \boldsymbol{\beta}_{(\tau)})\}$ that cuts the space into

Table 2: Deviance $D(y, \mu)$ for members of exponential families, where μ is the population mean of the distributions. Note that the variance σ^2 in the normal distribution and the shape parameter α in the Gamma distribution are assumed to be known

Distribution	Deviance
Normal	$(y-\mu)^2/\sigma^2$
Poisson	$2\{y(\log y - \log\mu - 1) + \mu\}$
Bernoulli	$-2\{y\log\mu + (1-y)\log(1-\mu)\}$
Gamma	$2\alpha\{-1 + y/\mu + \log y - \log\mu\}$

the proportions τ and $1-\tau$. We define this surface to a regression percentile, determined by $\boldsymbol{\beta}_{(\tau)}$. When $F_{Y|\mathbf{x}}$ is the cumulative distribution function of Y given $\mathbf{x}$, then the percentile rank is defined as $\tau = F_{Y|\mathbf{x}}\left\{\mu(\mathbf{x}, \boldsymbol{\beta}_{(\tau)})\right\}$.

Back to the formula (2.6), the $\hat{\boldsymbol{\beta}}_w$ can be put to the $\boldsymbol{\beta}$ to minimize the sum of the weighted deviance D_w. The first term is an observed response value, and the second is the expected value of the model,

$$\hat{\boldsymbol{\beta}}_w = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{n} D_w\{y_t, \mu(\mathbf{x}_t, \boldsymbol{\beta})\} \tag{2.7}$$

where $\mu(\mathbf{x}, \boldsymbol{\beta}) = g^{-1}(h(\mathbf{x}, \boldsymbol{\beta}))$. Then, the surface $\mathcal{L}_w = \{(\mathbf{x}, y) : y = \mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_w)\}$ cuts the points of the dataset. If $w = 1$, $\hat{\boldsymbol{\beta}}_w$ theoretically become the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$. In addition, the change in w determines where the obtained surface cuts the dataset. More points of data are located below the surface $\mathcal{L}_w$ compared with above the surface $\mathcal{L}_w$ if the value of w is greater than 1. Similarly, a smaller w implies a smaller number of data points under the surface (See Efron, 1991, Section 2). Because of the characteristic, $\hat{\boldsymbol{\beta}}_w$ can be an appropriate estimator for $\boldsymbol{\beta}_{(\tau)}$. Therefore, we can define the 100τth regression percentile as

$$\hat{\mathcal{L}}^{(\tau)} \equiv \left\{(\mathbf{x}, y) : y = \mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_w)\right\}, \quad \text{for} \quad \tau = \hat{\tau}_w, \tag{2.8}$$

where $\hat{\tau}_w = 1/n \sum_{t=1}^{n} I\{y_t \le \mu(\mathbf{x}_t, \hat{\boldsymbol{\beta}}_w)\}$, and $I(\cdot)$ is the indicator function. To estimate the parameters of the regression percentile, we must find the appropriate weight w for τ.

Since $\hat{\tau}_w$ is the proportion of points under the regression percentile surface, the greater w, the greater $\hat{\tau}_w$. To determine the weight, we first define the function $r(w) = \hat{\tau}_w - \tau$. Then the problem change to solving the equation $r(w) = 0$. The solution of the equation provides appropriate weight corresponding to τ. The bisection method is used to solve this equation. We assume that $r(w)$ is continuous on the interval $[w_L, w_U]$. The first step in algorithm is to find the points which values of $r(\cdot)$ are opposite sign. In this step, the tuning parameter k is used. After finding the points, the bisection algorithm is applied to detect the weight w. The whole procedure of the AML for generalized nonlinear percentile models is provided in **Algorithm 1**.

To estimate the confidence intervals for all percentile parameters, we utilize the parametric bootstrap method because we assume the link function, which implies the conditional distribution of the response. The parametric bootstrap needs the information of the population distribution. The parametric bootstrap method is more powerful and provides smaller variances than the nonparametric bootstrap if the distribution assumption is right and the sample sizes are relatively small sometimes. In particular, there are a large number of zero-response values, the model has a sparse response variable, at that time the parametric bootstrap provides stable results in comparison to the nonparametric bootstrap. The parametric bootstrap algorithm is provided in **Algorithm 2**.

Algorithm 1: AMLE for generalized nonlinear percentile model

Input : a data matrix, an initial weight w_0 and tuning parameter $k > 1$
Output: a estimated parameter $\hat{\boldsymbol{\beta}}_{(\tau)}$
repeat
  Calculate $\hat{\boldsymbol{\beta}}_w$ and $r_w = \hat{\tau}_w - \tau$;
  if $|r_w| < \epsilon$ **then**
    return $\hat{\boldsymbol{\beta}}_w$
  else if $r_w < 0$ **then**
    $w_L \leftarrow w$;
    $w_U \leftarrow kw$;
    $w \leftarrow w_U$;
  else if $r_w > 0$ **then**
    $w_L \leftarrow w/k$;
    $w_U \leftarrow w$;
    $w \leftarrow w_L$;
until $r_{w_u} \times r_{w_L} < 0$;
repeat
  $w \leftarrow (w_L + w_U)/2$;
  Calculate $\hat{\boldsymbol{\beta}}_w$ and r_w;
  if $r_w < 0$ **then**
    $w_L \leftarrow w$;
  else if $r_w > 0$ **then**
    $w_U \leftarrow w$;
until $|r_w| < \epsilon$;
return $\hat{\boldsymbol{\beta}}_{(\tau)} = \hat{\boldsymbol{\beta}}_w$

Algorithm 2: Computation of parametric bootstrap confidence intervals for generalized nonlinear percentile model

Input : a data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and response $\mathbf{y} = (y_1, \ldots, y_n)^T$
Output: a bootstrap $(1-\gamma) \times 100\%$ confidence interval for $\hat{\boldsymbol{\beta}}_{(\tau)}$
Compute the ML estimate $\hat{\boldsymbol{\beta}}$
for *bootstrap sample* $b = 1, \ldots, B$ **do**
  for *observed data* $i = 1, \ldots, n$ **do**
    Calculate $\hat{\mu}_i = \mu(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$;
    Using $\hat{\mu}_i$, generate bootstrap sample $y_i^{(b)}$;
  end
  Calculate the parameter of percentile model $\hat{\boldsymbol{\beta}}_{(\tau)}^{(b)}$ from the bootstrapped data $(\mathbf{X}, \mathbf{y}^{(b)})$;
end
$100(\tau/2)\%$, $100(1-\tau/2)\%$ percentiles of $(\hat{\boldsymbol{\beta}}_{(\tau)}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}_{(\tau)}^{(B)})$ are the $(1-\gamma) \times 100\%$ bootstrap confidence interval.;

3. Simulation studies

In this section, we conduct simulation studies for generalized nonlinear percentile regression. We consider the three conditional distributions (normal, exponential, and Poisson) of a response given the

covariates and three mean functions (one linear and two nonlinear models). The Monte Carlo runs of 1000 with sample sizes $N = 100, 300, 500$ are conducted, and the root mean squared error (RMSE) is considered to evaluate the performance of model fitting using the proposed method. The RMSE in this study is defined as,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left\{\mu\left(\mathbf{x}_t, \hat{\boldsymbol{\beta}}_{(\tau)}\right) - Q_{y_t}(\tau)\right\}^2}, \tag{3.1}$$

where $Q_{y_t|\mathbf{x}_t}(\tau)$ is the true percentile of a response mean function. In this simulation studies, we compute the Monte Carlo asymptotic values for $Q_{y_t|\mathbf{x}_t}(\tau)$.

3.1. Normal distribution

We assume that the error distribution follows the normal distribution with the standard deviation $\sigma = 1$, considering the three mean functions described below.

Linear : $E(Y|X) = \beta_0 + \beta_1 X$
Nonlinear 1 : $E(Y|X) = \beta_0 + X^{\beta_1}$
Nonlinear 2 : $E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 \exp(1 + \beta_2 X_2)$

For **Linear**, a covariate X is generated from a normal distribution with a mean of zero and a standard deviation of $\sigma = 2$. For **Nonlinear 1**, a covariate X is generated from a uniform distribution with a range of $(0, 1)$. Finally, for **Nonlinear 2**, we generate two covariates (X_1, X_2): X_1 is generated from a normal distribution with a mean of zero and a standard deviation of $\sigma = 2$, and X_2 is generated from an exponential distribution with scale parameter of 1. We set $(\beta_0, \beta_1) = (1, 0.2)$, $(\beta_0, \beta_1) = (1, 0.5)$ and $(\beta_0, \beta_1, \beta_2) = (1,\ 0.5,\ 0.2)$, corresponding to the mean functions, respectively.

Table 3 and Figure 1 are the results of the estimated percentile models in each mean function. Table 3 illustrates that the RMSE tends to be smaller as the sample size increases. Moreover, RMSE decreases as the quantile concentrates on $\tau = 0.5$, i.e., τ increases or decreases to 0.5 from $\tau = 0.2$ or $\tau = 0.8$. The lower (higher) percentile regression surface cuts the data points into the lower (higher) part of the points cloud.

In **Linear**, $\hat{\beta}_1$ is estimated to be almost identical to the real β_1 for all quantiles. Similar results appear in **Nonlinear 1 and 2**. Therefore, we can see that the lines of regression percentile are parallel in Figure 1. From these results, it can be seen that the percentiles of the mean functions are significantly influenced by the intercept term. The MLEs for all parameters are similar to the AMLEs when $\tau = 0.5$, which means that the distribution of residuals is symmetric based on X_2. However the RMSEs for the MLEs are smaller than those for the AMLEs with $\tau = 0.5$. It means that the conditional percentile estimator has higher variance comparing with conditional expectation estimator.

3.2. Exponential distribution

In the second numerical example, we consider the conditional distribution of the response y given the covariates as the exponential distribution. The three mean functions which were considered are as follows,

Linear : $\log E(Y|X) = \beta_0 + \beta_1 X$
Nonlinear 1 : $\log E(Y|X) = \beta_0 + X^{\beta_1}$
Nonlinear 2 : $\log E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1/(1 + \beta_2 X_2)$

Table 3: Monte Carlo RMSE when the response is normal. β's are represented as the Monte Carlo means of all estimated parameters. Note that ML is the MLE for the conditional mean function of response

τ	**Linear**			**Nonlinear 1**			**Nonlinear 2**			
	β_0	β_1	RMSE	β_0	β_1	RMSE	β_0	β_1	β_2	RMSE
$N = 100$										
$\tau = 0.20$	0.148	0.200	0.166	0.143	0.493	0.167	0.142	0.499	0.203	0.205
$\tau = 0.30$	0.479	0.201	0.153	0.477	0.494	0.152	0.475	0.500	0.202	0.188
$\tau = 0.40$	0.745	0.201	0.143	0.745	0.495	0.141	0.748	0.500	0.202	0.177
$\tau = 0.50$	0.999	0.201	0.139	0.998	0.496	0.138	1.004	0.500	0.201	0.172
$\tau = 0.60$	1.272	0.201	0.143	1.268	0.496	0.143	1.273	0.500	0.201	0.178
$\tau = 0.70$	1.530	0.200	0.152	1.525	0.497	0.150	1.533	0.500	0.201	0.188
$\tau = 0.80$	1.855	0.200	0.163	1.853	0.497	0.165	1.863	0.501	0.201	0.205
ML	0.999	0.201	0.003	0.997	0.499	0.004	1.000	0.500	0.202	0.159
$N = 300$										
$\tau = 0.20$	0.153	0.199	0.095	0.163	0.496	0.095	0.155	0.500	0.199	0.115
$\tau = 0.30$	0.473	0.199	0.085	0.477	0.496	0.086	0.475	0.500	0.199	0.104
$\tau = 0.40$	0.744	0.199	0.082	0.750	0.496	0.082	0.748	0.501	0.199	0.098
$\tau = 0.50$	1.001	0.199	0.083	1.004	0.496	0.081	1.000	0.501	0.199	0.097
$\tau = 0.60$	1.253	0.199	0.083	1.256	0.496	0.082	1.252	0.501	0.199	0.101
$\tau = 0.70$	1.524	0.199	0.087	1.526	0.496	0.085	1.522	0.501	0.199	0.106
$\tau = 0.80$	1.841	0.198	0.097	1.842	0.496	0.095	1.846	0.501	0.199	0.117
ML	0.998	0.199	0.003	1.003	0.498	0.003	1.000	0.501	0.199	0.090
$N = 500$										
$\tau = 0.20$	0.152	0.201	0.072	0.149	0.502	0.072	0.156	0.500	0.200	0.091
$\tau = 0.30$	0.474	0.201	0.066	0.470	0.502	0.065	0.474	0.500	0.200	0.081
$\tau = 0.40$	0.746	0.200	0.065	0.743	0.501	0.063	0.748	0.500	0.200	0.078
$\tau = 0.50$	1.001	0.200	0.063	0.997	0.501	0.063	1.002	0.500	0.200	0.076
$\tau = 0.60$	1.254	0.200	0.065	1.251	0.501	0.064	1.254	0.500	0.200	0.078
$\tau = 0.70$	1.527	0.200	0.069	1.523	0.500	0.068	1.524	0.500	0.200	0.083
$\tau = 0.80$	1.846	0.199	0.078	1.842	0.500	0.075	1.846	0.500	0.200	0.092
ML	1.000	0.200	0.000	0.996	0.502	0.003	0.999	0.500	0.200	0.071

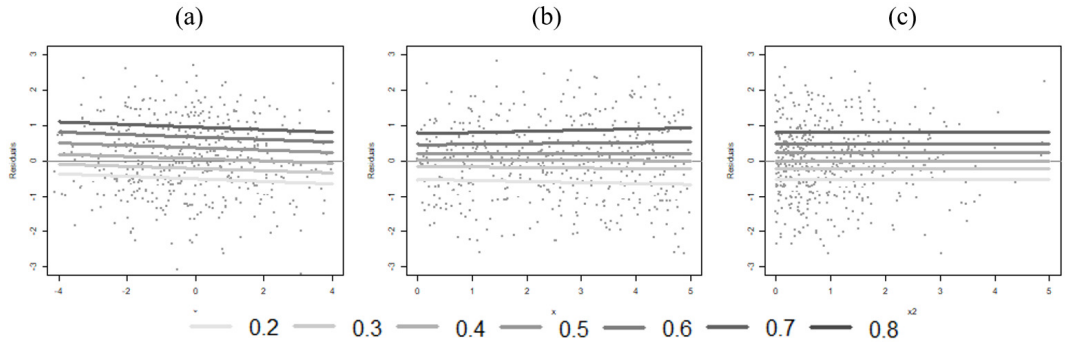

Figure 1: *Residual plots for estimated percentile lines based on X_2 with fixed $X_1 = 0$, corresponding to $\tau = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)$ for three models when the sample size is $n = 500$ and the response is normal: (a) Linear, (b) Nonlinear 1, and (c) Nonlinear 2.*

Table 4: Monte Carlo RMSE when the response is exponential. β's are represented as the Monte Carlo means of all parameters. Note that ML is the MLE for the conditional mean function of response

τ	**Linear**			**Nonlinear 1**			**Nonlinear 2**			
	β_0	β_1	RMSE	β_0	β_1	RMSE	β_0	β_1	β_2	RMSE
$N = 100$										
$\tau = 0.20$	−0.478	0.502	0.504	−0.480	0.183	0.473	−0.490	0.572	0.918	0.635
$\tau = 0.30$	−0.032	0.501	0.668	−0.029	0.186	0.630	−0.038	0.539	0.474	0.804
$\tau = 0.40$	0.338	0.500	0.851	0.334	0.187	0.782	0.329	0.532	0.428	0.998
$\tau = 0.50$	0.632	0.500	1.082	0.634	0.189	0.976	0.632	0.526	0.372	1.256
$\tau = 0.60$	0.933	0.500	1.422	0.928	0.191	1.201	0.932	0.523	0.356	1.609
$\tau = 0.70$	1.199	0.499	1.869	1.198	0.192	1.592	1.202	0.522	0.357	2.129
$\tau = 0.80$	1.491	0.500	2.537	1.489	0.192	2.162	1.494	0.524	0.377	2.960
ML	0.990	0.500	1.366	0.990	0.191	1.112	0.985	0.522	0.355	1.617
$N = 300$										
$\tau = 0.20$	−0.501	0.502	0.297	−0.500	0.194	0.264	−0.505	0.508	0.272	0.338
$\tau = 0.30$	−0.032	0.501	0.402	−0.033	0.194	0.345	−0.037	0.505	0.250	0.453
$\tau = 0.40$	0.325	0.501	0.507	0.326	0.195	0.437	0.323	0.504	0.242	0.579
$\tau = 0.50$	0.632	0.500	0.626	0.632	0.195	0.541	0.628	0.504	0.239	0.730
$\tau = 0.60$	0.914	0.500	0.787	0.912	0.195	0.671	0.906	0.504	0.237	0.925
$\tau = 0.70$	1.186	0.500	1.025	1.185	0.195	0.860	1.182	0.504	0.238	1.210
$\tau = 0.80$	1.476	0.500	1.430	1.475	0.194	1.145	1.476	0.504	0.238	1.689
ML	0.997	0.500	0.801	0.997	0.195	0.625	0.993	0.504	0.237	0.964
$N = 500$										
$\tau = 0.20$	−0.497	0.501	0.228	−0.498	0.198	0.201	−0.503	0.505	0.237	0.271
$\tau = 0.30$	−0.030	0.501	0.304	−0.029	0.199	0.264	−0.034	0.503	0.227	0.362
$\tau = 0.40$	0.327	0.501	0.399	0.327	0.200	0.333	0.325	0.503	0.224	0.467
$\tau = 0.50$	0.632	0.500	0.504	0.632	0.200	0.412	0.632	0.503	0.223	0.591
$\tau = 0.60$	0.910	0.500	0.633	0.911	0.200	0.504	0.909	0.503	0.223	0.754
$\tau = 0.70$	1.185	0.500	0.821	1.185	0.200	0.655	1.184	0.504	0.224	0.992
$\tau = 0.80$	1.476	0.500	1.135	1.474	0.199	0.863	1.474	0.505	0.227	1.399
ML	0.998	0.500	0.648	0.997	0.200	0.479	0.995	0.504	0.223	0.788

Linear and **Nonlinear 1** are similar to the mean functions in Section 3.1. For the exponential distribution, the logarithm link function is utilized. For **Linear**, we generate a covariate X from a normal distribution with a mean of zero and a standard deviation of $\sigma = 2$. For **Nonlinear 1**, a covariate X is generated by a Gamma distribution with scale parameter 2 and rate parameter 1. In **Nonlinear 2**, we generate two covariates, X_1 and X_2, where X_1 is a normal distribution with a mean of zero and standard deviation 2, and X_2 is a uniform distribution with a range of $(0, 1)$. We also take the model parameters to $(\beta_0, \beta_1) = (1, 0.2)$, $(\beta_0, \beta_1) = (1, 0.5)$ and $(\beta_0, \beta_1, \beta_2) = (1, 0.5, 0.2)$, corresponding to the mean functions, respectively.

Table 4 and Figure 2 show the numerical results, which demonstrate that the RMSE tends to be smaller for large N. However, we can see that the τ increases as it increases because the variance of the regression percentile tends to be larger in large τ because the variance of exponential distribution increases as the mean increases. In Figure 2, the residuals in (a) and (b) spread as values of X increase, the models are increasing functions for the covariate X. The MLEs for all the parameters in all the models are larger than the AMLEs with $\tau = 0.5$.

3.3. Poisson distribution

Finally, we assume that the response y is a Poisson random variable. The models are as follows,

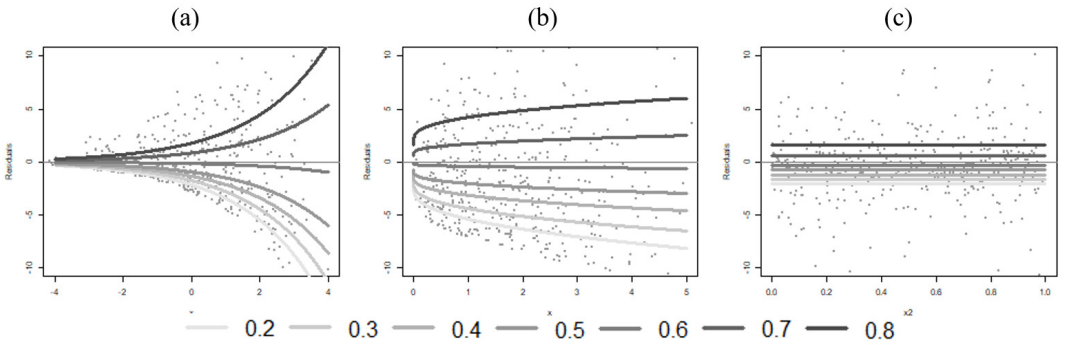

Figure 2: *Residual plots for estimated percentile lines based on X_2 with fixed $X_1 = 0$, corresponding to $\tau = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)$ for three models when the sample size is $n = 500$ and the response is exponential: (a) Linear, (b) Nonlinear 1, and (c) Nonlinear 2.*

Linear : $\log E(Y|X) = \beta_0 + \beta_1 X$
Nonlinear 1 : $\log E(Y|X) = \beta_0 + X^{\beta_1}$
Nonlinear 2 : $\log E(Y|X_1, X_2) = (\beta_0 + \beta_1 X_1)/(1 + \beta_2 e^{X_2})$

Linear and **Nonlinear 1** models are same as those in Section 3.2. However, we generate a covariate X for **Linear** from an exponential distribution with a mean of $\lambda = 1$. The covariate for **Nonlinear 1** is generated from the same distribution with **Linear**. In **Nonlinear 2**, two covariates (X_1, X_2) are generated from the exponential distribution with a mean of $\mu = 1$ and a Gamma distribution with shape parameter 2 and rate parameter 2, respectively. We set $(\beta_0, \beta_1) = (1, 0.5)$, $(\beta_0, \beta_1) = (1, 0.2)$ and $(\beta_0, \beta_1, \beta_2) = (1, 0.5, 0.2)$, corresponding to the three models, respectively.

Table 5 and Figure 3 illustrate that the RMSEs decrease as the sample size increases or as τ goes to 0.5 from $\tau = 0.8$ or $\tau = 0.2$. This is the same result as that of the numerical study for the normal response in Section 3.1. The MLEs for all the parameters in all the models are almost similar to the AMLEs with $\tau = 0.5$; however, the RMSEs for the MLEs are smaller than those for the AMLEs with $\tau = 0.5$.

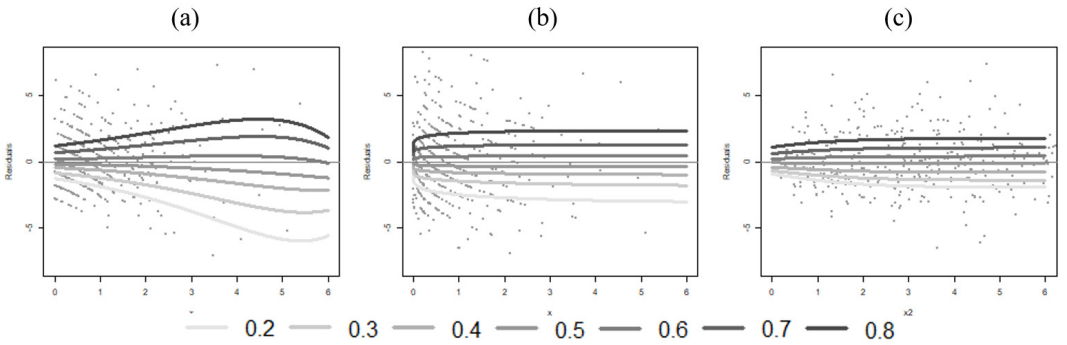

Figure 3: *Residual plots for estimated percentile lines based on X_2 with fixed $X_1 = 1$, corresponding to $\tau = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)$ for three models when the sample size is $n = 500$ and the response is Poisson: (a) Linear, (b) Nonlinear 1, and (c) Nonlinear 2.*

Table 5: Monte Carlo RMSE when the response is Poisson. β's are represented as the Monte Carlo means of all parameters. Note that ML is the MLE for the conditinoal mean function of response

τ	Linear			Nonlinear 1			Nonlinear 2			
	β_0	β_1	RMSE	β_0	β_1	RMSE	β_0	β_1	β_2	RMSE
$N = 100$										
$\tau = 0.20$	0.358	0.605	0.768	0.607	0.253	0.512	0.299	0.622	0.349	0.872
$\tau = 0.30$	0.611	0.560	0.636	0.760	0.231	0.495	0.587	0.567	0.280	0.727
$\tau = 0.40$	0.793	0.530	0.646	0.875	0.216	0.487	0.781	0.533	0.244	0.719
$\tau = 0.50$	0.967	0.502	0.598	0.978	0.204	0.484	0.962	0.502	0.220	0.690
$\tau = 0.60$	1.114	0.479	0.634	1.078	0.193	0.504	1.115	0.477	0.200	0.711
$\tau = 0.70$	1.255	0.458	0.683	1.168	0.183	0.527	1.258	0.454	0.184	0.766
$\tau = 0.80$	1.408	0.434	0.777	1.274	0.173	0.597	1.420	0.429	0.171	0.862
ML	1.003	0.499	0.474	1.000	0.202	0.329	0.998	0.499	0.216	0.583
$N = 300$										
$\tau = 0.20$	0.376	0.591	0.664	0.605	0.248	0.388	0.343	0.603	0.287	0.720
$\tau = 0.30$	0.608	0.555	0.534	0.755	0.228	0.378	0.600	0.560	0.256	0.574
$\tau = 0.40$	0.792	0.528	0.529	0.873	0.214	0.373	0.781	0.532	0.232	0.564
$\tau = 0.50$	0.964	0.503	0.475	0.975	0.203	0.373	0.963	0.504	0.210	0.515
$\tau = 0.60$	1.091	0.484	0.505	1.069	0.192	0.386	1.096	0.484	0.195	0.527
$\tau = 0.70$	1.244	0.463	0.562	1.163	0.183	0.401	1.249	0.461	0.179	0.586
$\tau = 0.80$	1.385	0.444	0.624	1.267	0.173	0.429	1.396	0.441	0.165	0.650
ML	1.000	0.500	0.344	0.999	0.201	0.199	1.000	0.500	0.207	0.395
$N = 500$										
$\tau = 0.20$	0.388	0.586	0.662	0.604	0.248	0.355	0.358	0.596	0.286	0.684
$\tau = 0.30$	0.612	0.553	0.534	0.756	0.228	0.347	0.605	0.557	0.256	0.535
$\tau = 0.40$	0.794	0.527	0.515	0.874	0.214	0.342	0.784	0.531	0.232	0.518
$\tau = 0.50$	0.965	0.503	0.454	0.975	0.203	0.345	0.965	0.504	0.210	0.469
$\tau = 0.60$	1.089	0.486	0.483	1.069	0.193	0.349	1.094	0.485	0.194	0.486
$\tau = 0.70$	1.243	0.465	0.543	1.164	0.183	0.360	1.245	0.464	0.177	0.552
$\tau = 0.80$	1.382	0.447	0.604	1.268	0.173	0.380	1.390	0.444	0.162	0.612
ML	1.000	0.500	0.309	1.000	0.201	0.153	0.999	0.501	0.207	0.339

4. Data applications

We apply generalized nonlinear percentile regression using AMLE to two data samples from a chemical study that investigated the relationship between the proportion of chlorine in the product (Draper and Smith, 1981; Smith and Dubey, 1964) in Section 4.1 and a radiation-associated epidemiological study, called the Life Span Study (LSS), of health effects in atomic bomb survivors in Japan Preston *et al.* (2007); Grant *et al.* (2017) in Section 4.2.

4.1. Chlorine dataset

Draper and Smith (1981) stated a problem due to Smith and Dubey (1964) about a certain product having 50% available chlorine at the time of manufacturing. When it reached the customer eight weeks later, the level of available chlorine had dropped to 49%. It is known that the level should stabilize at approximately 30%. To predict how long the chemical would last at the customer site, samples were collected at different times. The data includes 44 observations in which the response is the fraction of available chlorine, and the covariate is the length of time between when the product was produced and when it was used. The fraction of available chlorine in the product decreases with time. It was postulated that the following nonlinear model fits the data,

$$Y = \beta_0 + (0.49 - \beta_0)e^{-\beta_1(X-8)} + \varepsilon, \tag{4.1}$$

Table 6: Nonlinear percentile estimates using an identity link function for chlorine data along with $\tau = 0.25, 0.5, 0, 75$. The parentheses are the 95% bootstrap confidence intervals of each estimate. The ML is the MLE for each parameter

τ	β_0 (CI)	β_1 (CI)
0.25	0.388 (0.375,0.395)	0.128 (0.093,0.161)
0.50	0.390 (0.378,0.399)	0.102 (0.076,0.132)
0.75	0.388 (0.373,0.407)	0.075 (0.053,0.115)
ML	0.390	0.102

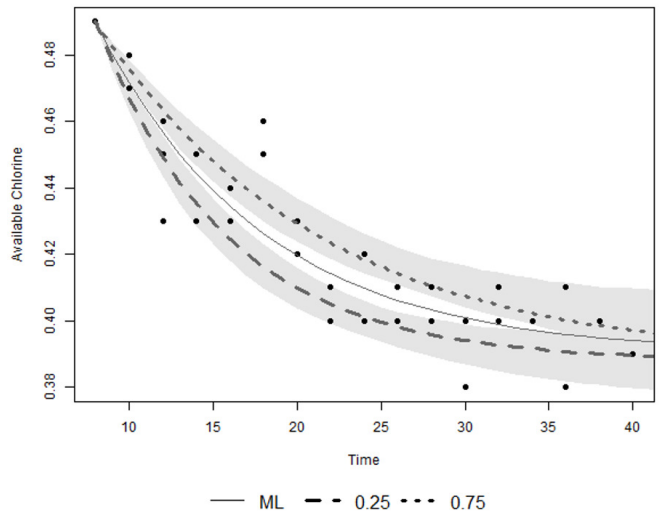

Figure 4: *Plot for nonlinear percentile estimates for chlorine data. The solid line represents the maximum likelihood estimate of β_1 on time, the dotted line represents the estimate of the 25% percentile model, and the dotted line represents the estimate of the 75% percentile model. The gray areas are the 95% bootstrap confidence intervals for each percentile model.*

where Y is the fraction of available chlorine, X is the length of time, and ε is a random error. We assume that the error is a Gaussian noise to apply the AML method to investigate the percentile relationship between the covariates and the response.

Table 6 and Figure 4 show that the MLEs of the parameters of interest have the same values as the parameter estimate of the 50% quantile for the percentile model. The result is similar to that explained in Section 3.1. Figure 4 illustrates that the line corresponding to the MLEs divides the entire data point by half. For the 25% and 75% percentile models, β_0 is estimated to be almost identical to MLE, but β_1 is changed. Thus, in this example, the length of time (X) seems to have more influence on estimating the percentile model of available chlorine fractions.

4.2. Life Span Study cohort

We also apply the proposed method to the LSS cohort data from the Radiation Effects Research Foundation (RERF), which has conducted health-related research among atomic bomb survivors in Hiroshima and Nagasaki, Japan, for more than 70 years. The following data is from the time period of 1958–01998 and includes 111,952 people with 2,939,361 person-year, and the data consists of cases,

Table 7: Poisson nonlinear percentile estimates using a log-link function for Life Span Study data along with $\tau = 0.75, 0.80, 0, 85, 0.90$. The parentheses are the 95% bootstrap confidence intervals of each estimate. The ML is the MLE for each parameter

τ	β_1	β_2	γ_{Male}	γ_{Female}	ω	δ
Linear						
0.75	0.000 (0.000,0.000)	-	−0.479 (−0.502, −0.397)	−0.039 (−0.073,0.104)	−1.577 (−1.611, −1.556)	−0.126 (−0.195,0.017)
0.80	0.274 (0.167,0.39)	-	−0.509 (−0.733, −0.296)	0.154 (−0.057,0.358)	−1.729 (−2.377, −1.359)	−0.341 (−0.518, −0.16)
0.85	0.995 (0.897,1.097)	-	−0.139 (−0.262, −0.069)	0.197 (0.131,0.314)	−1.458 (−1.835, −1.008)	−0.063 (−0.138,0.005)
0.90	1.521 (1.44,1.612)	-	0.101 (0.028,0.197)	0.437 (0.334,0.496)	−1.254 (−1.659, −0.894)	−0.015 (−0.077,0.046)
ML	0.510	-	−0.371	0.132	−1.621	−0.185
Linear-Quadratic						
0.75	0.000 (0.000,0.000)	0.000 (0.000,0.011)	−0.456 (−0.587, −0.305)	0.001 (−0.015,0.271)	−1.582 (−2.05, −1.573)	−0.155 (−1.44, −0.075)
0.80	0.180 (0.068,0.324)	0.061 (0.004,0.124)	−0.502 (−0.728, −0.283)	0.166 (−0.065,0.364)	−1.723 (−2.271, −1.368)	−0.354 (−0.531, −0.165)
0.85	0.980 (0.87,1.082)	0.000 (0.000,0.065)	−0.151 (−0.264, −0.061)	0.232 (0.139,0.33)	−1.442 (−1.839, −1.028)	−0.061 (−0.139,0.005)
0.90	1.509 (1.436,1.624)	0.000 (0.000,0.000)	0.122 (0.017,0.189)	0.436 (0.341,0.505)	−1.292 (−1.648, −0.904)	−0.010 (−0.074,0.049)
ML	0.451	0.035	−0.362	0.150	−1.617	−0.188

person-year, city, gender, attained age, age at exposure, total weighted $DS02$ colon dose estimate, ground distance, etc. (Preston *et al.*, 2007). The considered model is the excess relative risk (ERR) model,

$$\begin{aligned}\lambda(\mathbf{x}) &= \lambda_0(\mathbf{x})\{1 + \rho(d)\epsilon(s, e, a)\} \\ &= \lambda_0(\mathbf{x})\left\{1 + \rho(d)\exp\left(\gamma_s + \omega\log\frac{a}{70} + \delta\frac{e-30}{10}\right)\right\}\end{aligned} \tag{4.2}$$

where c is the city in which the person was located at the time of the bombing, s is gender, a is the age at cancer incidence, e is the age at the time of the bombing, d is the exposed dose of radiation at the time of the bombing, and $\mathbf{x}$ is the covariate vector contained (c, s, a, e, d). In this model, $\lambda_0(\mathbf{x})$ is the baseline (or background) incidence rate of cancer in the unexposed population with characteristics $\mathbf{x}$, and $\rho(d)$ is a dose-response function that represents the main effect of radiation on ERR. Y is the observed solid cancer counts in the group, and PY is the person-year. It is known that the Y follows a Poisson distribution with a mean of $\lambda(\mathbf{x})PY$. The Poisson nonlinear percentile regression is applied using the linear and linear-quadratic dose-response functions in the ERR model.

Table 7 and Figure 5 illustrate that when $\tau \leq 0.75$, the estimates of the dose-response functions for men and women are zero, and the 95% bootstrap confidence interval of 75% ERR also contains zero. It can be seen that radiation exposure has little effect on the risk of cancer in people with a general percentile of 0.75. In a linear-quadratic dose-response model, the 95% bootstrap confidence interval of $\hat{\beta}_2$ in Table 2 also contains zero at $\tau = 0.85$ and $\tau = 0.90$. This means that at $\tau = 0.85$ or 0.90 the linear dose-response model is more suitable than the linear-quadratic dose-response model. The percentile rank $\hat{\tau}_w$ for the ERR model corresponding to the conditional mean from the MLE is approximately 0.82, which can be considered a reliable result compared with that of the percentile

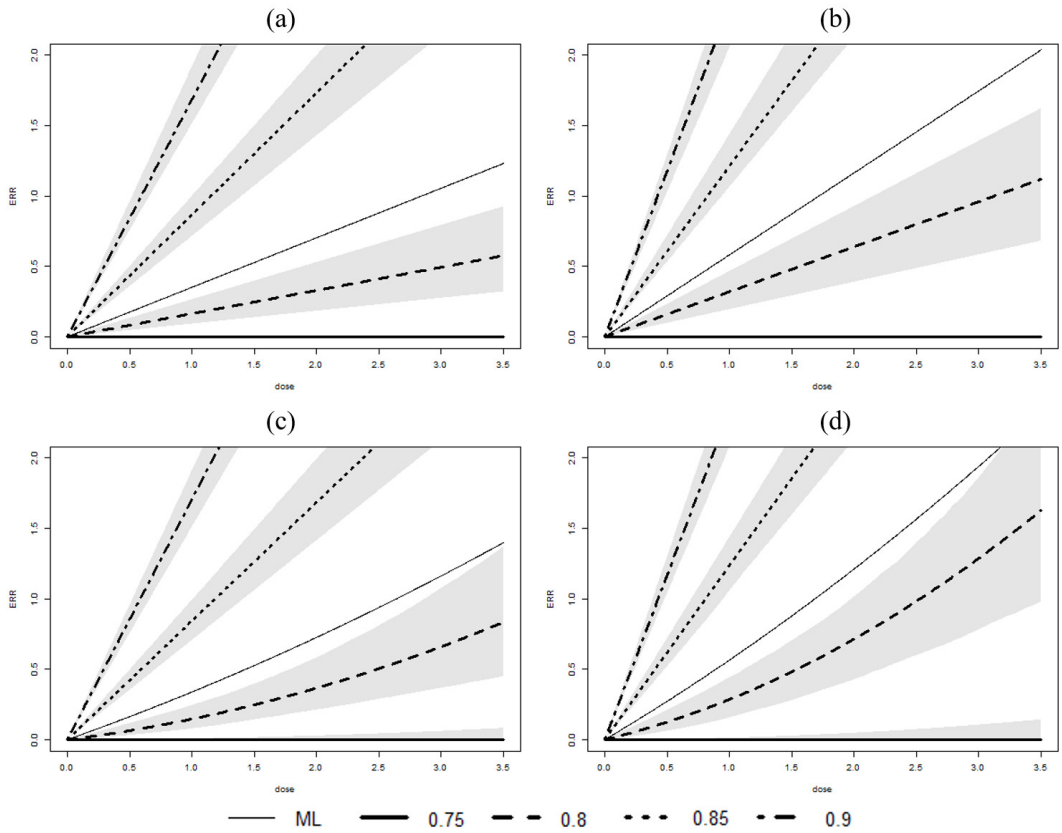

Figure 5: *Plots for nonlinear percentile estimates for Life Span Study data. The solid line represents the maximum likelihood estimate and the thick solid, dashed, dotted, and dot-dashed lines are the estimated excess relative risk (ERR) of each percentile models, such as* $\tau = 0.75, 0.80, 0.85, 0.90$. *The gray area is the 95% bootstrap confidence interval for each percentile model. The top two plots pertain to the linear dose-response function and the bottom plots pertain to the linear-quadratic dose-response function. The left two plots pertain to male participants and the right plots to female participants. (a) Estimated ERR for men with linear (b) Estimated ERR for women with linear (c) Estimated ERR for men with linear-quadratic (c) Estimated ERR for women with linear-quadratic.*

model.

5. Discussion and conclusions

We studied generalized nonlinear percentile regression using AMLE. We proposed estimating the percentile nonlinear parameters, and the algorithm of the parametric bootstrap was more powerful than the nonparametric bootstrap, which is advantageous for small samples which has sparsity. The simulation results show that the proposed method has comparable asymptotic performance, but the performance along with quantile τ depends on the variance of mean functions. In the chemical example, the length of time since the product was produced affects the estimation of the percentile of available chlorine fractions. In the LSS cohort example, we found the ERR percentile for atomic bombing survivors. The results illustrate that the ERR of women is higher than that of men.

The AMLE method is the estimation of percentile models minimizing the asymmetric version

of a deviance function. In this manuscript, we presented the algorithm for finding appropriate weight corresponding to the percentile rank τ. Since the proportion of the response values leads to the computation of the weight corresponding to τ, we must consider the possibility of estimating the regression percentile. The bisection method assumes the continuous function but $\hat{\tau}$ is the step function. It is an approximately smoothing function if there are enough data. However, if the conditional distribution of the response given the covariates is discrete, the algorithm might not converge. In particular, if the response has a Bernoulli distribution, there might be hard to optimize because the binary response has only two cases. Nonetheless, if there is another estimator of percentile which is robust at data size and the proportion of the zero-responses, our method will present better results in discrete type distribution.

The future topic related to the generalized percentile regression including the generalized nonlinear percentile regression is to investigate the approximate method for appropriate τ which can lead to overall optimization.

Acknowledgments

This research was supported by the Kyungpook National University Development Project Research Fund, 2018.

References

Austin PC, Tu JV, Daly PA, and Alter DA (2005). The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy, *Statistics in medicine*, **24**, 791–816.

Beyerlein A (2014). Quantile regression—opportunities and challenges from a user's perspective, *American journal of epidemiology*, **180**, 330–331.

Draper NR and Smith H (1981). *Applied Regression Analysis*, John Wiley & Sons.

Efron B (1991). Regression percentiles using asymmetric squared error loss, *Statistica Sinica*, **1**, 93–125.

Efron B (1992). Poisson overdispersion estimates based on the method of asymmetric maximum likelihood, *Journal of the American Statistical Association*, **87**, 98–107.

Eide E and Showalter MH (1998). The effect of school quality on student performance: A quantile regression approach, *Economics letters*, **58**, 345–350.

Geraci M and Bottai M (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution, *Biostatistics*, **8**, 140–154.

Geraci M (2019). Modelling and estimation of nonlinear quantile regression with clustered data, *Computational statistics & data analysis*, **136**, 30–46.

Grant EJ, Brenner A, Sugiyama H, *et al.* (2017). Solid cancer incidence among the life span study of atomic bomb survivors: 1958–2009, *Radiation Research*, **187**, 513–537.

He X, Ng P, and Portnoy S (1998). Bivariate quantile smoothing splines, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**, 537–550.

Karlsson A (2007). Nonlinear quantile regression estimation of longitudinal data, *Communications in Statistics-Simulation and Computation*, **37**, 114–131.

Koenker R and Bassett Jr. G (1978). Regression quantiles, *Econometrica: Journal of the Econometric Society*, **46**, 33–50.

Koenker R, Ng P, and Portnoy S (1994). Quantile smoothing splines, *Biometrika*, **81**, 673–680.

Koenker R and Park BJ (1996). An interior point algorithm for nonlinear quantile regression, *Journal*

of Econometrics, **71**, 265–283.
Koenker R and Hallock KF (2001). Quantile regression. *Journal of Economic Perspectives*, **15**, 143–156.
McCullagh P and Nelder JA (1989). *Generalized Linear Models(2ed.)*, London: Chapman and Hall.
Newey W and Powell J (1987). Asymmetric least squares estimation and testing, *Econometrica*, **55**, 819–47.
Preston DL, Ron E, Tokuoka S, *et al.* (2007). Solid cancer incidence in atomic bomb survivors: 1958–1998, *Radiation Research*, **168**, 1–64.
Rodrigo H and Tsokos C (2020). Bayesian modelling of nonlinear Poisson regression with artificial neural networks, *Journal of Applied Statistics*, **47**, 757–774.
Smith H and Dubey SD (1964). Some reliability problems in the chemical industry, *Industrial Quality Control*, **22**, 64–70.
Wang J (2012). Bayesian quantile regression for parametric nonlinear mixed effects models, *Statistical Methods & Applications*, **21**, 279–295.
Zietz J, Zietz EN, and Sirmans GS (2008). Determinants of house prices: a quantile regression approach, *The Journal of Real Estate Finance and Economics*, **37**, 317–333.

Received June 04, 2021; Revised September 01, 2021; Accepted October 04, 2021