International Conference on Sustainable Economics, Management, and Accounting (ICSEMA 2026)
Proceedings
Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337

THE EFFECT OF UNIVERSITY PRESENCE ON PER CAPITA GRDP IN JAVA AND KALIMANTAN USING PROPENSITY SCORE MATCHING

Arum Diyah Fatmawati[1], Hasna Aufa Al-amin[2*], Maulina Dwi Aprilia[3], Annisa Nur Fathonah[4], Alfitta Tiara Hapsari[5], Pratiwi Dwi Saputri[6]
[1-6]Department Regional Economic Development, Gadjah Mada University, Indonesia

[*]Email corresponding author: hasnaaufaal-amin2005@mail.ugm.ac.id

Article Information: Received: June, 25 2026, **Published: August, 1 2026**

Abstract

Regional economic development that is equitable remains one of the primary targets of Indonesia's national development policy. One factor presumed to contribute to regional economic growth is the presence of higher education institutions, particularly universities. Universities serve as centers for human resource development, technological innovation, and stimulants of local consumption and investment. This study aims to analyze the causal effect of university presence on per capita Gross Regional Domestic Product (GRDP) in districts/cities across Java and Kalimantan in 2023. Data were sourced from the Central Statistics Agency (BPS) and the Ministry of Education, Culture, Research, and Technology for the year 2023. Control variables include total population, poverty severity index, and open unemployment rate, with the treatment variable being a university presence dummy (1 = present, 0 = absent) and the outcome variable being per capita GRDP (in million rupiah). The method employed is Propensity Score Matching (PSM) to address selection bias inherent in non-experimental observational data. Estimation results indicate that districts/cities with universities have, on average, significantly higher per capita GRDP compared to those without universities, after controlling for relevant regional characteristics. These findings imply the importance of equitable access to higher education as an inclusive regional economic development policy strategy.

Keywords: Propensity Score Matching, Per Capita GRDP, Higher Education Institutions, University Presence, Regional Economic Development.

INTRODUCTION

Equitable economic development across regions is one of the primary priorities in Indonesia's national development agenda. This effort aims to ensure that economic progress is not concentrated in certain areas alone, but can reach all segments of society, including communities in underdeveloped, frontier, and outermost regions (3T). Through this equalization, every region is expected to have equal opportunities in accessing infrastructure, employment, and quality public services. Nevertheless, economic disparities among districts and cities remain a significant challenge. These inequalities are driven by various factors, ranging from the availability of supporting infrastructure and the quality of human resources to each region's capacity to attract investment. Based on data from the Central Statistics Agency (BPS) in 2023, per capita Gross Regional Domestic Product (GRDP) across regions in Indonesia shows significant gaps, particularly between districts/cities in Java and Kalimantan two islands that contribute the most to the national Gross Domestic Product (GDP).

Java currently dominates with a contribution of approximately 57% to the national GDP, making it the epicenter of economic activity, industry, and national education. In this region, the higher education sector is strongly linked to labor absorption by industry, thereby stimulating self-sustaining economic growth. On the other hand, Kalimantan is experiencing accelerated development through various strategic government programs, most notably the policy of relocating the National Capital (IKN) to East Kalimantan. This transformation is not merely an administrative shift but a concrete effort to create a new, Indonesia-centric economic center of gravity. The development dynamics of both islands make them highly relevant for examination, particularly in observing the relationship between the availability of higher education infrastructure and regional economic growth. While universities in Java have long served as established suppliers of human capital, Kalimantan faces the challenge of whether educational institutions can keep pace with physical infrastructure development such as IKN, so that economic growth can be fully absorbed by the local population rather than by migrant workers.

Higher education institutions are believed to play a strategic role in stimulating local economic growth. Through human capital development, universities contribute to increasing labor productivity, driving technological innovation, attracting investment, and stimulating economic activity through the consumption expenditure of their academic communities. The presence of universities is expected to catalyze a structural transformation away from resource-dependent economies toward knowledge-based economies. Several empirical studies across various countries have demonstrated that the presence of universities has a positive impact on regional economic growth (Abel & Deitz, 2012; Moretti, 2004; Drucker & Goldstein, 2007).

Despite this, the distribution of universities in Indonesia remains concentrated in major cities, while many districts in remote areas still lack adequate access to higher education. This raises a fundamental question regarding the causal effect of university presence on regional economic welfare as measured by per capita GRDP. This question is difficult to answer straightforwardly due to selection bias, where regions that have universities typically already possess superior socioeconomic characteristics from the outset. Without carefully isolating causal factors, there is a risk of overestimating the role of universities where progress that actually stems from basic infrastructure development or road access may be mistakenly attributed to the presence of educational institutions.

This study aims to analyze the causal effect of university presence on per capita GRDP in districts and cities across Java and Kalimantan in 2023, using Propensity Score Matching (PSM) as the primary analytical method. PSM operates by matching observational units between the treatment group (regions with universities) and the control group (regions without universities) based on propensity scores, the estimated probability of a region having a university derived from its regional characteristics (Rosenbaum & Rubin, 1983). Specifically, this study seeks to: (1) identify and analyze differences in per capita GRDP between regions with and without universities; (2) estimate the effect of university presence on per capita GRDP while controlling for population size, poverty severity index, and open unemployment rate; and (3) measure the magnitude of the Average Treatment Effect on the Treated (ATT) of university presence on per capita GRDP.

The novelty of this study lies in its application of PSM to rigorously address selection bias in the context of Indonesian regional development a methodological approach that has been underutilized in the domestic literature on higher education and economic growth. While most prior studies on the impact of universities tend to focus on developed countries with stable socioeconomic conditions, this research fills that gap by examining the interaction between educational institutions and economic dynamics in an archipelagic nation characterized by high levels of regional inequality.

From a theoretical standpoint, this study contributes to the body of knowledge on the impact of higher education on regional economic growth in developing countries, particularly Indonesia, providing a new perspective for understanding the application of human capital theory in explaining economic growth at the regional level across diverse contextual backgrounds. From a practical standpoint, the findings of this study are expected to serve as a reference for policymakers at both the central and regional levels in formulating higher education development strategies as instruments of economic equalization. By understanding the actual impact of universities, the government can more precisely determine development priorities whether to prioritize the establishment of new institutions in underdeveloped regions or to strengthen the synergy between existing institutions and local industries. This is crucial to ensure that development budget allocations are better targeted, so that state investment in education can deliver maximum impact on improving community welfare and reducing inter-regional disparities throughout Indonesia.

LITERATURE REVIEW AND HYPOTHESIS FORMULATION

Regional Economic Development

a. Regional Economic Development Concept

Regional economic development is a process that involves local governments and communities in managing and developing available resources to stimulate economic activity and improve the well-being of the community (Arsyad, 2016). More broadly, economic development is a multidimensional process that encompasses changes in economic, social, and institutional structures, and aims to improve the well-being of society by reducing poverty and inequality (Todaro & Smith, 2020). In the modern context, regional economic development also emphasizes the importance of sustainability that is, a balance between economic growth, social well-being, and environmental conservation.

b. Interregional Economic Disparities

Interregional economic disparities can be explained by the Williamson curve hypothesis, which suggests that in the early stages of economic development, interregional inequality tends to increase due to the concentration of economic activity in certain regions. However, in the long run, this inequality may decrease as economic activity spreads to other regions (McCann, 2016; Krugman in a recent regional study). Additionally, interregional disparities can also be explained through a cumulative approach, in which economic growth tends to be self-reinforcing. More developed regions will continue to attract resources such as labor and investment, thereby widening the gap with less developed regions (Martin & Sunley, 2022; Storper, 2018).

c. Economic Development Indicator: Per Capita GRDP

GRDP per capita is the gross value added generated by all economic activities in a region divided by the total population, and is thus used to describe the average economic capacity of the population (BPS, 2023). A higher GRDP per capita generally indicates a higher level of economic well-being in a region. However, this measure has limitations because it fails to reflect income distribution or non-economic aspects such as quality of life, health, and education (UNDP, 2020; Stiglitz et al., 2019).

Theories of Economic Growth

a. Classical and Neoclassical Growth Theories

The neoclassical growth theory developed by Solow explains that economic growth is determined by the accumulation of capital, labor, and exogenous technological progress. In this model, the economy will reach a long-run equilibrium (steady state), where per capita output growth is determined solely by technological progress, while the contribution of capital declines due to diminishing returns (Romer, 2019; Mankiw, 2019; IMF, 2020).

b. Endogenous Growth Theory

The endogenous growth theory emphasizes that knowledge and technology are factors of production generated within the economic system and exhibit characteristics of increasing returns, thereby driving long-term economic growth (Romer, 2019). Furthermore, the accumulation of human capital such as education and skills also plays a crucial role in enhancing labor productivity and economic growth (Mankiw, 2019; World Bank, 2020).

Human Capital Theory

a. The Concept of Human Capital

Investment in education plays a crucial role in improving the quality of human capital and driving economic growth by boosting labor productivity. Human capital encompasses various aspects, such as formal education, skills, experience, and the ability to innovate (World Bank, 2020; OECD, 2019).

b. Education as an Economic Investment

From an economic perspective, education is viewed as an investment in human capital aimed at improving the quality of human resources by enhancing individuals' knowledge, skills, and work capabilities. This investment benefits not only individuals but also the economy as a whole through increased productivity and economic growth (World Bank, 2020; UNESCO, 2021). Furthermore, education is a key factor in economic development as it plays a role in improving the quality of the workforce and fostering innovation (OECD, 2019). In the Indonesian context, improvements in educational quality are also correlated with increased societal well-being, as evidenced by rising incomes (BPS, 2023; ADB, 2020).

c. The Impact of Higher Education on the Economy

Higher education plays a crucial role in driving economic growth, particularly through improving the quality of the workforce and fostering innovation. According to BPS (2023), regions with higher levels of education tend to have higher per capita income, indicating a link between higher education and economic well-being. Additionally, Wicaksono (2018) states that the presence of universities can boost local economic activity through job creation and increased consumer spending.

Regional Disparities

a. Economic Concentration on the Island of Java

Java is the center of economic activity in Indonesia, contributing the largest share to the national GDP. This is evidenced by data from the Central Statistics Agency (BPS), which indicates that more than half of the national GDP comes from Java (BPS, 2024). This situation is influenced by the concentration of industry, more advanced infrastructure, and a higher level of urbanization compared to other regions (World Bank, 2020). Additionally, Java Island also holds an advantage in terms of human resource availability, supported by a greater number of educational institutions. This economic concentration can lead to inter-regional disparities due to differences in access to resources and economic opportunities between developed and underdeveloped regions (ADB, 2019; Kuncoro).

b. Economic Development in Kalimantan

Kalimantan is a region with significant natural resource potential, particularly in the mining and forestry sectors, which are the main contributors to the regional economy. However, the region's economic structure remains dominated by the primary sector, making its economic growth highly susceptible to fluctuations in commodity prices (BPS, 2024). This reliance on resource-based sectors makes economic growth more volatile compared to regions with a more diversified economic structure (World Bank, 2020). Additionally, infrastructure limitations and the quality of human resources also pose challenges in driving economic transformation in Kalimantan (ADB, 2021). Therefore, improving access to education and the quality of human resources are critical factors in supporting more sustainable economic development (UNDP, 2020).

Propensity Score Matching (PSM) Method

a. Definition and Purpose of PSM Propensity Score Matching (PSM)

Prepensity Score Matching (PSM) is a statistical method used to estimate the effect of a treatment on an outcome in observational data. This method is used because there are often differences in characteristics between the group receiving the treatment and the group not receiving the treatment, which can lead to bias in the analysis results. According to Gertler et al. (2016), PSM aims to reduce selection bias by matching units with similar characteristics so that comparisons between groups are more equitable.

b. The Concept of Propensity Score Matching

The key concept in Propensity Score Matching is a probability score that indicates the likelihood of a unit or individual receiving a treatment based on their characteristics, such as economic status, education, or other variables. According to Gertler et al. (2016), two units with the same propensity score are considered to have comparable characteristics, allowing them to be directly compared. The process in this method begins with calculating the propensity score, followed by matching the treatment group with the non-treatment group based on these scores. Next, a balance test is conducted to ensure that both groups share similar characteristics, and finally, an estimation of the treatment's effect on the outcome of interest is performed. This method is based on two main assumptions: the Conditional Independence Assumption (CIA), which states that no other variables influence the outcome besides those observed, and the Common Support assumption, which indicates that the two groups share similar characteristics.

RESEARCH METHODS

Research Design

This study adopts a quantitative research design grounded in a non-experimental, observational approach. The treatment variable of interest, namely the presence of a university within a given district or city, was not randomly assigned but emerged organically from historical developments and the structural socioeconomic characteristics inherent to each region. Given this observational nature, a direct comparison of outcomes between treated and untreated units would be susceptible to selection bias, rendering straightforward causal inference untenable. To address this methodological challenge, the study applies the Propensity Score Matching (PSM) framework, a quasi-experimental technique specifically designed to reduce bias in observational settings by constructing a credible counterfactual comparison group.

The empirical basis of this study consists of secondary cross-sectional data pertaining to the year 2023, capturing observations across multiple districts and cities at a single point in time. This design is well-suited to the research objective, which is to identify and quantify differences in regional economic performance, as measured by per capita Gross Regional Domestic Product (GRDP), between districts and cities that have universities and those that do not, within the same temporal framework. By applying PSM, this study seeks to approximate the conditions of a controlled experiment, thereby enabling a more rigorous and credible estimation of the causal effect of university presence on regional economic output.

Population and Sample

The population in this study comprises all districts/cities located on the islands of Java and Kalimantan. Java encompasses the provinces of DKI Jakarta, West Java, Banten, Central Java, the Special Region of Yogyakarta (DIY), and East Java, while Kalimantan covers the provinces of West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. These two islands were selected on the grounds that they represent the two largest contributors to Indonesia's national GDP yet exhibit markedly different patterns of economic development and higher education infrastructure. Java is characterized by a high concentration of universities and

Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337

advanced industrial activity, whereas Kalimantan is undergoing accelerated development, including the relocation of the national capital (IKN), but still faces limited access to higher education. This contrast makes the two islands highly relevant for examining the relationship between university presence and regional economic performance.

The sampling technique applied in this study is total population sampling, also referred to as a census approach, in which all members of the defined population are included as units of analysis. Accordingly, the sample consists of 175 districts and cities, with 119 located in Java and 56 in Kalimantan. No units were excluded, as complete data were available for all relevant variables from official statistical sources for the year 2023. This approach ensures that the findings are representative of the entire population of interest without the risk of sampling error.

Data Collection and Instrument Development Techniques

This study relies exclusively on secondary data collected through documentation of official statistical publications. Data was sourced from two institutional providers. The first is Statistics Indonesia (Badan Pusat Statistic/BPS), from which data on per capita Gross Regional Domestic Product (GRDP) at current market prices, total population, the Poverty Severity Index (P2), and the Open Unemployment Rate (OUR) were obtained through regional statistical publications and Provincial Statistical Yearbooks (Province In Figures) for the year 2023. The second source is the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek), from which the presence or absence of universities in each district/city was identified using the 2023 Indonesian Higher Education Directory.

The university presence variable was subsequently coded as a binary dummy variable, taking the value of 1 if at least one university is registered in the district/city, and 0 otherwise. All data were compiled and organized at the district/city level to form a single cross-sectional dataset comprising 175 observations. The complete operationalization of all research variables is presented in Table below.

Table 1. Research Variables

Variable	Notation	Operational Definition	Data Source
Per Capita GRDP	Y	The value of GRDP at Current Market Prices is divided by the total population of a district/city, expressed in millions of rupiah (IDR).	BPS, 2023
University Presence	T	A binary dummy variable equal to 1 if at least one university is present in the district/city, and 0 if none is present.	Kemendikbudristek, 2023
Total Population	X_1	The total number of residents in a district/city expressed in thousands of people.	BPS, 2023
Poverty Severity Index	X_2	An index measuring the degree of disparity between the average expenditure of the poor and the poverty line in a district/city.	BPS, 2023
Open Unemployment Rate (OUR)	X_3	The percentage of unemployed persons relative to the total labour force in a district/city, expressed as a percentage (%).	BPS, 2023

Source: BPS and Kemendikbudristek, 2023 (processed by the authors)

Data Analysis Techniques

Propensity Score Matching (PSM)

The primary analytical method employed in this study is Propensity Score Matching (PSM). This method is selected because the study is observational and non-experimental in nature, meaning that university presence as the treatment variable was not randomly assigned but rather emerged from historical processes and the structural characteristics of each region. Consequently, there is a substantial risk of selection bias districts and cities that have universities are likely to differ systematically from those without universities in terms of socioeconomic and demographic characteristics, even before any effect of the university is considered. A direct comparison of mean outcomes between the two groups would therefore yield biased and misleading causal estimates. All data processing and statistical analysis were carried out using Stata 17. The PSM procedure was implemented primarily through the teffects psmatch command, complemented by the psmatch2 package for additional matching specifications. Covariate balance was evaluated using the tebalance summarize command for numerical balance statistics and tebalance density for graphical inspection of propensity score distributions. Sensitivity analysis to assess the robustness of causal estimates against potential unobserved confounders was performed using the Rosenbaum Bounds procedure, executed via the rebounds command.

Stages of PSM Analysis

The PSM procedure in this study follows the systematic stages recommended by Caliendo and Kopeinig (2008):

a. Transformation Data

Prior to the main analysis, the GRDP per capita and total population variables were transformed into their natural logarithmic forms. This transformation is intended to reduce the skewness of the data distribution and stabilize the variance, as is common practice in empirical economics literature for variables with highly right-skewed distributions (Wooldridge, 2010). Operationally, the transformation is carried out using the following commands:

gen log_pdrbperkapita = ln(pdrbperkapita)

gen log_jumlahpendudukribu = ln(jumlahpendudukribu)

b. Pre-Matching Mean Comparison Test (Before Matching)

Prior to the matching procedure, a test is conducted to determine whether there are statistically significant differences in mean outcomes between districts/cities with and without universities. This test is performed using an independent two-sample t-test with the following command:

ttest log_pdrbperkapita, by(treatmentuniv)

The test statistic follows the standard formulation:

$$t = \frac{\bar{Y}_{treated} - \bar{Y}_{control}}{\sqrt{\frac{s^2 treated}{n_{treated}} + \frac{s^2 control}{n_{control}}}}$$

This stage is important both to confirm the presence of initial differences that need to be corrected through matching, and to establish a baseline for pre- and post-matching comparisons.

c. Propensity Score Estimation

Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337

The propensity score is defined as the conditional probability that a districts/cities receives the treatment given its observed covariate vector (Rosenbaum & Rubin, 1983):

$$p(X_i) = P(D_i = 1 \mid X_i) = \frac{e^{x_i^1}\beta}{1 + e^{x_i^1}\beta}$$

Where Di is the treatment status and Xi is the covariate vector comprising the natural logarithm of population, the poverty severity index, and the open unemployment rate (TPT). The propensity score for each observation is then predicted using:

predict ps0 ps1, ps

d. Matching

Once the propensity scores have been estimated, the matching procedure is carried out to pair each treated districts/cities with the control districts/cities that has the most similar propensity score value. Propensity score estimation and the matching process are conducted simultaneously using the following command:

teffects psmatch (log_pdrbperkapita) (treatmentuniv log_jumlahpendudukribu indexkeparahankemiskinan tpt), gen(match)

As an alternative and for robustness purposes, matching is also performed using the command:

psmatch2 treatmentuniv log_jumlahpendudukribu indexkeparahankemiskinan tpt, out(log_pdrbperkapita) at

The potential outcomes for each observation are also estimated through:

predict y0 y1, po

e. Treatment Effect Estimation (ATE/ATT)

The individual treatment effect for each observation is estimated using:

predict te

Two effect measures are then estimated. First, the Average Treatment Effect (ATE), which measures the mean treatment effect across the entire population:

$$ATE = E[Y(1) - Y(0)]$$

This is computed using:

su te

Second, the Average Treatment Effect on the Treated (ATT), which measures the mean treatment effect specifically for districts/cities that actually have universities:

$$ATT = E[Y(1) - Y(0) \mid D = 1]$$

This is computed using:

Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337

su te if treatmentuniv == 1

As a robustness check, propensity score estimation is also performed using a probit model as an alternative to the default logit model:

effects psmatch (log_pdrbperkapita) (treatmentuniv log_jumlahpendudukribu indexkeparahankemiskinan tpt, probit)

f. Result Evaluation

The distribution of matching weights in the control group is examined to detect whether any control units are used excessively (overweighted), which may indicate weak common support in the data:

tab _weight if !treatmentuniv

g. Balance Test

This stage constitutes a crucial step in validating that the matching procedure has successfully balanced the distribution of covariates between the treated and control groups. Balance is assessed using the standardized bias (SB), formulated as:

$$SB = \frac{\bar{X}_{treated} - \bar{X}_{control}}{\sqrt{\frac{1}{2}(V_{treated}\ (X) + (V_{control}\ (X)}} \times 100$$

The test is conducted using the following command:

pstest log_jumlahpendudukribu indexkeparahankemiskinan tpt, both

Rosenbaum and Rubin (1985) suggest that a standardised bias value below 10 percent after matching can be considered sufficient evidence of adequate covariate balance, thereby lending methodological credibility to the causal inferences drawn from the matching procedure

RESULTS AND DISCUSSION

Data and Variable Description

This study uses cross-sectional secondary data covering 175 districts/cities in Java and Kalimantan in 2023. The data were obtained from the Central Bureau of Statistics (BPS) through the Regional Statistics and Province in Figures 2023 publications, as well as university presence data from the Higher Education Directory of the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) 2023. The dependent variable is GDP per capita measured in thousands of rupiahs. This variable was transformed into its natural logarithm (ln) to address the right-skewed distribution and to allow the estimated coefficients to be interpreted directly as percentage changes in economic output. The treatment variable is university presence, coded as a binary dummy variable: 1 if a districts/cities has at least one university, and 0 if it has none. Three control variables are included in this study. First, the total population in thousands also transformed using the natural logarithm to capture the nonlinear relationship between population size and the probability of having a university. Second, the Poverty Severity Index (P2), which measures how far the average expenditure of the poor falls below the poverty line, reflecting the depth and severity of poverty in a region. Third, the Open Unemployment Rate (OUR), which measures the percentage

Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337

of the labor force that is unemployed but actively seeking work. These three control variables were selected because they theoretically influence both the regional economic outcome and the likelihood of a region having a higher education institution and therefore must be controlled in the PSM estimation.

Descriptive Statistic

Before conducting further analysis, it is necessary to first understand the general distribution of each variable in the study sample. Table 2 presents the descriptive statistics for all variables used.

Table 2. Descriptive Statistics of All Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
GDP per Capita (thousand IDR)	175	84,430.620	98,440.730	21,228.000	818,954.000
Log GDP per Capita	175	10.994	0.749	9.963	13.616
Total Population (thousands)	175	985.870	863.450	27.500	5,627.020
Log Total Population	175	6.480	0.813	3.314	8.636
Poverty Severity Index	175	10.857	44.319	0.010	395.000
OUR (%)	175	5.177	2.018	1.520	10.520
University Presence	175	0.577	0.495	0.000	1.000

Source: Stata output, processed by the author.

Table 2 shows that the average GDP per capita across districts/cities in the sample is IDR 84.4 million per year, with a very large standard deviation of IDR 98.4 million. This indicates a high degree of economic disparity across regions. The minimum value of IDR 21.2 million and the maximum of IDR 818.9 million reflects a nearly 38-fold gap between the poorest and wealthiest regions in the sample. This high dispersion is the primary reason why the GDP per capita variable was transformed into its natural logarithm for statistical analysis, making its distribution closer to normal. Of the 175 districts/cities in the sample, 101 regions (57.71%) have at least one university, and 74 regions (42.29%) have none. This proportion indicates that, while universities are relatively widespread, more than 40% of regions in Java and Kalimantan still lack a university-type higher education institution. Population size also shows substantial variation, ranging from 27,500 (very sparsely populated areas) to 5.6 million. The Poverty Severity Index similarly shows considerable disparity, with a maximum value of 395.00 that far exceeds the mean of 10.857, suggesting that some districts experience extremely severe poverty conditions. Meanwhile, the OUR ranges from 1.52% to 10.52% with a mean of 5.18%.

Pre-Matching Mean Difference Test (Before Matching)

The first step before applying PSM is to statistically verify that there is a significant difference between the treatment and control groups. A two-sample t-test was conducted on the log GDP per capita variable based on university presence status. Table 4 presents the results.

Table 3. Two-Sample T-Test Results Group Obs Mean Std. Err. S

Group	Obs	Mean	Std. Err.	Std. Dev.
No University	74	10.7970	0.0783	0.6733
University Present (T=1)	101	11.1380	0.0768	0.7722
Combined	175	10.9940	0.0566	0.7493
Difference (T=0 - T=1)	-	- 0.3414	0.1120	-

Source: Stata output, processed by the author.

Table 4. Summary Statistics of T-Test Results

t-statistic	df	H_0: diff = 0 (p-value, 2-tailed)	Conclusion
-3.047	173	0.0027 ***	Reject H_0 (Significant difference)

Source: Stata output, processed by the author.

The t-test results show that a t-statistic of -3.047 with 173 degrees of freedom yields a p-value of 0.0027, which is well below the 1% significance level. Accordingly, H_0 (no difference in mean log GDP per capita between the two groups) is strongly rejected. The mean log GDP per capita of regions with a university (11.138) is significantly higher by 0.341 points compared to regions without a university (10.797). These results confirm the presence of strong selection bias in the data, regions with universities are systematically different from those without, even before any intervention is considered. This difference may be driven by factors such as population size, poverty conditions, and labor market structure, rather than solely by the direct effect of a university. The existence of this selection bias provides strong methodological justification for using the Propensity Score Matching (PSM) method in this study.

Propensity Score Estimation

The propensity score was estimated using a probit model, with university presence (T) as the dependent variable and log total population (ln_X_1), poverty severity index (X_2), and OUR (X_3) as independent variables. The use of the probit model follows the procedure applied in the psmatch2 command in Stata. The estimation results are presented in Table 5.

Table 5. Propensity Score Model Estimation Results

Variable	Coefficient	Std. Error	z-stat	p-value	Conclusion
Log Total Population (ln X_1)	0.3179	0.1159	2.74	0.006	Significant
Poverty Severity Index (X_2)	-0.0033	0.0030	-1.12	0.262	Not Significant
OUR (X3)	0.2408	0.0592	4.07	0.000	Significant
Constant	-3.0348	0.7355	-4.13	0.000	Significant

Source: Stata output, processed by the author.

Log total population (ln X_1) has a positive and significant effect at the 99% confidence level, with a coefficient of 0.3179 (p = 0.006). This means that the larger a districts/cities population, the higher the probability that it has a university. This is theoretically consistent, as more populous regions have greater demand for higher education and provide a more viable market for the sustainable operation of a university. The Poverty Severity Index (X_2) has a negative coefficient (-0.0033), which is in line with theoretical expectations, but is not statistically significant (p = 0.262). The negative direction is nonetheless consistent with the logic that regions experiencing more severe poverty face greater limitations in providing the infrastructure and purchasing power needed to support higher education institutions, though the effect is not strong enough to be statistically detected in this sample. The OUR (X_3) shows a positive and significant coefficient (p = 0.000), which may appear counterintuitive at first glance. However, in the Indonesian context, this finding can be explained by the structural characteristics of urban areas. Districts/cities with universities tend to have more formal and complex labor market structures. In these areas, residents are more likely to be classified as active job seekers (and therefore counted in the OUR), whereas in rural areas without universities, unemployment tends to be disguised, taking the form of underemployment that is not captured in formal OUR statistics.

Matching Results and Treatment Effect Estimation

After estimating the propensity scores, the matching process was carried out using a 1:1 Nearest Neighbor Matching algorithm without replacement. All 175 observations fall within the common support region, meaning that every districts/cities in the treatment group could be matched to a counterpart in the control group, with no observations dropped due to a lack of overlap. Table 6 presents the estimated results for the three main treatment effect measures.

Table 6. Average Treatment Effect Estimates of University Presence on GDP per Capita

Effect Measure	Estimate (log)	Std. Error	t / z-stat	Approx. Effect (%)
ATT (psmatch2)	0.6399	0.1612	3.97	+89.6%
ATU (psmatch2)	0.4493	-	-	+56.7%
ATE (psmatch2)	0.5593	-	-	+74.9%
ATE (teffects - logit)	0.5774	0.1403	4.11	+78.1%
ATE (teffects - probit)	0.5593	0.1435	3.90	+74.9%

Source: Stata output, processed by the author.

The primary focus of this study is the Average Treatment Effect on the Treated (ATT). The estimation result from the psmatch2 command yields an ATT of 0.6399 with a t-statistic of 3.97, which is significant at the 1% level. In economic terms, since the dependent variable is the natural logarithm of GDP per capita, this coefficient can be converted using the formula (e^0.6399 - 1) x 100% = 89.6%. Therefore, districts/cities with a university have, on average, a GDP per capita approximately 89.6% higher than comparable districts/cities without a university, after controlling for differences in socioeconomic characteristics through the matching process. This represents an economically substantial effect.

In addition to the ATT, the Average Treatment Effect on the Untreated (ATU) of 0.4493 is also worth noting. The ATU measures the expected gain in GDP per capita for regions that currently do not have a university, were they are to receive one. This estimate is smaller than the ATT (0.6399), which indicates the presence of treatment effect heterogeneity — regions that currently have a university appear to be more responsive to university presence than those that do not, likely because they already possess the economic ecosystem and infrastructure needed to fully absorb the benefits of a university.

The Average Treatment Effect (ATE) from psmatch2 of 0.5593 reflects the average effect of university presence for the entire population of districts/cities regardless of their actual treatment status and falls between the ATT and ATU values.

Covariate Balance Test

The quality of the matching process was evaluated using the pstest command in Stata, which reports the Standardized Percentage Bias (%bias), variance ratio (V(T)/V(C)), and balance test statistics both before (Unmatched/U) and after (Matched/M) matching. The full results are presented in Table 7 and 8.

Table 7. Covariate Balance Test Results Before Matching

Variable	Sample	Mean Treated	Mean Control	%Bias	t-stat	p>|t|
Log Total Population	U	6.783	6.039	77.1	5.08	0.000
	M	6.783	6.771	1.3	0.09	0.926
Poverty Severity Index	U	4.280	19.833	33.2	-2.32	0.021
	M	4.280	4.053	0.5	0.07	0.943
OUR (%)	U	5.901	4.189	93.4	6.10	0.000
	M	5.901	5.883	1.0	0.07	0.947

Source: Stata output, processed by the author.

Table 8. Covariate Balance Test Results After Matching

Sample	Ps R2	LR chi2	p > chi2	Mean Bias	Med Bias	B Statistic
Unmatched	0.186	44.32	0.000	67.9%	77.1%	108.4 (>25%)
Matched	0.000	0.02	0.999	0.9%	1.0%	2.0 (<25%)

Source: Stata output, processed by the author.

The balance test results show a substantial improvement following the matching process. Before matching, all three control variables exhibited severe imbalance, the %bias for OUR reached 93.4%, log total population 77.1%, and the poverty severity index 33.2%, all far exceeding the 10% threshold. The overall Mean Bias before matching was 67.9%, and the B-statistic of 108.4% confirmed severe imbalance (threshold: B < 25%). After matching, all balance indicators improved substantially. The %bias for all variables dropped to below 2% (OUR: 1.0%, population: 1.3%, poverty: 0.5%), well below the 10% threshold. Mean Bias fell from 67.9% to only 0.9%, a reduction of 98.7%. The post-matching Ps R2 approached zero (0.000) with a chi2 p-value of 0.999, indicating that the probit model can no longer systematically distinguish between the treatment and control groups based on the covariate variables after matching. The post-matching B-statistic of 2.0 is also well below the 25% threshold. These results demonstrate that the matching process successfully balanced the characteristics of the treatment and control groups, such that the resulting ATT estimate can be considered free from selection bias arising from observed variables.

Robustness Test

To ensure that the estimation results are not sensitive to the choice of method or model specification, a series of robustness checks were conducted by comparing results across different estimation approaches. Table 9 summarizes these comparisons.

Table 9. Comparison of Estimation Results Across Specifications (Robustness Test)

Specification / Method	Estimator	Coefficient	Std. Error	z / t-stat	p-value
psmatch2 + probit (Nearest	ATT	0.6399	0.1612	3.97	< 0.001
teffects psmatch + logit	ATE	0.5774	0.1403	4.11	< 0.001
teffects psmatch + probit	ATE	0.5593	0.1435	3.90	< 0.001
Pre-matching ATT difference	Raw diff	0.3414	0.1120	3.05	0.003

Source: Stata output, processed by the author.

The robustness test results show strong consistency across all specifications. The estimated coefficients range from 0.5593 to 0.6399, with only minor variation across methods, and all are significant at the 1% level. Comparing the logit and probit models in the teffects command yields nearly identical ATE estimates (0.5774 vs. 0.5593), indicating that the choice of distributional function does not alter the substantive conclusion. The comparison between the pre-matching raw difference (0.3414, t = 3.05) and the post-matching ATT (0.6399, t = 3.97) is also noteworthy. The PSM estimate being larger than the raw difference is referred to as an upward bias correction, meaning that once treatment regions are matched with truly comparable control regions, the actual effect of university presence turns out to be larger than what is apparent in a simple descriptive comparison. This occurs because, before matching, the control group (regions without universities) includes many regions with very low socioeconomic conditions, which pulls the control group mean downward and makes the raw difference appear smaller than the true causal effect.

Economic Interpretation and Theoretical Linkage

The overall findings of this study provide strong and consistent evidence that university presence has a significant positive effect on GDP per capita across districts/cities in Java and Kalimantan. These findings can be explained through several transmission mechanisms grounded in the theoretical framework established in this study. First, the human capital accumulation mechanism. Consistent with the Human Capital Theory of Becker (1964) and Schultz (1961), universities serve as engines for producing skilled and educated workers. University graduates who are absorbed into the local labor market raise aggregate labor productivity, which in turn drives regional output and per capita income. In districts/cities in Java and Kalimantan, where the manufacturing, financial services, and knowledge-based sectors are growing, the availability of educated labor is a critical factor in regional economic competitiveness. Second, the economic agglomeration mechanism. The presence of a university attracts firms, particularly in the technology, consulting, and creative industries, to locate nearby to access educated labor and knowledge spillovers. This agglomeration process, as described by Marshall (1890), creates a positive feedback loop between university presence, industry location, the creation of quality employment, and GDP per capita growth. Overall, these findings are consistent with the work of Abel and Deitz (2012) in the United States, who found a positive contribution of higher education institutions to regional economic growth, as well as with the study by Drucker and Goldstein (2007), which identified multiple transmission channels through which universities affect local economies. In the Indonesian context specifically, these findings underscore the importance of policies aimed at distributing higher education institutions more equitably, not merely as a means of strengthening education, but as a regional development strategy with measurable economic impact.

CONCLUSION

This study provides causal evidence that university presence positively and significantly affects regional GDP per capita in districts and cities across Java and Kalimantan in 2023. Using Propensity Score Matching (PSM), the estimated Average Treatment Effect on the Treated (ATT) of 0.6399 indicates that regions with universities have a GDP per capita approximately 89.6% higher than comparable regions without universities. Robustness checks across logit and probit specifications confirm the stability of this finding. These results suggest that universities drive regional economic growth through human capital development, economic agglomeration, and knowledge spillovers. For policymakers, these findings support prioritizing higher education investment as a strategy for reducing interregional economic disparities. For future research, expanding to panel data, broader geographic coverage, and richer covariate sets would strengthen causal identification and generalizability. A key limitation of this study is its reliance on cross-sectional data from a single year and a restricted geographical scope, which limits the generalizability of the findings to other regions and time periods.

BIBLIOGRAPHY

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38. https://doi.org/10.1080/00031305.1985.10479383

Vol. 02 Year 2026, 428 – 442
ISSN: 3109-3337
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press

Badan Pusat Statistik. (2022–2024). *Produk domestik regional bruto (PDRB) per kapita atas dasar harga berlaku menurut kabupaten/kota di Pulau Jawa dan Kalimantan.*

Badan Pusat Statistik. (2022–2023). *Jumlah penduduk menurut kabupaten/kota di Pulau Jawa dan Kalimantan.*

Badan Pusat Statistik. (2023). *Tingkat pengangguran terbuka (TPT) menurut kabupaten/kota di Pulau Jawa dan Kalimantan.*

Badan Pusat Statistik. (n.d.). *Indeks keparahan kemiskinan (P2) menurut kabupaten/kota di Pulau Jawa dan Kalimantan.*

Badan Pusat Statistik. (2023). *Produk domestik regional bruto kabupaten/kota di Pulau Jawa dan Kalimantan.*

Gertler, Paul J.., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Pendekatan evaluasi dampak menggunakan propensity score matching.* Jurnal Kebijakan Ekonomi, 10(2), 95–120.

Kementerian Keuangan Republik Indonesia. (2020). *Evaluasi kebijakan publik.* Jakarta: Kementerian Keuangan RI.

Kuncoro, Mudrajad. (2010). *Analisis pertumbuhan dan ketimpangan ekonomi regional di Indonesia.* Jurnal Ekonomi Pembangunan Indonesia, 11(1), 1– 20.

Resosudarmo, Budy P.., Vidyattama, Y., & Hartono, D. (2014). *Dampak ketergantungan sumber daya alam terhadap pembangunan ekonomi daerah di Indonesia.* Jurnal Ekonomi dan Pembangunan Indonesia, 15(2), 120–145.

Suryadarma, Daniel. (2012). *Dampak pendidikan terhadap pendapatan dan produktivitas tenaga kerja di Indonesia.* Jurnal Ekonomi dan Pembangunan Indonesia, 13(1), 1–18.

Todaro, Michael P., & Smith, S. C. (2015). *Economic development* (11th ed.). Boston: Pearson.

Todaro, M. P., & Smith, S. C. (2015). *Economic development* (12th ed.). Boston: Pearson.

Wicaksono, Teguh Yudo. (2018). *Peran pendidikan tinggi dalam pembangunan ekonomi daerah.* Yogyakarta: UGM Press.

Arsyad, Lincolin. (2010). *Ekonomi pembangunan.* Yogyakarta: UPP STIM YKPN.