Panel Data Regression with Feasible Generalized Least Squares for Analyzing School Participation Rate Among Adolescents in Indonesia

Retno Mayapada[1*], Andi Harismahyanti[2], and A. Muthiah Nur Angriany[3]
[1]Author Affiliation: Statistics Study Program, Universitas Sulawesi Barat, Majene, Indonesia
[2]Author Affiliation: Statistics Study Program, Tadulako University, Palu, Indonesia
[3]Author Affiliation: Departemen of Statistics, Hasanuddin University, Makassar, Indonesia
*Corresponding author: retnomayapada@unsulbar.ac.id

Received: 17 June 2026
Revised: 11 August 2026
Accepted: 12 August 2026

ABSTRACT – **School participation among adolescents aged 16-18 years varies across provinces in Indonesia, reflecting disparities in educational access. This study examines the associations of the Unemployment Rate (UR), Average Hourly Earnings of Employees (AHE), Mean Years of Schooling (MYS), and Percentage of Poor Population (PPP) with the School Participation Rate (SPR) with a panel dataset comprising 34 provinces observed from 2018 to 2023. Selection between the panel specifications was based on the Chow and Hausman tests, with the results favoring the Fixed Effects Model (FEM). Diagnostic examination identified heteroskedasticity, serial correlation, and cross-sectional dependence. The model was therefore re-estimated using the modified Feasible Generalized Least Squares (FGLS) estimator proposed by Bai, Choi, and Liao, which accounts for these error dependencies. Results show that AHE, MYS, and PPP were positively associated with SPR, with significant relationships at the 5% level, whereas UR was positively but not significantly associated with SPR. These findings indicate that household economic capacity, educational attainment, and poverty levels are significantly associated with differences in school participation across provinces, while provincial unemployment conditions do not show a statistically significant association. The findings provide evidence on socioeconomic factors associated with school participation among adolescents aged 16-18 years in Indonesia.**

Keywords – **Autocorrelation, Feasible Generalized Least Squares, Heteroskedasticity, Panel Data Regression, School Participation Rate**

I. INTRODUCTION

Education plays a fundamental role in human capital development, particularly during the period when individuals transition from schooling to labor market participation. This transition represents a critical stage in adolescents' lives, as securing employment is often considered a key developmental milestone during this period [1]. For adolescents aged 16-18 years, the decision to continue their education or enter the workforce is influenced by various socioeconomic factors, including household income, parental educational attainment, and access to educational resources [2]. In Indonesia, despite the implementation of the 12-year compulsory education policy, substantial disparities in the School Enrolment Rate (SER; Angka Partisipasi Sekolah/APS) among upper secondary school-age adolescents persist across provinces. Data from [3] in 2023 indicate that the SER for individuals aged 16-18 years ranged from 64.15% to 91.17% across the country's 34 provinces. These disparities suggest unequal opportunities for adolescents to remain engaged in secondary education, potentially affecting their future educational attainment and labor market outcomes.

Adolescents' school participation is influenced not only by individual factors but also by the socioeconomic conditions of their surrounding environment. Economic instability and labor market pressures can affect adolescents' decisions to continue their education or enter the workforce prematurely. Research by [4] shows that parental employment status has a negative impact on children's educational continuity, particularly during adolescence, a period characterized by the transition from school to work. In addition to labor market conditions, wage policies may also influence adolescents' educational decisions. Changes in the minimum wage can alter the opportunity cost of education, thereby affecting whether adolescents choose to remain in school or enter the labor market. Research by [5] demonstrated that increases in the minimum wage can reduce the likelihood of school dropout, especially among adolescents from families with low socioeconomic status (SES). These findings suggest that wage policies have the potential to enhance educational continuity and promote school participation among adolescents.

Adolescents' school participation is influenced by various socioeconomic conditions that may differ across regions and change over time. This suggests that the effects of socioeconomic factors on school participation cannot be adequately examined by focusing on a single region or a specific period alone. Therefore, this study employs panel data, which combine cross-sectional and time-series data [6], allowing observations across provinces and over time simultaneously. The panel data are analyzed using panel regression to examine the effects of independent variables on the dependent variable. In panel data regression analysis, three modeling approaches are commonly used: the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The use of panel data has also been demonstrated in a study by [7], which investigated factors affecting the School Enrolment Rate (SER) of individuals aged 16-18 years in East Nusa Tenggara during 2017-2021 and identified the REM as the most appropriate model. Panel data enable the simultaneous analysis of changes in the SER among adolescents aged 16-18 years and the socioeconomic factors affecting it across provinces and over time.

Although panel data regression can be used to analyze the effects of socioeconomic factors on the School Enrolment Rate (SER) of individuals aged 16-18 years, such models may be subject to violations of classical assumptions, particularly heteroskedasticity and autocorrelation. Differences in characteristics across provinces, as well as changes in socioeconomic conditions over time, may lead to non-constant error variances and correlations among errors across periods. If these violations are ignored, the resulting parameter estimates may become inefficient. One method that can be employed to address these issues is the Feasible Generalized Least Squares (FGLS) estimator. By accounting for heteroskedasticity and autocorrelation in panel data, FGLS can yield more efficient parameter estimates than Ordinary Least Squares (OLS). [8]. The application of FGLS in panel data analysis has also been demonstrated in several studies, including those conducted by [9] and [10]. Pasaribu and Ariani in [9] analyzed the factors influencing the growth of Gross Regional Domestic Product (GRDP) across regencies and municipalities in East Nusa Tenggara Province. Meanwhile, Simarmata and Khoirunurrofik in [10] applied FGLS to analyze data on access to early childhood education services in Indonesia.

However, conventional FGLS may be limited when the panel error structure simultaneously exhibits heteroskedasticity, serial correlation, and cross-sectional dependence. To address these conditions, Bai et al. in [11] proposed a modified FGLS estimator that estimates the large error covariance matrix using a combination of thresholding and banding procedures. The thresholding procedure is used to accommodate cross-sectional correlations, while banding is used to account for serial correlations in the error structure [11][12]. This approach allows the covariance structure to be estimated without imposing a specific parametric form and is therefore suitable for panel data in which the dependence structure is not known a priori.

Although FGLS has been applied in various panel data studies, its application to the School Participation Rate (SPR), particularly among individuals aged 16-18 years in Indonesia, remains relatively limited. Previous research on SPR has predominantly relied on standard panel specifications, including CEM, FEM, and REM, while the joint presence of heteroskedasticity, serial correlation, and cross-sectional dependence has received limited attention. Therefore, this study examines the effects of the Unemployment Rate (UR; Tingkat Pengangguran Terbuka/TPT), Average Hourly Earnings of Employees (AHE; Upah Rata-Rata Per Jam Pekerja/UP), Mean Years of Schooling (MYS; Rata-rata Lama Sekolah/RLS), and Percentage of Poor Population (PPP; Persentase Penduduk Miskin/PPM) on the SPR of individuals aged 16-18 years in Indonesia using the modified FGLS estimator proposed by Bai et al. in [11].

II. LITERATURE REVIEW

A. *Panel Data Regression*

Panel data regression analysis is a statistical technique employed to investigate the impact of multiple predictor variables on a response variable within a panel data framework [13]. The general specification can be expressed as follows in Equation (1).

$$y_{it} = \beta_0 + \Sigma_{k=1}^{K}\beta_k x_{kit} + \varepsilon_{it} \tag{1}$$

where y_{it} denotes the dependent variable for unit i in period t. The term β_0 represents the constant term, while β_k denotes the coefficient associated with the kth explanatory variable. The notation x_{kit} refers to the value of explanatory variable k observed for unit i at time t, and ε_{it} represents the corresponding error term. Here, $i = 1,2,\ldots,N$ indexes the cross-sectional units, whereas $= 1,2,\ldots,T$ indicates the time periods.

Three alternative model specifications are generally considered in panel data analysis, namely CEM, FEM, and REM. The CEM combines data by disregarding differences in time and location, as illustrated in Equation (1). The FEM is used in panel data analysis to estimate differences in intercepts between cross-sectional units and time, assuming constant intercepts. In contrast, the REM is an estimation method in panel data analysis where the disturbance variables (error terms) may be correlated across time and individuals. The FEM and REM equations are presented in Equations (2) and (3), respectively [13].

$$y_{it} = \beta_{0i} + \Sigma_{k=1}^{K}\beta_k x_{it} + \varepsilon_{it} \tag{2}$$

$$y_{it} = \beta_0 + \sum_{k=1}^{k}\beta_k x_{kit} + u_{it}; \quad u_{it} = \mu_i + \varepsilon_{it} \tag{3}$$

μ_i denotes the error on the ith individual observation.

Several specifications can be considered in panel data regression, namely the CEM, FEM, and REM. These models differ in how they account for unobserved individual heterogeneity; therefore, selecting the appropriate specification requires a formal model comparison procedure. The Chow and Hausman tests are commonly applied to identify the specification supported by the data.

1. The Chow Test

The Chow test is a statistical technique employed to identify the optimal model in panel data analysis, specifically between FEM and CEM. This test aims to examine whether all provinces, as cross-sectional units, exhibit homogeneous behavior with respect to the observed variables (Abdussamad et al., 2024). The hypotheses used were as follows.

$H_0: \beta_{01} = \beta_{02} = \cdots = \beta_{0n}$ (CEM)

$H_1: \exists i \neq i^*$ such that $\beta_{0i} \neq \beta_{0i^*};\ i, i^* = 1,2,\cdots,n$ (FEM)

The test statistics used are [13]

$$F = \frac{(SSE_1 - SSE_2)/(n-1)}{SSE_2/(nT-n-k)} \sim F_{\alpha;(n-1),(nT-n-k)} \quad (4)$$

The notation n represents the number of individuals or locations observed and T indicates the number of observation periods. The number of predictor variables in the analysis is denoted as k. Furthermore, SSE_1 refers to the sum of squared errors from the estimation using CEM, whereas SSE_2 refers to the sum of squared errors from the estimation using FEM. If the value of F is greater than or equal to $F_{\alpha;(n-1),(nT-n-k)}$, or the p-value is less than α, H_0 is rejected, indicating that the FEM is more appropriate. Once the FEM is selected, the next step is to perform the Hausman test.

2. The Hausman Test

The Hausman test evaluates whether the unobserved individual effects are correlated with the explanatory variables in the regression model. When no such correlation is detected, the REM is preferred over the FEM. The hypotheses tested are formulated as follows:

$H_0: \mathrm{corr}(x_{it}, e_{it}) = 0$ (REM)

$H_1: \mathrm{corr}(x_{it}, e_{it}) \neq 0$ (FEM)

The test statistics for the hausman test are follows [13]:

$$\chi^2(k) = (\mathbf{b} - \boldsymbol{\beta})'[\mathrm{var}(\mathbf{b} - \boldsymbol{\beta})]^{-1}(\mathbf{b} - \boldsymbol{\beta}) \sim \chi^2_{\alpha;k} \quad (5)$$

When the test statistic $\chi^2(k)$ exceeds the critical value $\chi^2_{\alpha;k}$ or the p-value falls below the chosen significance level, the null hypothesis is rejected, indicating that the FEM is the preferred model.

B. Fixed Effect Model (FEM) Estimation

The Fixed Effects Model (FEM) is a panel regression specification in which the intercept is allowed to vary across individual units, while the slope coefficients of the independent variables remain constant across individuals. Although structurally different from CEM, this model is widely used in panel data analysis because it can capture individual characteristics that remain constant over time [14]. This model does not assume that individual characteristics are random but rather constant, so their effects must be eliminated to avoid contaminating the estimation of the structural parameters β_k. To eliminate the influence of β_{0i}, an individual transformation technique (within transformation) is used, which aims to subtract each observation value from the time average for each individual. The first step of this process is to calculate the average of each variable over time for individual i, which is formulated as follows:

$$\bar{y}_i = \frac{1}{T}\Sigma_{t=1}^{T} y_{it}, \qquad \bar{X}_{ki} = \frac{1}{T}\Sigma_{t=1}^{T} X_{kit}, \qquad \bar{\varepsilon}_i = \frac{1}{T}\Sigma_{t=1}^{T} \varepsilon_{it}$$

For each individual i, adjustments were made to the values $y^*_{it}, X^*_{it}, \varepsilon^*_{it}$, representing the deviations from their respective means.

$$y^*_{it} = y_{it} - \bar{y}_i, \qquad X^*_{kit} = X_{kit} - \bar{X}_{ki}, \qquad \varepsilon^*_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$$

Given that β_{0i} remains constant over time, it is eliminated through a transformation process. Consequently, the transformed model is expressed as,

$$y^*_{it} = \Sigma_{k=1}^{K} \beta_k X^*_{it} + \varepsilon^*_{it} \quad (6)$$

This equation (6) can be reformulated in matrix notation as:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad (7)$$

The parameter β is subsequently estimated using the Ordinary Least Squares (OLS) method applied to the transformed dataset, specifically:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{`*}\mathbf{X}^*)^{-1}\,(\mathbf{X}^*\mathbf{y}^*) \quad (8)$$

C. Feasible Generalized Least Square for Fixed Effect Model

Feasible Generalized Least Squares (FGLS) represents an advancement of the Generalized Least Squares (GLS) methodology, employed when the classical assumptions of Ordinary Least Squares (OLS) are not satisfied. A fundamental assumption of OLS is homoscedasticity, which requires the error term to have a constant variance and is mathematically denoted by $E(\varepsilon'\varepsilon) = \sigma^2 I$. When this assumption is breached, indicating that the variance may not be constant (i.e., heteroscedasticity is present), the GLS method is utilized to estimate the regression coefficients. The estimation of β in the GLS framework is achieved by initially transforming the linear regression model to align with OLS assumptions. The assumptions inherent in the GLS method are $E(\varepsilon) = 0$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Sigma = \sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a known, symmetric, positive-definite, and non-singular matrix of dimensions n × n. Consequently, $\boldsymbol{\Omega}$ can be decomposed as:

$$\boldsymbol{\Omega} = \mathbf{C}\Delta\mathbf{C}^{`} = \left(\mathbf{C}\Delta^{1/2}\right)\left(\Delta^{1/2}\mathbf{C}^{`}\right) = \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}' \quad (9)$$

the column vectors of C are the eigenvectors of $\boldsymbol{\Omega}$, $\boldsymbol{\Delta}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Omega}$, and $\Delta^{1/2}$ is a diagonal matrix whose i-th diagonal element is $\sqrt{\lambda_i}$. Based on Equation (9), the inverse of matrix $\boldsymbol{\Omega}$ can be expressed as:

$$\boldsymbol{\Omega}^{-1} = (\boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{B}}')^{-1} = \mathbf{C}\Delta^{-1/2}\Delta^{-1/2}\mathbf{C} = \mathbf{G}\mathbf{G}' \quad (10)$$

The transformed fixed effect panel data regression model in Equation (8), when applying the parameter estimation concept of GLS, can be articulated as in Equation (11):

$$\mathbf{G}\mathbf{y}^* = \mathbf{G}\mathbf{X}^*\boldsymbol{\beta} + \mathbf{G}\varepsilon^* \quad (11)$$

thus, the form of the fixed effect panel data regression model that satisfies the classical assumptions, based on Equations (8) and (11), can be expressed as:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta}^* + \tilde{\boldsymbol{\varepsilon}} \quad (12)$$

with $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}') = \mathbf{G}\boldsymbol{\sigma}^2\boldsymbol{\Omega}\mathbf{G}' = \sigma^2\mathbf{I}$. Since the variance now adheres to the assumption of homoskedasticity, the OLS method can be applied to the transformed regression model, yielding the following estimator for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{y}}) \quad (13)$$

[15]

In practical applications, the structure of the covariance matrix Ω is frequently unknown, rendering the direct application of the GLS method infeasible. Consequently, the FGLS method is employed, which constitutes a variant of GLS that incorporates an estimation of Ω. The FGLS method substitutes the unknown matrix Ω with an estimator that is consistent with its structure, denoted as $\Omega = \Omega(\beta)$. This method not only addresses heteroskedasticity but also effectively manages autocorrelation, a common issue in panel data [16]. Let $\hat{\beta}$ represent the consistent estimator of the parameter β; it can be formally expressed as:

$$\lim_{n\to\infty} \mathrm{P}\{|\hat{\beta} - \beta| < \delta\} = 1 , \delta > 0$$

It is indicating that the probability of $\hat{\beta}$ being within the δ interval and β approaches one as the sample size increases. Assuming that $\hat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$, it follows that asymptotically $\hat{\boldsymbol{\Omega}} \to \boldsymbol{\Omega}$, thereby allowing the FGLS estimator to approximate the performance of the ideal GLS estimator [15]. In the context of a fixed-effects panel data model, the within-transformation approach is initially applied to eliminate the influence of unobservable individual fixed effects. Subsequently, parameter estimation is conducted using FGLS in the following form:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\hat{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}'\hat{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{y}}) \quad (14)$$

D. Modified Feasible Generalized Least Square for Fixed Effect Model

Bai et al. in [11] proposed a modified Feasible Generalized Least Squares (FGLS) estimator for panel data models in which the error terms may exhibit heteroskedasticity, serial correlation, and cross-sectional correlation simultaneously. The main difficulty arises from the covariance matrix of the stacked error vector, which has a high-dimensional structure. Rather than imposing a specific parametric form on this covariance matrix, the proposed approach constructs a nonparametric estimator by exploiting the temporal dependence structure and the sparsity of cross-sectional correlations. Consider the linear panel model,

$$\mathbf{y}_{\{\mathrm{it}\}} = \mathbf{x}'_{\{\mathrm{it}\}}\boldsymbol{\beta} + \mathbf{u}_{\{\mathrm{it}\}} \quad (15)$$

or, after stacking the observations over individuals and time,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}. \quad (16)$$

Let the covariance matrix of the stacked error vector be,

$$\boldsymbol{\Omega} = \mathbf{E}(\mathbf{U}\mathbf{U}'). \quad (17)$$

Because $\boldsymbol{\Omega}$ has dimension $\mathrm{NT} \times \mathrm{NT}$, directly estimating all of its elements becomes computationally demanding when N and T are large. The modified FGLS procedure therefore estimates the covariance matrix in blocks according to the temporal lag.

For a given lag, the sample covariance between cross-sectional units i and j is estimated from the residuals as,

$$\hat{R}_{h,ij} = \frac{1}{T}\sum_{t=h+1}^{T} \hat{u}_{it}\hat{u}_{j,t-h} \quad (18)$$

An analogous expression is used for negative lags. These lag-specific covariance estimates form the basis for constructing the large covariance matrix.

To control the large number of cross-sectional covariance elements, Bai et al. in [11] employ a thresholding procedure. In particular, the soft-thresholding function is defined as,

$$s_{\{ij\}}(z) = \mathrm{sgn}(z)(|z| - \tau_{ij})_+ \quad (19)$$

where $(\boldsymbol{x})_+ = \boldsymbol{x}$ when $x \geq 0$, and $(\boldsymbol{x})_+ = \mathbf{0}$ otherwise. The threshold is allowed to vary across pairs of cross-sectional units and is specified as,

$$\tau_{ij} = M\gamma_T\sqrt{|\hat{R}_{0,ii}||\hat{R}_{0,jj}|} \quad (20)$$

where $\gamma_T = \sqrt{\frac{log(LN)}{T}}$

The thresholding operation reduces small cross-sectional covariance elements while retaining stronger dependencies. The resulting thresholded covariance matrix for lag h is denoted by,

$$\widehat{\boldsymbol{\Omega}}_{\boldsymbol{h}} = (\hat{\boldsymbol{\sigma}}_{\boldsymbol{h},\boldsymbol{ij}})_{N\times N} \quad (21)$$

This procedure allows the cross-sectional dependence structure to be sparse without requiring the researcher to specify the relevant clusters in advance. Serial correlation is incorporated through a banding procedure. For the (t,s)-th block, with $h = t - s$, the estimated covariance matrix is constructed as,

$$\widehat{\boldsymbol{\Omega}}_{t,s} = \begin{cases} \omega(|h|, L)\widehat{\boldsymbol{\Omega}}_{h}, & |h| \le L \\ \mathbf{0} & |h| > L \end{cases} \tag{22}$$

where L denotes the bandwidth and $\boldsymbol{\omega}(h, L)$ is a kernel function. Bai et al. in [11] use the Bartlett kernel, given by

$$\omega(h, L) = 1 - \frac{h}{L+1} \tag{23}$$

The banded and thresholded covariance blocks are then assembled to obtain the estimated large error covariance matrix,

$$\widehat{\boldsymbol{\Omega}} = (\widehat{\boldsymbol{\Omega}}_{t,s})_{t,s=1}^{T} \tag{24}$$

which has dimension $\mathrm{NT} \times \mathrm{NT}$. This estimator is constructed nonparametrically and does not require a prespecified parametric structure for the covariance matrix.

Given the estimated covariance matrix, the modified FGLS estimator of is obtained as,

$$\widehat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{Y}) \tag{25}$$

Thus, the estimated covariance structure is incorporated directly into the estimation through, which serves as the weighting matrix in the FGLS procedure. Bai et al. in [11] further establish the asymptotic validity of the proposed estimator by showing that the estimation error arising from replacing the population covariance matrix with its estimated counterpart is asymptotically negligible. This is expressed by,

$$\frac{\mathbf{1}}{\sqrt{\boldsymbol{NT}}}\boldsymbol{X}'(\widehat{\boldsymbol{\Omega}}^{-\mathbf{1}} - \boldsymbol{\Omega}^{-\mathbf{1}})\boldsymbol{U} = o_P(1) \tag{26}$$

Here $\boldsymbol{o_P}(\mathbf{1})$, denotes convergence to zero in probability as the panel dimensions increase.

III. METHODOLOGY

In this section, the authors state their data source, methodology of research, and step of analysis. This study employs panel data covering 34 provinces in Indonesia from 2018 to 2023, obtained from the Indonesian Central Statistics Agency (Badan Pusat Statistik, BPS) through https://www.bps.go.id/id. The study period was limited to 2023 for two main reasons. First, starting in 2024, statistical data for the Papua region were reported separately for six provinces, namely Papua, South Papua, Central Papua, Highland Papua, West Papua, and Southwest Papua. Consequently, the provincial units of observation are no longer consistent with those in the preceding years. Aggregating data from the newly established provinces into the original parent provinces is not methodologically feasible due to the characteristics of the variables used in this study, particularly the Average Hourly Earnings of Employees (AHE), which are measured separately for each province and therefore cannot be aggregated reliably. Second, using the most recent data following the regional expansion, namely the 2024-2025 period, would result in a relatively short time dimension (T = 2), which may be inadequate for panel data estimation. Therefore, the 2018-2023 period was selected because it provides a consistent set of cross-sectional units together with a sufficient time dimension for panel data analysis. The variables employed in this study are presented in Table 1.

Table 1 Research variables, definitions, and measurement unit

Variable	Description	Unit
Y	School Participation Rate (SPR; Angka Partisipasi Sekolah/APS) among adolescents aged 16-18 years	Percent (%)
X_1	Unemployment Rate (UR; Tingkat Pengangguran Terbuka/TPT),	Percent (%)
X_2	Average Hourly Earnings of Employees (AHE; Upah Rata-Rata Per Jam Pekerja/UP)	Indonesian Rupiah per hour (IDR/hour)
X_3	Mean Years of Schooling (MYS; Rata-rata Lama Sekolah/RLS)	Years
X_4	Percentage of Poor Population (PPP; Persentase Penduduk Miskin/PPM)	Percent (%)

The data analysis in this study was conducted through the following steps:

1. Conducting exploratory data analysis (EDA) to examine the trends and characteristics of the School Participation Rate (SPR), Unemployment Rate (UR), Average Hourly Earnings of Employees (AHE), Mean Years of Schooling (MYS), and Percentage of Poor Population (PPP) during the period 2018-2023.
2. Standardizing all variables using a global z-score transformation, with the mean and standard deviation calculated from all province-year observations, to ensure comparability among variables measured on different scales
3. Estimating panel data regression models, including the CEM, FEM, and REM.
4. Selecting the appropriate panel data regression specification involves the Chow and Hausman tests. The CEM and FEM are compared using the Chow test, whereas the FEM and REM are assessed using the Hausman test.
5. Performing diagnostic tests on the selected model, including the residual normality test, heteroskedasticity test, autocorrelation test, cross-sectional dependence test, and multicollinearity test.

6. Re-estimating the selected panel data model using the FGLS approach when violations of the error structure assumptions were detected. The modified FGLS estimator proposed by Bai et al. in [11] was employed to jointly account for heteroskedasticity, serial correlation, and cross-sectional correlations through estimation of the large error covariance matrix using banding and thresholding methods.
7. Evaluating the modified FGLS estimation by examining the resulting standard errors and calculating the within R^2 as a descriptive measure of model fit. The within R^2 was obtained from the fitted values based on the modified FGLS coefficient estimates.
8. Comparing the standard errors obtained from the FEM-OLS, standard FEM-FGLS, and modified FEM-FGLS estimations to assess changes in statistical uncertainty resulting from alternative specifications of the error covariance structure.
9. Interpreting the estimation results and drawing conclusions regarding the determinants of school participation among adolescents aged 16-18 years in Indonesia.

IV. RESULTS AND DISCUSSIONS

A. Exploratory Data Analysis

The first stage of this study involved an exploratory data analysis of the School Participation Rate (SPR) among adolescents aged 16-18 years in Indonesia from 2018 to 2023. The analysis incorporated several explanatory variables, namely the Unemployment Rate (UR), Average Hourly Earnings of Employees (AHE), Mean Years of Schooling (MYS), and Percentage of Poor Population (PPP), across 34 provinces in Indonesia.

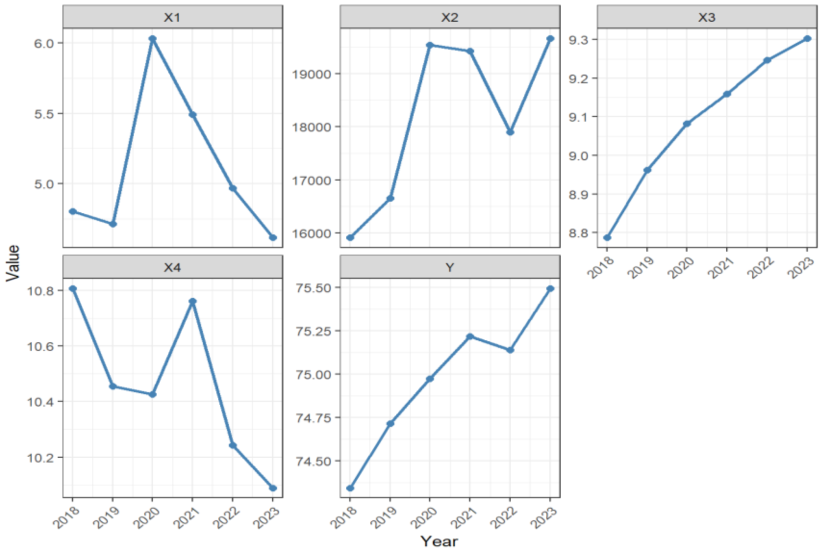

Figure 1 Average variable trends by year

The line graph in Figure 1 shows that the School Participation Rate (SPR; Y) increased from 74.3% in 2018 to 75.5% in 2021. This increase occurred relatively consistently from year to year. However, a slight decline was observed in 2022, with the SPR decreasing to 75.1%. One factor that may have contributed to this decline is the lingering impact of the COVID-19 pandemic, which continued to affect access to education, educational quality, and schooling continuity across various regions. Overall, the trend indicates that school participation among adolescents aged 16-18 years in Indonesia generally improved during the study period, although challenges remain that may affect the stability of these achievements.

In addition to SPR, Figure 1 also presents the trends of four predictor variables presumed to be associated with the school participation rate, namely the Unemployment Rate (UR; X_1) Average Hourly Earnings of Employees (AHE; X_2), Mean Years of Schooling (MYS; X_3), and Percentage of Poor Population (PPP; X_4). The AHE and MYS variables exhibited an upward trend throughout the observation period, consistent with the increasing pattern of SPR. In contrast, UR and PPP displayed more fluctuating patterns. UR increased sharply in 2020 before declining in subsequent years, whereas PPP decreased through 2020, rose in 2021, and declined again in 2022. Descriptively, these patterns suggest that improvements in community welfare and educational attainment tended to coincide with increases in SPR, while labor market conditions and poverty levels demonstrated more varied dynamics throughout the study period. Nevertheless, these relationships remain indicative and require further statistical analysis to examine the effect of each variable on SPR.

B. Panel Data Regression Model Selection

Panel data regression can be estimated using three alternative approaches, namely the CEM, FEM, and REM [17][18]. These models differ in how they account for unobserved heterogeneity among cross-sectional units, therefore, selecting the most appropriate model is essential to ensure consistency with the characteristics of the data. The CEM assumes that the data are homogeneous and does not account for individual or time-specific differences. In contrast, the FEM allows for unobserved individual heterogeneity by incorporating fixed effects assumed to remain constant over time. Meanwhile, the REM treats individual-specific differences as random effects uncorrelated with the explanatory variables.

The model selection procedure was carried out using several statistical tests. The Chow test was conducted to assess the choice between the FEM and CEM, while the Hausman test was applied to identify the preferred specification between the FEM and REM. In addition, the Breusch-Pagan Lagrange Multiplier (LM) test was used to evaluate the choice between the REM and CEM [19]. Based on the outcomes of these tests, the panel data regression model most suitable for this study can be selected.

For the SPR data, the Chow test resulted in an F-statistic of 594.16 and a p-value of 2.2×10^{-16}, which is below the 0.05 significance level. This finding suggests that the individual effects are statistically significant, hence, the null hypothesis that the CEM is more appropriate was rejected. Accordingly, the FEM was considered more suitable than the CEM for modeling the SPR data.

Following the selection of the FEM through the Chow test, the Hausman test was subsequently conducted to evaluate the choice between the FEM and the REM. The test produced a chi-square statistic of 21.52 with a p-value of 0.00025, which is lower than the 0.05 significance level. This led to rejection of the null hypothesis. This result indicates a correlation between the individual effects and the independent variables, suggesting that the REM estimator is inconsistent. Consequently, the FEM was selected as the most suitable model for the SPR data. As a result, the Breusch-Pagan LM test was not performed, since both the Chow and Hausman tests indicated that the FEM was the preferred model.

C. Estimation Results of the Selected Panel Data Regression Model

According to the model selection procedure presented in the preceding subsection, the FEM was selected as the most suitable specification for analyzing the School Participation Rate (SPR) data. Consequently, parameter estimation was performed using the within estimator under the Ordinary Least Squares (OLS) framework [20][21]. The FEM was applied to control for the influence of time-invariant unobserved characteristics at the province level, so that the resulting parameter estimates are not affected by factors that remain constant throughout the observation period but are not included in the model.

Prior to estimation, all variables were standardized to facilitate coefficient comparability. Therefore, the estimated coefficients reported in Table 2 can be interpreted as standardized effects, allowing direct comparison of the relative importance of the explanatory variables.

Table 2 Estimation Results of the Fixed Effect Model (FEM) Estimated by OLS

Variable	Coefficient	Standard Error	t-statistic	p-value	Decision
UR (X_1)	0.0082164	0.0118861	0.6913	0.49037	Not Significant
AHE (X_2)	0.0253022	0.0129478	1.9542	0.05236	Not Significant
MYS (X_3)	0.2772610	0.0283527	9.7790	$< 2 \times 10^{-16}$	Significant
PPP (X_4)	0.0988130	0.0658702	1.5001	0.13548	Not Significant

Table 2 shows that the Mean Years of Schooling (MYS) is the only variable with a statistically significant relationship with the School Participation Rate (SPR), with a standardized coefficient of 0.277 and a p-value below the 0.05 threshold. This result suggests that a one-standard-deviation increase in MYS corresponds to a 0.277-standard-deviation increase in SPR. The finding highlights the important role of educational attainment in explaining variations in school participation.

In contrast, UR, AHE, and PPP also exhibit positive coefficients of 0.0082164, 0.0253022, and 0.0988130, respectively, but their corresponding p-values of 0.49037, 0.05236, and 0.13548 exceed the 0.05 threshold. Hence, these three variables do not provide sufficient statistical evidence of a relationship with SPR under the selected significance criterion. Overall, MYS is the only explanatory variable in the model that shows a statistically supported positive relationship with SPR.

D. Panel Data Regression Assumption Tests

After identifying the FEM as the preferred panel data specification, the next step was to conduct diagnostic tests of the panel regression assumptions. These tests were performed to ensure that the selected model satisfies the underlying assumptions required for reliable parameter estimation and statistical inference. The diagnostic tests conducted in this study included the residual normality test, heteroskedasticity test, autocorrelation test, and multicollinearity test.

1) Residual Normality Test

   The residual normality assumption was examined using the Anderson-Darling test. The test yielded a A statistic of 0.6083 with a p-value of 0.1125, which exceeds the 5% significance level. With a p-value above 0.05, there is insufficient evidence to reject the null hypothesis, supporting the assumption of normally distributed residuals. Thus, the results provide no evidence of a violation of the normality assumption in the FEM.

2) Heteroskedasticity Test

   The heteroskedasticity assumption was assessed using the Breusch-Pagan (BP) test. The resulting BP statistic was 48.349, with a p-value of 7.98×10^{-10}, falling below the 0.05 criterion. This result provides sufficient evidence to reject the null hypothesis and indicates the presence of heteroskedasticity in the model. Consequently, the error variance is not uniform across observations, implying that the homoskedasticity assumption is violated.

3) Autocorrelation Test

Autocorrelation in the panel data model was examined using the Wooldridge first-difference test. The test yielded an F-statistic of 5.55 with a p-value of 0.0199, which is smaller than the 5% significance level. The test result provides sufficient evidence against the null hypothesis, suggesting that autocorrelation is present in the model residuals.

4) Cross-sectional Dependence Test

Cross-sectional dependence in the panel data model was examined using the Pesaran CD test. The test yielded a z-statistic of 2.702 with a p-value of 0.006891, which is below the 0.05 threshold. The findings provide sufficient evidence to reject the null hypothesis and suggest the presence of cross-sectional correlations in the model errors.

5) Multicollinearity Test

The multicollinearity test was conducted to detect the presence of high correlations among the independent variables. Variance Inflation Factor (VIF) values were used as the diagnostic measure, where a VIF value exceeding 10 indicates a potential multicollinearity problem. Based on the analysis, the VIF values for the independent variables were as follows: 1.50 for the Unemployment Rate (UR), 1.36 for the Average Hourly Earnings of Employees (AHE), 1.62 for the Mean Years of Schooling (MYS), and 1.22 for the Percentage of Poor Population (PPP). Since all VIF values are well below 10. As all VIF values remain substantially below the threshold of 10, the findings provide no evidence of multicollinearity among the explanatory variables.

Based on the results of the diagnostic tests, the FEM estimated using OLS satisfies the assumptions of residual normality and absence of multicollinearity. However, the model exhibits heteroskedasticity, serial correlation, and cross-sectional dependence. Therefore, parameter estimation was subsequently performed using the FGLS approach proposed by Bai et al. in [11], which is specifically developed to accommodate heteroskedasticity, serial correlation, and cross-sectional correlations in panel data. In their approach, serial correlations are addressed using a banding method, while cross-sectional correlations are accounted for using a thresholding method in estimating the large error covariance matrix. The proposed FGLS estimator is theoretically supported by a limiting distribution and is designed to improve estimation efficiency in the presence of heteroskedasticity, serial correlation, and cross-sectional correlations. Accordingly, this approach is employed to obtain parameter estimates while accounting for the identified dependence structure in the panel errors.

E. Fixed Effect Model Estimated Using Feasible Generalized Least Squares

Based on the diagnostic tests, the FEM was retained as the underlying model specification, while the estimation procedure was adjusted to account for the identified violations in the error structure. Specifically, the model exhibited heteroskedasticity, serial correlation, and cross-sectional dependence. Therefore, parameter estimation was subsequently performed using the FGLS estimator of Bai et al. in [11]. This approach jointly accommodates heteroskedasticity, serial correlation, and cross-sectional correlations through the estimation of a large error covariance matrix that combines banding and thresholding techniques. By explicitly accounting for these three sources of error dependence, the BCL-FGLS approach provides an estimation framework that is more appropriate for the error structure identified in this study than an FGLS specification that accounts only for heteroskedasticity and serial correlation.

Table 3 Estimation Results of the Fixed Effect Model Using the Modified FGLS Estimator of Bai et al. in [11]

Variable	Coefficient	Standard Error	t-statistic	p-value	Decision
UR (X_1)	0.0022585	0.0083050	0.2719	0.7860	Not Significant
AHE (X_2)	0.0224138	0.0090843	2.4673	0.0147	Significant
MYS (X_3)	0.2754109	0.0186659	14.7548	1.03×10^{-31}	Significant
PPP (X_4)	0.1023185	0.0390389	2.6209	0.0096	Significant

Table 3 reports the results obtained from the modified FGLS estimation. The signs of the estimated coefficients remain consistent with those obtained from the FEM estimation presented in Table 2. All explanatory variables retain positive coefficients under both estimation approaches. However, the statistical significance of some variables changes after accounting for heteroskedasticity, autocorrelation, and cross-sectional dependence in the error structure. This indicates that accounting for the dependence structure of the panel errors affects the estimated standard errors and, consequently, the statistical inference, while the direction of the estimated relationships remains unchanged. Since all variables were standardized prior to estimation, the estimated coefficients represent standardized effects. Based on the modified FGLS estimation, three variables have statistically significant positive effects on the School Participation Rate (SPR) at the 5% significance level, namely the Average Hourly Earnings of Employees (AHE), the Mean Years of Schooling (MYS), and the Percentage of Poor Population (PPP).

AHE has an estimated coefficient of 0.0224, and its p-value of 0.0147 falls below the 0.05 threshold, indicating a positive and statistically significant coefficient. With the other explanatory variables held constant, a one-standard-deviation increase in AHE corresponds to a 0.0224-standard-deviation increase in SPR. This finding suggests that higher employee earnings may contribute to greater school participation by enhancing household purchasing power and financial capacity to support educational expenses [22]. The findings reported by Widiasti [23] also lend support to this result, showing that higher provincial minimum wages were positively and significantly associated with upper-secondary school participation in Indonesia. Although the study in [23] employed a Random Effects Model based on data from 12 provinces, the consistency in the direction of the estimated effect with the present study which applies a Fixed Effect Model to data from 34 provinces suggests that higher earnings may serve as an indirect mechanism for improving access to education through enhanced household economic capacity.

Meanwhile, the MYS variable has a positive and statistically significant standardized coefficient of 0.2754 (p-value = 1.03×10^{-31}), implying that a one standard deviation increase in MYS is associated with a 0.2754 standard deviation increase in SPR. This finding indicates that provinces with higher educational attainment tend to exhibit higher levels of school participation. MYS reflects the overall educational attainment of the population [24] and serves as an indicator of educational development within a region. Provinces with higher MYS generally possess more developed educational systems, characterized by better school availability, higher-quality teachers, and more adequate supporting infrastructure. Such conditions are conducive to sustained school participation among adolescents.

Similarly, the PPP exhibits a positive and statistically significant association with the School Participation Rate (SPR), with a standardized coefficient of 0.1023 (p-value = 0.0096). This indicates that a one standard deviation increase in PPP is associated with a 0.1023 standard deviation increase in SPR, holding the other variables constant. This positive association contrasts with the conventional expectation that higher poverty constrains school participation, as reported by Rahmatin and Soejoto in [25]. However, its direction is consistent with Parawanza and Suharianto in [26] who reported a positive but statistically insignificant relationship. This positive association may reflect the role of educational and social assistance, such as scholarships, the Indonesia Smart Program (PIP), and targeted social assistance, in mitigating the adverse effects of poverty on school participation [26]. In addition, improved access to schools and greater public awareness of the importance of education may contribute to sustaining school participation among adolescents [26]. Thus, such interventions may help areas with relatively high poverty levels maintain relatively high school participation rates.

In contrast, the estimated coefficient of UR is positive, but the evidence (p-value > 0.05) is insufficient to establish a statistically significant effect on the SPR among adolescents aged 16–18 years. This result suggests that labor market conditions at the provincial level do not exert a strong influence on school participation decisions at the upper-secondary level. One possible explanation is that educational decisions among individuals aged 16-18 years are more strongly shaped by long-term structural factors, such as educational attainment as reflected by MYS and household economic capacity as proxied by AHE, rather than by short-term fluctuations in labor market conditions. Furthermore, provincial-level aggregate data may not be sufficiently sensitive to capture the influence of labor market conditions on individual schooling decisions, given the substantial heterogeneity in household characteristics within provinces.

The insignificance of the UR variable may also be associated with policies aimed at expanding access to secondary education in Indonesia. Under Permendikbud No. 80 of 2013, the Universal Secondary Education Program (Pendidikan Menengah Universal, PMU) was established, targeting adolescents aged 16-18 years, with the goal of achieving a gross enrollment ratio of 97 percent by 2020 [27]. Notably, the target age group of the PMU corresponds exactly to the age group examined in this study. The program was further supported by the Indonesia Smart Program (Program Indonesia Pintar, PIP), which aims to ensure that all citizens can pursue education up to the secondary level without financial barriers [28]. These policies are presumed to have reduced the influence of short-term economic conditions on the schooling decisions of adolescents aged 16-18 years, although the success of their implementation remains variable across regions in Indonesia.

In terms of model fit, the modified FGLS estimation yields a within R^2 of 0.5326. This indicates that approximately 53.26% of the within-province variation in the SPR over time is explained by the four explanatory variables, namely UR, AHE, MYS, and PPP, after accounting for province-specific fixed effects. Thus, the model captures a substantial proportion of the variation in SPR within provinces over the study period.

Table 4 Comparison of Standard Errors Across FEM-OLS, Standard FEM-FGLS, and Modified FEM-FGLS

Variable	FEM-OLS	FEM-FGLS	FEM-Modified FGLS
UR (X_1)	0.0118861	0.0077251	0.0083050
AHE (X_2)	0.0129478	0.0075610	0.0090843
MYS (X_3)	0.0283527	0.0293120	0.0186659
PPP (X_4)	0.0658702	0.0511636	0.0390389

Under the FEM estimated using OLS, the presence of heteroskedasticity, serial correlation, and cross-sectional dependence indicates that the error variance-covariance structure does not satisfy the classical assumptions. Consequently, the conventional standard errors may not provide reliable statistical inference, potentially affecting the validity of hypothesis tests [29]. Although these violations do not necessarily bias the coefficient estimates under the relevant exogeneity conditions ([30] as cited in [31]), they can affect the estimated standard errors and, consequently, statistical inference.

The standard FGLS approach explicitly models the error variance-covariance structure and can account for heteroskedasticity and serial correlation [11]. As shown in Table 4, the standard FGLS produces lower standard errors than the FEM-OLS estimation for UR, AHE, and PPP, while the standard error of MYS increases slightly. This increase does not by itself indicate a loss of overall estimator efficiency, since FGLS does not minimize the standard error of each coefficient individually. Rather, the change reflects the different weighting of observations resulting from the estimated error covariance structure. A notable change in statistical significance is observed for AHE, which becomes statistically significant at the 5% level following FGLS estimation. This change is attributable to the reduction in its standard error, resulting in a larger t-statistic.

To further account for the cross-sectional dependence identified in the diagnostic tests, the modified FGLS estimator of Bai et al. [11] was subsequently applied. This approach incorporates both banding to account for serial correlation and thresholding to account for cross-sectional correlation in the estimation of the error covariance matrix. The modified FGLS produces standard errors of 0.0083, 0.0091, 0.0187, and 0.0390 for UR, AHE, MYS, and PPP, respectively. Compared with the FEM-OLS estimation, the standard errors are lower for all four explanatory variables under the modified FGLS. This indicates that accounting for the heteroskedasticity, serial correlation, and cross-sectional dependence present in the panel errors changes the estimated uncertainty of the coefficients.

The modified FGLS estimation also changes the statistical inference for several variables. In particular, AHE, which was not statistically significant under the FEM-OLS estimation, becomes statistically significant at the 5% level under the modified FGLS, with its standard error decreasing from 0.0129 to 0.0091. PPP also becomes statistically significant, while MYS remains highly statistically significant. These changes illustrate that accounting for the dependence structure of the panel errors can affect statistical inference even when the direction of the estimated coefficients remains unchanged. Therefore, the modified FGLS provides statistical inference that accounts for the identified heteroskedasticity, serial correlation, and cross-sectional dependence.

V. CONCLUSIONS AND SUGGESTIONS

The model selection results favored the Fixed Effects Model (FEM) as the appropriate specification for the panel data analysis. Subsequent diagnostic testing revealed heteroskedasticity, serial correlation, and cross-sectional dependence. Therefore, the model was re-estimated using the modified FGLS estimator proposed by Bai et al. in [11]. The results show that AHE, MYS, and PPP have positive and statistically significant associations with SPR among adolescents aged 16-18 years, while UR is not statistically significant. The positive association between PPP and SPR should not be taken to mean that poverty itself increases school participation, as it may reflect the role of educational assistance, social protection, and access to education in maintaining school participation in provinces with relatively high poverty levels. These findings are limited to the provincial-level context and do not establish causal relationships.

Several limitations of this study should be taken into account in subsequent research. First, the modified FGLS estimator has specific theoretical requirements regarding the dimensions of panel data, which may affect the reliability of statistical inference in panels with a relatively short time dimension. Second, provincial-level aggregate data may not fully capture individual- and household-level factors underlying school participation, while the use of MYS may involve potential endogeneity and reverse causality. Future studies could incorporate micro-level data, lagged variables, or instrumental-variable approaches. Finally, spatial econometric models, such as SAR or SEM, could be considered to examine the spatial mechanisms underlying the observed cross-sectional dependence. Extending the observation period and incorporating additional socioeconomic, educational, and policy-related variables may also provide a broader understanding of the factors associated with school participation.

REFERENCES

[1] I. Grosemans and N. De Cuyper, “Career competencies in the transition from higher education to the labor market: Examining developmental trajectories,” *J. Vocat. Behav.*, vol. 128, p. 103602, Aug. 2021, doi: 10.1016/j.jvb.2021.103602.

[2] P. J. De Leon, “The influence of socioeconomic factors on career choices among senior high school graduates: A tracer study,” *International Journal of Progressive Research in Engineering Management and Science (IJPREMS)*, vol. 05, pp. 1252–1262, 2025, [Online]. Available: www.ijprems.com

[3] Badan Pusat Statistik, “Angka Partisipasi Sekolah (APS) menurut provinsi, 2023.” Accessed: May 15, 2026. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MzAxIzI=/angka-partisipasi-sekolah---a-p-s---.html

[4] H. Lehti, J. Erola, and A. Karhula, “The heterogeneous effects of parental unemployment on siblings’ educational outcomes,” *Res. Soc. Stratif. Mobil.*, vol. 64, p. 100439, Dec. 2019, doi: 10.1016/j.rssm.2019.100439.

[5] A. A. Smith, “The minimum wage and teen educational attainment,” *Labour Econ.*, vol. 73, p. 102061, Dec. 2021, doi: 10.1016/j.labeco.2021.102061.

[6] A. Ciarreta, M. P. Espinosa, and A. Zarraga, “Panel Data Analysis,” in *Encyclopedia of Law and Economics*, A. Marciano and G. Ramello, Eds., New York, NY: Springer New York, 2016, pp. 1–2. doi: 10.1007/978-1-4614-7883-6_360-1.

[7] E. W. Wadan, “Analisis regresi data panel terhadap faktor-faktor yang mempengaruhi angka partisipasi sekolah usia 16–18 tahun di Nusa Tenggara Timur tahun 2017-2021 ,” Undergraduate Thesis , Universitas PGRI Adi Buana Surabaya, 2023.

[8] C. Gazeloğlu and E. Erkılıç, “Economic development, renewable energy use, and international tourism: Strategic approaches to carbon emission reduction in SICA countries,” *Sustainability*, vol. 17, no. 5, p. 1995, Feb. 2025, doi: 10.3390/su17051995.

[9] E. R. N. Pasaribu and N. Ariani, "Pengaruh tingkat partisipasi angkatan kerja, belanja modal, dan pengguna internet terhadap tingkat pertumbuhan PDRB di kabupaten/kota Provinsi Nusa Tenggara Timur," *Journal of Development Economic and Digitalization*, vol. 2, no. 2, pp. 105–123, Aug. 2023, doi: 10.59664/jded.v2i2.6573.

[10] L. R. Simarmata and Khoirunurrofik, "Peranan bantuan operasional penyelenggaraan pendidikan anak usia dini terhadap akses layanan pendidikan anak usia dini di Indonesia," *Journal Publicuho*, vol. 7, no. 1, pp. 151–164, Feb. 2024, doi: 10.35817/publicuho.v7i1.343.

[11] J. Bai, S. H. Choi, and Y. Liao, "Feasible generalized least squares for panel data with cross-sectional and serial correlations (arXiv:1910.09004)," *ArXiv*, 2020, Accessed: Jun. 17, 2026. [Online]. Available: https://doi.org/10.48550/arXiv.1910.09004

[12] J. Bai, S. H. Choi, and Y. Liao, "Feasible generalized least squares for panel data with cross-sectional and serial correlations," *Empir. Econ.*, vol. 60, no. 1, pp. 309–326, Jan. 2021, doi: 10.1007/s00181-020-01977-2.

[13] I. F. Alamsyah, R. Esra, S. Awalia, and D. A. Nohe, "Analisis regresi data panel untuk mengetahui faktor yang memengaruhi jumlah penduduk miskin di Kalimantan Timur," in *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, 2022, pp. 254–266.

[14] N. F. Gamayanti and J. Junaidi, "Pemodelan hasil produksi padi di Provinsi Sulawesi Tengah menggunakan Fixed Effect Model (FEM)," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, pp. 347–354, Jun. 2021, doi: 10.30598/barekengvol15iss2pp347-354.

[15] C. A. Jacob, I. W. Sumarjaya, and M. Susilawati, "Analisis model Regresi Unbalanced panel data Komponen Galat Dua Arah dengan Penduga Feasible Generalized Least Square (FGLS)," *Jurnal Matematika*, pp. 22–38, 2014.

[16] M. D. Bertiani, N. N. Debataraja, and N. Imro'ah, "Penerapan model regresi unbalanced panels dengan metode Feasible Generalized Least Square (FGLS)," *BIMASTER: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 13, no. 1, 2024.

[17] M. Irwansyah, R. Ruliana, and M. K. Aidid, "Analisis regresi balanced panel dengan komponen galat dua arah pada kasus melek huruf masyarakat di Provinsi NTB," *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 3, no. 1, p. 10, Mar. 2021, doi: 10.35580/variansiunm14644.

[18] C. Liu, G. Ou, Y. Fu, C. Zhang, and C. Yue, "Application of a panel data quantile-regression model to the dynamics of carbon sequestration in Pinus kesiya var. langbianensis natural forests," *Forests*, vol. 13, no. 1, p. 12, Dec. 2022, doi: 10.3390/f13010012.

[19] V. Ratnasari, S. H. Audha, and A. T. R. Dani, "Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression," *MethodsX*, vol. 11, p. 102379, Dec. 2023, doi: 10.1016/j.mex.2023.102379.

[20] B. Huang, T.-H. Lee, and A. Ullah, "A combined random effect and fixed effect forecast for panel data models," *Journal of Management Science and Engineering*, vol. 4, no. 1, pp. 28–44, Mar. 2019, doi: 10.1016/j.jmse.2019.03.004.

[21] A. F. Jauhari, Y. K. Suprapto, and A. Mauludiyanto, "Data panel modelling with fixed effect model (FEM) approach to analyze the influencing factors of dengue hemorrhagic fever in Pasuruan Regency," in *Proceedings of the 2nd International Conference on Applied Science, Engineering and Social Sciences*, SCITEPRESS - Science and Technology Publications, 2019, pp. 224–232. doi: 10.5220/0009881702240232.

[22] M. N. Kurniawan and N. Imaningsih, "Pengaruh upah minimum kabupaten, rata-rata lama sekolah, dan tingkat partisipasi angkatan kerja terhadap produk domestik regional bruto di Kabupaten Gresik dan Lamongan," *Jambura Economic Education Journal*, vol. 7, no. 1, pp. 111–126, 2025.

[23] S. Widiasti, "Pengaruh upah minimum terhadap tingkat partisipasi sekolah di Indonesia ," Undergraduate Thesis , Universitas Gadjah Mada, 2020.

[24] R. Diana, "Faktor-faktor yang mempengaruhi siswa putus sekolah di Provinsi Sumatera Barat," *Jurnal Penelitian Kebijakan Pendidikan*, vol. 14, no. 2, Dec. 2021, doi: 10.24832/jpkp.v14i2.497.

[25] U. Z. Rahmatin and A. Soejoto, "Pengaruh tingkat kemiskinan dan jumlah sekolah terhadap angka partisipasi sekolah (APS) di Kota Surabaya," *JPEKA: Jurnal Pendidikan Ekonomi, Manajemen dan Keuangan*, vol. 1, no. 2, pp. 127–140, 2017.

[26] A. A. Parawanza and J. Suharianto, "Pengaruh PDRB dan kemiskinan terhadap angka partisipasi sekolah di Provinsi Sumatera Utara. ," *Jurnal Media Akademika (JMA)*, vol. 3, no. 6, pp. 1–14, 2025.

[27] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 80 Tahun 2013 tentang Pendidikan Menengah Universal*. 2013.

[28] N. A. Rahmi and L. K. Sari, "Variabel-variabel yang memengaruhi ketuntasan wajib belajar di Provinsi Papua tahun 2020," in *Seminar Nasional Official Statistics 2022*, 2022, pp. 917–926.

[29] A. Polselli, "Robust inference in panel data models: Some effects of heteroskedasticity and leveraged data in small samples (arXiv:2312.17676)," *ArXiv*, 2024, Accessed: Jun. 17, 2026. [Online]. Available: https://doi.org/10.48550/arXiv.2312.17676

[30] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006. doi: 10.1002/0470055464.

[31] A. F. Ernst and C. J. Albers, "Regression assumptions in clinical psychology research practice - a systematic review of common misconceptions," *PeerJ*, vol. 5, p. e3323, May 2017, doi: 10.7717/peerj.3323.