# Length-of-Stay of Hospitalized COVID-19 Patients Using Bootstrap Quantile Regression

Ferra Yanuar, Aidinil Zetra

***Abstract*—The purpose of this study is to identify the best model of the length-of-stay (LoS) hospitalized for patients with COVID-19 in West Sumatra, Indonesia. The LoS data is skewed to the right or violates linear model assumptions; thus, a quantile approach is employed. The asymptotic variance of quantile regression is estimated by constructing the confidence interval for the parameter of interest. This study will compare the result based on five different methods in resampling bootstrap. This study proves that wild bootstrap quantile tends to produce the shortest confidence interval. This study found that Diagnosis and Final outcome are statistically significant to give impact to the LoS hospitalized COVID-19 patients.**

***Index Terms*—Length-of-stay (LoS) hospitalized, quantile regression, wild bootstrap**

## I. Introduction

The Covid 19 pandemic has spread across the world since the beginning of 2020. In Indonesia, the COVID-19 confirmed cases had been reported rapidly increasing. The COVID-19 pandemic has also placed an unprecedented strain on health systems, with rapidly escalating healthcare demand in hospitals. As the pandemic increase, determining or predicting demand for hospital services (beds, staff, equipment) has become a key priority anywhere. Predicting demand for hospital services requires estimating how long each person will require hospital care and estimate the number of patients requiring hospitalization [1]. Therefore, it is important to determine the factors that affect the length-of-stay (LoS) hospitalized for patients with COVID-19. The most common approach for determining these causal effect is multivariate linear regression [2], [3]. However, this method is not the optimal approach due to some limitations such as requiring the normality assumption or homogeneity of data variance and inability to properly managed the outlier values [4], [5].

Manuscript received April 17, 2021; revised August 2, 2021. This research was funded by Directorate of Resources Directorate General of Higher Education, Ministry of Education, Culture, and Research and Technology of Indonesia, in accordance with Contract Number 104/E4.1/AK.04.PT/2021.

Ferra Yanuar (Corresponding Author) is a lecturer at Mathematics Department, Faculty of Mathematics and Natural Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. e-mail: ferrayanuar@sci.unand.ac.id

Aidinil Zetra, is a lecturer at Political Science Department, Faculty of Social and Political Sciences, Andalas University, Kampus Limau Manis, Padang, Indonesia, 25163. e-mail: aidinil@soc.unand.ac.id

Quantile regression is one of the methods that can overcome these limitations. The statistical procedure in this modeling technique is based on minimizing sums of asymmetrically weighted absolute residuals [6]. This method can be utilized to explore the relationship between quantiles of the response distribution and available covariates. Comparing such quantiles for a range of $\tau$ values enables researchers to obtain a more complete picture of the conditional distribution than mean regression [7]–[9]. Quantile regression allows researchers to explore a range of conditional quantile functions, thereby exposing various forms of conditional heterogeneity, and controlling for unobserved individual characteristics [10]. Many studies have been done to develop the asymptotic properties of these estimators. In particular, the asymptotic variance of quantile regression estimators depends on the density of a term which is not easy to compute in practice [7], [11], [12]. Several studies then considered to employ the Bootstrapping resampling methods to estimate the asymptotic variance for quantile regression that is easier to use. Yanuar et al [9] used the pairwise bootstrap to model the low birth weight based on a quantile approach. Parzen, Wei, and Ying [13] proposed a general and simple resampling method for inferences $\boldsymbol{\beta}$ based on pivotal estimating functions for quantile regression estimator. The quantile bootstrap using the Markov chain marginal bootstrap (MCMB) was proposed by He and Hu [14]. Meanwhile, Kocherginsky, He and Mu [15] did a modification of the Markov chain marginal bootstrap (MCMB) to construct confidence intervals in quantile regression. Bose and Chatterjee [16] proposed a generalized bootstrap technique for estimators obtained by minimizing functions that are convex in the parameter. They established the consistency of these schemes via representation theorems. Feng, He and Hu [17] proposes an adaptation wild bootstrap methods for quantile regression to achieve the best model.

In a preliminary analysis, we obtained that the distribution of error is not normal and not homogeneous due to any outliers in the data [18]. Therefore, we consider to implement the quantile regression method to describe some relationships between independent variables on selected quantiles of the response variable [5]–[7]. The asymptotic variance of the quantile regression estimator is estimated based on bootstrap resampling methods [20]. The smallest confidence interval for parameter estimated will be determined based on different five methods in bootstrap resampling methods. Accordingly, this study focuses on

modeling the LoS hospitalized for patient with COVID-19 using a comparison of the resampling bootstrap method based on quantile regression.

The rest of the paper is organized as follows. In Section 2, we present model formulation, data and methods implemented in this study. Section 3 contains the modeling of the LoS hospitalized for patients with COVID-19 in West Sumatera by employing five different methods of bootstrap quantile approaches. Finally, brief conclusions are given in Section 4.

## II. Materials and Methods

### A. Materials

The indicators for the LoS hospitalized are constructed based on previous studies. Epidemiological studies have shown that mortalities are higher in the elderly population, and the incidence is much lower in children [19], [20]. Yuki et al. [21] identified that elderly patients were more susceptible to longer LoS than younger. Many studies also investigated that the presence of hypertension, diabetes, and coronary artery disease were considered as hazard factors for confirmed COVID-19 patients [22], [23]. While, Gebhard et al. demonstrated that Covid-19 is deadlier for infected men than women [24]. Study by Cheda et al, [25] considered that gender, age and final outcome were keys for accurately forecasting the LoS in a case study in Galicia (Spain). Besides, Thai et al. [26] proved that patients with positive COVID-19 results tent to have longer length of hospital stay than patients had no symptoms.

Based on previous researchers, this present study assumes Age, Gender, Comorbidities, Diagnosis, and Final outcome as indicators for affecting the LoS hospitalized for patients with COVID-19. The response variable is LoS recorded in days. Age is a continuous type recorded in years. Comorbidities are how many diseases that each patient has represented in a continuous variable. Gender is recorded in males and females. Diagnosis is categorized into four types, i.e., Asymptomatic Person (denoted as Asymp), Person Under Supervision (denoted as PerUS), Patience Under Supervision (denoted as PaUS), and Confirmed. Meanwhile, the categories for Final outcome are Recovered, Died, Force discharge, Referred to other hospitals, and Outpatients.

This study uses the documented data obtained from several hospitals in West Sumatra, Indonesia, from March to October 2020. All 688 patients who were hospitalized which various Diagnosis-related COVID-19 were involved. The patients were diagnosed as Asymptomatic Person (10 patients or 1.5%), Person Under Supervision or PerUS (16 patients or 2.3%), Patience Under Supervision or PaUS (584 patients or 84.8%), and Confirmed (77 patients or 11.2%).

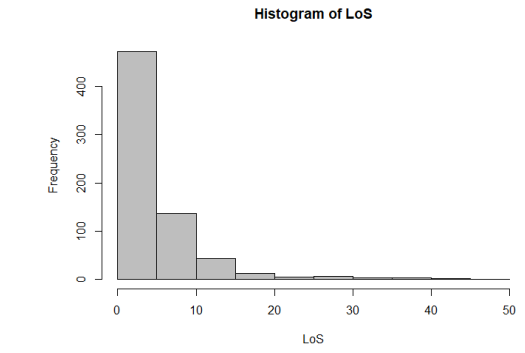
Based on gender, 347 patients, or 50.4%, are Male and 342 patients, or 49.6%, are Female. Based on the residential address, 218 of them (31.7%) are from a district in West Sumatera, 410 patients (59.6%) are living in a city in West Sumatera, and 60 patients (8.7%) among them came from West Sumatera but at the survey hold they stay not in West Sumatera. Characteristic patients based on occupation are Housewife (25.7%), Kids (18.2%), Entrepreneur (14.1%), Others (10.9%), Private and Students each are 10.5%, and 10.2% among patients are Civil servant. For Final outcome, 73.3% among them are Recovered, about 20.5% are Died, around 4.4% are Forced discharge, 1.3% are Referred to other hospital and 0.6% or four patients are Outpatients (controlled from home).

In our data set, the age of each patient is also documented. The average age is 36.14 years old, with under one year old as the youngest and 88 years old as the oldest patient. The information about comorbidities of each patient is recorded. The average comorbidities that they had is 2.79 or 3 diseases with the maximum comorbidities are ten diseases.

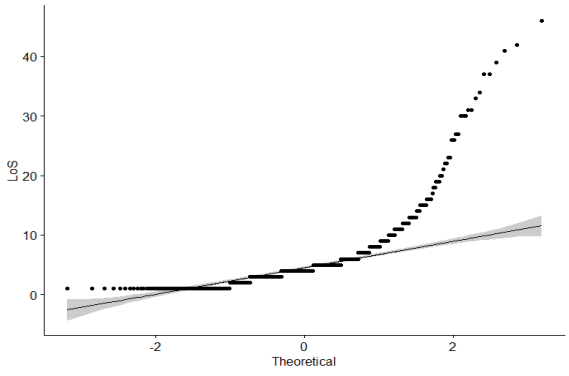
Table I. Summary statistics for continuous independent variables of LoS of COVID-19 Patients

| Variable | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|
| Age | 36.14 | 24.142 | 0 | 88 |
| Comorbidities | 2.79 | 1.723 | 0 | 10 |

Meanwhile, the response variable, LoS, is provided in Figure 1 (a) histogram of 688 patients. The distribution of data is skewed to the right. Figure 1 (b) demonstrates a normal Q-Q plot of the LoS data, indicating a violation of the normality assumption.




(a)




(b)

Fig 1. Length of hospital stay of COVID-19 patient in (a) Histogram and (b) normal QQ plot.

This study also did the heteroscedasticity test to the residuals of the regression model using the Durbin Watson test. After fitting the data based on the hypothesis model, this study resulted that the value of Durbin Watson test statistic is 1.834419 and the corresponding p-value is 0.032. Since this p-value is less than 0.05, it can be concluded that the residuals in this regression model are autocorrelated.

### B. Quantile Regression

Quantile regression is often used to explore the comprehensive relationship between response variable *y* and the explanatory variables *x*. Consider this following linear quantile regression model :

$$y_i = x_i^T \boldsymbol{\beta} + e_i, \quad i = 1,2,\dots,n, \qquad (1)$$

where $y_i$ is a response variable of interest which may represent the timing of the occurrence of some events such as disease recurrence or death, or some transformation of the time to the event, $x_i^T = (x_{1i}, x_{2i}, \dots, x_{pi})$ is the *i*th covariate in $\mathbb{R}^m$, and $e_i$ is an independent error terms. The error $e_i$ is restricted to have the τth quantile equal to zero, that is $\int_{-\infty}^{0} f_\tau(e_i) d\, e_i = \tau$. For $0 < \tau < 1$, let $Q_\tau(y_i|x_i)$ denote the $\tau$-th quantile regression function of $y_i$ with associated $p$ dimensional vector of covariates $x_i$. The quantile regression function is in the form of $Q_\tau(y_i|x_i) = x_i^T \boldsymbol{\beta}$ , for $i = 1,2,\dots,n$ , where $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients for indicator variables that depend on $\tau$. Then, quantile regression estimation $\boldsymbol{\beta}$ is obtained by minimizing

$$\min \textstyle\sum_i \rho_\tau(y_i - x_i^T \boldsymbol{\beta}), \qquad (2)$$

where $\rho_\tau(u)$ is the check function defined by

$$\rho_\tau(u) = u\big(\tau - I(u < 0)\big) \qquad (3)$$

Here *I*(.) is an indicator function that takes unity when *I*(.) is true and zeroes otherwise. However, this indicator function is not differentiable at zero, and explicit minimization problems are unobtainable [17], [31].

The goodness of fit for the quantile regression is measured using Pseudo-$R^2$ [9], [36]. The formula for Pseudo-$R^2$ is as follows:

$$\text{Pseudo} - \mathrm{R}^2 = 1 - \frac{RASW_\tau}{TASW_\tau}, \qquad (4)$$

where $RASW_\tau$ is the residual absolute sum of weighted differences between the observed dependent variable and the estimated quantile of conditional distribution in the more complex model. $TASW_\tau$ is the total absolute sum of weighted differences between the observed dependent variable and the estimated quantile of conditional distribution in the simplest model.

### C. Bootstrap Methods

The bootstrap resampling methods are designed to be employed when the innovations of a regression are not identically distributed. This method is fully nonparametric procedure and can be applied to a wide variety of models. The procedures in bootstrapping are based on XY pairs. In the pairs' bootstrap, the errors are not limited to *iid* in regression models, unlike resampling residuals. Instead of resampling the response, or residuals, bootstrapping possible to rescaled or centered. It bootstraps pairs consisting of an observation of the response along with the vector of indicator variables for that same observation. It is assumed implicitly that the pairs $(y_i, x_i)$ are independent. Although this is still a restrictive assumption, ruling out any form of dependence among observations, it does allow for arbitrary forms of heteroscedasticity of $y_i$ conditional on $x_i$ . The object resampled are *iid* drawings from the joint distribution of $y_i$ and $x_i$ . Each bootstrap sampling consists of some the original pairs once, some of them more than once and some of them not at all [37].

### D. Quantile Bootstrap Regression Methods

In particular, the asymptotic variance of quantile regression estimators depends on the density of the innovation term. The inference procedures and confidence interval can be greatly simplified by using bootstrap methods.

In general, the bootstrap resampling schemes for cross sectional data are as follows. Resampling Y and X with replacement from the cross-sectional dimension with probability $1/n$. Therefore, let $Y^* = (y_1, y_2, \dots, y_n)$ where each element is obtained by drawing with replacement from $i = 1, \dots, n$. The same vector of *i* is used to obtained $X^*$. Then fitting Eq. (1) to the data to obtain the parameter vector of $\hat{\beta}$ and the residual $\hat{e}_i$ for $i = 1, \dots, n$. Calculate the bootstrap sample as $y_i^* = x_i^T \boldsymbol{\beta} + e_i^*$. Then refit Eq. (1) to the bootstrap sample and denote the bootstrap estimated by $\widehat{\boldsymbol{\beta}}^*$. Repeat Steps 2 to 4 until *B* times and estimate the mean and the variance of the *B* copies of $\widehat{\boldsymbol{\beta}}^*$, denoted by $\hat{\beta}^b$ and Var ($\hat{\beta}^b$). This present study evaluates bootstrapping procedures for cross sectional quantile regression estimators. Five options of different bootstrap methods are implemented in this study.

First is the pairwise bootstrap, which many studies use, such as Yanuar et al. [9], also known as the XY-pair method. Then we implement a method based on a study by Parzen, Wei, and Ying [13], named is the PWY method. The PWY method proposed a general and simple resampling method for inferences $\boldsymbol{\beta}$ based on pivotal estimating functions. This study also compares the quantile bootstrap using the Markov chain marginal bootstrap (MCMB) based on the study of He and Hu [14] and Kocherginsky, He and Mu [15]. The difference between MCMB method from the usual bootstrap methods are in two important aspects: it involves solving only one-dimensional equations for parameters of any dimension and produces a Markov chain rather than a (conditionally) independent sequence. The fourth option uses WXY method based on the generalized bootstrap, a study proposed by Bose and Chatterjee [16]. Bose and Chatterjee proposed a generalized bootstrap technique for

estimators obtained by minimizing functions that are convex in the parameter. They established the consistency of these schemes via representation theorems. This study also employed the Wild bootstrap method which proposed by Feng, He and Hu [17].

## III. Result and Discussion

In this stage, firstly we did a comparison on several hypothesis model to obtain the reduced model. The reduced model means here is a model with significant variables only. The reduced model considers the simplicity of a model. After comparing, we found the significant indicator variables are Diagnosis and Final outcome.

We use 5000 Monte Carlo samples to estimate the parameter estimates' standard errors as the benchmark for comparison. We further compare the efficient method which is based on the smallest standard error. We also test the performance of 95% confidence intervals for each estimated parameter, i.e., the parameters for category Diagnosis are $\beta_1$(PerUS), $\beta_2$(PaUS), $\beta_3$ (Positive), and the parameter for category Final outcome are $\beta_4$ (Recovered), $\beta_5$ (Died), $\beta_6$(Outpatient) and $\beta_7$ (Referred). The estimated mean, the width of 95% confidence interval and standard errors for each parameter are presented in Table II for quantile $\tau = 0.10$, Table III for quantile $\tau = 0.25$, Table IV for quantile $\tau = 0.50$, Table V for quantile $\tau = 0.75$ and Table VI for quantile $\tau = 0.90$.

Table II. Bootstrap quantile methods at quantile $\tau = 0.10$

| Categories | Bootstrap Quantile Method ($\tau = 0.10$) | | | | |
|---|---|---|---|---|---|
| | XY | PWY | MCMB | WXY | Wild |
| PerUS ($\beta_1$) | | | | | |
| Estimate mean | -2.000 | -2.000 | -2.000 | -2.000*** | -2.000*** |
| Width of 95% CI | 5.570 | | 40.548 | 1.964 | **1.854** |
| Standard error | (1.421) | (NA) | (10.344) | (0.501) | (**0.473)** |
| Estimate mean | -1.000 | -1.000 | -1.000 | -1.000*** | -1.000*** |
| Width of 95% CI | 4.304 | | 37.490 | **0.548** | 1.128 |
| Standard error | (0.775) | (NA) | (9.564) | (**0.140)** | (0.288) |
| Positive ($\beta_3$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 3.822 | | 37.361 | 2.042 | **1.971** |
| Standard error | (0.975) | (NA) | (9.531) | (0.521) | (**0.503)** |
| Recovered ($\beta_4$) | | | | | |
| Estimate mean | 2.000 | 2.000 | 2.000 | 2.000*** | 2.000*** |
| Width of 95% CI | 0.611 | **0.388** | 4.558 | 0.548 | 0.709 |
| Standard error | (0.156) | (**0.099)** | (1.163) | (0.140) | (0.181) |
| Died ($\beta_5$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | **0.388** | | 4.558 | | 0.666 |
| Standard error | (**0.099)** | (NA) | (1.163) | (NA) | (0.170) |
| Outpatient ($\beta_6$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 8.165 | 4.496 | 7.063 | 7.020 | **2.842** |
| Standard error | (2.083) | (1.147) | (1.802) | (1.791) | (**0.725)** |
| Referred ($\beta_7$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 7.122 | **0.474** | 8.106 | 1.050 | 5.311 |
| Standard error | (1.817) | **(0.121)** | (2.068) | (0.268) | (1.355) |

***significant at level 1%,
** significant at level 5%,
* significant at level 10%
The smallest values are written in boldface.

Table II informs us that the estimated mean based on different methods result in similar value. The PWX method results in many NA (Not Available) values than others. The Wild method tends to result in the smallest standard errors among others. For this empirical case, at the 10th quantile, the wild method tends to yield shorter width of 95% CI and smaller standard error than other methods. Hold all else constant, at the 10th quantile of LoS for PerUS is shorter two days than at the 10th quantile of LoS for Asymptotic Person.

Based on Table III, holding all else constant, the 25th quantile of LoS for Positive is 3 days longer than the 25th quantile of LoS for Asymptomatic Person. The 25th quantile of LoS for Recovered is 3 days longer than the 25th quantile of LoS for Forced discharge, hold the other variables constant. The 25th quantile of LoS for outpatient is 4 days longer than the 25th quantile of LoS for Forced discharge, assume else constant. Table III informs us that the wild method is still the best method since it results in the shortest 95% confidence interval and the smallest standard error among others.

Table III. Bootstrap quantile methods at quantile $\tau$ = 0.25

| Categories | Bootstrap Quantile Method ($\tau$ = 0.25) | | | | |
|---|---|---|---|---|---|
| | XY | PWY | MCMB | WXY | Wild |
| PerUS ($\beta_1$) | | | | | |
| Estimate mean | -1.000 | -1.000 | -1.000 | -1.000 | -1.000 |
| Width of 95% CI | 5.535 | | 28.592 | **3.684** | 4.978 |
| Standard error | (1.412) | (NA) | (7.294) | (**0.940)** | (1.270) |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 4.480 | | 19.854 | **1.611** | 2.371 |
| Standard error | (1.143) | (NA) | (5.065) | (**0.411)** | (0.605) |
| Positive ($\beta_3$) | | | | | |
| Estimate mean | 3.000** | 3.000 | 3.000 | 3.000*** | 3.000*** |
| Width of 95% CI | 6.032 | | 19.883 | 4.370 | **4.186** |
| Standard error | (1.539) | (NA) | (5.072) | (1.115) | (**1.068)** |
| Recovered ($\beta_4$) | | | | | |
| Estimate mean | 3.000*** | 3.000*** | 3.000 | 3.000*** | 3.000*** |
| Width of 95% CI | 2.108 | 2.061 | 12.920 | 2.242 | **0.752** |
| Standard error | (0.538) | (0.526) | (3.296) | (0.572) | (**0.192)** |
| Died ($\beta_5$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 1.399 | 1.438 | 12.877 | 1.458 | **0.752** |
| Standard error | (0.357) | (0.367) | (3.285) | (0.372) | **(0.192)** |
| Outpatient ($\beta_6$) | | | | | |
| Estimate mean | 4.000* | 4.000** | 4.000 | 4.000** | 4.000*** |
| Width of 95% CI | 9.670 | 7.973 | 14.719 | 9.137 | **4.731** |
| Standard error | (2.467) | (2.034) | (3.755) | (2.331) | (**1.207)** |
| Referred ($\beta_7$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 6.601 | 2.524 | 35.354 | 3.504 | **1.834** |
| Standard error | (1.684) | (0.644) | (9.019) | (0.894) | (**0.468)** |

***significant at level 1%,
** significant at level 5%,
* significant at level 10%
The smallest values are written in boldface.

Moreover, for bootstrap quantile methods at quantile $\tau$ = 0.50, as provided by Table IV, WXY method tends to result the smallest standard error with the shortest confidence interval than others; mainly for estimated coefficient of PerUS, PaUS and Positive. Meanwhile, Wild method results the smallest estimated standard error for coefficient of Outpatient and Referred only. At this 50th quantile of LoS for Positive is 7 days longer than the 50th quantile of LoS for Asymptomatic Person, holding all else constant.

Table V presents the mean, width of 95% confidence interval and standard error of all five methods of Bootstrap quantile approach at quantile $\tau$ = 0.75. This table shows that the Wild method tends to result the shortest 95% confidence interval among others. Hold all else constant, the 75th quantile of LoS for Positive is 12 days longer than the 75th quantile of LoS for Asymptomatic Person.

Table VI mainly informs us that the WXY method is the best method for the Bootstrap quantile approach at quantile $\tau$ = 0.90. This table shows that the shortest confidence interval is resulted based on the WXY method. Hold all else constant, at this 90th quantile, the LoS for Positive is 28 days longer than the 90th quantile of LoS for Forced discharge. This value is increasing as the sequence of quantiles increases as well.

The analysis above informs us that the PWY and MCMB often result in NA (not available) estimated standard error values; hence, confidence intervals can't be estimated. The WXY resulted in a NA value at quantile 0.10, whereas this method tended to result in better-estimated values at higher quantiles. The XY method yielded a wider 95% confidence interval than others. In general, this study proved that Wild bootstrap quantile tends to be the best method since it produces the smallest estimated standard errors and the shortest 95% confidence interval.

The estimated mean values obtained in each selected quantile as provided in Table II to Table VI based on the Wild bootstrap method are also be presented in a plot. Meanwhile, Figure 2 presents plots of quantiles based on this Wild bootstrap estimated method only, as the best method resulted in this study.

Table IV. Bootstrap quantile methods at quantile $\tau$ = 0.50

| Categories | Bootstrap Quantile Method ($\tau$ = 0.50) | | | | |
|---|---|---|---|---|---|
| | XY | PWY | MCMB | WXY | Wild |
| PerUS ($\beta_1$) | | | | | |
| Estimate mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Width of 95% CI | 11.156 | | 9.243 | **7.048** | 7.671 |
| Standard error | (2.846) | (NA) | (2.358) | (**1.798)** | (1.957) |
| Estimate mean | 1.000 | 1.000 | 1.000 | 1.000*** | 1.000 |
| Width of 95% CI | 11.156 | | 6.573 | **0.811** | 6.224 |
| Standard error | (2.846) | (NA) | (1.677) | (**0.207)** | (1.588) |
| Positive ($\beta_3$) | | | | | |
| Estimate mean | 7.000*** | 7.000 | 7.000*** | 7.000*** | 7.000*** |
| Width of 95% CI | 10.909 | | 7.228 | **4.292** | 7.071 |
| Standard error | (2.783) | (NA) | (1.844) | (**1.095)** | (1.804) |
| Recovered ($\beta_4$) | | | | | |
| Estimate mean | 3.000*** | 3.000*** | 3.000*** | 3.000*** | 3.000*** |
| Width of 95% CI | 2.889 | 2.889 | **0.493** | 2.622 | 2.242 |
| Standard error | (0.737) | (0.737) | **(0.126)** | (0.669) | (0.572) |
| Died ($\beta_5$) | | | | | |
| Estimate mean | -1.000 | -1.000 | -1.000*** | -1.000 | -1.000* |
| Width of 95% CI | 2.779 | 2.826 | **0.486** | 2.512 | 2.242 |
| Standard error | (0.709) | (0.721) | **(0.124)** | (0.641) | (0.572) |
| Outpatient ($\beta_6$) | | | | | |
| Estimate mean | 3.000 | 3.000* | 3.000** | 3.000 | 3.000*** |
| Width of 95% CI | 8.201 | 6.518 | 4.135 | 7.742 | **3.477** |
| Standard error | (2.092) | (1.663) | (1.055) | (1.975) | (**0.887)** |
| Referred ($\beta_7$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 10.337 | 4.390 | 3.092 | 6.844 | **4.076** |
| Standard error | (2.637) | (1.120) | (0.789) | (1.746) | (**1.040)** |

***significant at level 1%,
** significant at level 5%,
* significant at level 10%
The smallest values are written in boldface.

Table V. Bootstrap quantile methods at quantile $\tau$ = 0.75

| Categories | Bootstrap Quantile Method ($\tau$ = 0.75) | | | | |
|---|---|---|---|---|---|
| | XY | PWY | MCMB | WXY | Wild |
| PerUS ($\beta_1$) | | | | | |
| Estimate mean | 5.000 | 5.000 | 5.000 | 5.000** | 5.000 |
| Width of 95% CI | 21.364 | | | **9.556** | 25.519 |
| Standard error | (5.450) | (NA) | (NA) | (**2.438)** | (6.510) |
| Estimate mean | 2.000 | 2.000 | 2.000 | 2.000*** | 2.000 |
| Width of 95% CI | 23.175 | | | **1.764** | 21.912 |
| Standard error | (5.912) | (NA) | (NA) | (**0.450)** | (5.590) |
| Positive ($\beta_3$) | | | | | |
| Estimate mean | 12.000** | 12.000 | 12.000 | 12.000*** | 12.000** |
| Width of 95% CI | 23.915 | | | **11.450** | 23.633 |
| Standard error | (6.101) | (NA) | (NA) | (**2.921)** | (6.029) |
| Recovered ($\beta_4$) | | | | | |
| Estimate mean | 2.000*** | 2.000*** | 2.000** | 2.000*** | 2.000*** |
| Width of 95% CI | 2.367 | 2.336 | 2.614 | 2.759 | **1.803** |
| Standard error | (0.604) | (0.596) | (0.667) | (0.704) | (**0.460)** |
| Died ($\beta_5$) | | | | | |
| Estimate mean | -2.000*** | -2.000*** | -2.000** | -2.000*** | -2.000*** |
| Width of 95% CI | 2.085 | 2.007 | 2.724 | 2.222 | **1.838** |
| Standard error | (0.532) | (0.512) | (0.695) | (0.567) | (**0.469)** |
| Outpatient ($\beta_6$) | | | | | |
| Estimate mean | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Width of 95% CI | 7.949 | | 8.843 | 8.067 | **7.585** |
| Standard error | (2.028) | (NA) | (2.256) | (2.058) | (**1.935)** |
| Referred ($\beta_7$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 14.061 | 14.617 | 11.164 | 14.782 | **5.315** |
| Standard error | (3.587) | (3.729) | (2.848) | (3.771) | (**1.356)** |

***significant at level 1%,
** significant at level 5%,
* significant at level 10%
The smallest values are written in boldface.

Table VI. Bootstrap quantile methods at quantile τ = 0.90

| Categories | Bootstrap Quantile Method ($\tau$ = 0.90) | | | | |
|---|---|---|---|---|---|
| | XY | PWY | MCMB | WXY | Wild |
| PerUS ($\beta_1$) | | | | | |
| Estimate mean | 8.000** | 8.000 | 8.000 | 8.000*** | 8.000 |
| Width of 95% CI | 15.993 | | | **8.314** | 51.791 |
| Standard error | (4.080) | (NA) | (NA) | (**2.121)** | (13.212) |
| Estimate mean | 6.000 | 6.000 | 6.000 | 6.000*** | 6.000 |
| Width of 95% CI | 14.896 | | | **3.296** | 50.113 |
| Standard error | (3.800) | (NA) | (NA) | (**0.841)** | (12.784) |
| Positive ($\beta_3$) | | | | | |
| Estimate mean | 28.000*** | 28.000 | 28.000 | 28.000*** | 28.000** |
| Width of 95% CI | 23.990 | | | **19.396** | 52.398 |
| Standard error | (6.120) | (NA) | (NA) | (**4. 948)** | (13.367) |
| Recovered ($\beta_4$) | | | | | |
| Estimate mean | 5.000*** | 5.000 | 5.000 | 5.000*** | 5.000*** |
| Width of 95% CI | 4.221 | 4.300 | 28.161 | **3.916** | 4.645 |
| Standard error | (1.077) | (1.097) | (7.184) | (**0.999)** | (1.185) |
| Died ($\beta_5$) | | | | | |
| Estimate mean | -2.000 *** | -2.000 | -2.000 | -2.000** | -2.000** |
| Width of 95% CI | 3.214 | 3.477 | 27.436 | **3.214** | 3.555 |
| Standard error | (0.820) | (0.887) | (6.999) | (**0.820)** | (0.907) |
| Outpatient ($\beta_6$) | | | | | |
| Estimate mean | 2.000 | 2.000 | 2.000 | 2.000 | 2.000 |
| Width of 95% CI | 9.945 | | | **7.557** | 14.578 |
| Standard error | (2.537) | (NA) | (NA) | (**1.928)** | (3.719) |
| Referred ($\beta_7$) | | | | | |
| Estimate mean | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Width of 95% CI | 11.905 | | | **10.540** | 10.862 |
| Standard error | (3.037) | (NA) | (NA) | (**2.689)** | (2.771) |

***significant at level 1%,
** significant at level 5%,
* significant at level 10%
The smallest values are written in boldface.

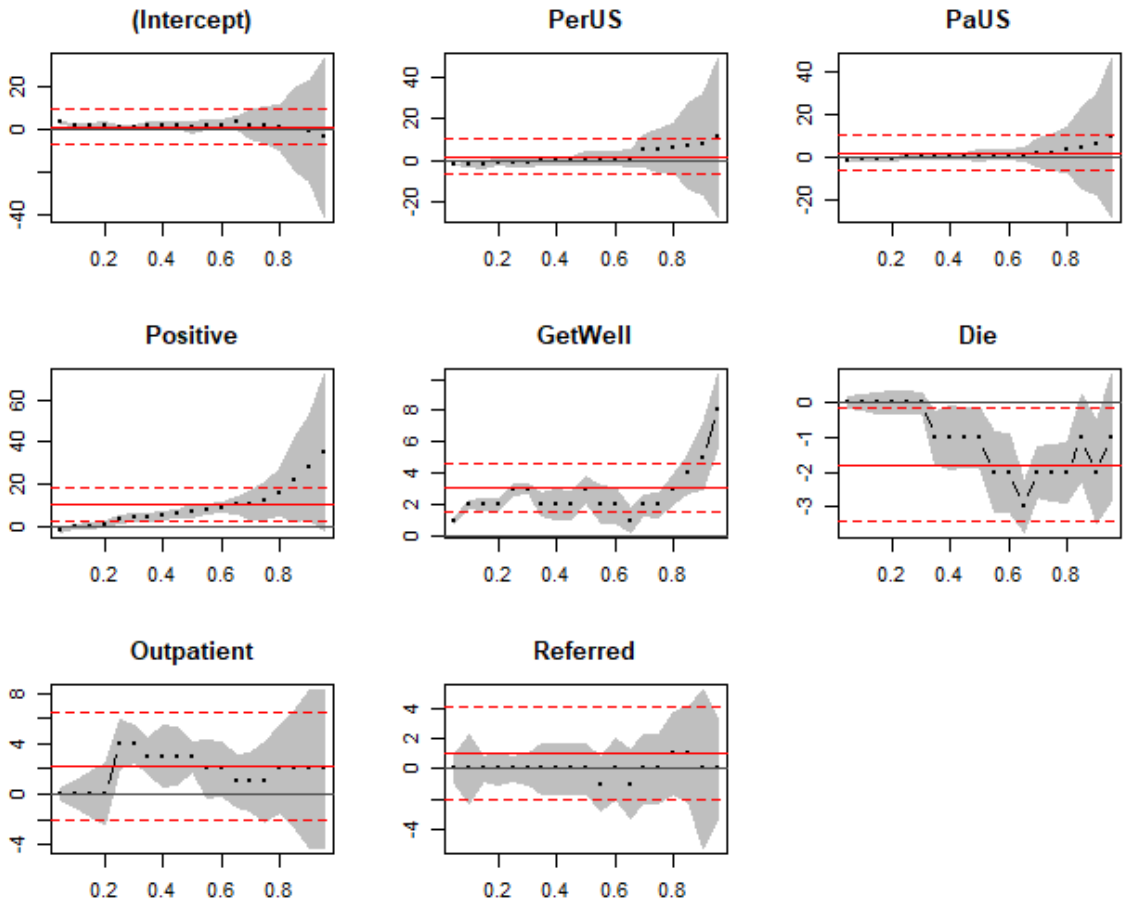



Fig 2. Estimate coefficients regression for each indicator of LoS and theirs trends along with the 95% CI over different quantiles indicated by grey area. The red straight line presents the OLS's mean and the red dash lines present the 95% CI for OLS

Figure 2 presents the OLS (ordinary least square) estimated mean indicated by the red straight line with its upper bound and lower bound for 95% confidence interval, indicated by the red dash line. Meanwhile, the quantile with CI estimated at any sequence quantiles is shown with the grey area. This figure informs us that the 95% confidence interval based on OLS estimated is wider than quantile, especially at lower quantiles. Based on the ANOVA test, we could conclude that there is no significant coefficient difference between OLS and quantiles, especially for lower quantile to higher quantile (such as at the 80th quantile). These facts could also be seen in Figure 2. The Positive plot shows no significant coefficient difference between OLS and quantiles at lower to higher quantile. In addition, comparisons among quantiles are also tested based on the ANOVA test as well. The quantile regression coefficients associated with Positive and Recovered are significantly different between the 75th quantile and the 25th quantile regressions ($p < 0.001$).

The next analysis is to measure the goodness of fit for all proposed models at all selected quantiles based on the Wild bootstrap method. The indicator for the goodness of fit in this study is based on the Pseudo-$R^2$ [9], [36].

Table VII presents the corresponding Pseudo-$R^2$ values for each selected quantile. It is informed that the 10th quantile is the best model since its Pseudo-$R^2$ value is the highest among other quantiles. This result is also in line with the values for the width of 95% CI, and standard error yielded in this quantile. All values for the width of 95% CI and standard error at this quantile are smaller than other quantiles for all seven categories (as provided in Table II, Table III, Table IV, Table V and Table VI).

Table VII. Pseudo-$R^2$ Based on the Wild Bootstrap Method

| Quantiles | Pseudo-$R^2$ |
|---|---|
| **0.10** | **0.839*** |
| 0.25 | 0.789 |
| 0.50 | 0.795 |
| 0.75 | 0.774 |
| 0.90 | 0.708 |

*The highest values.

## IV. Conclusions

Quantile regression has been utilized as an alternative tool to model distributional changes in LoS hospitalized. This study explores the heterogeneity in the relationship between Diagnosis and Final outcome of patients with COVID-19 suspect across the distribution of LoS hospitalized. This study found that in general, the patient with positive COVID-19 tends to have longer LoS hospitalized than other criteria of Diagnosis. While if the patient wants to have a Recovered status, they will have longer LoS hospitalized than other criteria in Final outcome.

The observed differences between quantile regression coefficients estimated at various percentiles compared with mean regression coefficients illustrate the influence methodological choices. It might have in the evaluation of factors influencing length of hospital stay which could impact to cost. Models estimating mean marginal effects may disguise significant changes that estimating effects across the range of the hospital days provides a more detailed, more thorough description of these relationships. This information may be useful in predicting the cost of care or evaluating interventions to increase appropriate hospital use.

## References


[1] E. M. Rees, "COVID-19 length of hospital stay: a systematic review and data synthesis," p. 22, 2020.

[2] M. W. Wachterman, E. R. Marcantonio, R. B. Davis, and E. P. McCarthy, "Association of Hospice Agency Profit Status With Patient Diagnosis, Location of Care, and Length of Stay," *JAMA*, vol. 305, no. 5, p. 472, Feb. 2011, doi: 10.1001/jama.2011.70.

[3] S. Wu *et al.*, "Understanding factors influencing the length of hospital stay among non-severe COVID-19 patients: A retrospective cohort study in a Fangcang shelter hospital," *PLoS ONE*, vol. 15, no. 10, p. e0240959, Oct. 2020, doi: 10.1371/journal.pone.0240959.

[4] C. Muharisa, F. Yanuar, and D. Devianto, "Simulation Study The Using of Bayesian Quantile Regression in Nonnormal Error," *CAUCHY*, vol. 5, no. 3, p. 121, Dec. 2018, doi: 10.18860/ca.v5i3.5633.

[5] O. D. Saputri, F. Yanuar, and D. Devianto, "Simulation Study The Implementation of Quantile Bootstrap Method on Autocorrelated Error," *CAUCHY*, vol. 5, no. 3, p. 95, Dec. 2018, doi: 10.18860/ca.v5i3.5349.

[6] H. Kozumi and G. Kobayashi, "Gibbs sampling methods for Bayesian quantile regression," *Journal of Statistical Computation and Simulation*, vol. 81, no. 11, pp. 1565–1578, Nov. 2011, doi: 10.1080/00949655.2010.496117.

[7] R. Alhamzawi and K. Yu, "Variable selection in quantile regression via Gibbs sampling," *Journal of Applied Statistics*, vol. 39, no. 4, pp. 799–813, Apr. 2012, doi: 10.1080/02664763.2011.620082.

[8] M.-S. Oh, E. S. Park, and B.-S. So, "Bayesian variable selection in binary quantile regression," *Statistics & Probability Letters*, vol. 118, pp. 177–181, Nov. 2016, doi: 10.1016/j.spl.2016.07.001.

[9] F. Yanuar, H. Yozza, F. Firdawati, I. Rahmi, and A. Zetra, "Applying bootstrap quantile regression for the construction of a low birth weight model," *msk*, vol. 23, no. 2, Aug. 2019, doi: 10.7454/msk.v23i2.9886.

[10] L. Wang, I. Van Keilegom, and A. Maidman, "Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors," *Biometrika*, vol. 105, no. 4, pp. 859–872, Dec. 2018, doi: 10.1093/biomet/asy037.

[11] Y. Yang, H. J. Wang, and X. He, "Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood: Bayesian Quantile Regression," *International Statistical Review*, vol. 84, no. 3, pp. 327–344, 2015, doi: 10.1111/insr.12114.

[12] M. Geraci and M. Bottai, "Linear quantile mixed models," *Stat Comput*, vol. 24, pp. 461–479, 2014.

[13] M. I. Parzen, L. J. Wei, and Z. Ying, "A resampling method based on pivotal estimating functions," *Biometrika*, vol. 81, no. 2, pp. 341–350, 1994.

[14] X. He and F. Hu, "Markov Chain Marginal Bootstrap," *Journal of the American Statistical Association*, vol. 97, no. 459, pp. 783–795, 2002, [Online]. Available: http://www.jstor.org/stable/3085721

[15] M. Kocherginsky, X. He, and Y. Mu, "Practical Confidence Intervals for Regression Quantiles," *Journal of Computational and Graphical Statistics*, vol. 14, no. 1, pp. 41–55, Mar. 2005, doi: 10.1198/106186005X27563.

[16] A. Bose and S. Chatterjee, "Generalized bootstrap for estimators of minimizers of convex functions," *Journal of Statistical Planning and Inference*, vol. 117, no. 2, pp. 225–239, Dec. 2003, doi: 10.1016/S0378-3758(02)00386-5.

[17] X. Feng, X. He, and J. Hu, "Wild bootstrap for quantile regression," *Biometrika*, vol. 98, no. 4, pp. 995–999, Dec. 2011, doi: 10.1093/biomet/asr052.

[18] C. Li and H. Hao, "Likelihood and Bayesian Estimation in Stress Strength Model from Generalized Exponential Distribution Containing Outliers," *IAENG International Journal of Applied Mathematics*, vol. 46, no. 2, pp. 155–159, 2016.

[19] H. A. Huskamp, D. G. Stevenson, D. C. Grabowski, E. Brennan, and N. L. Keating, "Long and Short Hospice Stays among Nursing Home Residents at the End of Life," *Journal of Palliative Medicine*, vol. 13, no. 8, pp. 957–964, Aug. 2010, doi: 10.1089/jpm.2009.0387.

[20] B. G. Kaufman, C. A. Sueta, C. Chen, B. G. Windham, and S. C. Stearns, "Are Trends in Hospitalization Prior to Hospice Use Associated With Hospice Episode Characteristics?," *Am J Hosp Palliat Care*, vol. 34, no. 9, pp. 860–868, Nov. 2017, doi: 10.1177/1049909116659049.

[21] K. Yuki, M. Fujiogi, and S. Koutsogiannaki, "COVID-19 pathophysiology: A review," *Clinical Immunology*, vol. 215, p. 108427, Jun. 2020, doi: 10.1016/j.clim.2020.108427.

[22] Y. Du *et al.*, "Clinical Features of 85 Fatal Cases of COVID-19 from Wuhan. A Retrospective Observational Study," *Am J Respir Crit Care Med*, vol. 201, no. 11, pp. 1372–1379, Jun. 2020, doi: 10.1164/rccm.202003-0543OC.

[23] S. Richardson *et al.*, "Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area," *JAMA*, vol. 323, no. 20, pp. 2052–2059, 2020, doi: 10.1001/jama.2020.6775.

[24] C. Gebhard, V. Regitz-Zagrosek, H. K. Neuhauser, R. Morgan, and S. L. Klein, "Impact of sex and gender on COVID-19 outcomes in Europe," *Biol Sex Differ*, vol. 11, no. 1, pp. 1–13, Dec. 2020, doi: 10.1186/s13293-020-00304-9.

[25] A. López-Cheda, M.-A. Jácome, R. Cao, and P. M. De Salazar, "Estimating lengths-of-stay of hospitalised COVID-19 patients using a non-parametric model: a case study in Galicia (Spain)," *Epidemiol. Infect.*, vol. 149, no. e102, pp. 1–8, 2021, doi: 10.1017/S0950268821000959.

[26] P. Q. Thai *et al.*, "Factors associated with the duration of hospitalisation among COVID-19 patients in Vietnam: A survival analysis," *Epidemiol. Infect.*, vol. 148, no. e114, pp. 1–20, 2020, doi: 10.1017/S0950268820001259.

[27] H. S., M. Subbiah, and M. R. Srinivasan, "Fitting length of stay in hospitals using transformed distributions," *Communications in Statistics: Case Studies, Data Analysis and Applications*, vol. 4, no. 1, pp. 1–8, Jan. 2018, doi: 10.1080/23737484.2018.1445979.

[28] H. Qiu, J. Wu, L. Hong, Y. Luo, Q. Song, and D. Chen, "Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study," *The Lancet Infectious Diseases*, vol. 20, no. 6, pp. 689–696, Jun. 2020, doi: 10.1016/S1473-3099(20)30198-5.

[29] Y. Shi *et al.*, "A quickly, effectively screening process of novel corona virus disease 2019 (COVID-19) in children in Shanghai, China," *Ann Transl Med*, vol. 8, no. 5, pp. 241–241, Mar. 2020, doi: 10.21037/atm.2020.03.22.

[30] W. Xia, J. Shao, Y. Guo, X. Peng, Z. Li, and D. Hu, "Clinical and CT features in pediatric patients with COVID-19 infection: Different points from adults," *Pediatr Pulmonol*, vol. 55, no. 5, pp. 1169–1174, May 2020, doi: 10.1002/ppul.24718.

[31] M. Sengupta, E. Park-Lee, R. Valverde, C. Caffrey, and A. Jones, "Trends in Length of Hospice Care From 1996 to 2007 and the Factors Associated With Length of Hospice Care in 2007: Findings From the National Home and Hospice Care Surveys," *Am J Hosp Palliat Care*, vol. 31, no. 4, pp. 356–364, Jun. 2014, doi: 10.1177/1049909113492371.

[32] N. S. Park, I. V. Carrion, B. S. Lee, D. Dobbs, H. J. Shin, and M. A. Becker, "The Role of Race and Ethnicity in Predicting Length of

Hospice Care among Older Adults," *Journal of Palliative Medicine*, vol. 15, no. 2, pp. 149–153, Feb. 2012, doi: 10.1089/jpm.2011.0220.

[33] H. N. Yeung *et al.*, "Palliative Radiation Before Hospice: The Long and the Short of It," *Journal of Pain and Symptom Management*, vol. 48, no. 6, pp. 1070–1079, Dec. 2014, doi: 10.1016/j.jpainsymman.2014.04.004.

[34] A. A. Butt *et al.*, "Hospital admission rates, length of stay, and in-hospital mortality for common acute care conditions in COVID-19 vs. pre-COVID-19 era," *Public Health*, vol. 189, pp. 6–11, Dec. 2020, doi: 10.1016/j.puhe.2020.09.010.

[35] N. Lapidus, X. Zhou, F. Carrat, B. Riou, Y. Zhao, and G. Hejblum, "Biased and unbiased estimation of the average length of stay in intensive care units in the Covid-19 pandemic," *Ann. Intensive Care*, vol. 10, no. 1, p. 135, Dec. 2020, doi: 10.1186/s13613-020-00749-6.

[36] C. Davino, M. Furno, and D. Vistocco, *Quantile Regression Theory and Applications*. Wiley Series, 2014.

[37] A. Galvao and G. Montes-Rojas, "On Bootstrap Inference for Quantile Regression Panel Data: A Monte Carlo Study," *Econometrics*, vol. 3, no. 3, pp. 654–666, Sep. 2015, doi: 10.3390/econometrics3030654.