



CPS Poster

A parametric quantile beta regression for modeling case fatality rates of COVID-19

Author: DRS Marcelo Bourguignon Pereira

Coauthors: Marcelo Bourguignon Pereira, Diego I. Gallardo, Helton Saulo

Submission ID: 361

Reference Number: 361

Presentation Image

**A parametric quantile beta regression for modeling case fatality rates of COVID-19**  
 Marcelo Bourguignon<sup>1</sup> Diego I. Gallardo<sup>2</sup> Helton Saulo<sup>3</sup>  
<sup>1</sup>Federal University of Rio Grande do Norte <sup>2</sup>University of Atacama <sup>3</sup>University of Brasilia

**An important problem**

The new coronavirus respiratory syndrome disease (COVID-19) pandemic has affected several countries around the world. In particular, the COVID-19 pandemic has hit Chile hard in the past few months. Some authors have studied different aspects of COVID-19 in Chile. In practice, research in this context can be useful in an epidemic, especially because they provide the basis for making decisions by policymakers, e.g. directing health care resources to certain areas or identifying how long social distancing policies may need to be in effect (Vieira-Sackey and Sark, 2021).

This work focuses on a measure of COVID-19 mortality risk (especially in Chile) known as case fatality rate (CFR). Such a measure is particularly important for health policies, and it is computed as the ratio between confirmed deaths and confirmed cases. It is one of the indicators that serve to monitor the severity of the pandemic; see <https://ourworldindata.org/>. In terms of stochastic representation, the CFR can be written as

$$Y = \frac{W}{W+V} \quad (1)$$

where  $W \in \mathbb{R}^1$  and  $V \in \mathbb{R}^1$  are two random variables representing the number of confirmed COVID-19-related deaths and COVID-19 cases with no death result, respectively. The sum  $W+V$  represents the number of confirmed COVID-19 cases, and the ratio (1) has support in the unit interval (0, 1). Nevertheless, it is well known that the widely popular mean regression model could be inadequate if the observed response variable follows an asymmetrical distribution, which is quite common for rates and proportions. In such a situation, quantile regression models (Koenker, 2005) can be more suitable. To the best of our knowledge, a specific parametric quantile regression model to describe data of the type  $Y = \frac{W}{W+V}$  at different levels has never been considered in the literature.

**Chilean COVID-19 data**

We analyze COVID-19 data from Chile and find that the CFR is explained by population density, positivity for tests, and the percentage of the population fully vaccinated. Chile is administratively divided into 348 communes, which belong to 10 regions. We considered the cases and deaths of COVID-19 reported by the Chilean Ministry of Health for each commune between May 24 and July 23, 2021, but only for communes with at least 50 cases and 1 death, totaling 335,727 cases and 8,492 deaths in 305 communes. The communes were divided into three zones: North, Center, and South (see Figure 1). This division is typically used in Chile and is supported by climatological characteristics from the different zones.

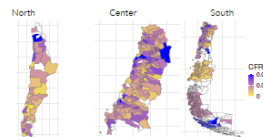


Figure 1. Averages of the observed CFR values in communes of the North, Center, and South zones of Chile between May 24 and July 23, 2021.

**The proposed quantile regression model**

A random variable  $Y$  follows the generalized beta distribution (Lobry and Novick, 1982) with three parameters:  $\alpha > 0, \beta > 0$ , and  $\lambda > 0$ , denoted by  $Y \sim GB3(\lambda, \alpha, \beta)$ , if its cumulative distribution function (CDF) is given by

$$F_Y(y; \lambda, \alpha, \beta) = I_{\lambda y(1+y)}(\alpha, \beta), \quad 0 < y < 1, \quad (2)$$

where  $I_x(a, b) = B_x(a, b) / B(a, b)$  is the incomplete beta function ratio,  $B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$  is the incomplete function and  $B(a, b) = B(a, b)$  is the usual beta function. The probability density function (PDF) associated with Equation (2) is

$$f_Y(y; \lambda, \alpha, \beta) = \frac{\lambda y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta) (1-\lambda y)^{\alpha+\beta}}, \quad 0 < y < 1. \quad (3)$$

We can see that Equation (3) reduces to the beta distribution when  $\lambda = 1$ . If  $X_1 \sim GA(\alpha, \theta)$  and  $X_2 \sim GA(\beta, \theta)$  are independent gamma distributions, then the random variable

$$Y \stackrel{d}{=} \frac{X_1}{X_1 + X_2} \quad (4)$$

is distributed according to a GB3 distribution, where  $\lambda = \theta/\beta$ . Note that, as  $I_x(a, b)$  corresponds to the CDF of the usual beta distribution, the  $r$ -th quantile of the GB3 distribution can be represented as

$$q(r; \lambda, \alpha, \beta) = \frac{t_{\alpha, \beta}(r)}{\lambda [1 - t_{\alpha, \beta}(r)] + t_{\alpha, \beta}(r)}, \quad 0 < r < 1, \quad (5)$$

where  $t_{\alpha, \beta}(r)$  denotes the  $r$ -th quantile of the beta distribution with parameters  $\alpha$  and  $\beta$ . To compute the quantile  $t_{\alpha, \beta}(r)$  as required, it is not necessary to implement such complex formulae as (5). Instead, one can use the `qbeta()` function in R software. To introduce the proposed quantile regression model, we shall reparameterize (5) in terms of the  $r$ -th quantile  $\mu = q(r; \lambda, \alpha, \beta)$  such that  $\lambda$  can be written from Eq. (4) as follows

$$\lambda = \frac{1-\beta}{\mu} \frac{t_{\alpha, \beta}(r)}{1 - t_{\alpha, \beta}(r)}, \quad 0 < \mu < 1.$$

Let  $Y_1, \dots, Y_n$  be independent random variables such that each  $Y_i, i = 1, \dots, n, Y_i \sim GB3(\mu_i, \theta_i, \beta_i)$ , for a fixed (known) probability  $r \in (0, 1)$  associated with the quantile of interest. Suppose  $\mu, \alpha$ , and  $\beta$  satisfy the following functional relations:

$$g(\mu(r)) = x_1^T \theta(r), \quad g(\alpha(r)) = x_2^T \eta(r), \quad \text{and} \quad g(\beta(r)) = w_1^T \eta(r), \quad (6)$$

where  $\theta(r) = (\theta_1(r), \dots, \theta_k(r))^T, \mu(r) = (\mu_1(r), \dots, \mu_p(r))^T$ , and  $\eta(r) = (\eta_1(r), \dots, \eta_m(r))^T$  are the vectors of the unknown regression coefficients; and  $x_1 = (x_{11}, \dots, x_{1k})^T, x_2 = (x_{21}, \dots, x_{2p})^T$ , and  $w_1 = (w_{11}, \dots, w_{1m})^T$  are the observations of the  $k, l$ , and  $m$  known regressors. The link functions  $g: (0, 1) \rightarrow \mathbb{R}, g: \mathbb{R}^1 \rightarrow \mathbb{R}$ , and  $g: \mathbb{R}^1 \rightarrow \mathbb{R}$  in (6) must be strictly monotone, positive, and at least twice differentiable.

**Analysis results**

We considered the following information for each commune:

- cfr: CFR (confirmed deaths/confirmed cases) between May 24 and July 23, 2021. Mean=0.026, Median=0.025, standard deviation=0.012, minimum=0.004, and maximum=0.099.
- dens: population density of the commune (in km<sup>2</sup>) according to the projected population of the Statistics National Institute of Chile for 2020.
- posit: average of positive tests between January 25 and May 24, 2021.
- vaccinat: percentage of the population fully vaccinated by May 24, 2021.

We considered modeling the  $r$ -th quantile of the CFR, say  $cfr(r)$ , such that  $cfr(r) \sim GB3(\mu(r), \alpha(r), \beta(r))$ . We considered the quantiles (0.1, 0.25, 0.50, 0.75, 0.90) and used the logit link (because it is the traditional link used). The method of maximum likelihood (ML) is used for estimating the model parameters. Considering that the terms related to the intercept,  $\log(\text{dens})$ ,  $\text{posit}$ , and  $\text{vaccinat}$ , and based on the results of AIC criteria, the final model was

$$g(\mu(r)) = \theta_0(r) + \theta_1(r) \log(\text{dens}) + \theta_2(r) \text{vaccinat},$$

$$g(\alpha(r)) = \eta_0(r) + \eta_1(r) \log(\text{dens}) + \eta_2(r) \text{posit}, \quad g(\beta(r)) = \eta_3(r).$$

Table 1 presents the ML estimates and standard errors (SEs) for the GB3 quantile model parameters. We observe that the estimates of  $\theta_j(r)$  increase across  $r$ , implying that the  $r$ -th quantile of CFR increases as  $r$  increases for the case when  $\log(\text{dens})$  and  $\text{vaccinat}$  are fixed at 0. On the other hand, the estimates of  $\eta_j(r)$  decrease across  $r$ , i.e., the  $r$ -th quantile of CFR decreases as  $\log(\text{dens})$  increases. This is expected because denser cities have greater and better access to health centers. Furthermore, the  $r$ -th quantile of CFR decreases as  $\text{vaccinat}$  increases.

Table 1. Estimated parameters and SEs (given in parentheses) for the GB3 quantile model.

Parameter	r = 0.10			r = 0.25			r = 0.50			r = 0.75			r = 0.90			
	Est.	SE	CI	Est.	SE	CI	Est.	SE	CI	Est.	SE	CI	Est.	SE	CI	
$\theta_0(r)$	-0.0000 (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)
$\theta_1(r)$	0.1109 (0.0211)	0.0276 (0.0077)	0.0619 (0.0089)	0.0476 (0.0089)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)	0.0178 (0.0087)
$\theta_2(r)$	0.0103 (0.0042)	0.0050 (0.0015)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)	0.0108 (0.0043)
$\eta_0(r)$	1.2410 (0.2228)	0.2243 (0.0722)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)	1.7041 (0.2412)
$\eta_1(r)$	0.1716 (0.0246)	0.0141 (0.0071)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)	0.1318 (0.0246)
$\eta_2(r)$	1.6053 (0.1782)	0.2080 (0.0778)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)	1.2882 (0.1782)
$\eta_3(r)$	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)

Figure 2 shows the predicted median in the two last months for the CFR in the three zones if the vaccination process is increased by 30% in each commune (labeled at 100%). Comparing Figures 1 and 2, we observe a large decrease in the predicted median for the average CFR in the two last months, showing that the vaccination is an effective way to reduce COVID-19 deaths.

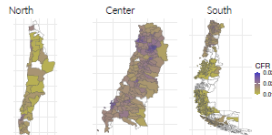


Figure 2. Estimated median for the average CFR in the two last months in the communes of the North, Center, and South zones of Chile if the vaccination process was augmented in 30%.

**References**

[1] D. I. Lobry and M. R. Novick. Multivariate generalized beta distributions with applications to ability assessment. *Journal of Educational Statistics*, 7, 1982.

[2] P. A. Vieira-Sackey and R. Sark. The impact of Covid-19 pandemic on the global economy: Emphasis on growth inhibition and economic growth. *The Economic and Financial Crisis*, 10(01):22-43, 2021.

Brazil 64th ISI World Statistics Congress - Ottawa, Canada m.pourguignon@gmail.com

Files/Uploads

Poster\_pdf

Brief Description

Motivated by the case fatality rate of COVID-19, we develop a fully parametric quantile regression model based on the generalized three-parameter beta (GB3) distribution.

We first reparameterize the GB3 distribution by inserting a quantile parameter and then we develop the new proposed quantile model.

We also propose a simple interpretation of the predictor-response relationship in terms of percentage increases/decreases of the quantile.

A real COVID-19 dataset from Chile is analyzed and discussed to illustrate the proposed approach.

Abstract

ISI - International Statistical Institute
ISI Permanent Office, P.O. Box 24070, 2490 AB The Hague, The Netherlands
info@isi2023.org

Motivated by the case fatality rate (CFR) of COVID-19, in this paper, we develop a fully parametric quantile regression model based on the generalized three-parameter beta (GB3) distribution. Beta regression models are primarily used to model rates and proportions. However, these models are usually specified in terms of a conditional mean. Therefore, they may be inadequate if the observed response variable follows an asymmetrical distribution, such as CFR data. In addition, beta regression models do not consider the effect of the covariates across the spectrum of the dependent variable, which is possible through the conditional quantile approach. In order to introduce the proposed GB3 regression model, we first reparameterize the GB3 distribution by inserting a quantile parameter and then we develop the new proposed quantile model. We also propose a simple interpretation of the predictor-response relationship in terms of percentage increases/decreases of the quantile. A Monte Carlo study is carried out for evaluating the performance of the maximum likelihood estimates and the choice of the link functions. Finally, a real COVID-19 dataset from Chile is analyzed and discussed to illustrate the proposed approach.