EKO-REGIONAL: JURNAL PEMBANGUNAN EKONOMI WILAYAH
Volume 20, Issue 2, September 2025, pp. 202-213
http://jp.feb.unsoed.ac.id/index.php/eko-regional/article/view/2953
DOI: https://doi.org/10.32424/1.erjpe.2022.17.2.2953

Heterogeneity of Social Returns to Education Across Employees Education Levels in Indonesia

By:
Ribut Nurul Tri Wahyuni[1*)], Retnaningsih[1)], Lia Yuliani[1)]
[1]Politeknik Statistika STIS
[*)]Corresponding Author: rnurult@stis.ac.id

Submission: April 25, 2025; Accepted: October 29, 2025

ABSTRACT: Agglomeration externalities can have various effects on wages depending on the educational level of employees. Regrettably, previous studies in Indonesia that addressed this topic failed to account for the endogeneity of agglomeration externalities and employed methodologies that were incapable of mitigating sample selection bias in employment data. To resolve these issues, this study applies the Instrumental Variable (IV) Heckman model with the historical population density and the number of universities as IVs. Using the 2020-2024 microdata, the result indicates that social returns to education from the local high-skill labor market, which is a part of agglomeration externalities, are positively significant and higher for high- and medium-skilled employees than for low-skilled employees. This finding demonstrates the efficacy of public education investments, especially Indonesian government policy that can enhance the quantity and quality of university graduates, thereby promoting the transfer of knowledge to all employees and eventually reducing wage inequality between regions.

Keywords: Agglomeration Externalities, Instrumental Variable Heckman Model, Social Returns to Education, Wage

ABSTRAK*: Dampak eksternalitas aglomerasi terhadap upah dapat bervariasi, tergantung pada tingkat pendidikan pekerja. Sayangnya, penelitian sebelumnya di Indonesia yang membahas permasalahan tersebut masih mengabaikan endogenitas pada variabel eksternalitas aglomerasi dan menggunakan metode yang tidak mampu mengatasi sample selection bias pada data ketenagakerjaan. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan model Heckman Variabel Instrumental (IV) dengan kepadatan penduduk historis dan jumlah universitas sebagai IV. Dengan menggunakan data mikro tahun 2020-2024, hasilnya menunjukkan bahwa manfaat sosial pendidikan dari pasar tenaga kerja lokal berketerampilan tinggi, salah satu bagian dari eksternalitas aglomerasi, positif signifikan dan lebih tinggi di kelompok pekerja berpendidikan tinggi dan menengah, dibanding pekerja berpendidikan rendah. Temuan ini membuktikan keberhasilan investasi pendidikan yang dilakukan oleh pemerintah, terutama kebijakan yang dapat meningkatkan kuantitas dan kualitas lulusan perguruan tinggi, sehingga mendorong adanya transfer pengetahuan kepada semua pekerja dan pada akhirnya dapat mengurangi ketimpangan upah antar wilayah.*

Kata Kunci: *Eksternalitas Aglomerasi, Model Instrumental Variabel Heckman, Manfaat Sosial Pendidikan, Upah*

INTRODUCTION

Agglomerated areas often attract highly educated employees from nearby regions (Groot & de Groot, 2020). Consequently, these areas typically display wage levels that exceed the national average due to a more favorable composition of the labor force. A rising concentration of educated employees can also generate positive knowledge spillovers that enhance the productivity of both employees with similar educational backgrounds and those with lower levels of education. These benefits may occur through the exchange of skills and information, search-related externalities, or endogenous technological change that favors skilled labor (Moretti, 2004). This phenomenon is widely known as the social returns to education (Amponsah, 2020; Cui & Martins, 2021; Joshi et al., 2019). On the other hand, a substantial increase in the supply of educated employees may also decrease wages. When the number of high-skilled employees rises without a corresponding increase in demand, their wages may decline. Furthermore, high-skilled employees may substitute for employees with lower skill levels, which can place additional downward pressure on wages for the less-educated group (Groot & de Groot, 2020). Thus, the overall direction of social returns to education remains uncertain.

Furthermore, in the last two decades, Indonesia has enacted significant policy reforms to enhance education, including a constitutional requirement to allocate 20 percent of the national budget to education, decentralizing certain educational functions to district and school levels, and implementing Law on Teachers and Lecturers No. 14/2005 to elevate teacher quality (World Bank, 2020). The Indonesian government has implemented the School Operational Assistance (BOS) program since 2005 (Edwards & Storen, 2023). As long as the students attend institutions that receive BOS, the government will cover the cost of education for elementary, junior, and senior high school students under this program. To augment the quantity of college graduates, the government also implements financial support programs, including the Indonesia Smart College Card (KIP Kuliah) and Educational Fund Management Institution (LPDP) scholarships, as well as equal access programs, such as the Higher Education Affirmation Program (ADik). By 2025, expenditures on education surpassed those of any other sector (Ministry of Finance Republic of Indonesia, 2025).

The sign and size of the social returns on education significantly influence the efficacy of Indonesia's public investments in education. This investment has the potential to increase the average years of schooling for the Indonesian population, resulting in social and private returns to education. Ultimately, Indonesian workers may experience a change in their wages as a result of education policy. Despite the theoretical literature's claim that there is the existence of knowledge spillovers, the empirical evidence regarding the extent of these spillovers in Indonesia remains limited. Joshi et al. (2019) have conducted a study on the social returns of education in Indonesia. The study, however, exclusively examined the impact of the average education on worker wages. Social returns to education in Indonesia's local high-skill labor market remain unexplored. This study aims to address this gap in the literature by estimating the social returns to education in Indonesia, categorized by the educational cohort of workers.

This study offers two main contributions. First, it examines whether differences in regional wages can be explained by the concentration of educated employees in more agglomerated areas. Because educated and less-educated employees are not perfect substitutes, the magnitude of social returns to education varies across educational groups (Moretti, 2004). By comparing these variations, the analysis provides insight into how much knowledge spillover each group receives and reveals whether public spending on higher education generates positive or negative heterogeneous effects on wages in Indonesia. Second, the estimation of agglomeration externalities is susceptible to endogeneity and sample selection bias. To address potential biases arising from these challenges, the study applies an Instrumental Variable (IV)–Heckman approach. Historical population density is used as an instrument for present-day agglomeration levels (Combes & Gobillon, 2015), while the number of university graduates serves as an instrument for the number of educated employees (Groot et al., 2014; Moretti, 2011).The remainder of this study is organized as follows. The next section provides methods. Section 3 describes results and discussions, including robustness checks. Lastly, section 4 reports conclusions.

METHODS

This research draws on secondary data from the August round of the National Labor Force Survey (Sakernas) conducted by BPS–Statistics Indonesia. The dataset spans the years 2020 to 2024 and collects information on household members aged five and above. The survey uses different household samples each year, with approximately 300,000 households surveyed annually. Sakernas provides detailed individual-level information. For the purposes of this study, the unit of analysis is restricted to household members aged 15 and older who work as employees and have worked at least one hour during the reference week. Employees in the public sector are excluded because their wages are directly influenced by government regulations. The study further categorizes employees into three educational groups: low, medium, and high. Individuals with no diploma or whose highest completed education is primary or junior secondary school are grouped under low education. Those who have completed senior high school are classified as medium education, while college graduates comprise the high-education category.

In the literature, many earlier studies (Champagne et al., 2017; Greenspon et al., 2021; Lazear, 2019) used average wages or value added to measure regional productivity. Such approaches overlook heterogeneity among employees, which can inflate estimated effects. Conversely, micro-level studies typically use Mincer-type wage regressions, where wages serve as a proxy for productivity but social returns to education are not taken into account (Verstraten et al., 2019). Addressing these limitations, the present study incorporates the share of highly educated employees directly into the employee-level wage equation and estimates the social returns to education to better capture the positive spillover effects.

Individual wages are shown in Equation 1 by a mix of individual characteristics (Bahl & Sharma, 2020; Cunha, et al., 2023; Patrinos et al., 2021), the natural logarithm of the measurement of agglomeration externalities, and the natural logarithm of the proportion of highly educated employees. Due to the probable significant variability in the association across various employee characteristics, we do distinct regressions for each educational attainment level. The male–female wage disparity (Dunusinghe, 2023; Thapa & Izawa, 2024) and the impact of experience, as shown by age (Gashi & Adnett, 2022), may differ across employees with varying educational backgrounds. To account for sectoral variability, this study incorporates a sectoral dummy (primary, secondary, and tertiary sectors). Additionally, this study also incorporates a series of year dummies, province dummies, white-collar dummies, and formal training dummies. The white-collar dummy variable is assigned a value of 1 if the employee is classified as a professional, technician, or similar role; as leadership and administrative personnel; or as administrative and analogous individuals. The value of the formal training dummy is equal to 1 if the employee has participated in training and obtained credentials. The study formulates the regression equation as follows.

$$lnw_{isjt} = \beta_0 + \beta_1 age_{isjt} + \beta_2 age^2_{isjt} + \beta_3 D^{male}_{isjt} + \sum_{m=1}^{3} \beta_{4m} D^{educ}_{isjt} + \beta_5 D^{white\_collar}_{isjt} + \beta_6 D^{training}_{isjt} + \sum_{k=1}^{3} \beta_{7k} D^{sector}_{isjt} + \sum_{l=1}^{5} \beta_{8l} D^{year}_{isjt} + \sum_{p=1}^{34} \beta_{9p} D^{prov}_{isjt} + \beta_{10} ln\,(specialization_{sjt}) + \beta_{11} ln\,(share\ high\ educ_{jt}) + e_{isjt} \quad \text{....(1)}$$

lnw_{ijt} is the natural logarithm of the hourly wage of employee *i* in sector *s* district *j* year t, D^* represent dummies variables, $ln\,(agglomeration_{jt})$ is the natural logarithm of the measurement of agglomeration externalities, and $ln\,(share\ high\ educ_{jt})$ is the natural logarithm of the proportion of university graduates as a proxy for highly educated employees. The inclusion of the natural logarithms of hourly wages (Moreno & Patrinos, 2020), the measurement of agglomeration externalities, and the proportion of highly educated employees is justified because the resulting estimations can be viewed as elasticities. In this study, we use specialization economies (Marshall-Arrow-Romer spillover) as a proxy for agglomeration externalities (Isnaeni & Khoirunurrofik, 2021; Xiong et al., 2023) with the formula as follows.

$$specialization_{jt} = \frac{E_{sjt}}{E_{jt}} \qquad (2)$$

E_{sjt} is the number of employees in sector *s* district *j* year t, while E_{jt} is the number of employees in district *j* year t. An increased agglomeration value correlates with a greater proportion of employees in sector *s* within district *j* during year t. Estimating the impacts of aggregated variables on individual-level observations can result in underestimated standard errors (Combes et al., 2008). The most appropriate way to address this issue is to apply robust standard errors clustered at the district level (Groot & de Groot, 2020).

RESULTS AND DISCUSSIONS

Descriptive Statistics

Table 1 provides descriptive statistics regarding the data in this study. A total of 633,114 samples are collected, comprising 218,358 samples from low-educated employees, 245,396 samples from medium-educated employees, and 169,360 samples from high-educated employees, representing 17.24%, 19.38%, and 13.38% of the total employee sample, respectively. A greater level of education correlates with significantly elevated hourly wages. Employees with higher education are more inclined to occupy white-collar positions and possess formal training certifications. Employees with medium education levels are, on average, younger than those with low education levels. Moreover, employees with higher education tend to be older than those with medium education. This does not imply that older cohorts possess higher education, as this study is restricted to workers with employee status. Younger cohorts with higher education might choose other job types, such as entrepreneurs, rather than employees.

Individuals possessing high education are more inclined to reside in regions with a significant concentration of similarly educated individuals. Moreover, sectors characterized by a substantial labor share tend to employ individuals with lower educational attainment. This indicates that individuals with low education will be engaged in sectors necessitating substantial labor in a specific region. Owing to their restricted skills, integrating them into sectors with a low workforce proportion is challenging. Conversely, a sector may engage highly educated individuals, even though its workforce share is low. Their high skills allow them to operate in any sector.

Table 1. Descriptive Statistics by Employee Education Group

Descriptive statistics	All	Low	Medium	High
Observations	633,114	218,358	245,396	169,360
Hourly wage (mean, Rp)	17,695.45	12,215.59	.5,423.99	28,031.49
Male (share, %)	62.57	72.82	69.15	39.88
Age (years)	37.14	40.67	33.27	38.18
White collar (share, %)	35.24	4.64	25.11	89.26
Formal training (share, %)	21.84	3.14	19.60	49.16
Share highly educated (mean, %)	20.57	18.50	20.72	23.03
Specialization (share, %)	26.78	31.00	27.26	20.64

Sources: Author's calculations

Benchmark Regression Results

We first perform Ordinary Least Square (OLS) estimation, and Table 2 presents the results. Column 1 in this table presents the regression results encompassing all samples. At a significance level of 1%, the return on education for employees with high education is 62.8% greater than that for employees with low education, but the return on education for employees with medium education is 26.5% more than that for employees with low education. Thus, enhancing employees' educational attainment can elevate their wages. The grouping regression findings indicate the presence of gender discrimination in pay. Male employees earn higher wages than female employees. This finding aligned with Cameron (2023), who asserted that Indonesia exhibits gender inequality, with a value surpassing that of several surrounding countries. Furthermore, the impact of age exhibits diminishing marginal returns (Kimura et al., 2022). A rise in age may result in higher employee wages; however, this increment will decline as the employee matures further. At the 1% significance level, the impact of agglomeration is statistically significant and beneficial. For instance, a 1% increase in the share of the labor market in the sector where low-educated employees work can elevate their wages by approximately 0.115%. Social returns to education are also positively significant in medium and high groups.

Table 2. OLS Estimates by Employee Education Group

Dep.: ln hourly wage	All	Low	Medium	High
Male dummy	0.260***	0.316***	0.235***	0.168***
	(0.0063)	(0.0090)	(0.0079)	(0.0055)
Age	0.044***	0.039***	0.043***	0.067***
	(0.0009)	(0.0008)	(0.0012)	(0.0028)
Age^2	-0.000***	-0.000***	-0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
White collar dummy	0.239***	0.157***	0.233***	0.218***
	(0.0057)	(0.0092)	(0.0061)	(0.0093)
Formal training dummy	0.140***	0.124***	0.112***	0.167***
	(0.0044)	(0.0093)	(0.0048)	(0.0058)
Medium educ dummy	0.265***			
	(0.0079)			
High educ dummy	0.628***			
	(0.0102)			
Ln share highly educated	0.090***	0.011	0.106***	0.200***
	(0.0253)	(0.0251)	(0.0274)	(0.0280)
Ln specialization	0.115***	0.124***	0.144***	0.041***
	(0.0114)	(0.0088)	(0.0156)	(0.0136)
Sectoral dummy	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes
Provincial dummy	Yes	Yes	Yes	Yes
R-square	0.2949	0.1903	0.2644	0.2981
Observations	633,114	218,358	245,396	169,360

Notes: Standard errors are clustered at the district level in parentheses. Significant coefficients at the 1%, 5%, and 10% significance levels are denoted by ***, **, and *, respectively.

Source: Author's calculations

Endogeneity Problems and Sample Selection Bias

A prevalent issue in literature is the endogeneity of agglomeration externalities. Agglomeration enhances employees' wages, and elevated wages subsequently attract additional employees, leading to increased agglomeration. According to Groot & de Groot, (2020); Moretti (2011), the assessment of social returns to education also entails endogeneity concerns. The reasons for these conclusions are similar: wages in a region may be high not only because there are a lot of highly educated workers (at a certain level of agglomeration), but also because the educational makeup of the local labor market may just reflect differences in how productive different types of work are across the region. In examining the social returns to education, we aim to ascertain the causal relationship between social education and wages. Unobserved individual characteristics, such as ability, and unobserved regional characteristics, including geographical location, industrial structure, climate, and amenities, may correlate with both wages and the proportion of educated employees in specific regions. It is conceivable that individuals with high unobserved ability gravitate towards more agglomerated regions and regions with a well-educated labor force. Neglecting unobserved individual characteristics and the reverse causation of agglomeration and wage can result in OLS estimates that are both biased and inconsistent (Cholezas & Kanellopoulos, 2024; Purnastuti et al., 2015). Ridhwan (2021) commonly uses the instrumental variable-two-stage least squares (IV-2SLS) method, and Table 3 presents the findings.

Table 3. IV-2SLS Estimates by Employee Education Group

Dep.: ln hourly wage	All	Low	Medium	High
Male dummy	0.260***	0.321***	0.234***	0.165***
	(0.0069)	(0.0124)	(0.0082)	(0.0054)
Age	0.043***	0.038***	0.042***	0.071***
	(0.0012)	(0.0013)	(0.0014)	(0.0031)
Age^2	-0.000***	-0.000***	-0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
White collar dummy	0.258***	0.163***	0.246***	0.266***
	(0.0124)	(0.0099)	(0.0116)	(0.0213)
Formal training dummy	0.140***	0.119***	0.114***	0.163***
	(0.0045)	(0.0095)	(0.0053)	(0.0062)
Medium educ dummy	0.247***			
	(0.0109)			
High educ dummy	0.620***			
	(0.0116)			
Ln share highly educated	0.256***	0.158***	0.254***	0.470***
	(0.0652)	(0.0589)	(0.0781)	(0.0777)
Ln specialization	0.355**	0.239	0.373**	0.398**
	(0.1646)	(0.1474)	(0.1809)	(0.1854)
Sectoral dummy	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes
Provincial dummy	Yes	Yes	Yes	Yes
Centered R-square	0.2706	0.1794	0.2396	0.2549
Underid.test: p-value	0.0094	0.0077	0.0124	0.0203
Weak id.test: Cragg-Donald Wald F stat.	3,710.609	885.080	1,954.395	1,026.485
Observations	633,114	218,358	245,396	169,360

Notes: Standard errors are clustered at the district level in parentheses. Significant coefficients at the 1%, 5%, and 10% significance levels are denoted by ***, **, and *, respectively.
Source: Author's calculations

This study adopts population density from 1961 as an instrumental variable for measuring agglomeration in the period 2020–2024, following approaches used by Graham et al. (2010) and Groot & de Groot (2020). Historical population density is unaffected by contemporary economic output, yet it remains strongly correlated with today's local economic concentration, making it a suitable instrument. Similarly, to instrument the share of highly educated employees, we apply a strategy aligned with Moretti (2011). Although graduates can freely choose where to work after completing their studies, the number of universities serves as a valid instrument because it influences the supply of highly educated employees within a district while not having a direct effect on wages. Accordingly, this study uses the number of universities at the district level in 2018 as the instrument. Data on historical population density and university counts are sourced from the 1961 Population Census and the 2018 Village Potential (Podes) Survey conducted by BPS–Statistics Indonesia.

As shown in Columns 1-4 of Table 3, the F statistics for the first stage regression for all equations are higher than 10. This shows that the number of local universities and the historical population density are strong instrumental variables for social education and agglomeration variables. In comparison to the OLS estimation in Table 1, the use of the instrumental variables does not alter the signs of returns to social education, which remains statistically significant. Nonetheless, the benefits of social education have escalated. Consequently, the endogeneity issue underestimates the returns to social education in OLS estimation.

Table 4. Heckman Estimates by Employee Education Group

Dep.: ln hourly wage	All	Low	Medium	High
Male dummy	0.238***	0.263***	0.189***	0.152***
	(0.0064)	(0.0103)	(0.0084)	(0.0056)
Age	0.043***	0.036***	0.040***	0.059***
	(0.0009)	(0.0009)	(0.0012)	(0.0028)
Age^2	-0.000***	-0.000***	-0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
White collar dummy	0.239***	0.158***	0.237***	0.215***
	(0.0057)	(0.0092)	(0.0061)	(0.0093)
Formal training dummy	0.140***	0.122***	0.112***	0.164***
	(0.0044)	(0.0094)	(0.0048)	(0.0058)
Medium educ dummy	0.253***			
	(0.0078)			
High educ dummy	0.603***			
	(0.0104)			
Ln share highly educated	0.085***	0.003	0.086***	0.185***
	(0.0252)	(0.0250)	(0.0273)	(0.0277)
Ln specialization	0.114***	0.122***	0.136***	0.033**
	(0.0113)	(0.0088)	(0.0155)	(0.0132)
IMR	-0.075***	-0.088***	-0.249***	-1.417***
	(0.0083)	(0.0086)	(0.0158)	(0.0699)
Sectoral dummy	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes
Provincial dummy	Yes	Yes	Yes	Yes
R-square	0.2949	0.1916	0.2680	0.3018
Observations	633,114	218,358	245,396	169,360

Notes: Standard errors are clustered at the district level in parentheses. Significant coefficients at the 1%, 5%, and 10% significance levels are denoted by ***, **, and *, respectively.

Source: Author's calculations

Besides the endogeneity problem, the sample selection in this study is also non-random because it only comprises workers with employee status. Individuals without income, freelancers, entrepreneurs, and workers who fail to report their wages are excluded. It may result in biased sample selection. Heckman's two-step methodology is the principal technique for resolving this issue (Dumauli, 2015; Nguyen et al., 2025). The process is divided into two phases. Initially, the probit model is employed to estimate the subsequent job involvement equation.

$$P(employee)_{it} = \alpha_0 + \boldsymbol{\alpha_1 Z_{it}} + \varepsilon_{it} \quad (3)$$

$P(employee)$ indicates the worker's status. If the worker is an employee, assign a value of 1; otherwise, assign a value of 0. ***Z*** comprises a set of characteristics affecting an individual's decision to work as an employee, including gender, age, marital status, residency status, presence of toddlers, and household member status (whether the individual is the household head or not). The estimation result of Equation 3 facilitates the computation of the inverse Mills ratio. Subsequently, we incorporate the derived inverse Mills ratio into Equation 2, utilize it as an explanatory variable, and proceed to estimate it via OLS.

Table 4 presents the estimation outcomes of the Heckman model. The coefficients of the inverse Mills ratio are negative and statistically significant at the 1% level, indicating the presence of selection bias in the samples. The Heckman model reduces returns to social education compared to the OLS estimation results. Due to sample selection bias, the OLS estimator overestimates returns to social education. Nonetheless, the disparity in the rate of returns to social education across the three groups diminished following the adjustment for sample selection bias.

Table 5. IV-Heckman Estimates by Employee Education Group

Dep.: ln hourly wage	All	Low	Medium	High
Male dummy	0.251***	0.280***	0.204***	0.154***
	(0.0095)	(0.0183)	(0.0130)	(0.0055)
Age	0.042***	0.036***	0.040***	0.066***
	(0.0011)	(0.0011)	(0.0012)	(0.0036)
Age^2	-0.000***	-0.000***	-0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
White collar dummy	0.258***	0.164***	0.249***	0.263***
	(0.0124)	(0.0099)	(0.0110)	(0.0218)
Formal training dummy	0.140***	0.117***	0.114***	0.162***
	(0.0045)	(0.0095)	(0.0053)	(0.0061)
Medium educ dummy	0.242***			
	(0.0092)			
High educ dummy	0.608***			
	(0.0130)			
Ln share highly educated	0.252***	0.146**	0.234***	0.451***
	(0.0666)	(0.0606)	(0.0829)	(0.0795)
Ln specialization	0.354**	0.234	0.369**	0.397**
	(0.1650)	(0.1488)	(0.1825)	(0.1864)
IMR	-0.033*	-0.067***	-0.161***	-0.942***
	(0.0188)	(0.0138)	(0.0539)	(0.1779)
Sectoral dummy	Yes	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes	Yes
Provincial dummy	Yes	Yes	Yes	Yes
Centered R-square	0.2710	0.1814	0.2429	0.2581
Underid.test: p-value	0.0091	0.0074	0.0116	0.0197
Weak id.test: Cragg-Donald Wald F stat.	3,709.485	879.822	1,947.313	1,029.509
Observations	633,114	218,358	245,396	169,360

Notes: Standard errors are clustered at the district level in parentheses. Significant coefficients at the 1%, 5%, and 10% significance levels are denoted by ***, **, and *, respectively.
Source: Author's calculations

To simultaneously address endogeneity issues and sample selection bias, we use the IV-Heckman model for estimates. Table 5 presents the findings. In the initial stage, the F statistics exceed 10, and the coefficients of the inverse Mills ratio are likewise substantial. The OLS estimation results are biased downward due to endogeneity and sample selection bias. The empirical results in this table indicate a significant disparity in returns to social education between low-educated and high-educated employees. A cohort of highly educated employees yields the greatest returns on social education. This group derives the greatest advantage from the positive spillover effect associated with a rising share of highly educated employees in districts where they work. Nonetheless, the positive spillover effect is also experienced by low- and medium-educated employees.

A variety of elements drive wage disparities across regions. According to Combes et al. (2008), regional wage differences stem primarily from three sources: the structure of the regional labor force, the presence of natural amenities, and the influence of agglomeration economies. The concentration of firms, the thickness of factor markets, and the spread of knowledge all play a role in strengthening the positive link between agglomeration and regional productivity. Larger and thicker labor markets allow for greater specialization. Because the tasks performed by highly educated employees tend to be more complex and specialized than those of less-educated employees, the productivity gains associated with agglomeration are likely to be stronger for higher-educated employees.

The existence of highly skilled employees is frequently regarded as a regional benefit, as evidenced by the work of Groot & de Groot (2020); Moretti (2011). Table 5 indicates that highly educated employees exert significant and higher positive influence on peers with the same educational backgrounds than on those with lesser education. This transpires due to their comparable knowledge bases and cognitive frameworks, which facilitate communication and knowledge transfer. Moreover, advanced abilities are typically synergistic, enabling highly educated colleagues to comprehend and promptly implement new ideas or innovations generated by other employees. The employment framework of highly educated individuals is often collaborative, facilitating greater connection and idea sharing. Conversely, positions requiring lower educational qualifications tend to be more repetitive and solitary, hindering the assimilation of new insights or expertise from highly educated personnel. Moreover, highly educated individuals generally inhabit similar social and professional networks, resulting in rapid dissemination of influence among themselves, while such impact is less pervasive among lower-educated employees.

Table 6. Heterogeneity Effect in IV-Heckman Estimates, by Employee Education Group and Generation

Dep.: ln hourly wage	All	Low	Medium	High
Generation Z				
Ln share highly educated	0.370***	0.080	0.276***	0.721***
	(0.0696)	(0.0578)	(0.0720)	(0.0995)
Ln specialization	0.395**	0.133	0.404**	0.435*
	(0.1658)	(0.1387)	(0.1728)	(0.2269)
Underid.test: p-value	0.0089	0.0109	0.0108	0.0241
Weak id.test: Cragg-Donald Wald F stat.	1,895.750	318.172	1,203.139	411.849
Observations	284,712	69,168	143,174	72,370
Millenial generation				
Ln share highly educated	0.211***	0.093	0.192*	0.418***
	(0.0709)	(0.0619)	(0.1096)	(0.0782)
Ln specialization	0.291	0.202	0.277	0.338*
	(0.1787)	(0.1718)	(0.2113)	(0.1928)
Underid.test: p-value	0.0109	0.0101	0.0137	0.0206
Weak id.test: Cragg-Donald Wald F stat.	1,030.756	294.837	471.517	321.267
Observations	201,626	77,756	65,042	58,828
Generation X and baby boomer				
Ln share highly educated	0.180***	0.256***	0.241**	0.149*
	(0.0661)	(0.0640)	(0.1042)	(0.0783)
Ln specialization	0.315*	0.339**	0.334	0.302*
	(0.1647)	(0.1654)	(0.2084)	(0.1744)
Underid.test: p-value	0.0086	0.0070	0.0115	0.0151
Weak id.test: Cragg-Donald Wald F stat.	977.805	325.172	369.618	345.243
Observations	178,507	83,395	47,459	47,653

Notes: Standard errors are clustered at the district level in parentheses. Significant coefficients at the 1%, 5%, and 10% significance levels are denoted by ***, **, and *, respectively.
Source: Author’s calculations

Robustness checks

We utilize the IV-Heckman model to report the heterogeneity effects of the local supply of high-skilled laborers in this section. The knowledge-sharing and receiving behavior of elder employees and younger employees are influenced by the level of generativity (desire to contribute to the next generation) and development striving (desire to develop), as demonstrated by Fasbender & Gerpott (2022). Young employees who possess a high level of development motivation are more likely to acquire knowledge from their elder colleagues. This argument is the foundation of our interest in examining the heterogeneous effects of local high-skill labor markets in relation to education and generation of employees.

The definitive criterion for defining the various user generations could be the date of birth. Nevertheless, there is a debate in the literature about the age range of each user generation. According to Vassilakaki (2016), generations are classified as follows: baby boomers (born between 1946 and 1965), Generation X (born between 1965 and 1979), Millennials (born between 1979 and 1990), and Generation Z (born from 1990 to 2010). The findings indicate that the highest social return to education is in Generation Z. In contrast, the lowest social return to education is in Generation X and baby boomers. The results of this study further support the idea that the younger generation is more receptive to the spillover knowledge from the highly educated workforce in their vicinity. We attribute this to their ease of adaptation to technological advancements, the rapidity of communication, and the fact that many of them are novice workers. The implication is that their social returns to education are greater than those of older generations.

CONCLUSIONS

This study aims to measure the social returns to education for highly skilled employees in Indonesia, both within their own educational group and across other groups. In addition, the analysis includes another dimension of agglomeration externalities—specialization economies—as a control variable to better evaluate the social returns to education. The findings indicate that the influence of agglomeration differs markedly depending on an employee’s education level. All education groups appear to benefit from higher economic density in local labor markets. Once observable and unobservable differences among employees are taken into account, the results show that university-educated employees receive substantially higher wages when employed in densely concentrated high-skill labor markets. Importantly, the study also reveals that the presence of universities—representing a form of public investment in education—positively affects regional wages. This advantage emerges as universities help cultivate agglomeration economies and reinforce the local stock of knowledge, which in turn boosts productivity at the regional level.

For achieving optimal social returns from education, the government must simultaneously increase the number of universities, the number of academics, the affordability of scholarships for high-achieving and underprivileged students, and the number of other supporting facilities. The objective of this policy is to enhance the availability of higher education, which will subsequently result in an increase in worker wages in Indonesia. In addition, the government also needs to consider aligning the curriculum with labor market needs and strengthening vocational pathways to minimize wage disparities between low- and high-educated employees.

While this study offers useful insights, it is constrained by its dependence on secondary data. This study used cross-sectional data, so it cannot handle unobserved individual characteristics. Subsequent research may utilize the Indonesian Family Life Survey (IFLS) to tackle this matter. Despite this survey being last performed in 2014, the resultant data constituted panel data. This study does not utilize long historical IVs, such as the population density in the 18th century, due to limitations in the available population census data. Additionally, other measures of agglomeration externalities, such as urbanization economies, can be included in the wage equation to assess the social returns on education.

ACKNOWLEDGEMENT

We are sincerely appreciative of the generous support provided by Politeknik Statistika STIS, which facilitated the completion of this study. The study's substantial success is attributable to the financial support we received for our undertaking.

REFERENCES

Amponsah, E. A. (2020). The Social Returns to Education: A Review. *The International Journal of Humanities & Social Studies*, *8*(3). https://doi.org/10.24940/theijhss/2020/v8/i3/HS2003-015

Bahl, S., & Sharma, A. (2021). Education–Occupation Mismatch and Dispersion in Returns to Education: Evidence from India. *Soc Indic Res*, *153*, 251–298. https://doi.org/10.1007/s11205-020-02483-9

Cameron, L. (2023). Gender Equality and Development: Indonesia in A Global Context. *Bulletin of Indonesian Economic Studies, 59*(2), 179-207. https://doi.org/10.1080/00074918.2023.2229476

Champagne, J., Kurmann, A., & Stewart, J. (2017). Reconciling the Divergence in Aggregate U.S. Wage Series. *Labour Economics, 49*, 27-41. https://doi.org/10.1016/j.labeco.2017.08.002

Cholezas, I., & Kanellopoulos, N. C. (2024). Returns to Education in Greece: Adjusting to Large Wage Cuts. *Education Economics, 32*(5), 599-612. https://doi.org/10.1080/09645292.2024.2353868

Combes, P.P., Duranton, G., & Gobillon, L. (2008). Spatial Wage Disparities: Sorting Matters! *Journal of Urban Economics*, *63*(2), 723–742. https://doi.org/10.1016/j.jue.2007.04.004

Combes, P.P., & Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics, 5*, 247–348. Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00005-2

Cui, Y., & Martins, P. S. (2021). What Drives Social Returns to Education? A Meta-Analysis. *World Development*, *148*. https://doi.org/10.1016/j.worlddev.2021.105651

Cunha, D. R., Saulo, H., Monsueto, S. E., & Divino, J. A. (2023). A Bivariate Approach to the Mincerian Earnings Equation. *Revista Brasileira de Economia, 77*. https://doi.org/10.5935/0034-7140.20230008

Dumauli, M. T. (2015). Estimate of the Private Return on Education in Indonesia: Evidence from Sibling Data. *International Journal of Educational Development*, *42*, 14–24. https://doi.org/10.1016/j.ijedudev.2015.02.012

Dunusinghe, P. (2023). Returns to Education: The Case of Nepal. *Journal of Economics and Development Studies,* *9*(1), 1-13. https://jeds.thebrpi.org/journals/jeds/Vol_9_No_1_March_2021/1.pdf

Edwards, B., & Storen, I. (2023). The World Bank and Education Governance in Indonesia: Influence Around and Beyond School-Based Management. In *Rethinking World Bank influence: Governance reforms and the ritual aid dance in Indonesia*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429054945-8/world-bank-education-governance-indonesia-brent-edwards-inga-storen

Fasbender, U., & Gerpott, F. H. (2022). Why Do or Don't Older Employees Seek Knowledge from Younger Colleagues? A Relation–Opportunity Model to Explain How Age-Inclusive Human Resources Practices Foster Older Employees' Knowledge Seeking from Younger Colleagues. *Applied Psychology*, *71*(4), 1385–1406. https://doi.org/10.1111/apps.12362

Gashi, A., & Adnett, N. J. (2022). Estimating the Returns to Education in A Chronically Depressed Labour Market: the Case of Kosovo. *International Journal of Development Issues*, 21(3), 321–335. https://doi.org/10.1108/IJDI-12-2021-0254

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., & Noland, R. B. (2010). Testing for Causality Between Productivity and Agglomeration Economies. *Journal of Regional Science*, *50*(5), 935–951. https://doi.org/10.1111/j.1467-9787.2010.00676.x

Greenspon, J., Stansbury, A. M., & Summers, L. H. (2021). Productivity and Pay in the US and Canada. *Working Paper Series (National Bureau of Economic Research), 29548*. http://dx.doi.org/10.3386/w29548

Groot, S. P. T., & de Groot, H. L. F. (2020). Estimating the Skill Bias in Agglomeration Externalities and Social Returns to Education: Evidence from Dutch Matched Worker-Firm Micro-Data. *De Economist*, *168*(1), 53–78. https://doi.org/10.1007/s10645-019-09354-w

Groot, S. P. T., de Groot, H. L. F., & Smit, M. J. (2014). Regional Wage Differences in the Netherlands: Micro Evidence on Agglomeration Externalities. *Journal of Regional Science*, *54*(3), 503–523. https://doi.org/10.1111/jors.12070

Isnaeni, N. F., & Khoirunurrofik, K. (2021). The Effect of Knowledge Spillovers and Human Capital Through Technological Intensity on Employment Growth in Indonesia. *Asia-Pacific Journal of Regional Science*, *5*(1), 21–39. https://doi.org/10.1007/s41685-020-00174-4

Joshi, R., Subramanian, C., & Swaminathan, S. (2019). Are There Social Returns to Education in Developing Countries? Evidence from Indonesia. *Economic Development and Cultural Change*, *67*(2), 315–332. https://doi.org/10.1086/698165

Kimura, T., Kurachi, Y., & Sugo, T. (2022). Decreasing Wage Returns to Human Capital: Analysis of Wage and Job Experience Using Micro Data of Workers. *Journal of the Japanese and International Economies, 66*. https://doi.org/10.1016/j.jjie.2022.101217

Lazear, E. P. (2019). Productivity and Wages: Common Factors and Idiosyncrasies Across Countries and Industries. *Working Paper Series (National Bureau of Economic Research), 26428*. http://dx.doi.org/10.3386/w26428

Ministry of Finance Republic of Indonesia. (2025). Buku II Nota Keuangan Beserta Anggaran Pendapatan dan Belanja Negara Tahun Anggaran 2025. Ministry of Finance Republic of Indonesia

Moreno,V. A., & Patrinos, H. A. (2020). Returns to Education in Azerbaijan : Some New Estimates. *Policy Research Working Paper Series (World Bank), 9117*. https://openknowledge.worldbank.org/server/api/core/bitstreams/3ba894b9-e5c2-5055-a3b1-8e85642fc4b8/content

Moretti, E. (2004). Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data. *Journal of Econometrics*, *121*, 175–212. https://doi.org/10.1016/j.jeconom.2003.10.015

Moretti, E. (2011). Local Labor Markets. In *Handbook of Labor Economics, 4b*. Elsevier. https://doi.org/10.1016/S0169-7218(11)02412-9

Nguyen, C. V., Chacaltana, J., Estupinan, X., & Huynh, P. (2025). Manuscript: Heterogenous Returns to Education in Labor Markets: New Evidence from Vietnam. *International Journal of Educational Development, 118*. https://doi.org/10.1016/j.ijedudev.2025.103407

Patrinos, H. A., Psacharopoulos, G., & Tansel, A. (2021). Private and Social Returns to Investment in Education: the Case of Turkey with Alternative Methods. *Applied Economics, 53*(14), 1638-1658. https://doi.org/10.1080/00036846.2020.1841086

Purnastuti, L., Salim, R., & Joarder, M. A. M. (2015). The Return to Education in Indonesia: Post Reform Estimates. *The Journal of Developing Areas, 49*(3), 183-204. https://www.jstor.org/stable/24737315

Ridhwan, M. M. (2021). Spatial Wage Differentials and Agglomeration Externalities: Evidence from Indonesian Microdata. *Economic Analysis and Policy, 71*, 573-591. https://doi.org/10.1016/j.eap.2021.06.013

Thapa, A., & Izawa, M. (2024). Returns to Education in Nepal An Analysis of Educational Attainment, Employability and Social Mobility. *Education Economics, 32*(5), 649-664. https://doi.org/10.1080/09645292.2024.2351882

Vassilakaki, E. (2016). Knowing your Users, Discovering your Library: An Overview of the Characteristics of User Generations. In *Digital Information Strategies: from Applications and Content to Libraries and People*, 215–224. Chandos Publishing. https://doi.org/10.1016/B978-0-08-100251-3.00015-9

Verstraten, P., Verweij, G., & Zwaneveld, P. J. (2019). Complexities in the Spatial Scope of Agglomeration Economies. *Journal of Regional Science*, *59*(1), 29–55. https://doi.org/10.1111/jors.12391

World Bank. (2020). Revealing How Indonesia's Subnational Governments Spend Their Money on Education: Subnational Education Public Expenditure Review 2020. World Bank. https://documents1.worldbank.org/curated/en/487071605565796167/pdf/Revealing-How-Indonesia-s-Subnational-Governments-Spend-Their-Money-on-Education.pdf

Xiong, N., Wei, Y. D., & Wu, Y. (2023). Tech Firm Births and Agglomeration Economies: (Un)related Variety, Specialization, and Spatial Externalities. *Cities, 138*. https://remote-lib.ui.ac.id:2075/10.1016/j.cities.2023.104349