DOI: https://doi.org/10.34123/jurnalasks.v17i1.793

POLITEKNIK STATISTIKA STIS

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 17 No. 1, June 2025
e- ISSN: 2615-1367, p-ISSN:2086-4132

Available online at https://jurnal.stis.ac.id/

Quantile Regression with Constrained B-Splines for Modelling Average Years of Schooling and Household Expenditure

Yoga Sasmita[1*], Muhammad Budiman Johra[2], Yogo Aryo Jatmiko[3], Deltha A. Lubis[4], Rizal Rahmad[5], Gama Putra Danu Sohibien[6]

[1]BPS-Statistics Central Kalimantan Province, Palangka Raya, Indonesia, [2]BPS-Statistics South Halmahera Regency, Bacan, Indonesia, [3]BPS-Statistics Indonesia, Jakarta, Indonesia,, [4]BPS-Statistics North Sumatera Province, Medan, Indonesia, [5]BPS-Statistics Pidie Regency, Pidie, Indonesia,[6]Politeknik Statistika STIS, Jakarta, Indonesia

**Corresponding Author: E-mail address:* yoga_sasmita@bps.go.id

ARTICLE INFO

Article history:
Received 30 August, 2024
Revised 29 October, 2024
Accepted 9 November, 2024
Published 24 February, 2025

Keywords:
Average Years of Schooling; B-Splines; COBS; Constrained B-Splines; Quantile Regression; Per Capita Household Expenditure

Abstract

Introduction/Main Objectives: Education serves as a driving force for the transformation of society to break the cycle of poverty. This study examines the relationship between average years of schooling and per capita household expenditure in Kalimantan Tengah Province in 2020. **Background Problems:** The method of estimating a regression model that is assumed to follow a certain form of regression equation such as linear, quadratic and others is called parametric regression. However, researchers often encounter difficulties in determining the model specification through data distribution, so the method used is nonparametric regression. **Novelty:** This research uses a quantile-based approach to explore how the impact of education on per capita expenditure varies across different levels of household education. This provides a more nuanced understanding of the relationship, showing not just whether education matters, but how its influence changes at different levels of educational attainment. **Research Methods:** The relationship between average years of schooling and per capita household expenditure is modeled using a quantile regression model with the constrained B-Splines method. **Finding/Results:** Based on the established classification, it can be concluded that an increase in the average years of schooling among household members tends to have a greater impact on raising per capita expenditure.

1. Introduction

Poverty is seen as an economic inability to fulfill basic food and non-food needs measured in terms of expenditure [1]. Expenditure on food and non-food consumption needs can reflect the level of the community's economic capacity, and the purchasing power of the community can provide an overview of the level of community welfare. The higher the purchasing power of the community, the higher the ability to fulfill their needs, which in turn will lead to an increase in community welfare.

Education serves as a driving force for the transformation of society to break the cycle of poverty. Education helps reduce poverty through its effect on labor productivity and through social benefit channels, so education is an important development goal for the nation [2]. Education is a means to gain insight, knowledge, and skills so that employment opportunities are more open and wages are also higher.

A person's education is one of the determinants of per capita consumption [3]. The average years of schooling, which shows the level of education of the community, can reduce the poverty rate in Indonesia [4]. Highly educated people will have skills and expertise so that they can increase their productivity. Increased productivity will increase company output, increase worker wages, and increase people's purchasing power so that it will reduce poverty. The education, especially an increase in the number of years of learning, is a prerequisite for this stage of economic development [5]. The higher a person's education, the better the quality of human resources and will affect productivity. And of course, higher productivity will increase income and expenditure.

Education is concerned with the development of knowledge as well as the expertise and skills of people and labor in the development process. Due to its enormous contribution to economic development, education is said to be human capital. Education is one of the investments in human resources to get a better life. A person with a higher education usually has greater access to higher-paying jobs, compared to individuals with lower levels of education [6]. Through adequate education, the poor will have a better chance of escaping poverty in the future [7]. This is in line with [8] that if education investment is made evenly, including in low-income communities, poverty will be reduced.

B-splines method has been used in several modelling applications by implementing constraints. The constrained smoothing B-splines (COBS) method nonparametrically estimates interest rate structures while meeting no-arbitrage constraints, such as monotonicity and positive rates, enhancing robustness against outliers. Balancing flexibility and constraint adherence, COBS occupies a middle ground between parametric and nonparametric methods, making it well-suited for markets with varying liquidity [9]. A method for constructing COBS wavelets by [10] using the lifting scheme, enabling multiresolution analysis with control over specific points and derivatives. This approach allows curve smoothing while preserving selected "feature points," seamless representation across different resolutions, and editing under constraints. The algorithm is optimized with linear time and storage complexity in the number of control points, making it highly efficient for large datasets. A method by [11] for designing optimal smoothing splines with derivative constraints, using a linear control system to generate the spline. Constraints on spline derivatives are formulated as controls on the system’s input and initial state, useful in applications like trajectory planning and convex shape-preserving splines. The method reduces the problem to convex quadratic programming, effectively handling pointwise constraints.

This study will look at the relationship between education level (average years of schooling) and poverty level (household expenditure per capita) in Kalimantan Tengah Province in 2020. Household expenditure per capita is a proxy for household income per capita, which is difficult to obtain in practice. Furthermore, the data was collected in March 2020 as we know that in that period the COVID-19 outbreak began. Average years of schooling is the number of years of study that the population aged 25 years and over has completed in formal education (excluding years repeated). Concerning household expenditure, the variable that has a significant effect is working/not working status. The reference population aged 18 years and above is used because, at the age point of 18 years and above, the proportion working is greater than those not working. Therefore, the reference population taken is the population aged 18 years and above. Kalimantan Tengah Province has the second lowest poverty rate in Kalimantan Island after Kalimantan Selatan Province. 5.36 percent of the population was recorded as poor in 2016 with an average monthly per capita expenditure of IDR 920,786. The average years of schooling in 2015 was recorded at 8.03 years. Nationally, per capita income was recorded at IDR.868,823 and the average years of schooling was 7.84. So that the higher the average years of schooling, the greater the expenditure/income, so that it will have an impact on poverty status.

At the household level, the relationship between education level (average years of schooling) and poverty level (household expenditure) can be shown based on a regression model. The method of estimating a regression model that is assumed to follow a certain form of regression equation such as linear, quadratic, and others is called parametric regression. However, researchers often encounter difficulties in determining the model specification through data distribution, so the method used is nonparametric regression. One of the estimation techniques in nonparametric regression is B-splines. B-splines is an estimation technique in regression curve fitting that takes smoothing into account. B-Splines are good at handling nonlinear relationships. Through movable knot locations that serve as anchor points where the curve can alter its behavior, they provide flexibility. Because they can describe both linear and complex nonlinear interactions, this flexibility is useful in situations that need for both smoothness and precision [12].

Furthermore, [13] proposes Constrained B-Splines to accomodates the constraines which can be monoton, convec or periodic based on the assumed of the form of curve regression so the regression curve will be more smooth by facilitates the addition of smoothing parameters. The addition of monotone constraints is often applied to estimate parameters where the relationship between the response variable and the predictor variables is assumed to be monotone [14]. The addition of monotone constraints has a smoothing effect on the estimated regression model [15].

2. Materials and Methods

2.1. Materials

The data in this study uses data sourced from the results of the National Socio-Economic Survey (Susenas) semester I 2020 in Kalimantan Tengah Province. The variables used are the Per Capita Expenditure variable as the response variable and the Average Years of Schooling per capita variable as the predictor variable. Household expenditure according to [1] is the cost incurred for consumption by all household members during the month, which consists of food and non-food consumption, regardless of the origin of the goods and is limited to consumption for business purposes or given to other parties. Per capita household expenditure is household expenditure divided by the number of household members in a household or in other words the average household expenditure for each household member.

Average Years of Schooling (RLS) is the number of years spent in formal education. The population included in the calculation of RLS is the population aged 25 years and over. However, based on the background discussed earlier, this study uses the limitation of RLS calculation on the population aged 18 years and above. Average Years of Schooling per capita is the average years of schooling of all household members aged 18 years and above in a household divided by the number of household members. RLS is calculated using the following formula [1]:

$$RLS = \frac{1}{P_{18+}} \sum_{i=1}^{P_{18+}} (LS_i) \tag{1}$$

P_{18+} : Total population aged 18 years and over

LS_i : years of schooling of the i-th population.

Years of schooling of the population aged 18 years and over at the last completed level of education using the following conversion [16]:

Table 1. Conversion highest education completed

No.	Highest education completed	Years
1.	No/never been to school	0
2.	Primary school/equivalent	6
3.	Junior high school/equivalent	9
4.	High school/equivalent	12
5.	Diploma I	13
6.	Diploma II	14
7.	Academy/ Diploma III	15
8.	Diploma IV/ Bachelor (S1)	16
9.	Magister (S2)	18
10.	Doctor (S3)	22

Source: BPS, 2011

2.2. Methods

2.2.1. Nonparametric Regression

Suppose Y is the response variable and X is the predictor variable with $\{(x_i, y_i), i = 1,2, \dots n\}$, $x_i \in X, y_i \in Y$. The relationship between x_i and y_i can be assumed to follow the regression model as follows:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{2}$$

where ε_i is the random error and $f(x_i)$ is the regression function.

2.2.2 B-Splines

B-splines is one of the methods used to estimate nonparametric regression functions. B-splines are defined as polynomial functions that have segmented properties at the interval x formed by knot points (piecewise polynomial) which are then locally estimated at these intervals for a certain polynomial degree [17]. To obtain B-splines of degree v with u knot points, additional knots of 2v are first defined so that a knot row $T = (t_1, \dots, t_v, t_{v+1}, \dots, t_{u+v}, t_{u+v+1}, \dots, t_{u+2v})$ with $t_1 = \cdots = t_v < t_{v+1} < \cdots < t_{u+v} < t_{u+v+1} = \cdots = t_{u+2v}$. Furthermore, the jth B-splines with $j = 1, \dots, u + v = m$ are recursively denoted by the following formula:

$$B_j(x;\upsilon) = \frac{x - t_j}{t_{j+\upsilon-1} - t_j} B_j(x;\upsilon - 1) - \frac{x - t_{j+\upsilon}}{t_{j+\upsilon} - t_{j+1}} B_{j+1}(x;\upsilon - 1) \tag{3}$$

where:

$$B_j(x;1) = \begin{cases} 1, & \text{if } t_j \le x \le t_{j+1} \\ 0, & \text{others} \end{cases} \tag{4}$$

From equations (3) and (4) it is obtained that on the interval $[t_v, t_{u+v+1}]$, $\sum_{j=1}^{m} B_j(x; v) = 1$ holds for every x.

The regression model (2) is a regression function of unknown shapes that will be approximated by the B-splines function. The B-splines function is formulated:

$$f(x) \approx \sum_{j=1}^{m} \alpha_j B_j(x;\upsilon) \tag{5}$$

From equation (5) above, the regression model (2) becomes:

$$y_i = \sum_{j=1}^{m} \alpha_j B_j(x_i;\upsilon) + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{6}$$

or we can denote it as

$$\mathbf{Y} = \mathbf{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{7}$$

In general, the objective function of B-spline regression is as follows:

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{m} \alpha_j B_j(x_i;\upsilon) \right)^2 \right\} \tag{8}$$

So that by using the matrix form (8), the estimator of the B-splines parameter is obtained as follows

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{B}\boldsymbol{\alpha})^T (\mathbf{Y} - \mathbf{B}\boldsymbol{\alpha}) = \mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\alpha}^T\mathbf{B}^T\mathbf{Y} + \boldsymbol{\alpha}^T\mathbf{B}^T\mathbf{B}\boldsymbol{\alpha} \tag{9}$$

The minimum value of $\boldsymbol{\varepsilon}^\mathbf{T}\boldsymbol{\varepsilon}$ is obtained if $\frac{\partial(\boldsymbol{\varepsilon}^\mathbf{T}\boldsymbol{\varepsilon})}{\partial\boldsymbol{\alpha}} = 0$, so $-2\mathbf{B}^\mathbf{T}\mathbf{Y} + 2\mathbf{B}^\mathbf{T}\mathbf{B}\hat{\boldsymbol{\alpha}} = 0$. So, the estimator of B-splines is

$$\hat{\boldsymbol{\alpha}} = (\mathbf{B}^T\mathbf{B})^{-1} \mathbf{B}^T\mathbf{Y} \tag{10}$$

Based on the results in (10), the estimator for the regression model (7) in matrix form is

$$\widehat{\mathbf{Y}} = \mathbf{A}\mathbf{Y} \tag{11}$$

where:

$$\mathbf{A} = \mathbf{B}(\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}\mathbf{B}^\mathbf{T}$$

2.2.3. Derivation and Monotonicity of B-Splines (Constrained B-Splines)

The first derivative of v-order B-splines with v>1 in equation (3) is

$$\frac{\partial B_j(x;v)}{\partial x} = \frac{v-1}{t_{j+v-1}-t_j} B_j(x; v-1) - \frac{v-1}{t_{j+v}-t_{j+1}} B_{j+1}(x; v-1), \tag{12}$$

while the first derivative for v = 1 is equal to 0 [17]. After obtaining the first derivative of B-splines (12) then the first derivative for the B-splines function (5) is

$$\frac{\partial f(x)}{\partial x} = (v-1)\left(\sum_{j=2}^{m} \frac{\alpha_j - \alpha_{j-1}}{t_{j+v-1} - t_j} B_j(x; v-1)\right). \tag{13}$$

From the above results, it is obtained that the value of $\frac{\partial f(x)}{\partial x}$ is affected by the value of $\alpha_j - \alpha_{j-1} = \delta_j$ for j = 2,…,m. So, it can be concluded that for

$$\alpha_j \geq \alpha_{j-1} \ (\delta_j \geq 0), \quad j = 2, \dots, m \tag{14}$$

the value of $\frac{\partial f(x)}{\partial x}$ is non-negative so the B-splines function is a monotonically increasing function. While for

$$\alpha_j \leq \alpha_{j-1} \ (\delta_j \leq 0), \quad j = 2, \dots, m \tag{15}$$

the value of $\frac{\partial f(x)}{\partial x}$ is non-positive so the B-splines function is a monotone-decreasing function.

The addition of monotone constraints as in (14) and (15) is often applied to estimate parameters where the relationship between response variables and predictor variables is assumed to be monotone [14]. The addition of monotone constraints provides a smoothing effect on the estimated regression model [15]. From the first derivative of B-splines in equation (12), the second derivative is

$$\begin{aligned} \frac{\partial^2 B_j(x;v)}{(\partial x)^2} = &\ (v-1)(v-2)\left[\frac{1}{(t_{j+v-1}-t_j)(t_{j+v-2}-t_j)} B_j(x;v-2) \right. \\ &- \left(\frac{1}{(t_{j+v-1}-t_j)} + \frac{1}{(t_{j+v}-t_{j+1})}\right)\frac{1}{(t_{j+v-1}-t_{j+1})} B_{j+1}(x;v-2) \\ &\left. + \frac{1}{(t_{j+v}-t_{j+1})(t_{j+v}-t_{j+2})} B_{j+2}(x;v-2)\right]. \end{aligned} \tag{16}$$

While the second derivative of the B-splines function is

$$\frac{\partial^2 f(x)}{(\partial x)^2} = (v-1)(v-2)\left(\sum_{j=3}^{m} \frac{\frac{\alpha_j - \alpha_{j-1}}{t_{j+v-1} - t_j} - \frac{\alpha_{j-1} - \alpha_{j-2}}{t_{j+v-2} - t_{j-1}}}{t_{j+v-2} - t_j} B_j(x; v-2)\right) \tag{17}$$

From the above results, the second derivative is obtained if the order of the B-splines is $v > 2$.

2.2.4. Quantile Regression

Quantile regression introduced by [18] is an extension of median regression, where quantile regression allows to estimate of quantile functions at various desired quantile values [19]. Suppose Y is a random variable that has a distribution center, denoted c, then the cumulative distribution function $F_Y(.)$ of c is written:

$$F_Y(c) = P(Y \leq c). \tag{18}$$

For $\tau \in [0,1]$, the τ-th quantile of Y which is based on the objective function L_1 (loss-function), indicates the specific locations of a distribution. The function L_1 is defined

$$q_\tau(Y) = {F_Y}^{-1}(\tau) = \inf\{c : F_Y(c) \geq \tau\}. \tag{19}$$

In general, the τ-th quantile of Y can be expressed by minimizing

$$q_\tau(Y) = \operatorname{argmin}_c \mathrm{E}[\rho_\tau(Y - c)], \tag{20}$$

with the function $\rho_\tau(.)$ referred to as the defined 'check-function':

$$\rho_t(z) = \begin{cases} \tau z, & if\ z > 0 \\ -(1-\tau)z, & others \end{cases} \tag{21}$$

Furthermore, from equations (20) and (21) obtained

$$\mathrm{E}[\,\rho_\tau(Y-c)] = (\tau-1)\int_{-\infty}^{c}(Y-c)dF_Y(y) + \tau\int_{c}^{\infty}(Y-c)dF_Y(y). \tag{22}$$

By minimising the first derivative of the function in equation (22) is obtained:

$$0 = (1-\tau)\int_{-\infty}^{c} dF_Y(y) - \tau\int_{c}^{\infty} dF_Y(y) = F_Y(c) - \tau \tag{23}$$

The function $F_Y(.)$ is monotone, so every element of $\{y: F_Y(y) = \tau\}$ minimizes the function (23). From equation (19), it is obtained that $c = F_Y{}^{-1}(\tau)$ is a unique solution. Suppose $Y_1, \dots, Y_n$ are random samples from Y such that $Y_1, \dots, Y_n$ are independently and identically distributed ($i.i.d$) with Y. The empirical cumulative distribution function of $Y_1, \dots, Y_n$ is written:

$$F_n(Y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le y), \tag{24}$$

where I(A) is the indicator of the set A that satisfies the conditions:

$$I(A) = \begin{cases} 1, A\ fulfilled \\ 0, A\ fulfilled \end{cases}$$

The function $F_Y(.)$ can be replaced by $F_n(Y)$ and $\hat{F}_Y{}^{-1}(\tau)$ which is the estimator of $F_Y{}^{-1}(\tau)$ can be obtained by minimizing

$$\mathrm{argmin}_c \int \rho_\tau\,(Y-c)dF_n(Y) = \mathrm{argmin}_c \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\,(Y_i - c). \tag{25}$$

$$Y = \beta_0 + \beta_1 X^{(1)} + \cdots + \beta_p X^{(p)} + \varepsilon = \mathbf{X}^{\boldsymbol{T}}\beta + \varepsilon, \tag{26}$$

with $\beta = (\beta_0, \dots, \beta_p)^T$, $\mathbf{X} = (1, X^{(1)}, \dots, X^{(p)})^T$ and ε is assumed to have a distribution with the notation F. In general, the τ-th quartile of the error (ε) which is

$$F^{-1}(\tau) = \inf\{u: P\{\varepsilon \le u\} \ge \tau\}, \tag{27}$$

with u being the error of the regression model (27). The quantile curve equation for Y conditional on X can be written

$$q_\tau(Y|\mathbf{X}) = [\beta_0 + F^{-1}(\tau)] + \beta_1 X^{(1)} + \cdots + \beta_p X^{(p)} = \mathbf{X}^{\boldsymbol{T}}\beta(\tau) \tag{28}$$

with $\beta(\tau) = \left(\left(\beta_0 + F^{-1}(\tau)\right), \dots, \beta_p\right)^T$. As in the previous discussion, the estimator of the parameter $\hat{\beta}(\tau)$ is obtained by minimizing

$$\min_{\beta}\ \mathrm{E}\left[\rho_\tau\left(Y - \mathbf{X}^{\boldsymbol{T}}\beta(\tau)\right)\right]. \tag{29}$$

Let $\left(X_1{}^{(1)}, \dots, X_1{}^{(p)}, Y_1\right), \dots, \left(X_n{}^{(1)}, \dots, X_n{}^{(p)}, Y_n\right)$ be random samples from $\left(X^{(1)}, \dots, X^{(p)}, Y\right)$ that are independently and identically distributed ($i.i.d$) so that the conditional quantile objective function in equation (28) becomes

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \mathbf{X}_{\boldsymbol{i}}{}^{\boldsymbol{T}}\beta(\tau)\right), \tag{30}$$

with $\mathbf{X}_{\boldsymbol{i}} = \left(1, X_i{}^{(1)}, \dots, X_i{}^{(p)}\right)^T$ being the ith observation of X.

2.2.5. Confidence Interval for Quantile Regression

One system of estimating population parameters based on samples is the confidence interval, which produces more representative parameter estimators than the point estimator system [20]. A confidence interval is an interval between two numbers, where the parameter value of the population lies within the interval. Since quantile regression was introduced, various methods have been used to estimate confidence intervals on quantile regression curves. One of the methods used is the direct method. The direct method is more efficient in estimating confidence intervals than other methods [21].

For $\tau \in [0,1]$ and $\alpha \in (0,1)$, the $(1-\alpha)$ percent confidence interval for the quantile regression curve equation (28) is

$$I_n = \left(\left(\mathbf{X}^{\boldsymbol{T}}\hat{\beta}(\tau - b_n)\right), \left(\mathbf{X}^{\boldsymbol{T}}\hat{\beta}(\tau + b_n)\right)\right) \tag{31}$$

with,

$$b_n = z_\alpha\sqrt{\mathbf{X}\mathbf{Q}^{-1}\mathbf{X}^{\boldsymbol{T}}\tau(\tau-1)}/\sqrt{n},$$

where z_α is the (1-α) standard normal percentile point and

$$\mathbf{Q} = \mathbf{n}^{-1}\left(\mathbf{X}_i\mathbf{X}_i{}^T\right), \tag{32}$$

Where **Q** is a positive definite matrix of size ((*p*+1)×(*p*+1)).

2.2.6. Quantile Regression Smoothing B-Splines

The quantile objective function for smoothing B-Splines in the form of a linear equation is:

$$\min\left\{\widehat{\mathbf{W}}^T\mathbf{u}+\widehat{\mathbf{W}}^T\boldsymbol{\upsilon} \mid \widehat{\mathbf{X}}\boldsymbol{\alpha}+\mathbf{u}-\boldsymbol{\upsilon}=\widehat{\mathbf{Y}},\left(\mathbf{u},\boldsymbol{\upsilon}\in\mathbb{R}_+^{(\mathbf{n}+\mathbf{u})}\right)\right\} \quad (33)$$

Where **u** and $\boldsymbol{\upsilon}$ are vectors of positive and negative parts of the regression residuals.

$$\widehat{\mathbf{W}}_{(n+u)\times 1}=\begin{pmatrix}\mathbf{W}\\ \mathbf{1}_{u\times 1}\end{pmatrix} \quad (34)$$

with $\mathbf{W}=\left(\rho_\tau\left(z_1\right),\ldots,\rho_\tau\left(z_n\right)\right)^T$ is the weight vector

$$\widehat{\mathbf{Y}}=\begin{pmatrix}\mathbf{Y}\\ \mathbf{0}_{u\times 1}\end{pmatrix} \quad (35)$$

$\widehat{\mathbf{Y}}_{(n+u)\times 1}$ is a pseudo response vector with $\mathbf{Y}=\left(y_1,\ldots,y_n\right)^T$

$$\widehat{\mathbf{X}}=\begin{pmatrix}\mathbf{B}\\ \boldsymbol{\lambda}\mathbf{C}\end{pmatrix} \quad (36)$$

$\widehat{\mathbf{X}}_{((n+u)\times m)}$ is a pseudo matrix design with:

$$\mathbf{B}_{n\times m}=\begin{bmatrix}B_1\left(x_1;\upsilon\right) & B_2\left(x_1;\upsilon\right) & \ldots & B_m\left(x_1;\upsilon\right)\\ B_1\left(x_2;\upsilon\right) & B_2\left(x_2;\upsilon\right) & \ldots & B_m\left(x_2;\upsilon\right)\\ \vdots & \vdots & \ddots & \vdots\\ B_1\left(x_n;\upsilon\right) & B_2\left(x_n;\upsilon\right) & \ldots & B_m\left(x_n;\upsilon\right)\end{bmatrix}$$

$$\mathbf{C}_{u\times m}=\begin{bmatrix}B_1'\left(t_{\upsilon+1};\upsilon\right)-B_1'\left(t_\upsilon;\upsilon\right) & \ldots & B_m'\left(t_{\upsilon+1};\upsilon\right)-B_m'\left(t_\upsilon;\upsilon\right)\\ \vdots & \ddots & \vdots\\ B_1'\left(t_m;\upsilon\right)-B_1'\left(t_{m-1};\upsilon\right) & \ldots & B_m'\left(t_m;\upsilon\right)-B_m'\left(t_{m-1};\upsilon\right)\end{bmatrix}$$

The objective function:

$$\widehat{\mathbf{W}}^T\mathbf{u}+\widehat{\mathbf{W}}^T\boldsymbol{\upsilon}$$

The control function:

$$\widehat{\mathbf{X}}\boldsymbol{\alpha}+\mathbf{u}-\boldsymbol{\upsilon}=\widehat{\mathbf{Y}}$$

2.2.7. Selection of Smoothing Parameters and Knots

The criterion for selecting the most optimum smoothing parameter (λ) uses the smallest Schawrz Information Criterion (SIC) value [13], with the formulation:

$$SIC(\lambda)=\log\left(\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-\sum_{j=1}^{m}\hat{\alpha}_j B_j(x_i;v)\right)\right)+\frac{1}{2}p_\lambda\frac{\log(n)}{n} \quad (37)$$

Where p_λ is the sum of the zero residuals for the fitted model.

The number of knots for B-Splines smoothing quantile regression is 20 knots where the location is chosen based on the unique value of the variable X [13]. The u-th knot point (tu) is obtained from:

$$t_u = \text{quantile of } -\left(\frac{u}{20}\right) \text{of the value for variable X;} \quad u = 1, 2, ..., 20 \tag{38}$$

2.2.8. Monotone Constraint Function on Linear Programmes for Quantile Regression

The addition of an increasing or decreasing monotone constraint function when estimating parameters provides a smoothing effect on the curve of a regression model. The criteria for checking monotone constraints on the objective function of quantile regression on Smoothing B-Splines are:

$\mathbf{H}\boldsymbol{\alpha} > \mathbf{0}$, for monotone increasing function

$\mathbf{H}\boldsymbol{\alpha} < \mathbf{0}$, for monotone decreasing function

With:

$$\mathbf{H} = \begin{bmatrix} B_1^{'}(t_\upsilon;\upsilon) & \cdots & B_m^{'}(t_\upsilon;\upsilon) \\ \vdots & \ddots & \vdots \\ B_1^{'}(t_m;\upsilon) & \cdots & B_m^{'}(t_{m-1};\upsilon) \end{bmatrix}$$

$\widehat{\boldsymbol{H}}\boldsymbol{x} > 0$, for monotone increasing function

$\widehat{\boldsymbol{H}}\boldsymbol{x} < 0$, for monotone decreasing function

With:

$$\widehat{\mathbf{H}} = \begin{pmatrix} \mathbf{H} & \mathbf{0}_{((u+2)\times 2(n+u))} \end{pmatrix}$$

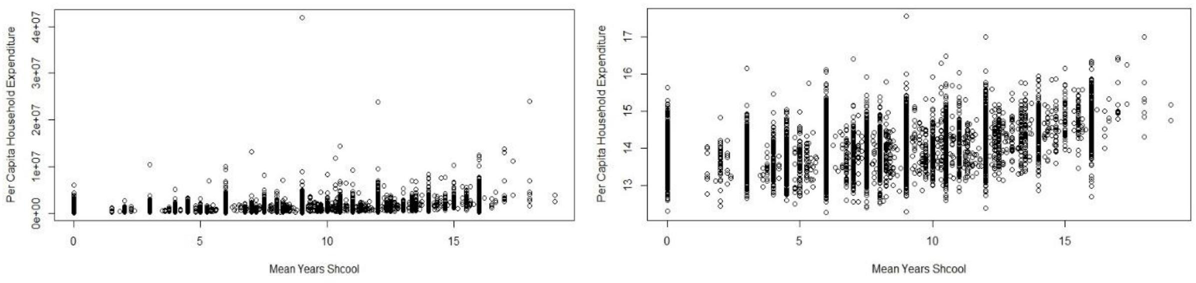
2.2.9. Research Steps

The steps in this study are:

1. Creating a scatter plot between the response variable and the independent variables
2. Performing model specification based on the scatter plot, in this case a B-Splines function approach is used.
3. Checking for outliers in the scatter plot results and if there are outliers then quantiles are used. Checking for outliers can also be done by looking at the distribution of errors with the mean as a measure of data concentration in the B-Splines function.
4. Determine the constraints of the relationship between the two variables, whether it is monotonically increasing or monotonically decreasing.
5. Determine the number of knots and smoothing parameter (λ). In this paper, the B-Splines smoothing function with the number of knots used is 20 knots, and the smoothing parameter (λ) is determined based on the smallest Schawrz Information Criterion (SIC) value.
6. Estimate the quantile regression curve based on the optimal value of smoothing parameter (λ) at several quantile points, namely at $\tau = 0.2; 0.4; 0.6; 0.8$.
7. Estimating the confidence interval for the quantile regression curve by the direct method

3. Results and Discussions

This chapter will explain the relationship between the variables of average years of schooling and average per capita household expenditure in Central Kalimantan Province in 2020 modelled by the COBS method. The reason for using the COBS method is that the data plot (Figure 1) shows a pattern that cannot be clearly specified but has an increasing trend, so it would be better to do the modelling nonparametrically. What is meant by ‘Constrained’ here is the assumption that the relationship between the two data is an increasing pattern which is further referred to as ‘Increase Constrained’. This can be interpreted that the average household expenditure per capita increases along with the average years of schooling of household members.



(a) (b)

Figure 1. Data plots of household expenditure per capita and average years of schooling: (a) original plot; (b) transformation plot

Addition to the irregularity of the data pattern, the next phenomenon is the presence of outlier data (Figure 2) from the residuals of the B-splines model presented in the mean regression. Thus, to capture the phenomenon of the existence of outlier residuals, quantile regression analysis is applied in estimating model parameters.

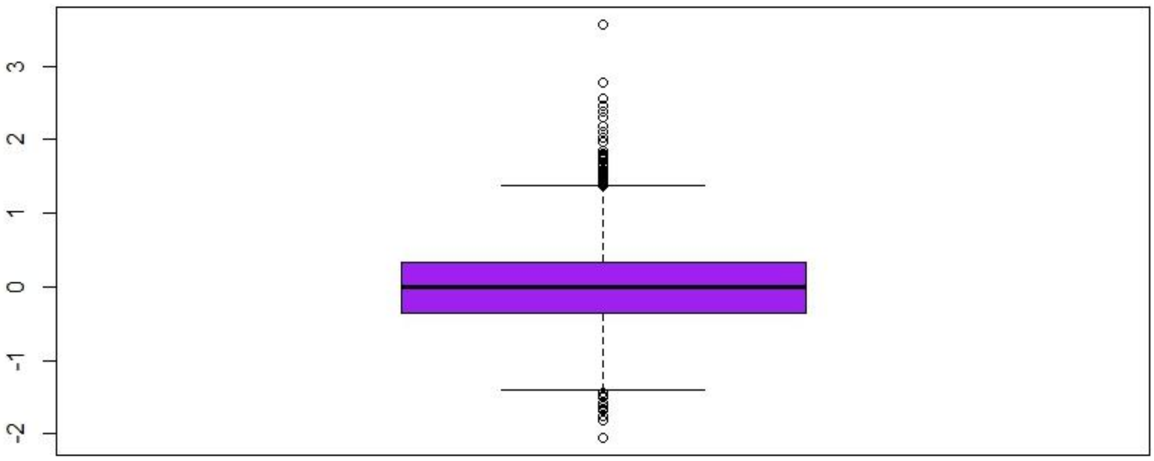


Figure 2. Boxplot of residuals from the b-splines model

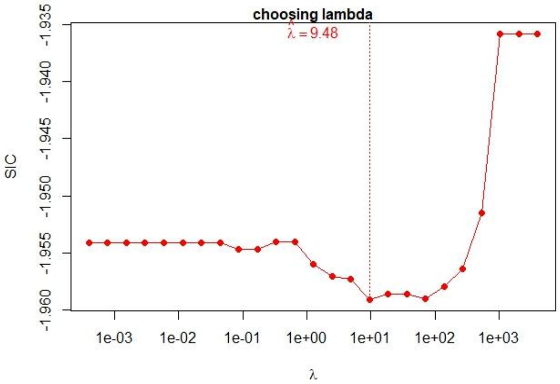
Furthermore, to divide households into groups with similar characteristics based on the average years of schooling of household members and per capita household expenditure, four quantile regression modelling will be applied with the boundaries of the 0.2nd quantile, 0.4th quantile, 0.6th quantile and 0.8th quantile. In quantile regression modelling, each quantile has the same number and location of knot points as presented in Table 2.

Table 2. Knot points at each quantile

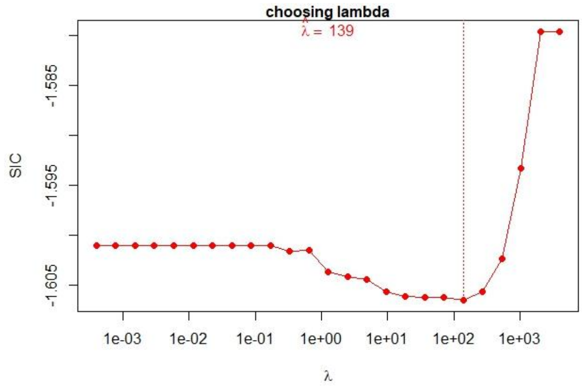
Point	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Knot	0	3.00	4.80	5.60	6.60	7.33	8.14	8.60	9.33	10.12

Point	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}
Knot	10.50	11.25	11.67	12.33	12.83	13.50	14.40	15.00	16.00	19.00

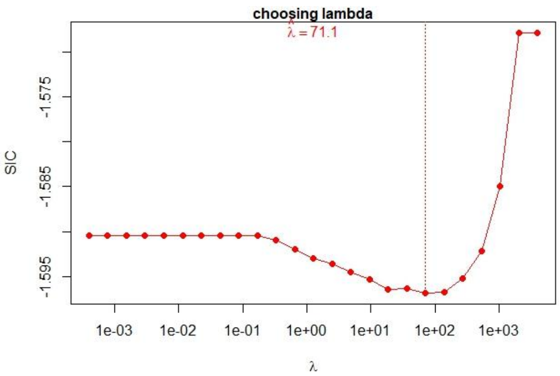
Table 2 shows the knot points calculated by the quantile method from the unique values of the average years of schooling variable as many as 126 values. Meanwhile, the optimum curve smoothing parameter (λ) in each quantile has different values as presented in Figure 3.



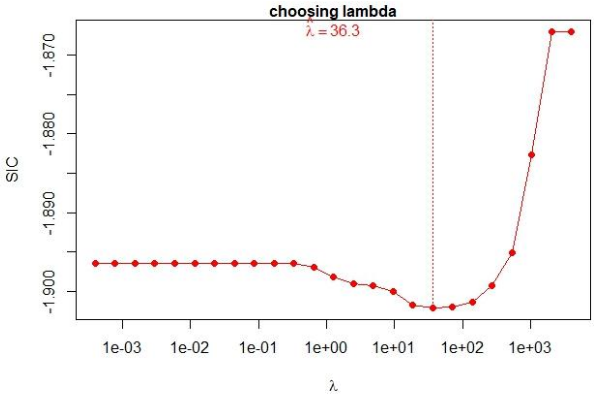
(a) Quantile (τ) = 0.2



(b) Quantile (τ) = 0.4



(c) Quantile (τ) = 0.6



(d) Quantile (τ) = 0.8

Figure 3. Optimum smoothing parameter (λ) based on the smallest SIC value in each quantile

Figure 3 (a) shows that at the 0.2th quantile, the optimum smoothing parameter (λ) is 9.48 with a minimum SIC of -1.9591. This indicates that, at the lower end of the distribution, a relatively smaller λ provides the best smoothing effect, resulting in a more accurate model with minimized information loss. Figure 3 (b) shows that at the 0.4th quantile, the optimum smoothing parameter (λ) is 139 with a minimum SIC of -1.6065. This suggests that, as we move towards the median of the data distribution, a much larger λ is required to achieve optimal smoothing, potentially due to increased variability in this middle range that requires more significant smoothing to reduce the SIC. Figure 3 (c) shows that at the 0.6th quantile, the optimum smoothing parameter (λ) is 71.1 with a minimum SIC of -1.5969. This result indicates a moderate level of smoothing is ideal for the upper-middle quantile, which is lower than that required for the 0.4th quantile but higher than at the 0.2th quantile.

This pattern might reflect changes in data variability or distribution characteristics that affect the model's performance at this quantile level. Figure 3 (d) shows that at the 0.8th quantile, the optimum smoothing parameter (λ) is 36.3 with a minimum SIC of -1.9021. Compared to the lower quantiles, the decrease in the optimum λ suggests less need for aggressive smoothing, possibly due to reduced variability or a different distribution pattern in the upper quantiles. Armed with the optimum knot points and smoothing parameters (λ) that have been obtained at each quantile, the quantile regression curve based on the COBS method for linear B-Splines smoothing with the assumption of monotonous increase (Increase Constrain) is presented in Figure 4.

Figure 4 shows the estimated household expenditure per capita based on the average years of schooling of household members at the 0.2 quantile, 0.4 quantile, 0.6 quantile and 0.8 quantile. At the 0.6 and 0.8 quantiles, an average year of schooling of around seven years and above drastically increases per capita expenditure. The estimated per capita household expenditure is obtained from the quantile regression model which has coefficients (α) as presented in Table 3.

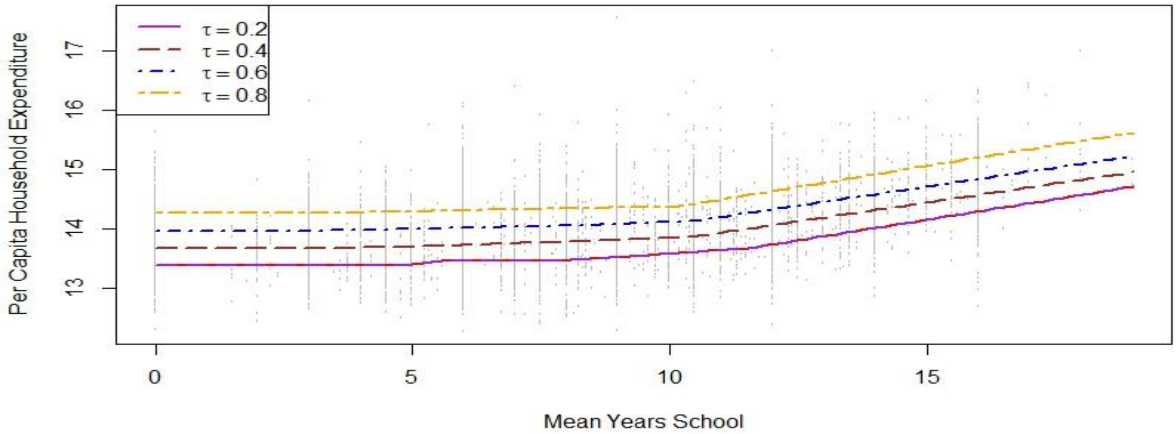


Figure 4. Quantile regression curve with COBS

Table 3 shows that the coefficient value in each quantile tends to increase, this is what makes the quantile regression curve in Figure 4 tend to rise. Thus, the relationship between the average years of schooling of household members and household expenditure per capita tends to increase in line with the assumption (Increase Constrain).

Table 3. Linear b-splines quantile regression coefficients at the 0.2nd, 0.4th, 0.6th and 0.8th quantiles

Knot i^{th} (t_i)	Coefficient (α)			
	τ = 0.2	τ = 0.4	τ = 0.6	τ = 0.8
1	13.3890	13.6773	13.9650	14.2611
2	13.3890	13.6773	13.9650	14.2611
3	13.3890	13.6921	13.9970	14.2906
4	13.4570	13.7163	14.0112	14.3037
5	13.4640	13.7466	14.0289	14.3201
6	13.4692	13.7688	14.0420	14.3321
7	13.4749	13.7933	14.0563	14.3454
8	13.5016	13.8071	14.0720	14.3529
9	13.5444	13.8293	14.0972	14.3649
10	13.5907	13.8533	14.1243	14.3779
11	13.6126	13.8753	14.1372	14.4274
12	13.6565	13.9705	14.2328	14.5329
13	13.6934	14.0234	14.2860	14.5916
14	13.7859	14.1081	14.3710	14.6854
15	13.8553	14.1716	14.4348	14.7558
16	13.9478	14.2563	14.5198	14.8497
17	14.0728	14.3706	14.6346	14.9764
18	14.1560	14.4468	14.7111	15.0608
19	14.2948	14.5738	14.8386	15.2016
20	14.7112	14.9549	15.2213	15.6239

Source: Susenas March 2020 BPS, processed

The quantile regression model with the coefficients presented in Table 3 is as follows:

a) Quantile Regression Model for the 0.2nd Quantile:

$$\begin{aligned}\hat{y} &= \sum_{j=1}^{20} \alpha_j B_j(x; v=2) \\ &= 13.39\left(\tfrac{x}{3.00} B_1(x;1) - \tfrac{x-4.80}{1.80} B_2(x;1)\right) + 13.39\left(\tfrac{x-3.00}{1.80} B_2(x;1) - \tfrac{x-5.60}{0.80} B_3(x;1)\right) + \\ &\quad 13.39\left(\tfrac{x-4.80}{0.80} B_3(x;1) - \tfrac{x-6.60}{1.00} B_4(x;1)\right) + \cdots + 14.16\left(\tfrac{x-15.00}{1.00} B_{18}(x;1) - \tfrac{x-19}{3.00} B_{19}(x;1)\right) + \\ &\quad 14.29\left(\frac{x-16.00}{3.00} B_{19}(x;1)\right) \\ &= 4.46x B_1(x;1) - (7.44x - 35.71) B_2(x;1) + \cdots + (4.76x - 76.21) B_{19}(x;1)\end{aligned}$$

b) Quantile Regression Model for the 0,4th Quantile:

$$\begin{aligned}\hat{y} &= \sum_{j=1}^{20} \alpha_j B_j(x; v=2) \\ &= 13.68\left(\tfrac{x}{3.00} B_1(x;1) - \tfrac{x-4.80}{1.80} B_2(x;1)\right) + 13.68\left(\tfrac{x-3.00}{1.80} B_2(x;1) - \tfrac{x-5.60}{0.80} B_3(x;1)\right) + \\ &\quad 13.69\left(\tfrac{x-4.80}{0.80} B_3(x;1) - \tfrac{x-6.60}{1.00} B_4(x;1)\right) + \cdots + 14.45\left(\tfrac{x-15.00}{1.00} B_{18}(x;1) - \tfrac{x-19}{3.00} B_{19}(x;1)\right) + \\ &\quad 14{,}57\left(\frac{x-16{,}00}{3{,}00} B_{19}(x;1)\right) \\ &= 4.56x B_1(x;1) - (7.60x - 36.48) B_2(x;1) + \cdots + (4.86x - 70.71) B_{19}(x;1)\end{aligned}$$

c) Quantile Regression Model for the 0,6th Quantile:

$$\begin{aligned}\hat{y} &= \sum_{j=1}^{20} \alpha_j B_j(x; v=2) \\ &= 13.96\left(\tfrac{x}{3.00} B_1(x;1) - \tfrac{x-4.80}{1.80} B_2(x;1)\right) + 13.96\left(\tfrac{x-3.00}{1.80} B_2(x;1) - \tfrac{x-5.60}{0.80} B_3(x;1)\right) + \\ &\quad 13.99\left(\tfrac{x-4.80}{0.80} B_3(x;1) - \tfrac{x-6.60}{1.00} B_4(x;1)\right) + \cdots + 14.71\left(\tfrac{x-15.00}{1.00} B_{18}(x;1) - \tfrac{x-19}{3.00} B_{19}(x;1)\right) + \\ &\quad 14.84\left(\frac{x-16.00}{3.00} B_{19}(x;1)\right) \\ &= 4.65x B_1(x;1) - (7.76x - 37.23) B_2(x;1) + \cdots + (4.95x - 79.15) B_{19}(x;1)\end{aligned}$$

d) Quantile Regression Model for the 0,8th Quantile:

$$\begin{aligned}\hat{y} &= \sum_{j=1}^{20} \alpha_j B_j(x; v=2) \\ &= 14.26\left(\tfrac{x}{3.00} B_1(x;1) - \tfrac{x-4.80}{1.80} B_2(x;1)\right) + 14.26\left(\tfrac{x-3.00}{1.80} B_2(x;1) - \tfrac{x-5.60}{0.80} B_3(x;1)\right) + \\ &\quad 14.29\left(\tfrac{x-4.80}{0.80} B_3(x;1) - \tfrac{x-6.60}{1.00} B_4(x;1)\right) + \cdots + 15.06\left(\tfrac{x-15.00}{1.00} B_{18}(x;1) - \tfrac{x-19}{3.00} B_{19}(x;1)\right) + \\ &\quad 15{,}20\left(\frac{x-16{,}00}{3{,}00} B_{19}(x;1)\right) \\ &= 4.75x B_1(x;1) - (7.92x - 38.03) B_2(x;1) + \cdots + (5.07x - 81.07) B_{19}(x;1)\end{aligned}$$

Next, we calculate the estimated per capita expenditure at several average years of schooling of a person indicating a certain level of education. Some of the education levels used in the estimation of per capita expenditure include 0 years (no/never been to school), 6 years (elementary school/equivalent), 9 years (junior high school/equivalent), 12 years (senior high school/equivalent), 13 years (Diploma I/II), 15 years (Academy/Diploma III), 16 years (Diploma IV/Bachelor's degree), 18 years (Master's/Secondary degree) and 22 years (Doctoral degree).

Table 4. Estimated value of household expenditure by average years of schooling at the 0.2, 0.4, 0.6 and 0.8 quantiles

Mean Years School (Years)	Estimated Expenditure Per Capita (IDR)			
	τ = 0.2	τ = 0.4	τ = 0.6	τ = 0.8
0	652,763	870,933	1,161,264	1,561,410
6	700,668	916,594	1,224,790	1,640,104
9	747,844	1,003,746	1,310,239	1,722,746
12	926,936	1,284,359	1,670,244	2,277,320
13	1,064,952	1,458,297	1,897,447	2,621,560
15	1,405,692	1,880,033	2,448,779	3,473,975
16	1,614,991	2,134,664	2,781,887	3,999,060
18	2,131,721	2,752,004	3,590,205	5,299,431
19	2,449,122	3,124,704	4,078,582	6,100,431

Source: Susenas March 2020 BPS

Table 4 informs that in the 0.2 quantile, household members with an average year of schooling of 0 years have a per capita expenditure of IDR 652,763, while household members with an average year of schooling of 6 years have a per capita expenditure of IDR 700,668, and so on until household members with an average year of schooling of 19 years have a per capita expenditure of IDR 2,449,122. In the 0.4th quantile shows members who have never been to school have per capita expenditure of IDR 870,933, household members with an average length of schooling of 6 years will have per capita expenditure of IDR 916,594, and so on until household members with an average length of schooling of 19 years have per capita expenditure of IDR 3,124,704. In the 0.6th quantile shows members who do not / have never been to school have per capita expenditure of IDR 1,161,264, household members with an average length of schooling of 6 years will have per capita expenditure of IDR 1,224,790, and so on until household members with an average length of schooling of 19 years have per capita expenditure of IDR 4,078,582. At the 0.8 quantile, members who have never been to school have a per capita expenditure of IDR 1,561,410, household members with an average year of schooling of 6 years will have a per capita expenditure of IDR 5,299,431, and so on until household members with an average year of schooling of 19 years have a per capita expenditure of IDR 6,100,431.

Based on the estimated per capita expenditure in each quantile, the classification of households based on average years of schooling and per capita household expenditure in Kalimantan Tengah Province in 2020 will be determined. The classification for each average years of schooling are as follows:

- 'Very poor' if the per capita expenditure is less than IDR 652,763.
- 'Poor' if the per capita expenditure is between IDR 652,763 and IDR 870,933.
- 'Middle' if the per capita expenditure is between IDR 870,933 and IDR 878,788.91.
- 'Rich' if the per capita expenditure is between IDR 878,788.91 and IDR 1,161,264; and
- 'Very rich' if the per capita expenditure is above IDR 1,161,264.

The complete classification of households with an average year of schooling of 6 years, 9 years, 12 years, 13 years, 15 years, 16 years, 18 years and 19 years is presented in Table 5.

Figure 5 illustrates the relationship between the average years of schooling on the x-axis and household expenditure on the y-axis at 95 percent significance level. The solid red line in Figure 5 (a) represents the estimated trend of household expenditure, while the dashed lines likely represent confidence intervals around the trend estimate of household espenditure in quantile = 0.2. The dashed yellow line in Figure 5 (b) represents the estimated trend of household expenditure, while the dashed lines likely represent confidence intervals around the trend estimate of household espenditure in quantile = 0.4. The dashed purple line in Figure 5 (c) represents the estimated trend of household expenditure, while the dashed lines likely represent confidence intervals around the trend estimate of household espenditure in quantile = 0.6. The dashed black line in Figure 5 (d) represents the estimated trend of household expenditure, while the dashed lines likely represent confidence intervals around the trend estimate of household espenditure in quantile = 0.8. The detail of convidence intervals each quantile presented in Table 6.

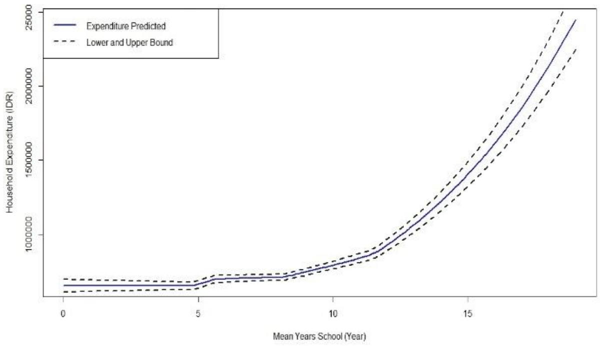
Table 5. Classification of households by average years of schooling and per capita expenditure

Mean Years School (Years)	Estimated Expenditure per Capita (IDR)				
	Very Poor	Poor	Midle	Rich	Very Rich
0	< 652,763	652,763 - 870,933	870,933 - 1,161,264	1,161,264 - 1,561,410	> 1,561,410
6	< 700,668	700,668 - 916,594	916,594 - 1,224,790	1,224,790 - 1,640,104	> 1,640,104
9	< 747,844	747,844 - 1,003,746	1,003,746 - 1,310,239	1,310,239 - 1,722,746	> 1,722,746
12	< 926,936	926,936 - 1,284,359	1,284,359 - 1,670,244	1,670,244 - 2,277,320	> 2,277,320
13	< 1,064,952	1,064,952 - 1,458,297	1,458,297 - 1,897,447	1,897,447 - 2,621,560	> 2,621,560
15	< 1,405,692	1,405,692 - 1,880,033	1,880,033 - 2,448,779	2,448,779 - 3,473,975	> 3,473,975
16	< 1,614,991	1,614,991 - 2,134,664	2,134,664 - 2,781,887	2,781,887 - 3,999,060	> 3,999,060
18	< 2,131,721	2,131,721 - 2,752,004	2,752,004 - 3,590,205	3,590,205 - 5,299,431	> 5,299,431
19	< 2,449,122	2,449,122 – 3,124,704	3,124,704 – 4,078,582	4,078,582 – 6,100,431	> 6,100,431

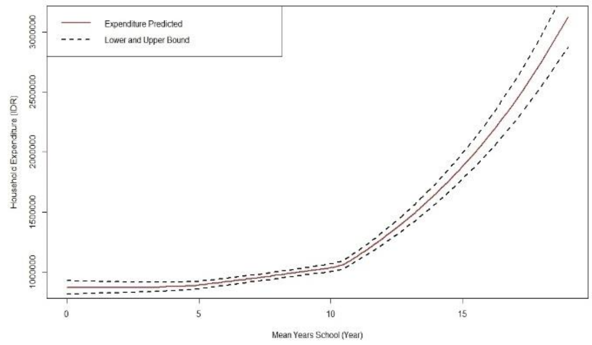
Source: Susenas 2020 March

The four graphs collectively illustrate a distinct pattern in the expenditure distribution of households based on their average years of schooling. At lower levels of education, particularly when the average year of schooling is between 0 to 6 years, the range of household expenditures is relatively broad. This indicates a high variability in spending among households with minimal education; some households may have very low expenditures, possibly due to limited income-earning opportunities, while others might still maintain moderate levels of expenditure despite lower education levels. This variability suggests diverse economic situations even among households with similarly low education.

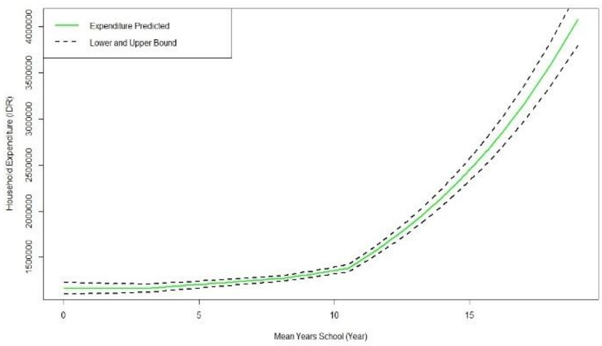
Overall, this pattern demonstrates that both low and high education levels are associated with greater variability in household expenditure, while households with moderate levels of education (6-13 years) exhibit more consistent expenditure levels. The findings imply that education significantly influences economic stability and expenditure behavior, with moderate education levels fostering a more uniform economic condition among households, while very low or very high education levels lead to a wider range of economic outcomes.



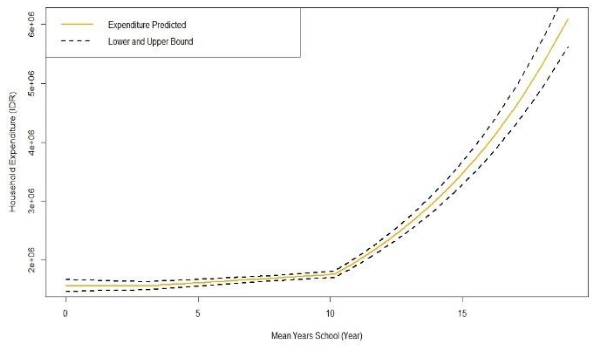
(a) Quantile (τ) = 0.2



(b) Quantile (τ) = 0.4



(c) Quantile (τ) = 0,6



(d) Quantile (τ) = 0,8

Figure 5. Interval of quantile regression curves with COBS for each quantile: (a) 0.2nd quantile; (b) 0.4th quantile; (c) 0.6th quantile; (d) 0.8th quantile

Table 6 presents the lower and upper bounds of household expenditure by average years of schooling in each quantile. All lower and upper bounds show an increasing trend both from the lowest average years of schooling (0 years) to the highest average years of schooling (19 years) and an increasing trend from the 0.2 to the 0.8 quantile. This indicates that there is an Increase Constrained assumption in the average years of schooling data that significantly affects the increase in per capita household expenditure.

Table 6. Lower and upper bound of per capita household expenditure by average years of schooling and quantiles

Mean Years School (Years)	Estimated Expenditure Per Capita (IDR)							
	Quantile (τ) = 0.2		Quantile (τ) = 0.4		Quantile (τ) = 0.6		Quantile (τ) = 0.8	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
0	610,644	697,788	816,621	928,857	1,098,394	1,227,733	1,465,592	1,663,509
6	677,720	724,393	887,594	946,532	1,191,222	1,259,304	1,589,055	1,692,792
9	725,474	770,905	974,744	1,033,612	1,277,455	1,343,878	1,673,771	1,773,153
12	889,345	966,117	1,234,048	1,336,734	1,613,522	1,728,942	2,189,536	2,368,625
13	1,016,179	1,116,077	1,393,766	1,525,831	1,824,634	1,973,146	2,507,408	2,740,909
15	1,324,864	1,491,451	1,775,584	1,990,646	2,330,702	2,572,812	3,284,024	3,674,913
16	1,512,100	1,724,883	2,003,227	2,274,703	2,633,173	2,938,971	3,756,770	4,257,020
18	1,968,613	2,308,365	2,548,435	2,971,805	3,359,457	3,836,765	4,913,613	5,715,487
19	2,245,772	2,670,884	2,873,870	3,397,397	3,793,957	4,384,517	5,618,414	6,623,801

Source: Susenas 2020 March

4. Conclusion

This research draws several conclusions regarding household expenditure and schooling through the quantile regression model formed by the COBS method. The model is divided into four quantiles—0.2, 0.4, 0.6, and 0.8—with all coefficients displaying an increasing trend both from the smallest knot to the largest knot and across quantiles from smallest to largest. Meanwhile, the estimated per capita household expenditure shows a similar upward trend along these quantiles and knots. Additionally, based on the estimated values of household expenditure per capita and average years of schooling, households can be classified economically as very poor, poor, middle class, rich, and very rich. This classification suggests that higher average years of schooling among household members significantly influence increased per capita expenditure. In other words, household members with a higher level of education have a higher level of welfare. Quantile regression modelling with COBS is still limited to using only one predictor variable, it is better to use other methods such as P-Spline which can accommodate the use of more than one predictor variable.

Acknowledgments

The authors would like to thank BPS Kalimantan Tengah Province, BPS North Sumatra Province, BPS South Halmahera District, Central Bureau of Statistics, Jakarta, BPS Pidie District, and Politeknik STIS.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available at BPS-Statistics Kalimantan Tengah Province.

Credit Authorship

Yoga Sasmita: Conceptualization, Data Collection, Formal Analysis, Writing–Original Draft. **Muhammad Budiman Johra:** Transformation Data. **Yogo Aryo Jatmiko:** Methodology. **Deltha A. Lubis:** Writing–Review. **Rizal Rahmad:** Editing**. Gama Putra Danu Sohibien:** Supervision.

References

[1] BPS, *Data and Poverty Information of Regencies/Cities in 2014*, 2014th ed., no. 112. Jakarta: Subdirectorate of Social Vulnerability Statistics, 2014.

[2] A. D. Steer, *The New Era of Poverty Reduction in Indonesia*. Jakarta: The World Bank, 2007.

[3] M. Andersson, A. Engvall, and A. Kokko, *Determinants of poverty in Lao DPR*, no. March, 2006.

[4] Suparno, *Analysis of Economic Growth and Poverty Reduction: A Study of Pro Poor Growth Policy in Indonesia*. Bogor: Institut Pertanian Bogor, 2010.

[5] E. A. Hanushek and L. Woessmann, "The Role of Cognitive Skills in Economic Development," *J. Econ. Lit.*, vol. 46, no. 3, pp. 607–668, 2008.

[6] BPS, *Expenditure and Consumption Pattern of Indonesian People 2013*. Jakarta: Household Statistics Subdirectorat, 2013.

[7] M. R. Montecel, "Education as Pathway Out of Poverty," *Idra Newsl.*, vol. XL, pp. 1–8, 2013.

[8] G. Psacharopoulos and H. Patrinos, "Returns to Investment in Education Result," *Economics*, no. April, p. 25, 2018.

[9] M. Poletti Laurini and M. Moura, "Constrained smoothing B-splines for the term structure of interest rates," *Insur. Math. Econ.*, vol. 46, no. 2, pp. 339–350, 2010, doi: 10.1016/j.insmatheco.2009.11.008.

[10] D. Li, K. Qin, and H. Sun, "Curve modeling with constrained B-spline wavelets," *Comput. Aided Geom. Des.*, vol. 22, no. 1, pp. 45–56, 2005, doi: 10.1016/j.cagd.2004.08.004.

[11] H. Fujioka and H. Kano, "Control theoretic B-spline smoothing with constraints on derivatives," *Proc. IEEE Conf. Decis. Control*, no. 1, pp. 2115–2120, 2013, doi: 10.1109/CDC.2013.6760194.

[12] Y. Zhao, M. Zhang, Q. Ni, and X. Wang, "Adaptive Nonparametric Density Estimation with B-Spline Bases," *Mathematics*, vol. 11, no. 2, pp. 1–12, 2023, doi: 10.3390/math11020291.

[13] X. He and P. Ng, "Cobs: Qualitatively constrained smoothing via linear programming," *Comput. Stat.*, vol. 14, no. 3, pp. 315–337, 1999, doi: 10.1007/s001800050019.

[14] K. Kagerer, "A Short Introduction to Splines in Least Squares Regression Analysis," no. 472, p. 51, 2013.

[15] P. Dierckx, "Curve and Surface Fitting with Splines," *Numer. Math. Sci. Comput.*, Mar. 1993, doi: 10.1093/oso/9780198534419.001.0001.

[16] BPS, *Encyclopedia of Social Economic Indicators*. Jakarta: BPS, 2011.

[17] J. R. and C. de Boor, "A Practical Guide to Splines.," *Math. Comput.*, vol. 34, no. 149, p. 325, 1980, doi: 10.2307/2006241.

[18] R. Koenker and G. Bassett, "Regression Quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[19] L. Hudoyo, "Modelling the Relationship between Average Years of Schooling and Household Expenditure Using Constrained B-Splines (COBS) in Quantile Regression," Padjadjaran University, 2017.

[20] E. Walpole, *Introduction to Statistics*, Third. London: MacMilan, 1982.

[21] K. Q. Zhou and S. L. Portnoy, "Direct Use of Regression Quantiles To Construct," *Ann. Stat.*, vol. 24, no. March 1994, pp. 287–306, 1996.