Link function in semiparametric quantile estimation

Abstract
This paper studies the estimation of the link function of a broad family of asymmetric densities known as a generalized quantile-based asymmetric family, since the link function is a crucial component in quantile estimation. We proposed a link function and derived its distribution. Again, we explore the proposed distribution's distributional features and parameter estimates for unconditional and conditional cases. The estimator's asymptotic properties are also discussed here. To demonstrate the proposed methods for estimating the quantile function, an actual data application including the proportion of daily SARS-Cov-2 infected persons tested for COVID-19 infection and meteorological factors such as temperature and humidity is included. We discovered that the amount of daily SARS-Cov-2 infected people tested for COVID-19 infection is significantly influenced by temperature and humidity.
Keywords: Generalized quantile-based asymmetric family, Link function, Quantile estimation, COVID-19.

1 Background

Regression is one of the fundamental statistical tools that determines the strength and nature of the relationship between a set of response variables and a set of covariates. The mean regression focuses on an average relationship between a set of response variables and a set of covariates. It provides a single characteristic of a conditional distribution. It performs better results with nice mathematical properties for symmetric distribution (e.g., normal distribution). It is also not suitable when the data comes from the skewed distribution (see, for example, Gijbels et al. (2019b)). In regard to parameter estimation (or in general statistical inference and asymptotic properties) for a given distributions, a convenient class of families is the exponential family where the response variable Y given a covariate X is as follows

$f_{Y|X}(y|X = x) = \exp$

_

$(y\_(x)-b(\_(x))$

$a(\_) + c(y; \_)$

_

, (1)

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot, \cdot)$ are measurable functions (see, for example, Fan and Gijbels (1996)). The parameter function $\_(\cdot)$ is called the canonical parameter and $\_$ is a scale parameter.

$m(x)\ \_\ E(Y|X = x) = b_0(\_(x))$

1

$\text{var}(Y|X = x)\ \_\ \_(\_)b_0(\_(x))$

The function $g(b_0)_{-1}$, which links the mean regression function to the canonical parameter function $(b_0)_{-1}(m) = \_$ is called the canonical link.

However, the mean regression is highly influenced by extreme values. It is not usable when the quantile of the conditional distribution is the main interest (see, for example, Koenker (2005)).

Therefore, Komunjer (2005) provided tick exponential family, whose role in the conditional quantile estimation is analog to the role of the linear exponential family (1) in the conditional mean estimation. The general form of the tick exponential family for y 2 R is given by

$f_{\_}(y; \_) = \_(1 - \_)g0(y)$

8<

:

$\exp[-(1 - \_)(g(\_) - g(y))]$ if $y\ \_\ \_$

$\exp[\_(g(\_) - g(y))]$ if $y > \_.$

(2)

It is noted that the tick-exponential family (2) is only used for the whole real line continuous variable. It is not useful for boundary response variables. Beside of this, the asymmetric Laplace is the only member of this family.

On the other hand, Koenker and Bassett Jr (1978) proposed quantile regression which minimizes the tick function. It provides full characteristics of the distribution. This quantile regression is actually nonparametric because it does not need the underlying parametric assumption.

It is more robust to outliers than mean regression. It is the only regression tool
which is used for finding the effect of the covariate on different quantile level of the response variables. A nice discussion of quantile regression is presented by Koenker (2005). There are many problems that arise in nonparametric quantile regression due to the unknown underlying distribution. For example, crossing problem in quantile curves which leads to invalid inference, less efficiency etc.
Recently, Gijbels et al. (2019b) proposed the generalized quantile-based asymmetric family for any continuous variable *Y* which takes the form

*f*g_ (*y*; _, _) = 2_(1 − _)*g*0(*y*)

_
8<
:
f

_
(1 − _)

_
g(_)−*g*(*y*)
_

__
if *y* _ _
f

_
_

_
g(*y*)−*g*(_)
_

__
if *y* > _,
(3)

where _ is the location parameter and the _th quantile of *Y* and _ is the scale parameter. The speciality of family (3) is the location parameter(_) is a specific quantile of this family. There are many members of the family available in the literature such as asymmetric normal, asymmetric Laplace, asymmetric *t* and at least three big families that are subset of this family e.g, tick exponential (see, Gijbels et al. (2019a)), asymmetric power family (see, Gijbels et al. (2019a)), quantile-based asymmetric family (see, Gijbels et al. (2019a)). For any _ 2 (0, 1), the _th-quantile of *Y* equals

{*F*g
_}−1(_; _, _) =
8<
:
g−1

_
g(_) + _
1−_*F*−1

_
_
2_

__
if _ _ _
g−1

_
g(_) + _
_*F*−1

_
1+_−2_
2(1−_)

__
if _ > _,

with in particular $\{F^{g}_{\_}\}^{-1}(\_; \_, \_) = \_$. In the regression setting, the family (3) can be written as

2

$$f^{g}_{\_,Y|X}(y; \_(x), \_(x)|X = x) = \frac{2\_(1 - \_)g'(y)}{\_(x)}$$

8<

:

$$f\left((1 - \_)\frac{g(\_(x)) - g(y)}{\_(x)}\right) \text{ if } y \_ \_(x)$$

$$f\left(\_\frac{g(y) - g(\_(x))}{\_(x)}\right) \text{ if } y > \_(x),$$

(4)

where $\_(x)$ and $\_(x)$ are now the function of the covariate(s) x. In the setting of (4), the _th-conditional quantile function of Y given X = x (with (0 < _ < 1)) is then

$$\{F^{g}_{Y|X,\_}\}^{-1}(\_; \_(x), \_(x)|x) = g^{-1}(g(\_(x)) + \_(x).C_{\_}(\_))$$

where

$$C_{\_}(\_) = \frac{1}{1 - \_}F^{-1}\left(\frac{1 + \_}{2\_}\right)I(\_ < \_) + \frac{1}{\_}F^{-1}\left(\frac{1 + \_ - 2\_}{2(1 - \_)}\right)I(\_ \_ \_).$$

With F^{-1} the quantile function associated with the reference symmetric density f. The quantity $C_{\_}(\_)$ is known as a constant and is a monotonic function of _. The family (3) depends on two vital elements:

- the reference symmetric density f and
- monotone strictly increasing link function g.

When the link function is identity (i.e., $g(y) = y$) then family tends to quantile-based asymmetric family given in Gijbels et al. (2019a). In this study, the reference symmetric density f is assumed to be known. So the main focus is to estimate the link function g.

The link function is a crucial element in semiparametric quantile regression under the generalized quantile-based asymmetric family (see, Gijbels et al. (2019b)). It allows to explain any type of continuous response in terms of covariates. Besides, estimating the maximum likelihood estimator for unconditional setting and local likelihood estimator for conditional setting, the link function should be known (see, Gijbels et al. (2019b)). Usually we assume that, the link function in semiparametric quantile regression is known. For example, identity link, logit link, log link, canonical link, reciprocal link etc. But in real life data application, the link function is unknown. So it is very important to estimate link function. Alternatively, logit-type link function could be a solution. So, in this research, we focus on the study of different link functions in semiparametric regression.

The main contribution of the study is to estimate the link function of the generalized quantile-based asymmetric family. Section 2 proposed a logit-type link function and derive its distribution. The parameter estimation by using maximum likelihood estimation and real data application are illustrated in this section. The estimation of the logit-type link function in conditional settings is described in Section 3. The real data application is added to demonstrate the proposed methodology. The concluding remarks are added in the final section.

2 Logit-type Link Function

Let the density of Y is a member of the generalized quantile-based family of distributions, and G is a distribution function of Y. Suppose g be a logit-type link function of Y depends on G such that

$$g(Y) = \text{logit}(G(Y)) = \ln\left(\frac{G(Y)}{1 - G(Y)}\right). \quad (5)$$

If we know the distribution function of G, we can easily derive the logit-type link function by using (5). The link function of different probability distributions is presented in Table 2 in Appendix. The graphical presentation of the some link functions under some real-valued random variable and semi-infinite supported random variables are depicted in Figure 1. From Table 2, it is noticed that the link function of the logistic distribution is an "identity" function, i.e., $g(\_) = \_$ and Figure 1 (a) also confirm this.

Since _ is the _th quantile of Y and g is the monotone strictly increasing link function, then $g(\_)$ is also the _th quantile of $Z = g(Y)$ (see for example, Koenker (2005)). By introducing the _th quantile parameter μ 2 R and a scale parameter _ > 0 in the density (7), we get

$$f_{\_}(z;\mu,\_) = 2\_(1-\_)\begin{cases} \dfrac{e^{-\_\left(\frac{z-\mu}{\_}\right)}}{\left(1+e^{-\_\left(\frac{z-\mu}{\_}\right)}\right)^2} & \text{if } z > \mu \\ \dfrac{e^{-(1-\_)\left(\frac{\mu-z}{\_}\right)}}{\left(1+e^{-(1-\_)\left(\frac{\mu-z}{\_}\right)}\right)^2} & \text{if } z \_ \mu, \end{cases} \quad (6)$$

where $F^{-1}_{\_}(\_) = \mu$. The density given in (6) is denoted by ALD(μ, _, _) and called quantilebased asymmetric logistic density (ALD) proposed in Gijbels et al. (2019a). The graphical presentation of this distribution is displayed in Figure 2. From Figure 2, it is seen that the curves are unimodal and the densities are right-skewed (left-skewed) for the value of _ < 0.50 (for the value of _ > 0.50). The density is symmetric if and only if the index parameter

_ = 0.50.

2.1 Distribution of Logit-Type Link Function

We now want to find the distribution of $Z = g(Y)$ when Y _ G. We get

Pr(Z _ z) = Pr

_

ln

_

$G(Y)$

$1 - G(Y)$

_

_ z

_

= Pr

_

$G(Y)$

$1 - G(Y)$ _ e^z

_

= Pr

_

$G(Y)$ _

e^z

$1 + e^z$

_

[since, $G(Y)$ _ $U(0, 1)$]

$= e^z$

$1 + e^z = 1$

$1 + e^{-z}$; $-1 < z < 1$.

4

−10 −5 0 5 10

−40 −20 0 20 40

y

g(y)

Normal

cauchy

logistic

t dist

gumble

Laplace

0 2 4 6 8 10

−2 0 2 4 6 8 10

y

g(y)

exponential

gamma

weibull

lognormal

chi square

uniform

Figure 1: *Link function curve for (a). the real-valued random variable; (b). the semi-infinite supported random variable.*

This is the cumulative distribution function of a standard logistic distribution. The probability density function of Z can be written as

$f_Z(z) = d$

dz

_ 1

$1 + e^{-z}$

_

$= e^{-z}$

$(1 + e^{-z})^2$. (7)

The mean and variance of Y _ ALD(μ, _, _) are following:

$E(Z) = \mu + 2$_

h 1 − 2_

_(1 − _)

i

ln(2),

var(Z) = $\_^2$

_2(1 − _)2
h
(1 − 2_)2
n_2
3 − 4(ln(2))2
o
+ _2_(1 − _)
3
i
.
We can easily find the cumulative distribution function and the quantile function of *Z* which respectively are
*F*_(*z*) =
8>>><
>>>:
2_
1+exp
n
−(1−_)
□
z−*μ*
_
_o ; if *z* < *μ*
2_ − 1 + 2(1−_)
1+exp
n
−_
□
z−*μ*
_
_o ; if *z* _ *μ*.
,
and
F−1
_ (_) =
8<
:
μ − _
1−_ ln(2_
_ − 1) ; if _ < _
μ − _
_ ln(1−_
_−2_+1) ; if _ _ _.
Figure 3 depicts the cumulative distribution function (left panel) and the quantile function (right panel) when *Z* _ ALD(*μ*, _, _), for two values of _. Recall that *F*_(*μ*; *μ*, _) = _ and
F−1
_ (_) = *μ*.

−30 −20 −10 0 10 20 30
0.00 0.02 0.04 0.06 0.08 0.10 0.12
z
fα(z;0,1)
α = 0.15
α = 0.25
α = 0.5
α = 0.75
α = 0.85
Figure 2: The density plots of a quantile-based asymmetric logistic distribution with _ = (0.15, 0.25, 0.50, 0.75, 0.85)th Quantile of *μ* = 0 and _ = 1.

2.2 Maximum Likelihood Estimation

Based on an independent and identically distributed (i.i.d.) sample $Z_1, \dots, Z_n$ from Z _ ALD(μ, _, _) the likelihood function of _ = (μ, _, _)^T is defined by

$$L_n(\mu, \_, \_) = \left[\frac{2\_(1-\_)}{\_} \right]^n \prod_{i=1}^n \left[\frac{e^{-(1-\_)(\frac{\mu - Z_i}{\_})}}{\left(1+e^{-(1-\_)(\frac{\mu-Z_i}{\_})}\right)^2}\right]^{I(Z_i \_ \mu)} \times \left[\frac{e^{-\_(\frac{Z_i-\mu}{\_})}}{\left(1+e^{-\_(\frac{Z_i-\mu}{\_})}\right)^2}\right]^{I(Z_i>\mu)}.$$

And the log-likelihood function for _ = (μ, _, _)^T can be written as

$$\ln[L_n(\mu,\_,\_)] = n\ln[2\_(1-\_)] - n\ln(\_) - (1-\_)\sum_{i=1}^n \left(\frac{\mu - Z_i}{\_}\right) I(Z_i \_ \mu) - 2\sum_{i=1}^n \ln\left[1+e^{-(1-\_)(\frac{\mu-Z_i}{\_})}\right] I(Z_i \_ \mu) - \_\sum_{i=1}^n \left(\frac{Z_i-\mu}{\_}\right) I(Z_i>\mu) - 2\sum_{i=1}^n \ln\left[1+e^{-\_(\frac{Z_i-\mu}{\_})}\right] I(Z_i>\mu).$$

The MLE of μ, _ and _ is obtained from the optimization problem $\max_{\_2\_} \ln[L_n(\mu, \_, \_)]$.
We now can easily estimate _ by using the inverted link function _ = $g^{-1}(\mu)$ and hence can easily estimate the quantile function by using (1).

−10 −5 0 5 10 15 20 25
0.0 0.2 0.4 0.6 0.8 1.0

y
$F_{\alpha}(y)$
$F_{0.50}(y)$
$F_{0.25}(y)$
0.0 0.2 0.4 0.6 0.8 1.0
−10
−5
0
5
10
15
20
β
F_{α}
$^{-1}(\beta)$
$F_{0.50}$
$^{-1}(\beta)$
$F_{0.25}$
$^{-1}(\beta)$

Figure 3: *Cumulative distribution function (left) and quantile function (right) for* $\mu = 0$*,* _ = 1 *and* _ = (0.25*,* 0.50)*.*

Notably, this log-likelihood function is nonlinear and complex as well. It is also a nondifferentiable function at $y = \mu$. In this situation, we can estimate the parameter by using an algorithm offered by (Gijbels et al. (2019a)). They also implemented this algorithm in the R package QBAsyDist.

2.3 Real Data Application

For illustrative purposes, we consider the daily proportion of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infected people who have tested for Coronavirus disease (COVID-19) infection from August 3, 2020 to February 12, 2021 in Bangladesh. The number of daily new SARS-CoV-2 infected cases and daily new tested peoples are reported by the Institute of Epidemiology Disease Control and Research (IEDCR), Dhaka, Bangladesh. The data are available on the website with web-link https://covid19.who.int/. Notice that the daily proportion of SARS-CoV-2 infected people (Y) is a bounded variable with support [0*,* 1]. It is observed that on average each day, 13.61% of peoples are infected who have tested for COVID-19 infection.

z
Densities
−4 −2 0 2 4
0.0 0.1 0.2 0.3 0.4 0.5 0.6
0.0 0.2 0.4 0.6 0.8
−3 −2 −1 0
β
Quantile function of Z

Figure 4: *(a). Histogram and fitted density estimate (solid red line) of the uniform logittype transformation of the proportion of daily SARS-CoV-2 infected people (left); (b). the estimated quantile function (right) the uniform logit-type transformation of the proportion of daily SARS-CoV-2 infected people.*

For the bounded random variable, we can not directly compute the quantile function of

the distribution. Therefore, many authors including Bottai et al. (2010) and Columbu and Bottai (2016) used the uniform logit-type link function in quantile estimation. That is the link function is $Z = \text{logit}(G(Y))$, where $G(y) = (y - a)/(b - a)$. We also consider this link function to estimate the quantile function of the proportion of daily SARS-CoV-2 infected cases among the people who have tested for COVID-19 infection on that day. In Section 2, we have shown that the distribution of Z is a quantile-based asymmetric logistic distribution given in (6).

For the uniform logit-type link function for this data set, we consider a as the minimum proportion of infected people minus k and b as the maximum infected people plus k, where k is very small number. In this case, we use $k = 0.01$. To add (subtract) a small value of k to b (a) to avoid the zero value of denomination (numerator) in the logit-type link function. The resulting link function is $z = g(y) = \ln$

$\frac{y-a}{b-y}$

for y 2 (a, b). Using this link function $Z = g(Y)$, we estimate parameter _ = (μ, _, _)^T of the distribution of Z via the method of maximum likelihood estimation. The maximum likelihood estimates of b_ are $(-0.1467, 0.1824, 0.7234)^T$. Using these maximum likelihood estimates, we draw estimated density and estimate quantile function. The histogram of Z with estimated density is presented in Figure 4 (a). From the histogram, it is clear that the variable Z is left-skewed which also confirm by getting b_ = 0.7234 (>0.5). Based on the estimated density, the estimated quantile function is also depicted in Figure 4 (b). We now can easily estimate the quantile function of *Y* using the link function. In this case, $Y = (a + be^Z)/(1 + e^Z)$ and for any _ 2 (0, 1), the estimated _th-quantile of *Y* equals

{ b *Fg*
_}-1(_; _, _) =
8>>>>>><
>>>>>>:
(*a*+*b* exp
_
b*μ*- b_
1-b_
ln(2b_
_ −1)
_
)
1+exp
_
b*μ*- b_
1-b_
ln(2b_
_ −1)
_ if _ _ b_
a+*b* exp
_
b*μ*-b_b
_
ln(1-_
_-2b_+1
)
_
1+exp
_
b*μ*-b_b
_
ln(1-_
_-2b_+1
)
_ if _ > b_.

The estimated quantile curve of *Y* and the QQ-plot are presented in Figure 5. The QQplot looks like a curve, but actually, it is very close to a straight line. Because the scale of both axes is tiny, therefore it seems like a curve. Otherwise, it is very close to the 45_ line. This is confirmed by looking at the linear correlation coefficient of theoretical quantile and sample quantile which is 0.9863, indicating a nearly perfect relationship between theoretical quantile and sample quantile.

3 Estimation of Logit-type Link Function for Conditional Setting

Let the density of Z_i is the link function of quantile-based asymmetric family of logistic distribution and F is distribution function of response y_i.

Z_i = ln
 F(y_i)
1 − *F*(y_i)
_
; *where,*Z_i _ ALD(*μ*, _, _)

8

0.0 0.2 0.4 0.6 0.8
0.00 0.05 0.10 0.15 0.20 0.25
β
Quantile function of Y
0.00 0.05 0.10 0.15 0.20 0.25
0.00 0.05 0.10 0.15 0.20
Theoritical Quantiles of Y
Sample Quantiles of Y

Figure 5: *(a). Estimated quantile function of Y ; (b). The QQ-plot (right).*

In conditional setting, the link function Z_i is now function of covariate X_i takes the form,
$Z_i|X_i$ _ ALD($\mu(x)$, _(x), _(x)); *where,* $Z_i|X_i$ = ln
 $F(y|X)$
$1 - F(y|X)$
_
.
Where,
$\mu(X)$ 2 R, _(X) 2 R and _ 2 (0, 1).
The parameters are defined as,
_1(**X**) = **X**_**1** = μ(**X**) 2 R
_2(**X**) = **X**_**2** = ln _(**X**) 2 R
_3(**X**) = **X**_**3** = ln _(**X**)
1−_(**X**)
Where,
z =
2
6666664
Z_1
Z_2
...
Z_n
3
7777775
$n\times1$
; **X** =
2
6666664
1 X_{11} X_{21} . . .X_{p1}
1 X_{12} X_{22} . . .X_{p2}
...
...
...
...
1 X_{1n} X_{2n} . . .X_{pn}
3
7777775
$n\times(p+1)$
and _j =
2
6666666664
_$j0$
_$j1$
_$j2$
...
_jp
3
7777777775
$(p+1)\times1$
; j = 1, 2, 3.
9
We now turn to the regression setting involving one covariate. For conditional density of Z given X=x we consider the density $f_{Z|X}$(·; _1(x), _2(x)) in (6) and allow _1, _2 and index parameter _ depend on x. This leads to the conditional density
$f_{Z|X,\_3(\mathbf{x})}$(Z; _1(**X**), _2(**X**)) = C
8>>>>>><
>>>>>>:
exp

h
−
□
$e\_3(\mathbf{X})$
$1+e\_3(\mathbf{X})$
_□
$Z-e\_1(\mathbf{X})$
$e\_2(\mathbf{X})$
_i
h
1+exp
□
−(e_
3(X)
1+e_3(X))(Z−e_1(X)
e_2(X))
_i2 if $Z > e_{\_1}(\mathbf{X})$
exp
h
−
□
1
1+e_3(X)
_□
e_1(X)−Z
e_2(X)
_i
h
1+exp
□
−(1
1+e_3(X))(e_1(X)−Z
e_2(X))
_i2 if Z _ $e_{\_1}(\mathbf{X})$,
(8)
where,
$C = 2e_{\_3(x)}$
$(1 + e_{\_3(x)})2e_{\_2(x)}$
The conditional likelihood function for $\_(x) = (\mu(x), \_(x), \_(x))^T$ can be written as
$L(\_) =$
$\prod_{i=1}^{n}$
$f_{Z|X,\_3(\mathbf{x})}(Z; \_1(\mathbf{X}), \_2(\mathbf{X}))$
The conditional likelihood function for $\_(x) = (\mu(x), \_(x), \_(x))^T$ can be written as
$\ln L =$
$\sum_{i=1}^{n}$
$\log f_{Z|X,\_3(\mathbf{x})}(Z; \_1(\mathbf{X}), \_2(\mathbf{X}))$.
For simplification we use, $l = \ln L$
$l(\_1(\mathbf{X}), \_2(\mathbf{X}), \_3(\mathbf{X}); Z_i) = n \ln$
□ $2e_{\_3}(\mathbf{X})$
$1 + e_{\_3}(\mathbf{X})$
_
$- n\_2(\mathbf{X}) -$
1
$1 + e_{\_3}(\mathbf{x})$
$\sum_{i=1}^{n}$
$e_{\_1}(\mathbf{x}) - Z_i$
$e_{\_2}(\mathbf{x})$
$\mathbf{I}(Z_i$ _ $e_{\_1}(\mathbf{x})) - 2$
$\sum_{i=1}^{n}$
ln
_

1 + exp
□
−
$e_1(\mathbf{x}) - Z_i$
$(1 + e_3(\mathbf{x}))e_2(\mathbf{x})$

$\mathbf{I}(Z_i \_ e_1(\mathbf{x}))$ −
$e_3(\mathbf{x})$
$1 + e_3(\mathbf{x})$
$\sum_{i=1}^{n}$
$Z_i - e_1(\mathbf{x})$
$e_2(\mathbf{x})$
$\mathbf{I}(Z_i > e_1(\mathbf{x}))$
− 2
$\sum_{i=1}^{n}$
ln

1 + exp
□
−
$(Z_i - e_1(\mathbf{x}))e_3(\mathbf{x})$
$(1 + e_3(\mathbf{x}))e_2(\mathbf{X})$

$\mathbf{I}(Z_i > e_1(\mathbf{x}))$.
$l(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3,) = n \ln$
□ $2e^{\mathbf{x}_3}$
$1 + e^{\mathbf{x}_3}$

$- n\mathbf{X}_2 -$
1
$1 + e^{\mathbf{x}_3}$
$\sum_{i=1}^{n}$
$e^{\mathbf{x}_1} - Z_i$
$e^{\mathbf{x}_2}$
$\mathbf{I}(Z_i \_ e^{\mathbf{x}_1})$
− 2
$\sum_{i=1}^{n}$
ln

1 + exp
□
−
$e^{\mathbf{x}_1} - Z_i$
$(1 + e^{\mathbf{x}_3})e^{\mathbf{x}_2}$

$\mathbf{I}(Z_i \_ e^{\mathbf{x}_1})$
−
$e^{\mathbf{x}_3}$
$1 + e^{\mathbf{x}_3}$
$\sum_{i=1}^{n}$
$Z_i - e^{\mathbf{x}_1}$
$e^{\mathbf{x}_2}$

I(Z_i > ex_1)
− 2
X$_n$
i=1
ln
_
1 + exp

−
(Z_i − ex_1)ex_3
(1 + ex_3)ex_2
__
I(Z_i > ex_1).

3.1 Asymptotic Properties of Estimators

Differentiating the liklihood function with respect to _1, _2 and _3 we get,

_l
__1
= −K
X$_n$
i=1
ex_1 − Z_i
ex_2
− 2 _
__1
X$_n$
i=1
ln
_
1 + exp

−
ex_1 − Z_i
(1 + ex_3)ex_2
_
− Kex_3
X$_n$
i=1
Z_i − ex_1
ex_2
− 2 _
__1
X$_n$
i=1
ln
_
1 + exp

−
(Z_i − ex_1)ex_3
(1 + ex_3)ex_2
__
,
_l
__2
= −nX + K
X$_n$

$i=1$
$e^{x_1} - Z_i$
e^{x_2}
$- 2$ _
__2
X_n
$i=1$
ln
_
1 + exp
□
−
$e^{x_1} - Z_i$
$(1 + e^{x_3})e^{x_2}$
_
$+ Ke^{x_3}$
X_n
$i=1$
$Z_i - e^{x_1}$
e^{x_2}
$- 2$ _
__2
X_n
$i=1$
ln
_
1 + exp
□
−
$(Z_i - e^{x_1})e^{x_3}$
$(1 + e^{x_3})e^{x_2}$
__
,
_l
__3
$= 2Kn + Ke^{x_3}$
$(1 + e^{x_3})$
X_n
$i=1$
$e^{x_1} - Z_i$
e^{x_2}
$- 2$ _
__3
X_n
$i=1$
ln
_
1 + exp
□
−
$e^{x_1} - Z_i$
$(1 + e^{x_3})e^{x_2}$
_
$+ Ke^{x_3}$
$(1 + e^{x_3})$
X_n
$i=1$
$Z_i - e^{x_1}$

ex_2
− 2 _
__3
Xn
i=1
ln
_
1 + exp

−
(Zi − ex_1)ex_3
(1 + ex_3)ex_2
—
,
Now, the partial derivatives with respect to _1, _2 and _3 is given by,
_2l
__1__2
= KX
Xn
i=1
ex_1 − Zi
ex_2
− 2 _
__1_2
Xn
i=1
ln
_
1 + exp

−
ex_1 − Zi
(1 + ex_3)ex_2
_
+ KXex_3
Xn
i=1
Zi − ex_1
ex_2
− 2 _
__1_2
Xn
i=1
ln
_
1 + exp

−
(Zi − ex_1)ex_3
(1 + ex_3)ex_2
—
,
_2l
__1__3
= −KXex_3
1 + ex_3
Xn
i=1

$e^{x_1} - Z_i$
e^{x_2}
-2 _
___1_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$e^{x_1} - Z_i$
$(1 + e^{x_3})e^{x_2}$
_
$+ K\mathbf{X}e^{x_3}$
$1 + e^{x_3}$
$\sum_{i=1}^{n}$
$Z_i - e^{x_1}$
e^{x_2}
-2 _
___1_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$(Z_i - e^{x_1})e^{x_3}$
$(1 + e^{x_3})e^{x_2}$
__
,
_l
___2___3
$= -Ke^{x_3}$
$1 + e^{x_3}$
$\sum_{i=1}^{n}$
$e^{x_1} - Z_i$
e^{x_2}
-2 _
___2_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$e^{x_1} - Z_i$
$(1 + e^{x_3})e^{x_2}$
_
−
$K(e^{x_3})^2$
$1 + e^{x_3}$
$\sum_{i=1}^{n}$

$Z_i - e^{x_1}$
e^{x_2}
-2 _
___2_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$(Z_i - e^{x_1})e^{x_3}$
$(1 + e^{x_3})e^{x_2}$
__
,
11
_2l
___2**1**
$= -K\mathbf{X}$
$\sum_{i=1}^{n}$
$e^{x_1} - Z_i$
e^{x_2}
-2 _
___2_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$e^{x_1} - Z_i$
$(1 + e^{x_3})e^{x_2}$
_
$- K\mathbf{X}e^{x_3}$
$\sum_{i=1}^{n}$
$Z_i - e^{x_1}$
e^{x_2}
-2 _
___2_3
$\sum_{i=1}^{n}$
ln
_
1 + exp
□
−
$(Z_i - e^{x_1})e^{x_3}$
$(1 + e^{x_3})e^{x_2}$
__
,
_2l
___2**2**
$= -K\mathbf{X}$
$\sum_{i=1}^{n}$

ex_1 − Zi
ex_2
− 2 _
__22
Xn
i=1
ln
_
1 + exp

−
ex_1 − Zi
(1 + ex_3)ex_2
_
− KXex_3
Xn
i=1
Zi − ex_1
ex_2
− 2 _
__22
Xn
i=1
ln
_
1 + exp

−
(Zi − ex_1)ex_3
(1 + ex_3)ex_2
__
,
_2I
__23
= 2KnX
1 + ex_3 +
_KXex_3
1 + ex_3 + KX(ex_3)2
(1 + ex_3)2
_Xn
i=1
ex_1 − Zi
ex_2
− 2 _
__22
Xn
i=1
ln
_
1 + exp

−
ex_1 − Zi
(1 + ex_3)ex_2
_
+
_KXex_3

1 + e^{x_3} + $K\mathbf{X}(e^{x\_3})^2$
$(1 + e^{x\_3})^2$

_
$\sum_{i=1}^{n}$
$Z_i - e^{x\_1}$
e^{x_2}
− 2 _
__2**2**
$\sum_{i=1}^{n}$
ln

_
1 + exp
□
−
$(Z_i - e^{x\_1})e^{x\_3}$
$(1 + e^{x\_3})e^{x\_2}$

__
,
where, $K = x$
$1+e^{x\_3}$.
It can be shown that, for large sample
2
6664
_ˆ1
_ˆ2
_ˆ3
3
7775
_ N(_,_−1),
where,
_ =
2
6664
_2l
__21
_2l
__1__2
_2l
__1__3
_2l
__1__2
_2l
__22
_2l
__2__3
_2l
__1__3
_2l
__2__3
_2l
__23 3
7775
.
After differentiating the conditional log liklihood function, we have noticed that the derivatives are non linear and complex. That's why, for estimating parameter we have used a R package *QBAsyDist* which is introduced by Gijbels et al. (2019a). In Section 3.2 we have applied a real data for illustrative purpose.

3.2 Real Data Application

Data described in Section 2.3 are also applied in this section. The daily proportion (Y) of

SARS-CoV-2 infected people who have tested positive for COVID-19 infection is considered a response variable, and the daily temperature (X_1) and humidity (X_2) are considered covariates from August 3, 2020 to February 12, 2021 for illustrating the proposed method. The data

12

of daily temperature humidity is available on the website https://www.timeanddate.com/weather/bangladesh/dhaka. In this case, the parametric functions can be written as

$$\_{1}(\mathbf{X}) = \mathbf{X\_1} = \_{10} + \_{11}X_1 + \_{12}X_2,$$

$$\_{2}(\mathbf{X}) = \mathbf{X\_2} = \exp(\_{20} + \_{21}X_1 + \_{22}X_2),$$

$$\_{3}(\mathbf{X}) = \mathbf{X\_3} = \frac{\exp(\_{30} + \_{31}X_1 + \_{32}X_2)}{1 + \exp(\_{30} + \_{31}X_1 + \_{32}X_3)}.$$

Table 1: The summary statistics of the estimators for estimating $\mu(X; \_1)$, $\_(X; \_2)$ and $\_(X; \_3)$ and the p-values obtained by using Bootstrapping.

$\mu(X; \_1)$			$\_(X; \_2)$			$\_(X; \_3)$		
$\_1$	b_1 ($\overline{se}(b\_1)$)	$\overline{P}$-value	$\_2$	b_2 ($\overline{se}(b\_2)$)	$\overline{P}$-value	$\_3$	b_3 ($\overline{se}(b\_3)$)	$\overline{P}$-value
$\_{10}$	2.3414(0.3994)	<0.005	$\_{20}$	1.0759(0.3160)	0.006	$\_{30}$	0.3947(0.5239)	<0.0001
$\_{11}$	0.7312(0.6943)	<0.0001	$\_{21}$	0.02031(0.2076)	<0.0001	$\_{31}$	0.0641(0.2431)	<0.0001
$\_{12}$	0.7514 (0.4129)	0.008	$\_{22}$	-0.0901(0.2836)	0.064	$\_{32}$	0.0166(0.2844)	0.424

The estimated parametric functions can be written as

$$b\mu(X_i; b\_1) = 2.3414 + 0.7312X_1 + 0.7514X_2,$$

$$b\_(X_i; b\_2) = \exp(1.0759 + 0.02031X_1 - 0.0901X_2),$$

$$b\_(X_i; b\_3) = \frac{\exp(0.3947 + .0641X_1 + 0.0166X_2)}{1 + \exp(0.3947 + 0.0641X_1 + 0.0166X_3)}.$$

Table 1 shows the summary statistics of the estimated models. Regression coefficients for the temperature and humidity significantly impact the daily proportion of infected cases for the estimated function $b\mu(X, b\_1)$. For estimated $b\_(X, b\_2)$, we see the regression coefficients for only temperature is statistically significant. Similarly, for $b\_(X, b\_3)$, we observe that temperature is highly statistically significant but humidity is not at 5% level of significance.

4 Concluding Remarks

In this research, we study the theory of quantile regression using a generalized quantile-based asymmetric family of densities. We provide the theory of logit-type link function for estimating quantile function in unconditional and conditional settings. For unconditional settings, the distribution and properties of the logit-type link function are addressed. The proposed theory is illustrated in real data application.

In conditional settings, we consider the response variable is the proportion of daily SARSCov-2 infected people tested for COVID-19 infection and two covariates: temperature and

13

humidity. Again, this research covers the quantile estimation by using logit-type link function in conditional settings. A real-world data application employing the fraction of daily SARSCov-2 infected people tested for COVID-19 infection is included to demonstrate the proposed methods for calculating the quantile function. We noticed that the temperature and humidity have a significant impact on the proportion of daily SARS-Cov-2 infected persons tested for COVID-19 infection.

Declarations

Data and code availability

Sources of data are mentioned in the text. Codes and data will be provided if needed.

Funding statement

This work was funded by Faculty of Mathematical and Physical Sciences, Jahangirnagar University.

Conflicts of interest

The authors declare no conflict of interest.

References

Bottai, M., Cai, B. and McKeown, R. E. (2010). Logistic quantile regression for bounded outcomes, *Statistics in Medicine* **29**(2): 309–317.

Columbu, S. and Bottai, M. (2016). Logistic quantile regression to model cognitive impairment in Sardinian cancer patients, *in* T. Di Battista, M. E. and R. W. (eds), *Topics on Methodological and Applied Statistical Inference*, Springer, pp. 65–73.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall/CRC.

Gijbels, I., Karim, R. and Verhasselt, A. (2019a). On quantile-based asymmetric family of distributions: Properties and inference, *International Statistical Review* **87**(3): 471–504.

Gijbels, I., Karim, R. and Verhasselt, A. (2019b). Quantile estimation in a generalized asymmetric distributional setting, *in* A. Steland, E. Rafajlowicz and O. Okhrin (eds), *Stochastic Models, Statistics and their Application*, Springer, pp. 13–40.

Koenker, R. (2005). *Quantile Regression*, New York: Cambridge University Press.

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

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles, *J of Econometrics* **128**(1): 137–164.

5 Appendix

Table 2: Link function for different distribution

Distribution
name
Link function Support
Normal
g(*y*) = log
_1 + erf(*y* /
p2)
1 − erf(*y* /
p2)
_ *y* 2 R
Exponential
g(*y*) = log
□1 − exp(−*y*)
exp(−*y*)
_ *y* 2 [0,1)
Gamma
g(*y*) = log
□ □*y*()
− □*y*()
_ *y* 2 (0,1)
Cauchy
g(*y*) = log
_0.5 −
arctan(*y*)
_
0.5 −
arctan(*y*)
_
_ *y* 2 (−1,1)
Weibull
g(*y*) = log
□1 − exp(*y*)
exp(*y*)
_ *y* 2 [0,1)
Gumble

$g(y) = \log$
exp[−exp(−y)]
1 − exp[−exp(−y)]
_ y 2 R
Lognormal
$g(y) = \log$
_ _(ln(y) / _
_
1 − _
ln(y) / _)
_ y 2 (0,1)
Logistic
$g(y) = y$
y 2 (−1,1)
Laplace
$g(y) =$
8<
:
log
1
2 exp(y)
1−1
2 exp(y)
_
if y _ 0
log
1−1
2 exp(y)
12
exp(y)
_
if y _ 0
y 2 R
15
Chi-square
$g(y) = \log$
_ 1
12
(12
, y
2)
1 − 1
12
(1
2 , y
2)
_
y 2 (−1,1) if
k=1, otherwise y 2
[−1,1)
Student-t
$g(y) = \log$
_12
+ y
12
_ 2F1(12
,12
;32

;−y2)
p
_ □____________(12
)
12
− y□
□12
_ 2F1(12
,12
;32
;−y2)
p
_ □(12
)
_
y 2 (−1,1)
Uniform
g(y) = log
□
y / (1 − y)
_ 0 _ y _ 1
16