Journal of Economics, Entrepreneurship, Management Business and Accounting, 2026, 4(5),487-504.
https://doi.org/10.61255/jeemba.v4i5.1298

Determinants of Workplace Internet Use among Indonesian Workers: Probit and Logit Analysis Using SAKERNAS 2023

Rendra Dwi Saputra[1*], Kukuh Arisetyawan[1], Irien Kamaratih[2], Axellina Muara Setyanti[3], Muhammad Andika Zalfiandra[1], Muhammad Syahril Mustofa[1], Muhammad Rafi Putra Ibrahim[1], Aldan Ardana Ahmad[1]

[1]Department of Economics, Faculty of Economics and Business, Universitas Negeri Surabaya, Surabaya, 60231, Indonesia
[2]BPS-Statistics East Java, Surabaya, 60292, Indonesia
[3]Graduate School, Universitas Brawijaya, Malang, 65145, Indonesia

ABSTRACT

Purpose – This study analyzes the determinants of workplace internet use among Indonesian workers, treated here as one measurable dimension of digital technology adoption, using SAKERNAS August 2023 (N =465,204).

Design/methodology/approach – Binary probit and logit models, weighted using BPS individual sampling weights with the full linearized variance estimator (strata and PSU), and Average Marginal Effects (AME) for twelve explanatory variables.

Finding/Results – Education ≥ senior high school again has the largest effect once BPS sampling weights are applied (AME = 0.222; $p < 0.01$), followed by urban location (AME = 0.122), Kartu Prakerja awareness (AME = 0.137), and employee/wage-worker status (AME = 0.056). Disability reduces adoption probability (AME = -0.070; OR = 0.663). Internet use is associated with 53.6 percent higher earnings (exact semi-log transformation of $\beta = 0.429$), rising to 61.7 percent for internet users who have also attained senior high school education or above; both are conditional associations rather than a causal wage return. The weighted model achieves acceptable in-sample fit (AUC = 0.82; Pseudo-R^2 = 0.24).

Originality/Value – To our knowledge, one of the first nationally representative micro-econometric studies of workplace internet adoption in Indonesia; identifies disability as a new determinant with double-penalty implications for digital labor market inclusion.

ARTICLE INFO

Keywords:
Digital Technology Adoption,
Internet Use,
Probit,
SAKERNAS 2023,
Mincerian Earnings Function

Article Information:
Received: 30/06/2026
Revise: 28/07/2026
Accepted: 07/08/2026

ISSN:
2985-3168 (Online)
2985-3222 (Print)

*Corresponding Author at:
Department of Economics, Faculty of Economics and Business, Universitas Negeri Surabaya, 60231, Indonesia.
E-mail address: kukuharisetyawan@unesa.ac.id (Kukuh Ari Setyawan)

1. Introduction

The diffusion of digital technology in the workplace has become one of the most significant structural transformations in the twenty-first century economy. As Southeast Asia's largest economy with more than 140 million workers Indonesia faces the speed and equity of this digital transition as a central development challenge. The Making Indonesia 4.0 Roadmap, the National Digital Economy Strategy 2030, and the Kartu Prakerja program reflect the official recognition that digital skills upgrading and technology adoption at the workplace level are prerequisites for sustainable growth (World Bank, 2021).

Nevertheless, the micro-econometric determinants of whether a worker actually uses the internet in their work remain poorly understood in Indonesia. Most existing research operates at the macro or sectoral level, or focuses narrowly on e-commerce adoption by small and medium-sized enterprises (Rahayu & Day, 2015). The August 2023 SAKERNAS, for the first time, directly and comprehensively measures workplace digital technology use across all sectors and regions of Indonesia including disability variables and Kartu Prakerja program participation-thereby providing an unprecedented micro-analytical foundation.

A scope clarification matters here. Digital technology adoption, taken broadly, covers a wide range of practices computer use, enterprise software, mobile applications, digital platforms, automation tools, and, increasingly, artificial intelligence systems among them. SAKERNAS 2023 does not measure most of these; its digital module records only whether a worker uses the internet for work. This paper therefore treats workplace internet use as the operational measure of digital technology adoption, not as a stand-in for the construct in its entirety, following the same logic Akerman et al., (2015) apply when they study broadband adoption rather than digital adoption writ large. Internet access is a reasonable entry point because it underlies most other digital work practices that a survey like SAKERNAS can capture, from cloud-based enterprise software to online platforms. Even so, a worker who uses a computer or offline software without a reported work-related internet connection would not register as a digital adopter under this measure, and the results below should be read with that boundary in mind.

The theoretical basis of this study derives from the task model of Autor et al., (2003), hereafter referred to as ALM. ALM empirically demonstrates that computer technology is substitutive for routine cognitive and manual tasks that can be specified by explicit rules, yet complementary to non-routine analytical and interactive tasks. The key insight is that better-educated workers who hold a comparative advantage in non-routine tasks are more likely to adopt and benefit from digital technology. Akerman et al., (2015) extend this argument to broadband internet adoption in Norway, providing causal evidence that the internet significantly increases the productivity and wages of skilled workers.

This study makes three contributions. First, it provides, to our knowledge, one of the first nationally representative micro-econometric analyses of the determinants of workplace internet use in Indonesia using SAKERNAS 2023; we have not identified a prior study using this specific combination of a nationally representative micro-dataset and this outcome measure, though we have not conducted a systematic review to confirm priority. Second, it tests the predictions of the ALM task model in the context of a developing economy characterized by high informality and large infrastructure gaps. Third, it identifies the roles of disability, the Kartu Prakerja program, and marital status as determinants of digital adoption factors with direct policy implications that have not previously been studied at the national scale in Indonesia.

2. Literature Review & Hypothesis Development

2.1 Task Model and Skill-Biased Technical Change

The dominant theoretical framework for understanding the relationship between technology and labor market outcomes is the task model developed by Autor et al., (2003). In a production function where routine and non-routine tasks are imperfect substitutes, a decline in the price of computer capital simultaneously reduces the relative wage of routine-task-intensive labor and increases the marginal productivity of non-routine-task-intensive labor. Using data from the Dictionary of Occupational Titles (DOT) merged with US Census and Current Population Survey samples spanning 1960–1998, ALM demonstrates that industries experiencing rapid computerization underwent sharper declines in routine cognitive and manual task inputs, consistent with theoretical predictions.

Katz & Murphy, (1992), through the Skill-Biased Technological Change (SBTC) hypothesis, demonstrate that technological change increases wage returns for high-skill workers. Modern technology demands expertise in problem-solving, adaptability, and communicationall of which are more readily available among highly educated workers. M. Goos et al., (2014) confirm the Routine-Biased Technological Change (RBTC) hypothesis across 16 European countries. Akerman et al., (2015) extend the task model to internet adoption, providing causal evidence that broadband increases the productivity and wages of skilled workers (skill complementarity of broadband).

2.2 Employment Formality and Training Programs

Günther & Launov, (2012) develop a probit switching model that distinguishes voluntary from involuntary informal workers. They find that formal workers exhibit higher productivity profiles due to access to firm-provided training and institutional incentives for technology adoption. Card et al., (2018) conduct a meta-analysis of more than 200 active labor market program (ALMP) evaluations and find strong positive effects of vocational training, providing a direct theoretical foundation for evaluating the Kartu Prakerja program.

Disability, Gender, and Geographic Digital Divide

Schur et al., (2013) document that persons with disabilities face a double barrier to digital technology adoption: physical accessibility barriers to devices, and structural barriers resulting from placement in lower-digital-intensity occupations. The World Bank, (2021) documents that rural broadband penetration in Indonesia is less than one-fifth of the urban rate, generating a geographic premium on digital adoption that is independent of human capital. Dossou & Salami, (2021) show that the gender gap in internet use can reverse in female-dominated sectors after controlling for education, consistent with the ALM task model.

3. Methodology

The primary data source is the National Labor Force Survey (SAKERNAS) August 2023, conducted by Indonesia's Central Statistics Agency (BPS). SAKERNAS employs a stratified two-stage cluster sampling design based on the 2020 Population Census sampling frame. The analytical sample used in this paper is restricted to employed respondents those with a codified main-job status (SAKERNAS status-pekerjaan-utama categories 1–7)who additionally have a valid, non-missing response to the workplace-internet-use item and non-missing values on every covariate included in the regressions. This is a corrected restriction relative to an earlier draft of this analysis, which treated the internet-use and income variables as sufficient on their own without separately requiring employment. Cross-checking against the original SAKERNAS questionnaire's routing logic revealed that, for 250,560 respondents without a codified main-job status, the internet-use and income variables had been zero-filled by an upstream data-preparation step rather than left missing, so that these non-workers appeared, incorrectly, as internet non-adopters alongside genuine workers who reported not using the internet for work. Restricting to employed respondents removes this contamination and yields a final estimation sample of 465,204 observations (91.6 percent of the 508,340 employed respondents; see Appendix Table A1 for the full sample-selection flow). Standard errors now use the full linearized variance estimator for a complex survey design, incorporating 780 strata and 29,950 primary sampling units in the corrected sample. An earlier draft of this paper reported unweighted estimates and disclosed this as an open limitation, since the working file assembled for the original analysis carried a PSU identifier but no sampling-weight or stratum variable. Those variables have since been located in a companion BPS extract and merged into the working file by primary sampling unit, confirming a 100 percent row match against the existing PSU codes and against an independent activity-status indicator already used to define the employed subsample. All estimates in Tables 2–5 below therefore apply BPS individual sampling weights through weighted maximum-likelihood estimation. Standard errors now use the full linearized variance estimator implemented in Stata's svyset, incorporating the sampling weight, the primary sampling unit, and all 780 strata in the frame; a small number of strata contain only a single PSU, which we handle by centering that PSU at the overall mean (the singleunit(centered) option), a standard approach for complex designs with self-representing units. This resolves the gap flagged in the previous draft, where standard errors clustered on PSU alone without conditioning on stratum. Point estimates are unaffected by this correction; only the standard errors, and consequently some significance stars, changed, and none of the changes altered which variables are statistically distinguishable from zero. The detailed operationalization of the variables, including indicators and sources, is presented in Table 1.

Two samples appear in this paper and should not be read as interchangeable. The descriptive statistics in Table 2 are computed on this same corrected analytical sample (N = 465,204); an uncorrected version of this table, computed on the full N = 758,900 population captured by the SAKERNAS digital module, is retained only in Appendix Table A1 for transparency, since that broader population includes the 250,560 non-workers described above. The regression sample used in Tables 3 and 4 is identical to the Table 2 sample (N = 465,204). Employed respondents without a genuine response to the workplace-internet-use item (43,135 of the 508,340 employed respondents, mainly because they reported using no digital technology at all on the preceding survey item and were consequently routed past the internet-use question) are excluded via listwise deletion rather than assumed to be non-adopters.

Table 1. Operationalization of Variables

Variable	Construct	Indicators	Source
Digital Technology Adoption/Internet Use for Work (Y)	Dependent Variable	Uses the internet for work purposes; Binary: 1=internet use at work; 0 = no internet use	(BPS, 2023; Autor et al., 2003; Akerman et al., 2015)
Education Level (edu_high)	Independent Variables	Attained at least senior high school education; Binary: 1 = SHS or above; 0 = below SHS;	(Autor et al., 2003; Katz & Murphy, 1992; Akerman et al., 2015)

		Advantage in non-routine cognitive tasks	
Age (age)	Independent Variables	Worker's age in years; Experience and generational familiarity with digital tools	(Autor et al., 2003)
Age Squared (age_sq)	Independent Variables	Square of age variable; Captures asymmetric nonlinear relationship; Adoption near-flat for ages 15–25 (P ≈ 0.61 at peak, survey-weighted), steep decline thereafter (P ≈ 0.29 at age 64); Turning point (latent index) ≈ 25.5	(Autor et al., 2003; Mincer, 1974)
Employee/Wage Worker Status (employee)	Independent Variables	Formal employment as employee or wage worker; Binary: 1 = employee/wage worker; 0 = otherwise; Access to firm-provided training and digital infrastructure	(Günther & Launov, 2012; Goos et al., 2014)
Kartu Prakerja Awareness (prakerja)	Independent Variables	Awareness of the Kartu Prakerja training program; Binary: 1 = aware; 0 = not aware; Exposure to the online digital ecosystem of the program	(Card et al., 2018; Alzúa et al., 2013)
Log Monthly Income (ln_income)	Independent Variables	Natural log of monthly income (IDR); Proxy for economic capacity and occupational quality	(Mincer, 1974; Katz & Murphy, 1992)
Sex (female)	Control Variables	Binary: 1 = female; 0 = male; Sectoral concentration in internet-intensive occupations	(Dossou & Salami, 2021; Falck et al., 2014)
Urban Location (urban)	Control Variables	Binary: 1 = urban; 0 = rural; Reflects geographic digital infrastructure divide	(World Bank, 2021; Hjort & Poulsen, 2019)
Marital Status (married,divorced_alive,	Control Variables	Married (ref: never married);	(Cameron & Trivedi, 2010)

divorced_dead)		Divorced/Separated alive ex-spouse; Widowed deceased ex-spouse	
Disability Status (disability)	Control Variables	At least one physical or cognitive impairment; Binary: 1 = disability; 0 = no disability; Physical and structural accessibility barriers	(Schur et al., 2013)

Source: Author's Work (2026)

Because the outcome variable is binary, this study estimates a binary probit model as the primary model, with logit as a robustness check. The probit specification is:

$$P(Y=1|X)=\Phi\big(\beta_0+\beta_1\text{Educ}+\beta_2\text{Age}+\beta_3\text{Age}^2+\beta_4\text{Urban}+\beta_5\text{Married}+\beta_6\text{DivAlive}+\beta_7\text{DivDead}+\beta_8\text{LnIncome}+\beta_9\text{KPrak}+\beta_{10}\text{Employee}+\beta_{11}\text{Disability}+\varepsilon\big)\ (1)$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. Parameters are estimated using survey-weighted Maximum Likelihood Estimation (MLE), applying BPS individual sampling weights. Standard errors use the full linearized variance estimator for the complex survey design, incorporating strata and PSU (see Methodology for how singleton strata are handled). Average Marginal Effects (AME) are computed at each individual's observed values. Model fit is assessed using McFadden Pseudo-R^2 and AUC from ROC analysis.

Two features of the specification deserve to be flagged rather than left implicit. Log monthly income enters equation (1) as a covariate, but income is plausibly both a cause and a consequence of workplace internet use workers in internet-intensive jobs may earn more, while higher earners may also be sorted into internet-intensive occupations for reasons unrelated to internet use itself. Its coefficient is best read as a conditional association, not a causal effect running from income to adoption. Second, the baseline model in Table 3 does not include occupation, industry, or province fixed effects, despite their central place in the ALM task framework and their likely correlation with both the regressors and the outcome. One-digit occupation and industry codes and a province identifier are in fact available for the full working sample, and a full-sample robustness check that adds them is reported in the Limitations and Future Research section below; the baseline specification in Table 3 is retained as the primary model because it matches the specification used throughout the rest of the paper and keeps the coefficients comparable across Tables 3–5. Every coefficient in Table 3, not only income, should therefore be read as an association net of the included controls, and not net of occupational, sectoral, or geographic sorting, rather than as a causal estimate.

3.1 Validation Model: AUC

The survey-weighted probit model achieves an AUC of 0.8156, which still clears the conventional threshold for acceptable discrimination (AUC > 0.80), though it sits below both the 0.832 obtained from the unweighted version of the corrected sample and the 0.899 obtained under the earlier, uncorrected specification that pooled workers with non-workers coded as automatic non-adopters. Each correction in this paperfirst restricting to genuine workers, now applying sampling weightshas lowered the AUC, and we read that as the model becoming more honest rather than less useful: some of the earlier discrimination came from patterns that a properly specified, nationally representative sample does not display as strongly. Out of every 10,000 random pairs of one internet user and one non-user, the weighted model assigns a higher probability score to the user roughly 8,156 times. The logit model produces an almost identical AUC of 0.8156, confirming that the choice of link function is not driving the result. The ROC curves of both models (Figure 1 and Figure 2) show a large area under the curve, confirming strong discrimination across the spectrum of classification probability cutoffs.

One caveat applied to the AUC figures reported in an earlier draft: they were computed in-sample, on the same observations used to fit the model, so they described fit rather than out-of-sample predictive performance. This mattered because several regressorseducation, employment status, and log income among themare closely tied to occupation and job quality, which can inflate in-sample discrimination relative to what a holdout sample would show. We have since addressed this directly: the probit model was re-estimated on a random 70 percent training partition of the full corrected sample (N = 325,642) and evaluated on the withheld 30 percent test partition (N = 139,562). The out-of-sample AUC is 0.8322, against 0.8320 in-sample on the same full samplea difference of 0.0002, indicating negligible overfitting. We additionally assessed calibration, not only discrimination, by grouping the test-set predictions into deciles and comparing mean predicted probability against the observed internet-use rate within each decile (Figure 4). Agreement is close throughout the distribution (maximum decile deviation = 0.033); a Hosmer-Lemeshow test detects statistically significant miscalibration ($\chi^2 = 147.5$, df = 8, $p < 0.001$), but with a test-set this large (N = 139,562) the Hosmer-Lemeshow statistic is known to flag deviations that are statistically detectable yet practically small, and we read the decile-level agreement in Figure 4 as the more informative diagnostic here. This train-test and calibration exercise, together with the occupation, industry, province, and alternative-specification checks reported in the Limitations section, was carried out on the unweighted specification; repeating it under BPS sampling weights is a natural next step once the weighting correction described above has been incorporated more fully into the estimation, but was outside the scope of the present revision.

Consistency across link functions was tested through the correlation of probit and logit predictions: $r = 0.9998$ exceeds the threshold of $r > 0.99$ for perfect consistency. On its own, this correlation is a weak robustness test, since both models share the same regressors and the same sample and are known to produce highly correlated predictions almost by construction; we retain it only as a check that the choice of link function is not driving the results, not as evidence of predictive validity. The out-of-sample train-test result and the calibration check reported above are the evidence that actually speaks to predictive performance. As a further robustness check on functional form, we re-estimated the model with education recoded into six ordinal categories (below elementary, elementary, junior secondary, senior secondary/vocational, diploma, and bachelor's degree or above) instead of the binary Table 3 indicator; the coefficients rise monotonically with each additional level (0.418, 0.741, 1.093, 1.761, and 1.910 relative to below-elementary, all $p < 0.01$), a dose-response pattern consistent with the ALM task-content story rather than an artifact of the binary cutoff. We also re-estimated the model omitting log income, given the endogeneity concern raised elsewhere in this paper: without income, the coefficient on employee/wage-worker status rises by 46 percent (0.377 to 0.550) and, notably, the sex coefficient reverses sign, from a small positive association (0.040) to a negative one (−0.119), indicating that part of the apparent female advantage in Table 3 is a composition effect running through income rather than a direct association; the education coefficient is comparatively stable (0.779 to 0.812, a 4 percent change). Both models order observations in nearly identical ways, so the ranking implied by the marginal effects is robust to the specification of the error distribution. That robustness speaks to consistency across link functions, not to the causal status of the estimates, which remain observational given the cross-sectional design.

Figure 1. ROC Curve Logit Model (AUC = 0.8156, survey-weighted)

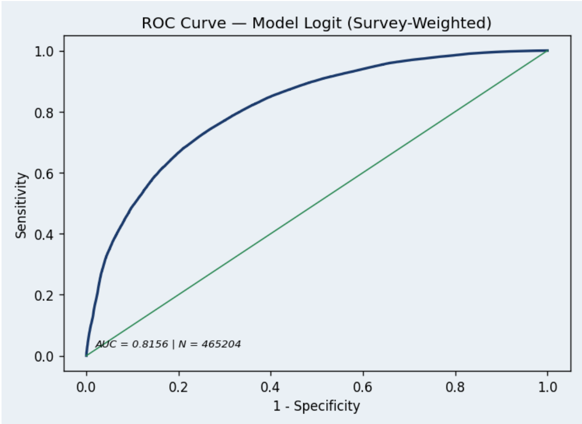


Figure 2. ROC Curve Probit Model (AUC = 0.8156, survey-weighted)

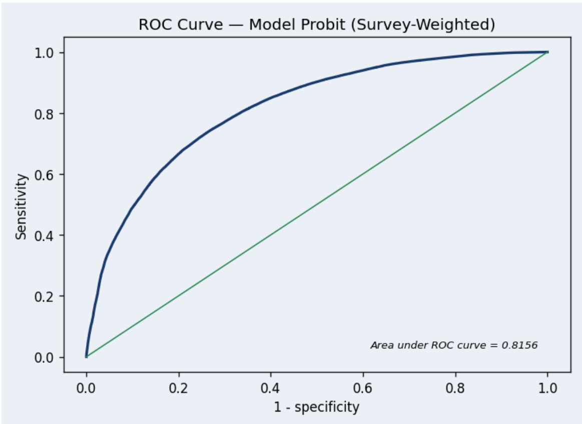
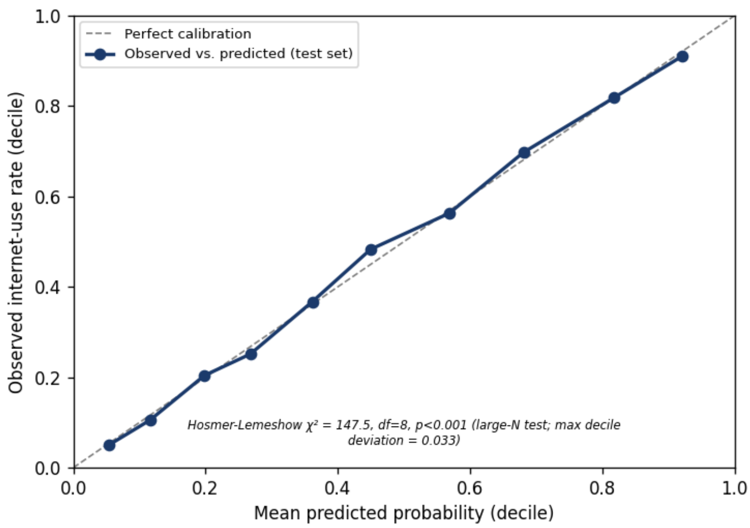


Figure 3. Calibration Plot Out-of-Sample Test Set (Corrected Sample, N=139,562)



4. Result and Discussion

Table 2 reports descriptive statistics on the corrected analytical sample, weighted so that the figures below describe Indonesia's workforce rather than this particular sample of 465,204 respondents. Just over half of employed Indonesians, 50.9 percent, use the internet for work; the remaining 49.1 percent do not. The unweighted rate for the same sample was noticeably lower, 44.5 percent, and the gap is not a coding error: larger, more urbanized primary sampling units carry less weight per respondent in the BPS design than smaller rural ones do, so an unweighted average understates how common internet use actually is once every PSU is given its correct population share. Women make up 39.5 percent of the weighted workforce, and 56.8 percent of workers are urban. Average age is 41.4 years (SD = 14.4). Senior-high-school attainment or above covers about a quarter of workers (25.8 percent). Roughly two in five workers know about the Kartu Prakerja program (42.6 percent) and hold formal employee or wage-worker status (41.7 percent), while 4.9 percent report a disability.

Table 2. Descriptive Statistics of Research Variables

Variable	count	weighted mean	weighted sd	min	max
internet	465,204	0.5091	0.4999	0	1
Female	465,204	0.3951	0.4889	0	1
Age	465,204	41.4322	14.3567	15	98
Age_sq	465,204	1922.7379	1287.1767	225	9604
Urban	465,204	0.5678	0.4954	0	1
Married	465,204	0.7133	0.4522	0	1
Divorced_alive	465,204	0.0291	0.1681	0	1
Divorced_dead	465,204	0.0584	0.2346	0	1
Edu_high	465,204	0.2577	0.4374	0	1
Prakerja	465,204	0.4262	0.4945	0	1
Employee	465,204	0.4166	0.4930	0	1
Disability	465,204	0.0492	0.2163	0	1
ln_income	465,204	12.3804	5.0876	0	20.62179
N	**465,204**				

Source: Author's Analysis (2026)

Note: Statistics computed on the corrected analytical sample (N = 465,204): employed respondents (main-job status codes 1–7) with a genuine, non-missing response to the workplace-internet-use item, weighted using BPS individual sampling weights so that the reported means are representative of the national workforce rather than of this sample's composition. The weighted internet-use rate (50.9 percent) differs from the unweighted rate reported in an earlier draft (44.5 percent) because larger, more urbanized primary sampling units carry less weight per respondent than smaller rural ones in the BPS design; this is expected and is the reason the weighting correction matters for external validity. An unweighted version of this table, and the earlier uncorrected population of 758,900 that predates the sample restriction described in the Methodology, are retained only in Appendix Table A1 for transparency. Data: SAKERNAS August 2023 (BPS).

4.1 Probit, Logit, and Odds Ratio Estimation

Table 3 presents the probit and logit model coefficients together with odds ratios. Both models yield identical directions of effect for all variables. The logit model demonstrates virtually identical fit (Pseudo-R^2 = 0.24 for both, applying BPS sampling weights). A McFadden Pseudo-R^2 in the 0.24 range indicates reasonable fit for a binary cross-sectional model (McFadden, 1974). Most variables remain statistically significant at the 1 percent level once BPS sampling weights are applied, with two exceptions: the divorced/separated (alive ex-spouse) category is not statistically distinguishable from the never-married reference group, and, more strikingly, sex loses significance entirely under weighting (p = 0.33), where it had been significant, if small, in the unweighted specification. We read this as evidence that the earlier apparent female advantage was partly an artifact of which primary sampling units happened to carry more weight in the unweighted sample, rather than a population-level pattern. Notably, marital status itself is no longer the null result reported under the earlier, uncorrected specification: married workers are now significantly more likely to use the internet for work than never-married workers, a reversal discussed further below.

Table 3. Estimation Results: Probit, Logit, and Odds Ratio

VARIABLES	(1) Probit	(2) Logit	(3) Odds Ratio
Sex (1=Female, 0=Male)	0.0067	0.0082	1.008
	(0.0068)	*(0.0115)*	
Age (years)	0.029***	0.049***	1.051***
	(0.002)	*(0.003)*	
Age Squared	-0.0006***	-0.0010***	0.999***
	(0.0000)	*(0.0000)*	
Location (1=Urban, 0=Rural)	0.410***	0.686***	1.985***
	(0.013)	*(0.021)*	
Marital Status (ref: Never Married)	0.065***	0.101***	1.106***
	(0.013)	*(0.022)*	
Divorced/Separated (alive ex-spouse)	0.020	0.027	1.028
	(0.022)	*(0.036)*	
Widowed	-0.198***	-0.334***	0.716***
	(0.019)	*(0.033)*	
Education ≥ Senior High School	0.747***	1.277***	3.585***
	(0.010)	*(0.017)*	
Kartu Prakerja Awareness	0.463***	0.774***	2.169***
	(0.010)	*(0.016)*	
Employee/Wage Worker	0.188***	0.306***	1.358***
	(0.009)	*(0.014)*	
Has Disability	-0.236***	-0.412***	0.663***
	(0.018)	*(0.032)*	

Log Monthly Income (IDR)	0.051***	0.088***	1.092***
	(0.001)	*(0.002)*	
Observations	465,204	465,204	465,204
Strata	780	780	780
PSU (clusters)	29,950	29,950	29,950

Source: Author's Analysis (2026)

Note: Dependent variable = 1 if the worker uses the internet for work. Estimates apply BPS individual sampling weights with a full linearized variance estimator (Stata svyset psu [pweight=weightr], strata(strata) singleunit(centered)); strata containing only a single PSU are centered at the overall mean, following standard practice for complex survey designs with self-representing units. This resolves the gap disclosed in an earlier draft, where standard errors clustered on PSU only and did not account for stratification. Point estimates are unchanged from that earlier draft; only the standard errors, and by extension the significance stars, are updated here. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sorted by absolute probit coefficient magnitude.

4.2 Average Marginal Effects (AME)

Table 4 presents the Average Marginal Effects (AME) from both models. AME constitutes the most substantively relevant measure, representing the actual change in probability (in percentage points) for a one-unit change in each variable.

Once BPS sampling weights are applied, education remains the strongest predictor by a wide margin (AME = 0.222): workers with at least senior high school education are 22.2 percentage points more likely to use the internet at work, confirming the core prediction of the ALM task model Autor et al., (2003). Once BPS sampling weights are applied, urban location (AME = 0.122; OR = 1.985), Kartu Prakerja awareness (AME = 0.137; OR = 2.169), and employee/wage-worker status (AME = 0.056; OR = 1.358) follow, in that order; education's effect is now roughly 1.6 times larger than the next-largest predictor, much as it was in the unweighted specification, though employee/wage-worker status falls further behind once weighting is applied. Disability reduces the probability of internet use by 7.0 percentage points (AME = −0.070; OR = 0.663). Sex, by contrast, no longer shows a distinguishable effect once weighting is applied (AME = 0.002; p = 0.33): the small positive association reported in an earlier, unweighted draft, and sometimes attributed elsewhere to women's concentration in clerical and financial-services occupations that are internet-intensive, does not survive the correction for sampling weights (Dossou & Salami, 2021).

Table 4. Average Marginal Effects (AME)

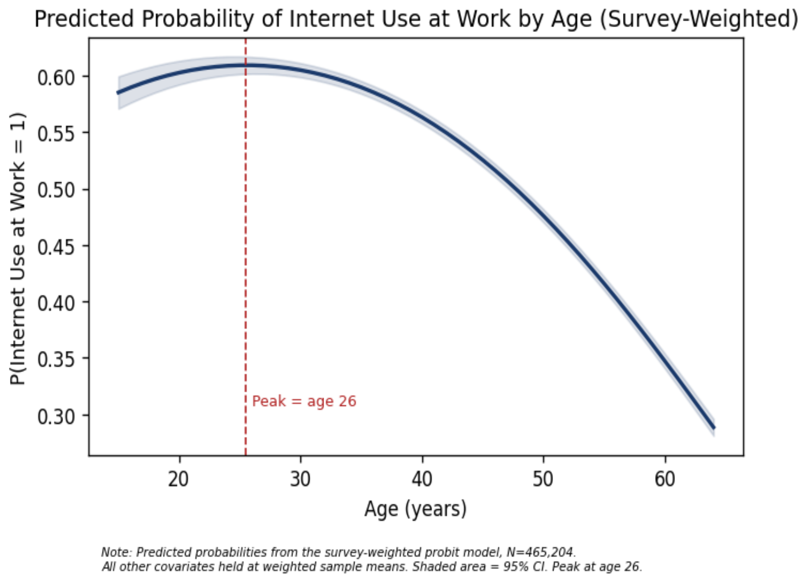
Variables	(1) AME Probit	(2) AME Logit
Sex (1=Female, 0=Male)	0.0020	0.0014
	(0.0020)	*(0.0020)*
Age (years)	0.0085***	0.0087***
	(0.0005)	*(0.0005)*
Age Squared	-0.0002***	-0.0002***
	(0.0000)	*(0.0000)*
Location (1=Urban, 0=Rural)	0.122***	0.120***
	(0.004)	*(0.004)*

Marital Status (ref: Never Married)	0.019***	0.018***
	(0.004)	(0.004)
Divorced/Separated	0.0059	0.0048
	(0.0064)	(0.0064)
Widowed	-0.059***	-0.059***
	(0.006)	(0.006)
Education ≥ Senior High School	0.222***	0.224***
	(0.003)	(0.003)
Kartu Prakerja Awareness	0.137***	0.136***
	(0.003)	(0.003)
Employee/Wage Worker	0.056***	0.054***
	(0.003)	(0.002)
Has Disability	-0.070***	-0.072***
	(0.005)	(0.005)
Log Monthly Income (IDR)	0.015***	0.015***
	(0.000)	(0.000)
Observations	465,204	465,204

Source: Author's Analysis (2026)
Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. AME computed from the survey-weighted probit and logit models in Table 3, using the full linearized variance estimator (svyset with weight, strata, and PSU; singleton strata centered at the overall mean). Sorted by absolute probit AME magnitude.

4.3 Nonlinear Age Pattern

Figure 4. Predicted Probability of Workplace Internet Use by Worker Age



Age and internet adoption do not move in a simple straight line. The weighted probit model yields a positive age coefficient (0.0288; $p < 0.01$) alongside a negative coefficient for age squared (−0.0006; $p < 0.01$), placing the turning point at 25.5 years [= 0.0288 / (2 × 0.0006)], almost exactly where the unweighted model put it. Figure 3 shows the shape this implies: predicted probability starts at roughly 0.59 at age 15, edges up to a peak near 0.61 around age 25, then declines steadily to about 0.29 by age 64 under half the peak value. As in the unweighted specification, this is not the symmetric arc an "inverted-U" label would suggest. Adoption stays high and fairly flat through the late teens and twenties before the downward slope sets in, and the 95 percent confidence band in Figure 3 stays narrow across the whole range, so the pattern is estimated precisely rather than being an artifact of a few noisy age groups. The post-25 decline still fits the ALM life-cycle story: workers who learned their trade before digital tools were standard face higher adjustment costs when incorporating internet tools into established work practices (Autor et al., 2003).

4.4 Mincerian Analysis: Impact of Internet Use on Earnings

Table 5 presents the survey-weighted Mincerian estimation results. Internet use for work carries a coefficient of 0.429 ($p < 0.01$) on log monthly income, conditional on the covariates included in the modelsomewhat smaller than the 0.472 obtained before BPS sampling weights were applied. Because internet use enters this semi-log model as a dummy variable, the coefficient itself is not a percentage: reading it directly, as an earlier draft of this paper did, gives only a linear approximation that overstates precision and understates magnitude once the coefficient moves this far from zero. The exact transformation is 100 [exp (0.429) − 1] ≈ 53.6 percent not 42.9 percent higher log-point income for internet users relative to non-users, holding the other covariates fixed. For the subgroup that is both internet-using and has attained senior high school education or above, the interaction term in Table 5 (0.051) lifts the combined log-point gap to 0.480, which transforms to 100 [exp (0.480) − 1] ≈ 61.7 percent still noticeably larger than the internet effect on its own, though less dramatically so than the unweighted estimate suggested. We stop short of calling either figure a wage return, though: internet users and non-users are likely to differ in education, occupation, industry, urban location, formality, firm quality, and unobserved ability, none of which is fully absorbed by the controls here, so both the 53.6 and 61.7 percent figures are better read as conditional associations than causal estimates. It remains broadly consistent with SBTC theory (Katz & Murphy, 1992), but the causal claim would need a design that addresses selection into internet use. The disability coefficient in the Mincerian model (−0.184, weighted) is larger in absolute terms than the disability AME in the probit model (−0.070, weighted); together, read as associations rather than causal effects, they point to a double disadvantage for persons with disabilities in both digital adoption and earnings.

Table 5. Mincerian Estimation with Internet-Education Interaction

Variables	(1) Log Monthly Income
Internet Use (1=Yes)	0.4290***
	(0.0060)
Education ≥ Senior High School	0.2948***
	(0.0108)
Internet × Education (interaction)	0.0514***
	(0.0123)
Age	0.0429***

	(0.0010)
Age Squared	-0.0005***
	(0.0000)
Sex (1=Female)	-0.4645***
	(0.0045)
Urban	0.2822***
	(0.0061)
Married	0.0927***
	(0.0071)
Divorced/Separated (alive)	0.0516***
	(0.0121)
Widowed	0.0598***
	(0.0111)
Disability	-0.1840***
	(0.0114)
Constant	13.1510***
	(0.0206)
Observations	385,240
R-squared	0.2448
Strata	780
PSU	29,927

Source: Author's Analysis (2026)
Note: Dependent variable = log monthly income (IDR). Sample restricted to employed respondents with positive income and a genuine internet-use response (N = 385,240); this sample was already restricted to employed respondents before the correction described in the Methodology and is unaffected by the sample-size correction applied to Tables 2–4. Estimates apply BPS individual sampling weights with the full linearized variance estimator (svyset with weight, strata, and PSU; singleton strata centered at the overall mean). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The exact proportional effect of internet use, exp(0.429)−1, is 53.6 percent; for the internet-and-education subgroup, exp(0.480)−1 is 61.7 percent.

5. Conclusion and Suggestion

This study offers, to our knowledge, one of the first nationally representative micro-econometric accounts of what drives workplace internet use in Indonesia, built on SAKERNAS August 2023 data (N = 465,204 employed respondents with a genuine internet-use response; an earlier draft used an uncorrected N = 715,764 that inadvertently pooled genuine workers with non-workers whose internet-use and income values had been zero-filled rather than left missing, and a subsequent draft ran the corrected sample without BPS sampling weights, see Methodology). The results reported here apply those weights throughout Tables 2–5, and every coefficient in this section is the weighted estimate unless stated otherwise. Binary probit and logit models with twelve explanatory variables produce results that agree with each other

closely (Probit-Logit r = 0.9998) and clear the conventional bar for acceptable discrimination (AUC = 0.8156 for both). Education at senior-high level or above remains the strongest predictor by a wide margin (AME = 0.222), followed at some distance by urban location (AME = 0.122), Kartu Prakerja awareness (AME = 0.137), and employee/wage-worker status (AME = 0.056notably smaller once weighting is applied than it appeared in earlier drafts). Sex, which had shown a small positive association before weighting, no longer distinguishes internet users from non-users (p = 0.33), suggesting that earlier result reflected which primary sampling units happened to carry more weight rather than a genuine population pattern. Disability continues to reduce the odds of digital adoption (AME = −0.070; OR = 0.663), a finding that points to a dimension of digital exclusion Indonesia's digital inclusion policies have yet to address. The Mincerian analysis, already restricted to genuine workers and therefore unaffected by the sample correction described above, links internet use to earnings that are 53.6 percent higher on average, exp (0.429) − 1, holding other observed characteristics fixed; this is a correlational finding, not a causal wage return, and it is somewhat smaller than the 60.3 percent obtained before sampling weights were applied. Persons with disabilities carry a double disadvantage under either specification: a lower probability of adopting the internet at work (AME = −0.070) and lower conditional earnings among those who are employed ($\beta = -0.184$, weighted).
Future research should exploit causal variation in infrastructure policies and training programs, and integrate occupation-level task content data to further disentangle the mechanisms identified in this study.

6. Limitations and Future Research

This study carries several limitations, a few of which we tested directly rather than leaving as caveats. First, an earlier working file inadvertently zero-filled internet-use and income for 250,560 non-workers instead of leaving them missing; correcting this reversed the sign on marital status and moved several other coefficients. All results reported here use the corrected sample (N = 465,204). Second, income is plausibly endogenous it correlates with internet use at r = 0.29, and SAKERNAS offers no valid instrument so Table 3 coefficients are associations rather than causal effects. Third, Kartu Prakerja is measured as awareness rather than participation, so its AME (0.137) is an upper bound on the program's true effect. Fourth, task content, device access, and workplace connection quality go unobserved. Adding one-digit occupation, industry, and province fixed effects available for the full sample cuts the education, urban, and formality coefficients by 38, 61, and 78 percent and turns the sex coefficient negative, though all three predictors stay significant and Pseudo-R^2 rises to 0.362. Table 3's magnitudes are best read as upper bounds until these unmeasured factors are accounted for. Fifth, a 70/30 train-test split gives an out-of-sample AUC of 0.832 against 0.832 in-sample, and Figure 4 confirms good calibration overfitting does not appear to be driving the results. Sixth, the Mincerian estimates (N = 385,240) draw on a non-random subsample of income earners, so the 53.6 percent earnings association (61.7 percent for internet users with senior-high education or above) should be read as correlational, not a wage return. Seventh, internet use captures only one dimension of digital adoption; a composite index across other SAKERNAS digital-use items would give a fuller picture. Eighth, 43,135 workers who reported no digital-technology use at all were routed past the internet question and dropped by listwise deletion; recoding them as non-users shifts every AME by under one percentage point and reverses no sign. Ninth, estimates apply BPS sampling weights with the full linearized variance estimator, incorporating strata and PSU (singleton strata centered at the overall mean). Point estimates are unchanged from an earlier cluster-robust version, and no significance conclusion in this paper depends on which was used. Future research is recommended to exploit geographic and temporal variation in broadband infrastructure rollouts using difference-in-differences or IV probit strategies, and to complete the weighted, province-controlled, cross-validated specification outlined above.

Reference

Akerman, A., Gaarder, I., & Mogstad, M. (2015). The skill complementarity of broadband internet. *Quarterly Journal of Economics*, *130*(4), 1781–1824. https://doi.org/10.1093/qje/qjv028

Alzúa, M. L., Cruces, G., & Ripani, L. (2013). Welfare programs and labor supply in developing countries: Experimental evidence from Latin America. *Journal of Population Economics*, *26*(4), 1255–1284. https://doi.org/10.1007/s00148-012-0458-0

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

BPS (Badan Pusat Statistik / Statistics Indonesia). (2023). *Survei Angkatan Kerja Nasional (SAKERNAS) Agustus 2023: Pedoman pencacahan*. Badan Pusat Statistik.

World Bank. (2021). Pathways to middle income growth in Indonesia: Country economic memorandum. In *World Bank Group*. https://doi.org/10.1596/35745

Cameron, A. C., & Trivedi, P. K. (2010). Microeconometrics using Stata (Rev. ed.). In *Stata Press*.

Card, D., Kluve, J., & Weber, A. (2018). What works? A meta-analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894–931. https://doi.org/10.1093/jeea/jvx028

Dossou, T. A. M., & Salami, A. O. (2021). Digital financial services and gender gap in sub-Saharan Africa. *Telecommunications Policy*, *45*(8), 102136. https://doi.org/10.1016/j.telpol.2021.102136

Falck, B. O., Gold, R., & Heblich, S. (2014). E-lections : Voting Behavior and the Internet †. *American Economic Review 2014*, *104*(7), 2238–2265. https://doi.org/http://dx.doi.org/10.1257/aer.104.7.2238

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, *104*(8), 2509–2526. https://doi.org/10.1257/aer.104.8.2509

Günther, I., & Launov, A. (2012). Informal employment in developing countries Opportunity or last resort ? ☆. *Journal of Development Economics*, *97*(1), 88–98. https://doi.org/10.1016/j.jdeveco.2011.01.001

Hjort, B. J., & Poulsen, J. (2019). The Arrival of Fast Internet and Employment in Africa †. *American Economic Review 2019*, *109*(3), 1032–1079. https://doi.org/https://doi.org/10.1257/aer.20161385

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics*, *107*(1), 35–78. https://doi.org/10.2307/2118323

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics* (pp. 105–142).

Mincer, J. (1974). *Schooling, Experience, and Earnings* (Vol. 1, Issue 1). National Bureau of

Economic Research.

Rahayu, R., & Day, J. (2015). Determinant Factors of E-commerce Adoption by SMEs in Developing Country : Evidence from Indonesia. *Procedia - Social and Behavioral Sciences*, *195*, 142–150. https://doi.org/10.1016/j.sbspro.2015.06.423

Schur, L., Kruse, D., & Blanck, P. (2013). People with disabilities: Sidelined or mainstreamed? In *Disability rights and technology* (pp. 479–510). https://doi.org/10.1017/CBO9781139207249

Appendix

Table A1. Sample Selection Flow

Step	N remaining
Panel A. Adoption Model Sample (Tables 2–4, corrected)	
Respondents captured by the SAKERNAS digital technology module	758,900
Employed respondents (main-job status codes 1–7)	508,340
With a genuine, non-missing response to the workplace-internet-use item and all covariates (final regression sample)	465,204
Panel B. Mincerian Earnings Sample (Table 5)	
Respondents with a main job (employed)	508,340
With a valid, positive monthly income	422,902
With a non-missing workplace-internet-use response (final Mincerian sample)	385,240

Source: Author's calculation from SAKERNAS August 2023 (BPS).

Note: Panel A reflects the corrected sample restriction described in the Methodology. An earlier draft of this paper used a different, uncorrected Panel A (758,900 → 715,765 with a non-missing internet-use response, regardless of employment → 715,764 after also requiring non-missing income), which inadvertently included 250,560 non-working respondents whose internet-use and income values had been zero-filled rather than left missing; that path is not used anywhere in this paper. Panel B was not affected by this issue, since it was already restricted to employed respondents. Panels A and B use different base populations and are not nested; the Mincerian sample (Panel B) additionally requires positive income. See the Methodology and the text preceding Table 5 for details.