JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 4, October 2024, pp. 1202-1216

Analysis of Factor Affecting Tuberculosis Cases in West Java Province Using Panel Data Regression Approach

Toha Saifudin[1], Arlisya Shafwan Aisyah[1], Irma Ayu Indrasta[1], Dita Amelia[1]
[1]Statistics Study Program, Departement of Mathematics, Universitas Airlangga, Indonesia
tohasaifudin@fst.unair.ac.id

Article History:
Received : 25-06-2024
Revised : 06-10-2024
Accepted : 13-10-2024
Online : 15-10-2024

Keywords:
Tuberculosis;
West Java;
Regression;
Panel;
Fixed Effect Model:
SDGs.

ABSTRACT

Tuberculosis (TB) is a disease that can cause death with the largest number of sufferers after COVID-19. In Indonesia, the number of TB cases reached 724.309 cases in 2022 with the highest number 184.406 cases in West Java Province. Given this situation, Indonesia must try to achieve the health target from SDGs, namely ending the TB epidemic by 2030. Therefore, this research aims to analyze the factors that have a significant influence on the incidence of TB in Indonesia, especially in West Java Province. The research focuses on four variables: percentage of poverty, number of diabetics, number of HIV/AIDS patients, and population density. To provide a more informative analysis, this research uses a combination of cross-section and time series data from 27 regions between 2020 and 2022. So, the method used according to the type of data is panel data regression including common effect, fixed effect, and random effect models. Based on statistical tests, namely through the chow test, hausman test, and lagrange multiplier test, it was found that the best model was fixed effect with an R-squared value of 90%. The research revealed that all the studied factors significantly influence the incidence of TB cases in West Java. The results of this study are expected to help the West Java government in an effort to reduce the number of TB cases and formulate policies by reducing the percentage of poverty and population density in West Java. By ensuring the availability of health facilities such as establishing health centers in densely populated areas and counseling programs also need to be conducted to underscore the importance of TB control in West Java.

Crossref
https://doi.org/10.31764/jtam.v8i4.24795

—————————— ◆ ——————————

A. INTRODUCTION

Tuberculosis, commonly referred to as TB, is an infectious disease that is caused by the bacteria Mycobacterium tuberculosis. Transmission of TB disease occurs when the patient sneezes, coughs, or sings which causes TB germs to contaminate the air and be inhaled by others (Maison, 2022). The World Health Organization (WHO) reports that by 2022, TB has reached 10,6 million cases in 2022. This figure is the second highest cause of death after COVID-19, with a mortality rate of 52 per 100,000 population. In Indonesia, TB cases in 2022 reached 724,309, with the 45-54 age group being the most affected and West Java Province recording the highest number with 184,406 cases, far above the estimated incidence of only 148,069 (Kementerian Kesehatan RI, 2023). In general, the cause of the increase in tuberculosis cases is caused by factors that influence the spread of Mycobacterium tuberculosis bacteria, namely environmental factors, individuals, and socioeconomic conditions (Kementerian Kesehatan RI, 2020). According to Suherni & Maduratna (2013), in their research using clustering, they stated that of the 18 variables studied, the most influential variables on TB cases in Surabaya were population density, the number of HIV patients, and the number of businesses in the trade sector.

The increase in tuberculosis cases is caused by factors that influence the spread of Mycobacterium tuberculosis bacteria, namely environmental factors, individuals, and socioeconomic conditions (Kementerian Kesehatan RI, 2020). According to Suherni & Maduratna (2013), in their research using clustering, they stated that of the 18 variables studied, the most influential variables on TB cases in Surabaya were population density, the number of HIV patients, and the number of businesses in the trade sector. Based on other research conducted by Brilliant & Fakhriyana (2023), the utilization of geographically weighted negative binomial regression revealed that the influential factor in TB cases in West Java in 2021 was the percentage of poor people and the percentage of people who have had complaints recently. Also, another study conducted by Purwanto (2020), using a multiple linear regression test stated in households that do not implement clean and healthy living behavior within 1 year there are 18,8% of family members suffering from TB in Sidoarjo regency. Based on other research conducted by Brilliant & Fakhriyana (2023), the utilization of geographically weighted negative binomial regression revealed that the influential factor in TB cases in West Java in 2021 was the percentage of poor people and the percentage of people who have had complaints recently. Also, another study conducted by Purwanto (2020), using a multiple linear regression test stated in households that do not implement clean and healthy living behavior within 1 year there are 18,8% of family members suffering from TB in Sidoarjo regency.

In previous studies, researchers often focused on areas with lower tuberculosis cases and collected data for only one year. So to fill the void, this study will update by analyzing the factors that influence the number of tuberculosis cases in West Java as the province with the highest number of tuberculosis cases in Indonesia, in 2020 to 2022 based on the percentage of poverty, the number of diabetics, the number of HIV/AIDS patients, and population density in West Java. This study is important because TB remains a significant public health problem in Indonesia, especially in areas with the highest number of cases such as West Java.

This study uses panel data, which combines cross-sectional and time series data (Liao & He, 2018). Cross-section data is data obtained by observing many subjects at one time or a fixed number of periods (Karavias et al., 2023). Meanwhile, time series data is data obtained from observing one object from several periods (Dewi Anggraeni Chairunnisa & Fauzan, 2023). So the appropriate method used in this study is panel data regression, which is a statistical method used to explain the correlation between variables under examination (Uyanık & Güler, 2013). Panel data regression has several methods for estimating regression models, including common effect, fixed effect, and random effect (Muhith et al., 2022). By analyzing data for three-year period from 2020 to 2022, this study aims to identify more accurate current conditions related to the impact of the COVID-19 pandemic which may have a distribution on changes in TB cases, so as to reflect the overall dynamics of TB cases.

The results of this study can support the implementation of SDGs policies in West Java by providing data and analysis of factors affecting TB cases. This is based on supporting the realization of the third point regarding good health and welfare which is based on research to reduce the number of TB cases in Indonesia, especially in the province of West Java. The first point regarding the elimination of poverty, in this study the variable percentage of poverty people was analyzed to determine whether the percentage of poverty people in West Java could affect the inequality of health facilities obtained by the population. The eleventh point regarding sustainable cities and communities, this research is expected to create sustainable cities with good population health. Thus, this research is expected to support more effective health policies.

B. METHODS

This research involves regression analysis, which combines cross-section and time series data. The methods use the common effect model, fixed effect model, and random effect model. In examining the factors affecting the number of tuberculosis cases in West Java in 2020-2022, panel data regression analysis was used, following these research steps. In this study, EViews software will be used to conduct the analysis.

1. Classical Regression Assumptions

a. Detecting Collinearity

Multicollinearity is an assumption in a multiple linear regression that indicates a strong correlation between multiple predictor variables (Ismaeel et al., 2021). A good regression model is that the predictor variables do not correlate (Marcoulides & Raykov, 2019). One method for detecting multicollinearity involves examining the VIF (Variance Inflation Factor). The VIF statistical test can be formulated as follows (Prunier et al., 2015):

$$VIF = \frac{1}{1-R_k^2}; k = 1,2,..p \tag{1}$$

the coefficient of determination is denoted as R_k^2 represents the extent of the relationship between the k_{th} variable and the other predictor variables. A VIF value > 10 indicates collinearity among the predictor variables within the regression model (Shrestha, 2020).

b. Heteroskedasticity Test

The heteroskedasticity test is used to determine whether the residuals within the regression model maintain a constant variance (Utari et al., 2020). The heteroskedasticity test is crucial because it can result in inefficient least squares estimates and an inconsistent covariance matrix estimate if the error terms are incorrectly assumed to be homoscedastic (Feng et al., 2020). The heteroskedasticity test used in this research is the glejser test. Hypothesis testing to evaluate the residual variance of data can be written as follows (Djalic & Terzic, 2021): H_0 is All x_k are not significant (homoscedasticity); H_1 is There is at least one x_k significant, $k = 1,2,\dots,p$ (heteroskedasticity). The heteroskedasticity statistical test can be formulated as follows (Djalic & Terzic, 2021):

$$|e| = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + v_i \tag{2}$$

where $|e|$ is residual absolute error from the estimated model for the predictor variables, x_k is the value of the k_{th} predictor variable for the i_{th} observation, β_k is the slope or coefficient at the k_{th} predictor variable, and v_i is variance error. With the test criteria H_0

is rejected if the p-value < α, this means that there is heteroskedasticity within the regression model.

2. Estimating Model

a. Common Effect Model (CEM)

The common effect model is a statistical model that employs the OLS (Ordinary Least Square) technique to estimate the parameters by mixing time series and cross-sectional data (Biørn, 2016). The estimation of parameter values in linear regression equations is commonly done through the OLS method (Grigore et al., 2022). This method is widely used for this purpose. The CEM equation is formulated as follows (Ratnasari et al., 2023):

$$Y_{it} = \alpha + \boldsymbol{X}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3}$$

where Y_{it} is the value of the response variable for the i_{th} observation and t_{th} time, α is intercept, $\boldsymbol{X_{it}}$ or $\left(X_{1it}, X_{2it}, \ldots, X_{pit}\right)$ is the value of the k_{th} predictor variable for the i_{th} observation and t_{th} time, $\boldsymbol{\beta}$ or $\left(\beta_1, \beta_2, \ldots, \beta_p\right)'$ is slope or coefficient at the k_{th} predictor variable, ε_{it} is regression model error at the i_{th} observation and t_{th} time with $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, i are 1,2,3,…, n, and t are 1,2,3,…, T.

b. Fixed Effect Model (FEM)

The fixed effect model involves the estimation of panel regression parameters by adding dummy variables (Wardhana & Indawati, 2021). This technique is also referred to as the least square dummy variable model. The slope coefficient in FEM is generally assumed to remain constant, while the intercept is not regarded as constant. The FEM equation can be denoted as follows (Ratnasari et al., 2023):

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{4}$$

where α_i is fixed effect for the i_{th} observation.

c. Random Effect Model (REM)

A random effect model estimates panel data with correlated residuals across individuals and time. It assumes that each subject has a changing intercept over time (Bell et al., 2019). In this model, differences in individual and time characteristics are accommodated in the error of the model (Piratla & Singh, 2023). The REM equation can be formulated as follows (Ratnasari et al., 2023):

$$Y_{it} = \alpha + \boldsymbol{X}_{it}\boldsymbol{\beta} + u_i + \varepsilon_{it} \tag{5}$$

where u_i is the individual residual value of the i_{th} observation unit with $u_i \sim N(0, \sigma_u^2)$.

3. Determining The Best Estimation Model

a. Chow Test

The Chow test is a statistical method used to determine the suitability of the common effect model or the fixed effect model for panel data regression analysis (Ratnasari et al., 2023). The test hypothesis is as follows (Ratnasari et al., 2023): H_0 is $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$; and H_1 is there is at least one $\alpha_i \neq 0; i = 1,2,..n$. The F test statistics used are as follows:

$$\boldsymbol{F} = \frac{\left(\boldsymbol{R^2_{LSDV}} - \boldsymbol{R^2_{pooled}}\right)/(\boldsymbol{N}-\mathbf{1})}{\left(\mathbf{1} - \boldsymbol{R^2_{LSDV}}\right)/(\boldsymbol{NT}-\boldsymbol{N}-\boldsymbol{k})} \tag{6}$$

where R^2_{LSDV} is a coefficient determination for the FEM model, R^2_{pooled} is a Coefficient determination for the CEM model, N is the amount of unit cross-section, T is the amount of time series, k is the amount of predictor variable. H_0 is rejected if the value of $F > F_{(\alpha,N-1,NT-N-k)}$ or p-value < α, so that the fixed effect model is the selected model.

b. Hausman Test

The Hausman test is a statistical method used to choose the suitable model for panel data regression, which could be either the fixed effect model or the random effect model (Pratiwi, 2022). This test is also used to determine whether there exists a correlation between the model errors in the model and one or more predictor variables. The test hypothesis is as follows (Ratnasari et al., 2023): H_0 is $E(u_i|X_{it}) = 0$; H_1 is $E(u_i|X_{it}) \neq 0$. The test statistics used are as follows:

$$\boldsymbol{m} = \left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{CV}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS}}\right)'\left[\boldsymbol{var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{CV}}) - \boldsymbol{var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS}})\right]^{-1}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{CV}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{GLS}}\right) \tag{7}$$

where $var(\widehat{\boldsymbol{\beta}}_{CV})$ is the variance of the matrix from the estimated slope parameter vector of the FEM, $var(\widehat{\boldsymbol{\beta}}_{GLS})$ is the variance of the matrix from the estimated slope parameter vector of the REM, $\widehat{\boldsymbol{\beta}}_{CV}$ is the estimated slope parameter vector of the FEM, and $\widehat{\boldsymbol{\beta}}_{GLS}$ is the estimated slope parameter vector of the REM. H_0 is rejected if the value of $m > \chi^2_{(\alpha,k)}$ or p-value < α, so that the best model is fixed effect model.

c. Lagrange Multiplier Test

The lagrange multiplier test is conducted when the chow test and hausman test do not provide the optimal model conclusion. The Lagrange Multiplier test is a statistical method used to choose the suitable model for panel data regression, which could be either the common effect model or the random effect model (Puspita et al., 2021). During the test, the hypothesis is tested as follows (Ratnasari et al., 2023): H_0 is $\sigma_u^2 = 0$; H_1 is $\sigma_u^2 \neq 0$. The test statistics used are as follows:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{n}\left(\sum_{i=1}^{n} \varepsilon_{it}\right)^2}{\sum_{i=1}^{n}\sum_{i=1}^{n} \varepsilon_{it}{}^2} - 1\right]^2 \tag{8}$$

the test criterion is if the value of LM > $\chi^2_{(\alpha,N-1)}$ or p-value < α then the decision to reject H_0, so that the REM model is better than CEM.

4. Hypothesis Testing of Parameter Significance

a. Simultaneous F-Test

The simultaneous F test is a statistical hypothesis test of the regression coefficient simultaneously (Hsiao, 2014). In general, the F-test hypothesis can be written as follows (Ratnasari et al., 2023): H_0 is $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ (There is no effect of variable X on Y); and H_1 is There is at least one $\beta_k \neq 0\,; k = 1, \dots, p$ (There is an influence of variable X on Y). The F test statistic:

$$F = \frac{R^2/(N+p-1)}{(1-R^2)/(NT-N-p)} \tag{9}$$

where R^2 is coefficient of determination and p is a number of predictor variables. H_0 is rejected if the value of $F > F_{(\alpha,N+p-1,NT-N-p)}$ or p-value < α which can be concluded that between all predictor variables and the response variable have a significant effect.

b. Partial t-Test

The partial t-test is a hypothesis test used to determine whether the predictor variable exerts a significant effect on the response variable (Hsiao, 2014). This test is conducted on β_k (population regression coefficient). In general, the t-test hypothesis can be written as follows (Ratnasari et al., 2023): H_0 is $\beta_k = 0; k = 1, \dots, p$ (There is no effect of variable X on Y); H_1 is There is at least one $\beta_k \neq 0; k = 1, \dots, p$ (There is an influence of variable X on Y). The t test statistic:

$$t = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \tag{10}$$

where $\hat{\beta}_k$ is predictor variable regression coefficient and $se(\hat{\beta}_k)$ is standard error predictor variable regression coefficient. The t-test testing criteria are H_0 rejected if $|t| > t_{(\frac{\alpha}{2},NT-k-1)}$ or p-value < α, the number of estimated parameters, denoted as K, indicates that the predictor variable has a significant impact on the response variable.

C. RESULT AND DISCUSSION

1. Research Variable

The study employed secondary data obtained from BPS West Java Province covering the period from 2020 to 2022. In this research, the variables used are presented in Table 1.

Table 1. Research Variables

Variables	Symbol	Indicator	Unit
Response (Y)	Y	The number of tuberculosis cases	Cases
Predictor (X)	X_1	Percentage of poverty	Percent
	X_2	The number of diabetics	Cases
	X_3	The number of HIV/AIDS patients	Cases
	X_4	Population density	People/sq.km.

2. Descriptive Statistics

The data used in this study has a total sample size of 81. Table 2 below is a presentation of the data concentration used in completing this research.

Table 2. Descriptive Statistics

Variables	Mean	Standard Deviation	Minimum	Region of Minimum Value	Maximum	Region of Maximum Value
Y	3403,86	2713,83	256	Banjar City	12153	Bogor Regency
X_1	87,85	27,06	24,5	Depok City	131,3	Tasikmalaya City
X_2	33280,11	37731,33	2340	Banjar City	242169	Bekasi Regency
X_3	236,88	155,06	4	Pengandaran Regency	690	Bogor City
X_4	3878,98	4546,39	383	Pengandaran Regency	14957	Cimahi City

Based on the table above, Bogor Regency has the highest number of tuberculosis cases. Based on research conducted by (Fitria & Hartono, 2023), there is a significant correlation between temperature and the prevalence of tuberculosis in West Java Province in 2021. This research suggests that the temperature can affect the presence of bacteria that cause tuberculosis to grow optimally. In addition, Bogor District also has the highest number of HIV/AIDS cases, with 690 cases, which may increase susceptibility to TB infection. Tasikmalaya City has a high percentage of poor people, reaching 131,1%, which contributes to limited access to health services and increases the risk of TB. Bekasi Regency, with the highest number of diabetics at 14.957, shows how chronic health conditions can exacerbate susceptibility to TB. Meanwhile, Cimahi City has a high population density, with 242.169 people, which increases the potential for disease transmission. This data indicates that socioeconomic and health factors play an important role in TB prevalence in different regions.

3. Test of Classical Regression Assumptions

a. Detecting Collinearity

The VIF values of variables are presented in Table 3 below:

Table 3. Multicollinearity Detection Result

Label	Variables	VIF	Description
X_1	Percentage of poverty	1,74	No multicollinearity
X_2	The number of diabetics	1,19	No multicollinearity
X_3	The number of HIV/AIDS patients	1,13	No multicollinearity
X_4	Population density	1,76	No multicollinearity

According to the multicollinearity detection results in Table 3, the VIF value for each variable percentage of poverty (1,74), the number of diabetics (1,19), the number of HIV/AIDS patients (1,13), and population density (1,76) is less than 10. Therefore, it can be concluded that the model does not experience multicollinearity problems.

b. Heteroskedasticity Test

In this analysis, the heteroskedasticity test uses glejser method with the results presented in Table 4.

Table 4. Heteroskedasticity Test Result

Label	Variables	P-Value	Description
X_1	Percentage of poverty	0,2178	Not significant
X_2	The number of diabetics	0,3891	Not significant
X_3	The number of HIV/AIDS patients	0,3338	Not significant
X_4	Population density	0,3112	Not significant

According to the Heteroskedasticity test in Table 4, the probability value for each variable percentage of poverty (0,2178), the number of diabetics (0,3891), the number of HIV/AIDS patients (0,3338), and population density (0,3112) is greater than alpha (0,05). This means that all variables are not significant, it can be inferred that there is no heteroskedasticity present.

4. Panel Data Regression Model

a. Common Effect Model (CEM)

Common effect model test is a type of panel data model that combines cross-sectional and time series data. This model assumes that the behavior of the data across different cross-sections is consistent across different periods. The results of the common effect models are summarized in Table 5.

Table 5. Parameter Estimation in CEM

Variables	Estimated Value	Standard Deviation	t	P-Value
Intercept	3971,408	1463,622	2,713411	0,0082
X_1	-26,41333	12,91082	-2,045830	0,0442
X_2	0,015596	0,007644	2,040311	0,0448
X_3	5,586238	1,816412	3,075425	0,0029
X_4	-0,023030	0,077200	-0,298308	0,7663

The Common Effect Model (CEM) test results show a small R-squared value of 0.276. In addition, the population density variable has a p-value of 0.7663, which is greater than alpha (0.05), so it is not significant. In contrast, the significant variables are the percentage of poverty (X_1) with a p-value of 0,0442, the number of diabetics (X_2) with 0,0448, and the number of HIV/AIDS patients (X_3) with 0,0029. This indicates that the percentage of poverty, the number of diabetics, and the number of HIV/AIDS patients have a significant influence on tuberculosis cases.

b. Fixed Effect Model (FEM)

The assumption made by the fixed effect model is that differences among individuals can be accounted for by variances in intercepts. The results of the fixed effect models are summarized in Table 6.

Table 6. Parameter Estimation in FEM

Variables	**Estimated Value**	**Standard Deviation**	**t**	**P-Value**
Intercept	-13224,16	5749,893	-2,299897	0,0257
X_1	113,9799	45,49859	2,505130	0,0155
X_2	-0,011056	0,004979	-2,220542	0,0309
X_3	-4,209627	1,844810	-2,281875	0,0268
X_4	2,057155	0,976115	2,107492	0,0401

The FEM test results revealed an R-squared value of 0,90 and all variables have a p-value less than alpha (0,05). This suggests that all variables have significant values, which means that the number of tuberculosis cases in West Java is influenced by the percentage of poverty (X_1), the number of diabetics (X_2), the number of HIV/AIDS patients (X_3), population density (X_4) by 90% and the rest is influenced by other factors.

c. Random Effect Model (REM)

The test for the random effect model is utilized to calculate panel data that may have disturbance variables that are correlated between individuals and time. The results of the random effect models are summarized in Table 7.

Table 7. Parameter Estimation in REM

Variables	**Estimated Value**	**Standard Deviation**	**t**	**P-Value**
Intercept	4978,675	1833,217	2,715813	0,0082
X_1	-19,45322	17,00672	-1,143855	0,2563
X_2	-0,003443	0,004696	-0,733263	0,4657
X_3	0,058258	1,536371	0,037919	0,9699
X_4	0,060585	0,106453	0,569125	0,5709

The REM test results show that the R-squared value is very small at 0,03. In additional, the p-value for all predictor variables is greater than alpha (0,05). The percentage of poverty (X_1) with a p-value of 0,2563, the number of diabetics (X_2) with 0,4657, the number of HIV/AIDS patients (X_3) with 0,9699, and the population density (X_4) with 0.5709. So, it can be concluded that in the REM test there are no significant predictor variables.

5. Determining The Best Model of Panel Data Regression

In the pursuit of determining the best panel data regression model, researchers use two model estimation techniques, the chow test and the hausman test.

a. Chow Test

The results of the chow test analysis are summarized in Table 8 below:

Table 8. Chow Test Result

Effect Test	Statistics	Df	P-Value
Cross-section F	12,122429	(26;50)	0,0000

According to the results of the chow test in Table 8, the p-value for the cross-section F is 0,0000. This implies that the value is less than alpha (0,05). Therefore, the decision is rejected H_0 or accept H_1, and it can be inferred that the model chosen is the fixed effect model. Furthermore, the process of conducting the hausman test can be initiated.

b. Hausman Test

The results of the hausman test are typically presented in a table, such as Table 9.

Table 9. Hausman Test Result

Effect Test	Statistics	Df	P-Value
Cross-section random	32,210706	4	0,0000

According to the result of the hausman test in Table 9, the p-value for cross-section random is 0,0000. This implies that the value is less than alpha (0,05). Therefore, the decision to reject H_0 or accept H_1, and it can be inferred that the model chosen is the fixed effect model. The best model obtained is the fixed effect model, so there is no need to continue with the lagrange multiplier test.

6. Hypothesis Test of The Selected Model

The results obtained from both the Chow test and the Hausman test indicate that the model selected for the analysis is a fixed effect model.

a. Simultaneous F-Test

This study, the simultaneous F test was conducted using alpha (0,05) to identify significant differences between the means of groups or treatments in statistical analysis.

Table 10. Simultaneous F Test Results

F Test	P-Value
15,15991	0,000000

Based on the results of the simultaneous F test in Table 10 above, the p-value is 0,000000 or less than alpha (0,05), so the F test decision is obtained, namely reject H_0 with the conclusion that there is an influence of X on Y. This means that there is at least one predictor variable that affects the number of tuberculosis cases in West Java.

b. Partial t-Test

In this study, testing was carried out using a significance level (alpha) of 0,05 for the four predictor variables that were thought to have a significant effect on the response variable. Based on the partial t-test results in Table 6 above, it is obtained that the significant probability value for each variable X_1, X_2, X_3, and X_4 has a value less than alpha (0,05), so that the H_0 is rejected for all predictor variables. It can be concluded that the variable percentage of poverty, the number of diabetics, the number of HIV/AIDS patients, and population density each partially affect the response variable, namely the variable number of tuberculosis cases in West Java.

7. The Best Model of Panel Data Regression Interpretation

In the tests that have been carried out, it is obtained that the best model for this data is the fixed effect model estimate, so the formation of the regression model from this data is based on the output of the fixed effect model estimate. The formulation used to determine the relationship of this regression model is as Equation (11):

$$\hat{Y}_{it} = \hat{\alpha}_i + \boldsymbol{X_{it}}\widehat{\boldsymbol{\beta}} \tag{11}$$

Based on the previous FEM test result, the value for the formulation of the regression model is as follows:

$$\hat{Y}_{it} = \hat{\alpha}_i - 13224{,}16 + 113{,}9799X_1 - 0{,}011056X_2 - 4{,}209627X_3 + 2{,}057155X_4 \tag{12}$$

The interpretation of the regression model is that if the percentage of poverty (X_1) increases by 1%, the number of tuberculosis cases (Y) increases by 113 cases with other predictor variables considered constant. This is supported by research conducted by (Wong et al., 2013) using poisson regression that household poverty is positively associated with the sputum-positive TB CNR in Cambodia.

If the number of diabetics (X_2) increases by 100 cases, the number of tuberculosis cases (Y) decreases by 1 case with other predictor variables held constant. Based on the research conducted by (Ernest Yorke, 2017) there is a bidirectional relationship between TB and diabetes, and they both impact each other. So, it is not inline that the increase in the number of diabetics should increase the number of tuberculosis cases, this is due to factors that are not involved in the calculation of the number of diabetics, so it does not support the increase in the number of tuberculosis cases. If the number of HIV/AIDS patients (X_3) increases by 1 case, the number of tuberculosis cases (Y) decreases by 4 cases with other predictor variables held constant. It is unrealistic that the increase in the number of HIV/AIDS patients should increase the number of tuberculosis cases, this is due to factors that are not included in the calculation of the number of HIV/AIDS patients, so it does not support the increase in the number of tuberculosis cases.

The fact that an area can have a high prevalence of HIV and diabetes but a low incidence of TB can be explained by several factors. First, effective health programs, including prevention and treatment for HIV/AIDS and diabetes, often include efforts to address TB, thereby improving access to health services. In addition, high levels of education and public awareness encourage individuals to seek treatment early and implement preventive measures. Environmental factors, such as good sanitation and decent housing, also play a role in reducing TB risk. In addition, social support and active community networks help individuals manage their health, while genetic factors may make some people more resistant to TB. If the population density (X_4) increases by 10 people/sq. km. Then the number of tuberculosis cases (Y) increases by 20 people with other predictor variables held constant. This is supported by research conducted by (Asemahagn et al., 2021) using spatial-temporal clustering which states that TB high-risk clusters are increasingly emphasized by the effect of high population density.

The results of this study are in line with the findings of another study conducted by (Rahmawati et al., 2024) in East Java, where the variables of poverty and population density were also found to have a significant effect on tuberculosis (TB) cases. Research in East Java showed that in 2021, the number of poor people contributed significantly to the number of TB cases in all districts/cities, while population density had an effect in most areas, except in some districts such as Pacitan, Bondowoso, and Situbondo. These findings reinforce the argument that socioeconomic factors, such as poverty and population density, strongly influence the spread of TB in regions that share similar demographic and geographic characteristics. Therefore, policies aimed at reducing poverty through more equitable job creation are expected to reduce the number of TB cases. This strategy is important so that people are not concentrated in one area, so that population density can be controlled and the risk of disease transmission can be minimized.

D. CONCLUSION AND SUGGESTIONS

According to the data, the highest number of tuberculosis cases is in the Bogor Regency. Among the CEM, FEM, and REM models conducted, the best model to analyze the factors percentage of poverty, the number of diabetics, the number of HIV/AIDS patients, and population density towards tuberculosis cases in West Java is the FEM. Based on the results of this research data analysis and hypothesis testing of the FEM parameters, the conclusions obtained from this study are that all predictor variables have a significant effect on the variable number of TB cases in West Java. This means that the variables of the percentage of poverty, the number of diabetics, the number of HIV/AIDS patients, and population density each partially affect the response variable, namely the number of TB cases in West Java.

Recommendations that can be given by the West Java government from this research should be followed up by the government in order to reduce the number of TB cases in West Java in the following years. The government can reduce TB rates by reducing the percentage of poverty and population density in West Java. This can be accomplished by ensuring the availability of easily accessible healthcare facilities for the underprivileged and establishing health centers in densely populated areas. Additionally, it is recommended that public awareness campaigns and extension programs be initiated to underscore the importance of TB control in West Java. Furthermore, it is crucial to continue and enhance existing programs aimed at managing TB cases in West Java, including providing widespread healthcare access and offering vaccines to prevent TB.

Then for further research, you can add variables other things that might influence economic growth such as temperature, humidity, or variables that have a linear influence on TB cases in east Java. According to WHO, TB bacteria can survive for hours in a dark and humid environment, but they perish immediately when exposed to sunlight. From previous research in the health sector, TB susceptibility is potentially affected by temperature and humidity. This is because TB spreads through the air, and TB bacteria can survive longer in a humid environment (Amjani & Cusmarih, 2023). The weather conditions in West Java are characterized by cold temperatures and high humidity. These conditions may have an impact on the prevalence of TB cases in West Java.

ACKNOWLEDGEMENT

The authors express their gratitude to BPS for providing the data, the Statistics Study Program, the Faculty of Science and Technology, and Universitas Airlangga.

REFERENCES

Amjani, A., & Cusmarih, C. (2023). Hubungan Lingkungan Rumah Terhadap Kejadian Tuberkulosis Paru di RSUD Kabupaten Bekasi. *MAHESA : Malahayati Health Student Journal*, *3*(10), 3131–3140. https://doi.org/10.33024/mahesa.v3i10.11007

Asemahagn, M. A., Alene, G. D., & Yimer, S. A. (2021). Spatial-temporal clustering of notified pulmonary tuberculosis and its predictors in East Gojjam Zone, Northwest Ethiopia. *PLoS ONE*, *16*(1), 1–19. https://doi.org/10.1371/journal.pone.0245378

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality and Quantity*, *53*(2), 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Biørn, E. (2016). *Econometrics of Panel Data: Methods and Applications*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198753445.001.0001

Brilliant, I. I., & Fakhriyana, D. (2023). Modeling The Number Of Tuberculosis Cases In West Java Using The Negative Binomial Approach. *Consilium Sanitatis: Journal of Health Science and Policy*, *1*(2), 107–119. https://doi.org/10.56855/jhsp.v1i2.282

Dewi Anggraeni Chairunnisa, A., & Fauzan, A. (2023). Implementation of Panel Data Regression in the Analysis of Factors Affecting Poverty Levels in Bengkulu Province in 2017-2020. *EKSAKTA: Journal of Sciences and Data Analysis*, *4*(1), 40–45. https://doi.org/10.20885/eksakta.vol4.iss1.art5

Djalic, I., & Terzic, S. (2021). Violation of the assumption of homoscedasticity and detection of heteroscedasticity. *Decision Making: Applications in Management and Engineering*, *4*(1), 1–18. https://doi.org/10.31181/dmame2104001d

Feng, S., Li, G., Tong, T., & Luo, S. (2020). Testing for heteroskedasticity in two-way fixed effects panel data models. *Journal of Applied Statistics*, *47*(1), 91–116. https://doi.org/10.1080/02664763.2019.1634682

Grigore, G. E., Musetescu, R. C., Nicolae, S., & Vladut, O. (2022). A panel data analysis in estimating the economic growth for oil-producing countries.Evidence from the caspian region. *Economic Computation and Economic Cybernetics Studies and Research*, *56*(4), 225–242. https://doi.org/10.24818/18423264/56.4.22.14

Hsiao, C. (2014). Analysis of panel data: Third edition. *Analysis of Panel Data: Third Edition*, 1–539. https://doi.org/10.1017/CBO9781139839327

Ismaeel, S. S., Midi, H., & Sani, M. (2021). Robust multicollinearity diagnostic measure for fixed effect panel data model. *Malaysian Journal of Fundamental and Applied Sciences*, *17*(5), 636–646. https://doi.org/10.11113/MJFAS.V17N5.2391

Karavias, Y., Narayan, P. K., & Westerlund, J. (2023). Structural Breaks in Interactive Effects Panels and the Stock Market Reaction to COVID-19. *Journal of Business and Economic Statistics*, *41*(3), 653–666. https://doi.org/10.1080/07350015.2022.2053690

Kementerian Kesehatan RI. (2020). *Pedoman Nasional Pelayanan Kedokteran Tata Laksana Tuberkulosis*. https://tbindonesia.or.id/download/9840/?tmstv=1718644405

Kementerian Kesehatan RI. (2023). *Laporan Program Penanggulangan Tuberkulosis Tahun 2022 Kementerian Kesehatan Republik Indonesia Tahun 2023*. https://tbindonesia.or.id/wp-content/uploads/2023/09/Laporan-Tahunan-Program-TBC-2022.pdf

Liao, N., & He, Y. (2018). Exploring the effects of influencing factors on energy efficiency in industrial sector using cluster analysis and panel regression model. *Energy*, *158*, 782–795. https://doi.org/10.1016/j.energy.2018.06.049

Maison, D. P. (2022). Tuberculosis pathophysiology and anti-VEGF intervention. In *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* (Vol. 27). Elsevier Ltd. https://doi.org/10.1016/j.jctube.2022.100300

Marcoulides, K. M., & Raykov, T. (2019). Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educational and Psychological Measurement*, *79*(5), 874–882. https://doi.org/10.1177/0013164418817803

Muhith, A., Susanto, I. H., Rahmalia, D., Adzkiya, D., & Herlambang, T. (2022). The Analysis of Demand and Supply of Blood in Hospital in Surabaya City Using Panel Data Regression. *Nonlinear Dynamics and Systems Theory*, *22*(5), 550–560.

Piratla, S., & Singh, B. K. (2023). Airport passenger movement, revenue and expenditure at non-major airports using panel data regression. *Social Sciences and Humanities Open*, *8*(1), 1–18. https://doi.org/10.1016/j.ssaho.2023.100658

Pratiwi, K. D. (2022). E-Commerce and Economic Growth in Indonesia: Analysis of Panel Data Regression. *Gema Publica*, *7*(1), 171–186. https://doi.org/10.14710/gp.7.1.2022.171-186

Prunier, J. G., Colyn, M., Legendre, X., Nimon, K. F., & Flamand, M. C. (2015). Multicollinearity in spatial genetics: Separating the wheat from the chaff using commonality analyses. *Molecular Ecology*, *24*(2), 263–283. https://doi.org/10.1111/mec.13029

Purwanto, E. (2020). Analysis of The Effect of Healthy and Healthy Life and Health (PHBS) Behavior of Tuberculosis Events. *Journal of Public Health Science Research*, *1*(1), 5. https://doi.org/10.30587/jphsr.v1i1.1179

Puspita, S. N., Maryani, S., & Purwantho, H. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Penyerapan Tenaga Kerja Di Provinsi Jawa Tengah. *Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, *13*(2), 141. https://doi.org/10.20884/1.jmp.2021.13.2.4546

Rahmawati, N., Karno, F., & Hermanto, E. M. P. (2024). Analisis Penyakit Tuberkulosis (TBC) pada Provinsi Jawa Timur Tahun 2021 Menggunakan Geographically Weighted Regression (GWR). *Indonesian Journal of Applied Statistics*, *6*(2), 116. https://doi.org/10.13057/ijas.v6i2.78593

Ratnasari, V., Audha, S. H., & Dani, A. T. R. (2023a). Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression. *MethodsX*, *11*. https://doi.org/10.1016/j.mex.2023.102379

Ratnasari, V., Audha, S. H., & Dani, A. T. R. (2023b). Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression. *MethodsX*, *11*(July), 102379. https://doi.org/10.1016/j.mex.2023.102379

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Suherni, N., & Maduratna. (2013). Analisis Pengelompokan Kecamatan di Kota Surabaya Berdasarkan Faktor Penyebab Terjadinya Penyakit Tuberkulosis. *Jurnal Sains Dan Seni ITS*, *2*(1), 2337–3520.

Utari, D. T., Yuliana, T., & Hendradewa, A. P. (2020). *A Panel Data Analysis of Rice Production in Ngawi Regency, East Java*. *474*, 212–217. https://doi.org/10.2991/assehr.k.201010.031

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Wardhana, F. A., & Indawati, R. (2021). Panel Data Regression Analysis For Factors Affecting Infant Mortality Rate In East Java 2013-2017. *Indonesian Journal of Public Health*, *16*(3), 437–448. https://doi.org/10.20473/ijph.v16i3.2021.437-448

Wong, M. K. ai, Yadav, R. P., Nishikiori, N., & Eang, M. T. an. (2013). The association between household poverty rates and tuberculosis case notification rates in Cambodia, 2010. *Western Pacific Surveillance and Response Journal : WPSAR*, *4*(1), 25–33. https://doi.org/10.5365/WPSAR.2013.4.1.002