Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Digital Transformation and New Forms of Exclusion: Rural Peripheries, Gender and Income Inequality

Angelo Leogrande

leogrande.cultore@lum.it

LUM University Giuseppe Degennaro

Ada Spiru

Universiteti Metropolitan Tirana

Alberto Costantiello

LUM University Giuseppe Degennaro

Maria Buonarota

University of Foggia

Carlo Drago

University Niccolò Cusano

Research Article

Keywords: digital divide, income inequality, broadband infrastructure, rural–urban disparities, machine learning

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11284958/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Digital Transformation and New Forms of Exclusion: Rural Peripheries, Gender and Income Inequality

Abstract

Digital transformation can foster inclusive growth, but it may also deepen income and territorial inequalities. Using World Bank data for 155 countries over 2000–2023, we test three hypotheses derived from a stylised model of low-cost adoption and skill-complementary infrastructure. We combine pooled, fixed-effects, random-effects, between, weighted, dynamic and instrumental-variable panel models with six clustering algorithms and explainable and causal machine learning. Internet use reduces inequality: a ten-point increase in Internet users lowers the Gini index by 0.4–0.6 points within countries, and the fixed-effects estimate survives region-specific shocks, external instruments, spatial models and five alternative inequality measures. Fixed broadband is associated with lower inequality across countries and higher inequality within them, but the within-country association reflects regional trajectories rather than an established causal effect. Skill-based amplification is not supported, and no robust non-linearity emerges. A persistent rural digital periphery combines weak connectivity, incomplete electrification and low female financial inclusion; the few countries that leave it first expanded electrification and mobile access. Simulations suggest that closing the periphery's adoption gap would lower its Gini index by about 4 points. Policies should pair infrastructure with affordability, digital skills and gender-sensitive inclusion.

Keywords: digital divide; income inequality; broadband infrastructure; rural–urban disparities; machine learning

JEL codes: D63; O33; R11; C23; C38

1. Introduction

Digital transformation is reshaping how income is generated and distributed. Yet the benefits of digital technologies depend on complementary skills, infrastructure and institutions, which are themselves unequally distributed (Bauer, 2018; Hilbert, 2016). Whether digitalisation narrows or widens income inequality is therefore an empirical question, and it is central to the policy debate on inclusive and territorially balanced development.

The evidence remains contested. Several cross-country studies find that Internet and mobile diffusion reduce income inequality (Canh et al., 2020; Lee et al., 2026; Asongu & Odhiambo, 2019), whereas others find effects that depend on the technology, the inequality measure and the level of development (Richmond & Triplett, 2018; Nguyen, 2022). Fixed broadband is especially ambiguous: it reduces inequality in some settings (Houngbonon & Liang, 2021) but widens income gaps through skill- and location-biased diffusion in others (Cao et al., 2025; He et al., 2025). Individual adoption and fixed infrastructure may therefore affect income distribution in opposite directions.

Digital inequalities also have a marked territorial and gender dimension. Rural and peripheral areas suffer from weaker infrastructure, lower digital skills and limited electrification (Myovella et al., 2021;

Morris et al., 2022), core–periphery patterns persist even in advanced economies (Rubiera-Morollón et al., 2026), and rural and low-income women remain excluded from meaningful digital use (Mariscal et al., 2019). Yet global panels rarely test whether rurality moderates the distributional effect of digitalisation.

A methodological gap compounds these empirical ones. Average panel effects can conceal persistent digital regimes (Śledziewska & Łebkowska, 2026), and machine learning, which could detect non-linearities and interactions, remains mostly predictive in this field (Hoepner et al., 2021; Mulenga et al., 2026).

This article addresses these gaps with an integrated three-stage design applied to World Bank data for 155 countries over 2000–2023. A stylised model of two digital technologies, low-cost adoption and skill-complementary infrastructure, first generates three testable hypotheses. We then estimate static, dynamic and instrumental-variable panel models of the Gini index, using regional diffusion and a shift-share instrument based on historical fixed-telephone networks, and test whether the estimates survive region-specific common shocks. Second, we compare six clustering algorithms on eleven validity indices to identify digital–territorial regimes that combine digital, territorial, gender and inequality indicators. Third, we use machine-learning regression not for prediction but to validate the econometric results without parametric assumptions.

The results show that digital transformation has two distributional faces, but only one of them is robust. A ten-percentage-point rise in Internet use lowers the Gini index by about 0.6 points within countries, and this effect survives instrumental variables, region-by-period effects, alternative inequality measures and spatial models. Fixed broadband is associated with lower inequality across countries and higher inequality within them, but the latter association disappears once regional shocks are controlled for. Rurality does not weaken the Internet effect, yet it defines a persistent rural digital periphery where low connectivity, weak electrification and limited female financial inclusion overlap.

The article makes three contributions: it separates adoption and infrastructure within a single identification framework that includes external instruments and falsification tests for regional shocks; it brings territorial and gender dimensions into a global panel; and it shows how clustering and explainable machine learning can strengthen econometric inference. Section 2 reviews the literature and develops the model, Section 3 presents the data and methodology, Sections 4–6 report the panel, clustering and machine-learning results, Section 7 discusses them, Section 8 derives policy implications, Section 9 discusses limitations and Section 10 concludes. Supplementary Appendices A–E report additional analyses.

2. Literature Review

The relationship between digital transformation and income inequality remains unsettled, because ICTs interact with labour-market, institutional and policy conditions (Bauer, 2018). We organise the evidence into six macro-themes, synthesised in Supplementary Table E1 together with an extended review (Supplementary Appendix E).

2.1. ICT diffusion and income inequality: equalising, skill-biased or non-linear?

Several cross-country panels find that Internet and mobile diffusion reduce inequality (Canh et al., 2020; Lee et al., 2026; Sohag et al., 2024), with consistent system-GMM evidence for Africa (Asongu & Odhiambo, 2019; Tchamyou et al., 2019) and instrumental-variable confirmation based on lightning density (Liu, 2017). Others show that the sign depends on the technology, the inequality measure and the level of development, often with U-shaped trajectories (Richmond & Triplett, 2018; Nguyen, 2022; Baffour Gyau et al., 2025; Zhang, Li, et al., 2025). In Europe, digital skills rather than broadband reduce inequality (Au, 2024), and micro-evidence points to skill-biased returns to Internet use (Hua & Zhang, 2023).

2.2. Infrastructure versus adoption: the ambiguous role of broadband

Quasi-experimental studies of the "Broadband China" programme diverge: some report lower urban–rural, opportunity and employment inequality (Li et al., 2024; Liang et al., 2025; Zhang & Weng, 2024), others a wider income gap through skill-biased technical change (He et al., 2025; Cao et al., 2025). Chen et al. (2025) show that expanding access reduced deprivation, whereas speed upgrades for those already connected increased inequality, and broadband reduced inequality across French cities through manual-job creation (Houngbonon & Liang, 2021). Whether infrastructure is equalising thus depends on who is connected first, which is the core of our infrastructure–adoption distinction.

2.3. Territorial divides: rural–urban and core–periphery gaps

In China, the digital economy has widened the urban–rural income gap (Deng et al., 2023; Wang & Liu, 2025), although household evidence shows that digital literacy and Internet use reduce disparities among farmers (Wang, Qu, et al., 2025; Wei & Li, 2024). Rural exclusion, electricity infrastructure and regulatory quality shape digital access in developing countries,

with strong spatial spillovers (Bhandari & Shirazi, 2022; Myovella et al., 2021), and core–periphery gaps persist across EU regions, Indian states and Indonesian districts (Rubiera-Morollón et al., 2026; Arora & Sapre, 2026; Kartiasih et al., 2023a).

2.4. Gender digital divides and digital financial inclusion

Access alone does not close the gender digital gap, because affordability, skills and social norms constrain meaningful use (Mariscal et al., 2019). Mobile ownership and Internet access raise women's participation and earnings (Amber & Chichaibelu, 2023; Tarafdar & Swathysree, 2026), and mobile money reduced women's financial exclusion (Kim, 2022). However, digital credit and FinTech can widen gender gaps (Johnen & Mußhoff, 2023; Olaoye et al., 2025), and minimum ICT thresholds condition the gains in female employment (Asongu & Odhiambo, 2020).

2.5. Typologies and clustering of digital regimes

Digital development forms distinct regimes and convergence clubs rather than a single gradient (Pick et al., 2021; Saba & David, 2020). Śledziewska and Łebkowska (2026) find strongly persistent digital readiness regimes, and Xing et al. (2026) combine clustering, panel regressions and machine learning. Yet existing typologies seldom compare alternative algorithms on a broad battery of validity indices, and they rarely include inequality and gender inclusion jointly among the clustering variables (Bilozubenko et al., 2025; Gomes & Dias, 2025; L'Abbate et al., 2026).

2.6. Machine learning as a complement to econometrics

Machine learning has entered this field mainly as a predictive tool (Blumenstock et al., 2015), but methodological work stresses its complementarity with econometrics: predictive accuracy must not replace significance and explainability (Hoepner et al., 2021; Simonian, 2020), and tree ensembles and cross-validation must be adapted to panel data (Chen, 2021; Baylis et al., 2021). SHAP values make flexible models interpretable (Herrera et al., 2023), while random splits of pooled country-years yield only internal validation (Mulenga et al., 2026), which motivates the country-grouped cross-validation adopted here.

2.7. Research gap

Three gaps emerge. Most cross-country studies treat ICT as a single construct and seldom compare adoption with fixed infrastructure within the same identification framework; the territorial moderation of digital effects is rarely tested in global panels; and few studies

triangulate panel, instrumental-variable and dynamic estimators with clustering and explainable machine learning. This article addresses all three gaps.

2.8. Theoretical framework and testable hypotheses

To discipline the empirical analysis, we formalise the distinction between adoption and infrastructure in a stylised two-group model. A high-skill group H, with population share s, and a low-skill group L, with share 1 − s, can use two technologies. Internet and mobile adoption (A) has a low entry cost and widens access to information, markets, jobs and financial services. Fixed broadband infrastructure (B) requires complementary skills, capital and dense networks, so its returns accrue mainly to those who already possess human capital (Acemoglu and Autor, 2011; Hilbert, 2016). Group incomes are

$$y_L = w(1 + \alpha \cdot a_L) \tag{M1}$$

$$y_H = \theta \cdot w(1 + \alpha \cdot a_H + \beta \cdot B) \tag{M2}$$

where w is a common wage, $\theta > 1$ is the skill premium, a_j is the adoption rate of group j, $\alpha > 0$ is the return to adoption and $\beta > 0$ measures the complementarity between infrastructure and skills. Because high-skill households adopt first, aggregate adoption $A = s \cdot a_H + (1 - s) \cdot a_L$ grows, beyond the early stage of diffusion, mainly through low-skill adopters, so that $da_L/dA > da_H/dA$. Inequality is measured by the log income ratio

$$I = \ln(y_H / y_L) = \ln \theta + \ln(1 + \alpha \cdot a_H + \beta \cdot B) - \ln(1 + \alpha \cdot a_L) \tag{M3}$$

Differentiating (M3) with respect to the two technologies gives

$$\partial I/\partial A = \alpha[(da_H/dA)/(1 + \alpha \cdot a_H + \beta \cdot B) - (da_L/dA)/(1 + \alpha \cdot a_L)] < 0 \tag{M4}$$

$$\partial I/\partial B = \beta/(1 + \alpha \cdot a_H + \beta \cdot B) > 0 \tag{M5}$$

Inequality (M4) holds whenever diffusion occurs mainly among low-skill households, because the high-skill group also starts from a higher income ($1 + \alpha \cdot a_H + \beta \cdot B > 1 + \alpha \cdot a_L$), and (M5) holds for any $\beta > 0$. This yields the first hypothesis.

> *H1 (opposite effects). Within countries, adoption reduces inequality, whereas skill-complementary infrastructure increases it.*

Across countries, infrastructure is not randomly distributed. Suppose that both broadband and the skill premium depend on a persistent development and institutional factor D_c, so that $B_c = \varphi \cdot D_c + b_{ct}$ and $\ln \theta_c = \ln \theta_0 - \gamma \cdot D_c$, with $\varphi, \gamma > 0$: richer economies deploy more broadband and also compress skill premia through their education and redistribution systems. A between-country regression of I on B then estimates

$$\operatorname{plim} \hat{b}_{between} = \partial I/\partial B - \lambda \cdot \gamma/\varphi, \quad \lambda = \varphi^2 \cdot \mathrm{Var}(D)/\mathrm{Var}(B) \in (0, 1] \tag{M6}$$

which is negative whenever the institutional component dominates the direct effect ($\lambda \cdot \gamma / \varphi > \partial I / \partial B$), while country fixed effects remove D_c and recover $\partial I / \partial B > 0$. Adoption, whose direct effect is already negative, remains equalising in both dimensions.

> *H2 (sign reversal). Broadband is negatively associated with inequality across countries and positively within countries, whereas adoption is equalising in both dimensions.*

Finally, if the complementarity parameter is proportional to the human-capital advantage of the high-skill group, $\beta = \beta_0 \cdot (h_H - h_L)$, then $\partial^2 I / \partial B \partial (h_H - h_L) = \beta_0 (1 + \alpha \cdot a_H) / (1 + \alpha \cdot a_H + \beta \cdot B)^2 > 0$.

> *H3 (amplification). The unequalising effect of broadband is stronger where human capital is more concentrated.*

H1 and H2 are tested by comparing between- and within-country and instrumental-variable estimators (Sections 4.2–4.3), across regimes (Section 5) and without parametric assumptions (Section 6); H3 is tested in Section 4.4 by interacting broadband with educational attainment, where it implies a negative interaction.

3. Data and Methodology

3.1. Data

The analysis relies on World Bank World Development Indicators for 217 economies over 2000–2023. The dependent variable is the Gini index, with five alternative distributional measures as robustness checks. Digital transformation is measured through individual adoption (Internet users, % of population), fixed infrastructure (fixed broadband subscriptions per 100 people) and mobile access (mobile subscriptions per 100 people). Territorial structure is captured by the rural population share and rural electricity access, gender inclusion by women's account ownership, and development and policy by ln GDP per capita (PPP) and public education expenditure; Eurostat data by degree of urbanisation and NUTS-2 region support a European case study (Eurostat, 2026). Because household surveys are irregular, the panel uses eight non-overlapping three-year averages, yielding 712 observations for 155 countries; the clustering uses 2015–2023 averages for 123 countries.

3.2. Why three methods: an integrated design

The design combines panel econometrics, unsupervised clustering and supervised machine-learning regression because each method answers a different question and compensates for a limitation of the others (Table 1).

Table 1. Analytical design: methods, questions and rationale

Stage	Method	Research question	Why this method	Output used downstream
1	Panel econometrics: POLS, FE, RE, BE, WLS, FE-2SLS, system GMM	Does digitalisation change inequality within countries?	Controls unobserved heterogeneity; separates between and within variation; addresses endogeneity and persistence	Coefficients and specification used in Stages 2–3
2	Clustering: six algorithms, eleven validity indices	Do countries form distinct digital-territorial regimes?	Uncovers heterogeneity hidden by average effects; robust algorithm choice	Regimes for cluster-specific FE estimation
3	ML regression: six algorithms, grouped CV, SHAP	Are the econometric results robust to functional form?	Relaxes linearity and additivity; tests interactions without pre-specification	Validation of signs, linearity and moderation

Note: POLS = pooled OLS; FE = fixed effects; RE = random effects; BE = between; WLS = weighted least squares; FE-2SLS = fixed-effects instrumental variables; CV = cross-validation.

Panel econometrics is the core of the design because the question is whether digitalisation changes inequality within countries over time. Comparing pooled OLS, fixed effects, random effects, between and weighted least squares shows whether a relationship is driven by differences across countries or by changes within them, which is decisive for broadband. Reverse causality is addressed by fixed-effects instrumental variables based on regional diffusion and on a shift-share instrument built on 1990–1995 fixed-telephone density (Liu, 2017; Czernich et al., 2011), complemented by region-by-period effects and sensitivity tests (Conley et al., 2012), and persistence by system GMM.

Clustering is used because average effects can hide persistent regimes and convergence clubs (Pick et al., 2021; Śledziewska & Łebkowska, 2026). Because typologies may depend on the algorithm, we compare six algorithms on eleven validity indicators and select the best by its overall ranking; the resulting clusters feed back into the econometrics through regime-specific estimates.

Machine-learning regression is used for validation, not prediction (Hoepner et al., 2021; Simonian, 2020). Six learners are compared under country-grouped nested cross-validation, which avoids information leakage (Mulenga et al., 2026), and the best one is interpreted with SHAP values in levels and within countries, to test whether a model free of functional-form assumptions reproduces the fixed-effects results. Temporal hold-out tests, double machine learning and causal forests extend this validation (Chernozhukov et al., 2018; Wager & Athey, 2018). The full rationale for the design is given in Supplementary Appendix E.

The three stages are therefore sequential and mutually reinforcing. The panel models estimate effects, the clustering locates where these effects differ, and the machine-learning stage verifies that both results survive the relaxation of parametric assumptions.

4. Panel Data Econometric Analysis

4.1. Data and empirical specification

We use an unbalanced panel of 155 countries (2000–2023) averaged over eight three-year periods to reduce survey noise (Forbes, 2000), which yields 712 observations. The baseline specification is

$$Gini_{it} = \alpha_i + \lambda_t + \beta_1 Internet_{it} + \beta_2 Broadband_{it} + \beta_3 Mobile_{it} + \beta_4 Rural_{it} + \beta_5 (Internet_{it} \times Rural_{it}) + \gamma' Z_{it} + \varepsilon_{it} \quad (1)$$

Here α_i and λ_t are country and period effects, and Z_{it} contains ln GDP per capita and public education expenditure. Internet use and rurality are mean-centred, so β_1 is the effect of Internet adoption at average rurality and β_5 captures territorial moderation. Equation (1) is estimated by pooled OLS, fixed effects (FE), random effects (RE), between and feasible weighted least squares (WLS); a dynamic version is estimated by two-step system GMM (Arellano and Bond, 1991; Blundell and Bond, 1998). Standard errors are clustered by country. Table 2 summarises the variables.

Table 2. Variable definitions and descriptive statistics (estimation sample, N = 712)

Variable	WDI code	Role	Mean	SD (overall)	SD (within)	Min	Max
Gini index (0–100)	SI.POV.GINI	Dependent	37.26	8.11	2.46	23.80	65.00
Internet users (% of population)	IT.NET.USER.ZS	Digital adoption	45.68	31.66	19.81	0.15	99.35
Fixed broadband subscriptions (per 100 people)	IT.NET.BBND.P2	Digital infrastructure	12.34	13.64	7.45	0.00	48.19
Mobile cellular subscriptions (per 100 people)	IT.CEL.SETS.P2	Digital access	92.15	42.55	32.60	0.58	206.11
Rural population (% of total)	SP.RUR.TOTL.ZS	Territorial	39.03	20.37	2.62	1.02	90.22
ln GDP per capita, PPP (constant 2021 int. $)	NY.GDP.PCAP.PP.KD	Control	9.64	1.09	0.17	6.90	11.80
Government expenditure on education (% of GDP)	SE.XPD.TOTL.GD.ZS	Control	4.49	1.66	0.57	0.35	15.05

Source: World Bank, World Development Indicators. Three-year period averages, 2000–2023. SD (within) is the standard deviation of deviations from country means.

Table 2 shows that within-country variation in the Gini index (SD 2.46) is less than one-third of the overall dispersion (8.11), and rurality varies even less over time (2.62 against 20.37), whereas the digital indicators change substantially within countries. Figure 1 traces these dynamics by income group.

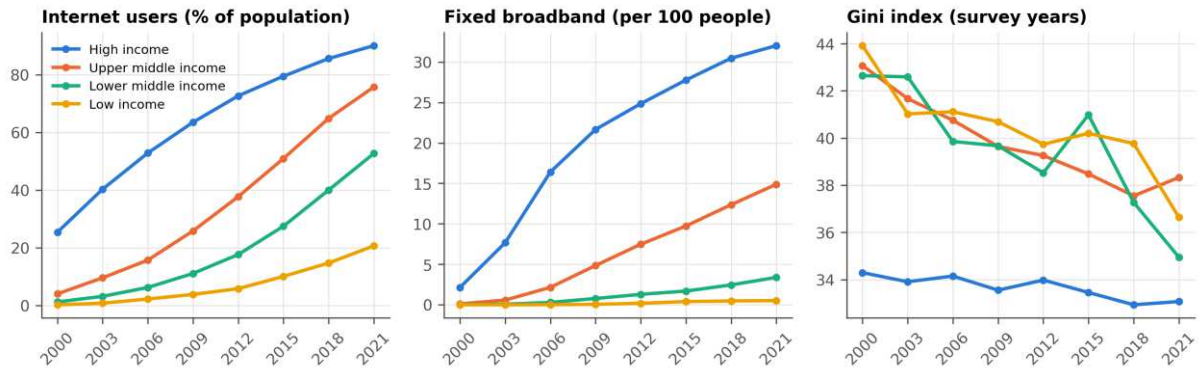


Figure 1. Digital diffusion and income inequality by World Bank income group, 2000–2023

Note: Unweighted group means of three-year period averages. The composition of countries with available Gini surveys changes across periods.

Figure 1 reveals rapid but stratified diffusion: by 2021–23, Internet use exceeds 90% in high-income countries but remains near 20% in low-income economies, and fixed broadband is almost absent outside the two richest

groups. Meanwhile, average inequality declined in low- and middle-income groups. Table 3 tests whether these co-movements survive controls for country heterogeneity and common shocks.

4.2. Baseline estimates

Table 3 reports the six static and dynamic estimators side by side.

Table 3. Static and dynamic panel estimates of the Gini index

Variable	Pooled OLS	Fixed effects	Random effects	Between	WLS	System GMM
Gini (t−1)	—	—	—	—	—	0.690***
						(0.090)
Internet users (%)	-0.130***	-0.063***	-0.066***	-0.100	-0.123***	0.030
	(0.040)	(0.016)	(0.016)	(0.068)	(0.017)	(0.201)
Fixed broadband (per 100)	-0.182***	0.095**	0.064*	-0.165*	-0.097***	0.079
	(0.062)	(0.042)	(0.037)	(0.084)	(0.028)	(0.187)
Mobile subscriptions (per 100)	-0.014	-0.028***	-0.024***	0.030	-0.011*	-0.021
	(0.016)	(0.009)	(0.009)	(0.030)	(0.007)	(0.026)
Rural population (%)	-0.185***	-0.070	-0.034	-0.078*	-0.188***	-0.050
	(0.038)	(0.072)	(0.037)	(0.041)	(0.020)	(0.035)
Internet × Rural (/100)	0.006	-0.007	0.008	0.111	0.029	0.092
	(0.069)	(0.048)	(0.045)	(0.094)	(0.044)	(0.093)
ln GDP per capita (PPP)	-0.359	2.145	-0.464	-0.457	-1.288***	-2.580
	(1.074)	(1.619)	(0.767)	(1.192)	(0.439)	(2.575)
Education expenditure (% GDP)	-0.191	-0.348	-0.404**	-0.085	-0.456**	-0.449
	(0.366)	(0.240)	(0.198)	(0.403)	(0.185)	(0.543)
Period dummies	Yes	Yes	Yes	No	Yes	Yes
Country effects	No	Yes	Random	—	No	System (FD + levels)
Observations	712	712	712	155	712	356
Countries	155	155	155	155	155	155
R^2	0.310	0.316	0.156	0.226	0.632	—
AR(2) / Hansen p-value						0.117 / 0.318

Notes: Cluster-robust standard errors (by country) in parentheses; robust standard errors for the between estimator; Windmeijer-corrected two-step standard errors for system GMM. R^2 is within for FE (including period dummies), overall for RE and weighted for WLS. Internet users and rural population are mean-centred. System GMM uses collapsed instruments: lags 2–3 of the Gini index, lag 2 of Internet users and broadband (treated as endogenous) and the remaining regressors as standard instruments. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3 distinguishes digital adoption from digital infrastructure. Internet use lowers inequality in the pooled OLS, FE, RE and WLS models: in the FE model, a 10 percentage-point rise in Internet users reduces the Gini index by 0.63 points ($p < 0.001$), and mobile subscriptions have a smaller equalising effect (−0.028, $p < 0.01$). Fixed broadband changes sign. In the estimators dominated by cross-country variation it proxies for advanced economies with redistributive institutions and is negative, whereas within countries ten additional subscriptions per 100 people raise the Gini index by 0.95 points ($p = 0.024$), consistent with skill- and location-biased diffusion (Acemoglu and Autor, 2011; Hilbert, 2016). The Internet × Rural interaction is insignificant everywhere, and in the dynamic model inequality is persistent ($\rho = 0.690$). This is the pattern predicted by H1 and H2. Table 4 reports the tests that guide the choice of estimator.

Table 4. Specification and diagnostic tests

Test	Comparison / scope	Statistic	p-value
F test for country effects (poolability)	POLS vs FE	35.21	<0.001
Breusch–Pagan LM (Baltagi–Li, unbalanced)	POLS vs RE	983.01	<0.001
Hausman (conventional)	FE vs RE	122.79	<0.001
Mundlak / correlated RE (cluster-robust Wald)	FE vs RE	24.51	<0.001
Wooldridge serial correlation	AR(1) in idiosyncratic errors	178.72	<0.001
Breusch–Pagan heteroskedasticity	FE residuals	58.25	<0.001
Pesaran CD (unbalanced)	Cross-sectional dependence	0.05	0.956
Arellano–Bond AR(1)	System GMM	-3.03	0.002
Arellano–Bond AR(2)	System GMM	-1.57	0.117
Hansen J (overidentification)	System GMM, 19 instruments	4.71	0.318

Notes: The Mundlak test is a cluster-robust Wald test on the country means of the time-varying regressors in a correlated random-effects model. The Pesaran CD statistic is computed on FE residuals with pairwise-available periods.

The F and Breusch–Pagan LM tests reject pooled OLS, while the Hausman and cluster-robust Mundlak tests reject RE (Hausman, 1978; Mundlak, 1978), so FE is the preferred static estimator; the Pesaran CD test shows no cross-sectional dependence ($p = 0.956$). For system GMM, the AR(2) ($p = 0.117$) and Hansen ($p = 0.318$) tests support instrument validity (Roodman, 2009). Because FE does not address reverse causality, Table 5 reports instrumental-variable estimates.

4.3. Endogeneity: panel instrumental-variable estimates

We re-estimate Equation (1) by FE two-stage least squares (FE-2SLS), treating Internet use, broadband and the interaction as endogenous. IV-A uses leave-one-out regional diffusion, i.e. average penetration among the other countries of the same World Bank region and period, and IV-B adds previous-period values. Because regional diffusion may also transmit common regional shocks, IV-C uses an external shift-share instrument (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022) whose exposure share is fixed-telephone density in 1990–1995, before the commercial diffusion of the Internet, and whose shifters are the global penetration rates of Internet use and broadband. Legacy copper networks lowered the cost of dial-up and DSL access, so countries with denser networks could adopt faster (Czernich et al., 2011). IV-D combines IV-A and IV-C, so that instruments with different exclusion restrictions can be tested against each other. Table 5 reports the results.

Table 5. Fixed-effects instrumental-variable (FE-2SLS) estimates with alternative instruments

Variable	FE (OLS)	IV-A	IV-B	IV-C	IV-D
Panel A. Country and period effects					
Internet users (%)	−0.063*** (0.016)	−0.141*** (0.048)	−0.087*** (0.019)	−0.088** (0.035)	−0.116*** (0.024)
Fixed broadband (per 100)	0.095** (0.042)	0.238*** (0.077)	0.091* (0.048)	0.155*** (0.060)	0.188*** (0.056)
Internet × Rural (/100)	−0.007 (0.048)	0.032 (0.060)	−0.019 (0.048)	−0.052 (0.087)	0.015 (0.052)
Mobile subscriptions (per 100)	−0.028*** (0.009)	−0.009 (0.012)	−0.017* (0.009)	−0.018* (0.010)	−0.014 (0.010)
Rural population (%)	−0.070 (0.072)	−0.099 (0.081)	−0.065 (0.079)	−0.106 (0.076)	−0.091 (0.076)
ln GDP per capita (PPP)	2.145 (1.619)	2.608 (1.690)	1.083 (1.780)	2.722* (1.650)	2.464 (1.612)
Education expenditure (% GDP)	−0.348 (0.240)	−0.240 (0.247)	−0.414 (0.277)	−0.256 (0.246)	−0.245 (0.239)
Instruments	—	Regional LOO (3)	Regional LOO + lagged ICT (6)	Telephony 1990–95 × global diffusion (3)	Regional LOO + telephony (6)
SW first-stage F: Internet users	—	23.5	204.5	27.9	53.5
SW first-stage F: Fixed broadband	—	34.8	174.6	35.7	42.9
SW first-stage F: Internet × Rural	—	143.2	513.2	14.3	85.2
Over-identification J (p-value)	—	Exactly identified	5.24 (0.155)	Exactly identified	4.04 (0.257)
Anderson–Rubin χ^2 (p-value)	—	12.26 (0.007)	29.28 (<0.001)	20.09 (<0.001)	28.30 (<0.001)
Endogeneity test (p-value)	—	6.44 (0.092)	15.34 (0.002)	8.30 (0.040)	13.49 (0.004)
Observations / countries	712 / 155	712 / 155	628 / 150	705 / 153	705 / 153
Panel B. Country and region × period effects					
Internet users (%)	−0.042*** (0.015)	n.i.	n.i.	−0.020 (0.060)	n.i.
Fixed broadband (per 100)	0.001 (0.047)	n.i.	n.i.	−0.024 (0.114)	n.i.
Internet × Rural (/100)	−0.106** (0.053)	n.i.	n.i.	−0.217 (0.146)	n.i.
Minimum SW first-stage F	—	—	—	12.4	—
Observations / countries	712 / 155	—	—	705 / 153	—

Notes: Cluster-robust standard errors (by country) in parentheses. Panel A includes country and period effects; Panel B replaces period effects with region × period effects, under which the leave-one-out regional instruments of IV-A, IV-B and IV-D are not identified (n.i.). LOO = leave-one-out regional mean (World Bank regions). Telephony = fixed-telephone subscriptions per 100 people averaged over 1990–1995, interacted with the global mean of Internet users and of fixed broadband in each period. SW = Sanderson–Windmeijer conditional first-stage F-statistic. The over-identification statistic is Wooldridge's robust score test. The Anderson–Rubin statistic tests the joint null that all endogenous coefficients are zero. The endogeneity test is a cluster-robust control-function Wald test on the first-stage residuals (3 d.f.). *** p < 0.01, ** p < 0.05, * p < 0.1.

Panel A of Table 5 shows that all four instrument sets are relevant, with Sanderson–Windmeijer F-statistics of at least 14.3 (Sanderson and Windmeijer, 2016), and the weak-instrument-robust Anderson–Rubin test rejects the joint null in every column ($p \leq 0.007$; Andrews et al., 2019). With IV-C, a ten-point rise in Internet use lowers the Gini index by 0.88 points ($p = 0.012$), whereas ten additional broadband subscriptions per 100 people raise it by 1.55 points ($p = 0.010$). In the over-identified IV-D model the estimates are −0.116 and 0.188, and the over-identification test does not reject the joint validity of the two instruments ($p = 0.257$). The IV coefficients for Internet use exceed the FE estimate, consistent with attenuation due to measurement error, and the control-function test rejects exogeneity in IV-B, IV-C and IV-D.

Panel B addresses the main threat to these instruments, namely common regional shocks. With region × period effects, the FE estimate for Internet use falls to −0.042 but remains significant ($p = 0.005$), whereas the broadband coefficient collapses to 0.001 ($p = 0.987$), and the IV-C estimates become imprecise although the instrument remains relevant (minimum SW F = 12.4). The within-country unequalising effect of broadband therefore reflects regional trajectories rather than country-specific responses to broadband expansion. The Internet × Rural interaction, by contrast, becomes negative and significant (−0.106, $p = 0.047$), so Internet use is, if anything, more equalising in more rural countries.

Further checks, reported in Supplementary Appendix A.2, qualify the causal interpretation. Region-specific linear trends reproduce the region × period pattern, but the shift-share instruments jointly predict the Gini index

of the previous period ($\chi^2(3) = 18.5$, $p < 0.001$), and leave-one-region-out estimates lose significance without Europe and Central Asia or Latin America and the Caribbean. Plausibly exogenous bounds (Conley et al., 2012) show that the Internet effect remains significant at 5% as long as the direct effect of the instrument is below about one-third of its reduced-form coefficient. The instrumental-variable evidence therefore supports the equalising effect of Internet use but does not establish a causal unequalising effect of broadband. Figure 2 summarises the coefficients across all eight estimators.

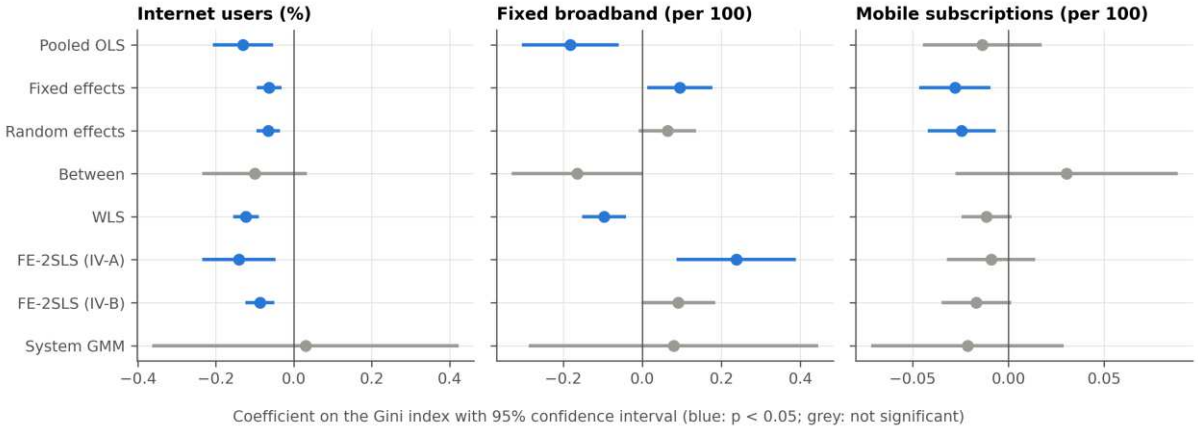


Figure 2. Coefficients of the digital variables across eight panel estimators

Note: Point estimates with 95% confidence intervals. Blue markers are significant at the 5% level.

Figure 2 shows that the Internet effect is significant at 5% in six of eight estimators, whereas broadband is negative in estimators driven by cross-country variation and positive in all within-country estimators. The mobile effect is significant only in FE and RE. Figure 3 examines whether rurality moderates the Internet effect.

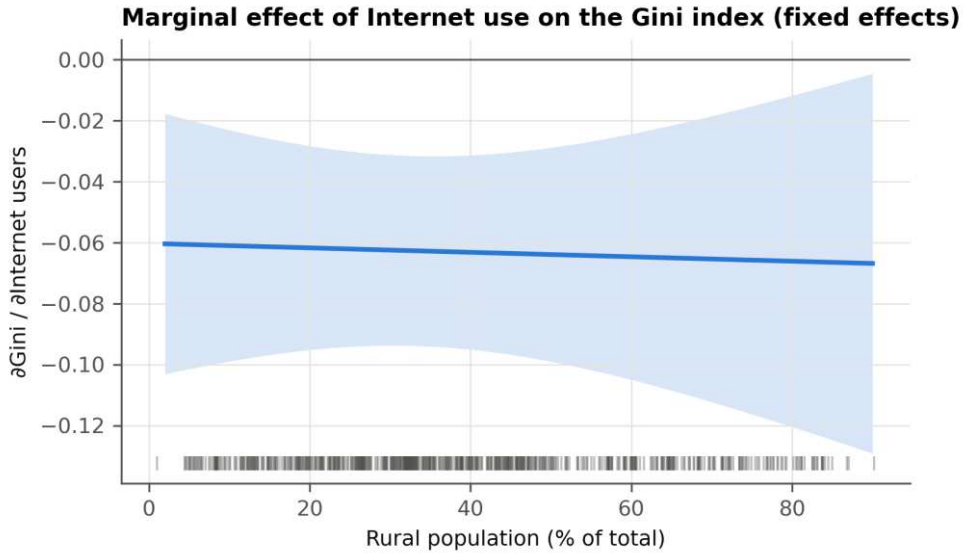


Figure 3. Marginal effect of Internet use on the Gini index across levels of rurality (FE model)

Note: Marginal effect $\beta_1 + \beta_5(\text{Rural} - \text{mean})$ with 95% cluster-robust confidence band. Tick marks show the distribution of the rural population share in the sample.

The marginal effect in Figure 3 is stable at about −0.06 and its confidence band excludes zero at every level of rurality, so the hypothesis that rurality weakens the equalising effect of Internet adoption is not supported at the national level. Driscoll–Kraay standard errors and further robustness checks confirm this result (Supplementary Table A3). Under region × period effects (Table 5, Panel B), the Internet × Rural interaction

becomes negative and significant, so rurality, if anything, reinforces rather than weakens the equalising effect of Internet use.

4.4. Robustness, transmission channels and extensions

Five further analyses, reported in full in Supplementary Appendices A.3–A.7, test the scope of these results. First, with the top 10% and bottom 40% income shares, the Palma and S80/S20 ratios and relative poverty as dependent variables, Internet use is equalising in the FE model for all six measures ($p \leq 0.019$) and remains significant under region × period effects, while the FE broadband association disappears with region × period effects in every case ($p \geq 0.31$). The negative Internet × Rural interaction holds only for the Gini index and the top 10% share and is therefore tentative.

Second, interactions with six human-capital and labour-market moderators, with Holm-adjusted p-values, and a bootstrap mediation analysis (Imai et al., 2010) give no support to the human-capital channel predicted by H3. Internet use is more equalising in service-intensive economies (−0.218, $p = 0.009$), but this pattern falls short of significance after the Holm correction (adjusted $p = 0.101$), and no indirect effect through employment structure is significant.

Third, spatial lag of X and spatial Durbin models (Halleck Vega & Elhorst, 2015; LeSage & Pace, 2009) confirm a negative and significant direct effect of Internet use under every weights matrix, whereas the indirect effects are imprecise, so digitalisation has no robust cross-border effect on inequality.

Fourth, a European case study of poverty by degree of urbanisation (34 countries) and NUTS-2 region (206 regions) finds negative but insignificant Internet × rural interactions, so rural areas do not benefit less from adoption.

Fifth, quadratic, panel threshold (Hansen, 1999) and quantile models (Machado & Santos Silva, 2019) detect no robust non-linearity: the sup-F tests do not reject linearity (bootstrap $p \geq 0.166$), and the Internet coefficient is stable across quantiles (−0.072 to −0.054).

4.5. Discussion

Digital transformation thus has two distributive faces that differ in robustness. Internet use, and less robustly mobile telephony, reduces inequality, whereas the within-country association of broadband with higher inequality vanishes once common regional shocks are absorbed and should be read as a regional pattern consistent with urban- and skill-biased deployment (Morris et al., 2022; Tewathia et al., 2020). Further estimates and diagnostics are reported in Supplementary Appendix A.

5. Machine-Learning Clustering of Digital-Territorial Regimes

5.1. Data, algorithms and validation strategy

The panel estimates identify average effects but cannot reveal whether countries form distinct digital–territorial regimes. We therefore cluster 123 countries on 2015–2023 averages of nine standardised indicators covering digital adoption (Internet users, mobile subscriptions), infrastructure (fixed broadband), gender inclusion (female account ownership), territorial structure (rural population, rural electricity access), development (ln GDP per capita), education expenditure and the Gini index.

Six algorithms are compared: K-Means (MacQueen, 1967), hierarchical clustering with Ward linkage (Ward, 1963), Fuzzy C-Means (Bezdek, 1981), a model-based Gaussian mixture (Fraley and Raftery, 2002), density-

based DBSCAN (Ester et al., 1996) and random-forest proximity clustering (Breiman, 2001; Shi and Horvath, 2006). Each partition is evaluated with eleven indicators: R^2, AIC, BIC, silhouette (Rousseeuw, 1987), maximum diameter, minimum separation, Pearson γ (Hennig and Liao, 2013), Dunn (Dunn, 1974), entropy, Calinski–Harabasz (Caliński and Harabasz, 1974) and the Herfindahl–Hirschman index of cluster sizes (Supplementary Table B1). The best algorithm is the one with the lowest mean rank. The comparison is made at k = 4, where the K-Means silhouette (0.306) exceeds its k = 3 value and the relative BIC of the Gaussian mixture flattens (Supplementary Figure B1); DBSCAN selects its own partition (two clusters and 26 noise points). Table 6 reports the results.

Table 6. Comparison of six clustering algorithms on eleven validity indicators (k = 4)

Algorithm	k	R^2	AIC	BIC	Silh.	Max diam.	Min sep.	Pearson γ	Dunn	Entropy	CH	HHI	Mean rank	First places
K-Means	4	0.648	213.7	314.9	0.306	4.889	0.798	0.602	0.163	1.286	73.1	0.291	2.00	6
Hierarchical (Ward)	4	0.621	223.1	324.3	0.303	5.442	1.006	0.640	0.185	1.187	64.9	0.340	2.73	2
Fuzzy C-Means	4	0.629	220.2	321.4	0.218	6.480	0.891	0.530	0.137	1.376	67.3	0.255	3.09	2
Model-based (GMM)	4	0.616	224.5	325.8	0.231	6.210	0.574	0.539	0.092	1.363	63.6	0.261	4.09	0
Density-based (DBSCAN)	2 + noise	0.474	245.3	321.2	0.285	8.936	1.129	0.624	0.126	0.948	54.1	0.442	4.09	1
Random Forest (proximity)	4	0.479	262.0	363.3	0.177	7.108	1.037	0.565	0.146	0.885	36.5	0.454	5.00	0

Notes: Indicators computed on standardised data (123 countries, 9 variables). For DBSCAN, noise points form a separate group. Mean rank is the average rank across the eleven indicators (1 = best); first places count the indicators on which an algorithm ranks best. Ranks are computed among the six algorithms, so they differ slightly from the five-algorithm ranks in Supplementary Table B3. Lower is better for AIC, BIC, maximum diameter and HHI; higher is better for the others.

Table 6 shows that K-Means dominates the comparison, with a mean rank of 2.00 and first place on six of the eleven indicators, including R^2 (0.648), silhouette (0.306), Calinski–Harabasz, AIC and BIC. Ward clustering ranks second (2.73), whereas DBSCAN leaves 26 countries unclassified and random-forest proximity clustering ranks last. Figure 4 shows that this ranking is not an artefact of the chosen k.

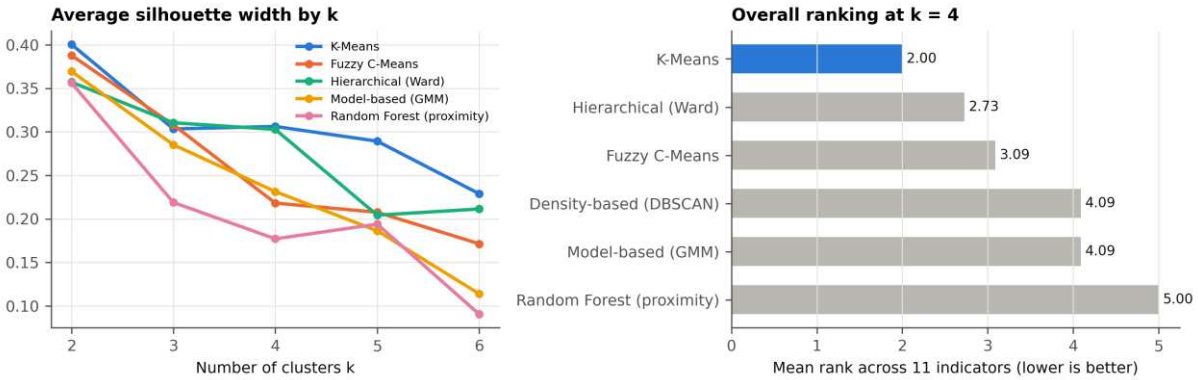


Figure 4. Selection of the number of clusters and overall ranking of the algorithms

Note: Left panel, average silhouette width for k = 2–6 (DBSCAN excluded, since it selects k internally). Right panel, mean rank across the eleven indicators at k = 4.

K-Means achieves the highest or near-highest silhouette at every k and the best or tied-best mean rank for every k from 2 to 6 (Supplementary Table B3), so it is used with k = 4 for the final clustering, whose profiles are reported in Table 7.

5.2. Digital-territorial regimes

Table 7. Cluster profiles: mean values of the clustering variables, K-Means (k = 4)

Variable (2015–2023 average)	C1 Digitally advanced	C2 Digital catching-up	C3 Unequal digital transition	C4 Rural digital periphery
Internet users (%)	86.69	63.62	55.82	21.46
Fixed broadband (per 100)	34.19	12.51	6.97	0.63
Mobile subs. (per 100)	126.03	121.38	118.20	79.15
Female account ownership (%)	92.78	57.45	60.50	34.19
Rural electricity access (%)	99.88	94.88	70.99	31.80
Rural population (%)	22.33	38.52	39.22	63.50
ln GDP per capita	10.81	9.61	9.34	8.05
Education exp. (% GDP)	5.01	3.92	6.66	3.14
Gini index	31.96	36.47	51.30	38.72
Countries (n)	41	40	10	32

Note: Unstandardised means of 2015–2023 averages. Female account ownership refers to accounts at a financial institution or with a mobile-money provider (% of women aged 15+).

Table 7 reveals four regimes. C1, "digitally advanced" (41 countries, mostly high-income European), combines near-universal Internet use (86.7%), dense broadband, complete rural electrification and female financial inclusion, and the lowest inequality (Gini 32.0). C2, "digital catching-up" (40 countries, mainly upper-middle-income), has intermediate adoption (63.6%). C3, "unequal digital transition" (10 countries in Latin America and Southern Africa), reaches similar adoption but extreme inequality (Gini 51.3) despite the highest education spending. C4, "rural digital periphery" (32 countries, 27 in Sub-Saharan Africa), combines a mostly rural population (63.5%) with low Internet use (21.5%), negligible broadband, limited rural electrification (31.8%) and female account ownership of only 34.2%. Figure 5 projects these regimes onto the first two principal components.

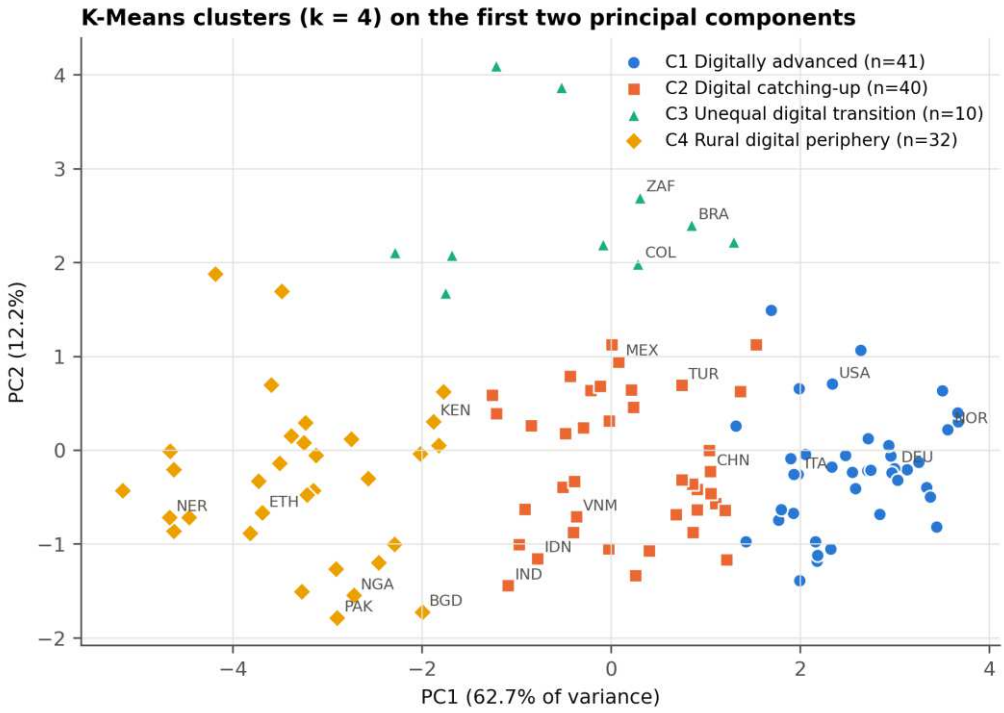


Figure 5. K-Means clusters on the first two principal components

Note: Principal components of the nine standardised variables. Selected countries are labelled with ISO3 codes.

Figure 5 shows that the first component (62.7% of the variance) is a digital-development gradient, while the second (12.2%), driven by the Gini index and education spending, isolates C3 as a distinct high-inequality branch. Figure 6 makes this structure explicit through standardised centroids.

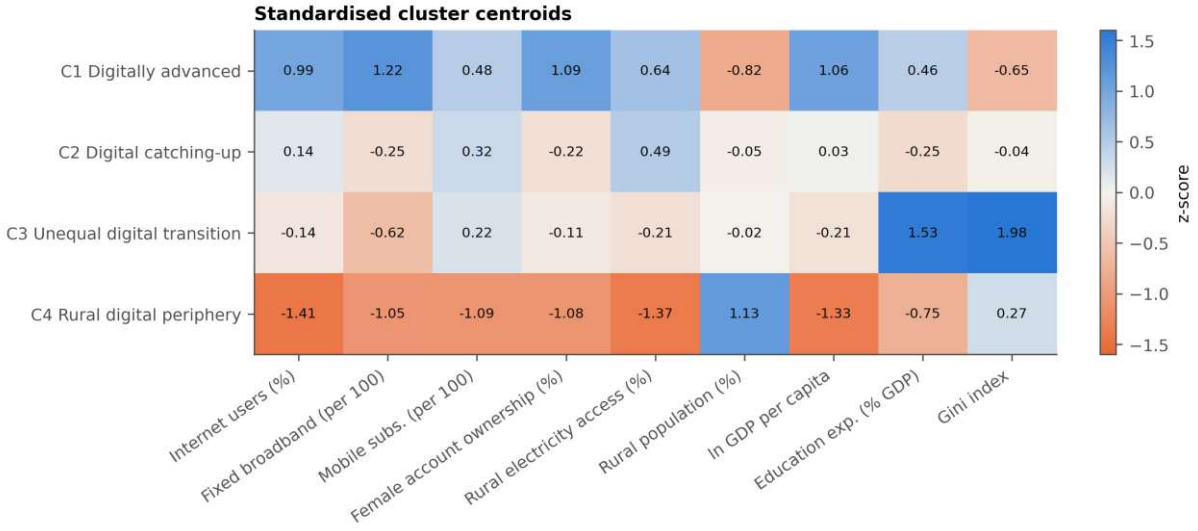


Figure 6. Standardised cluster centroids

Note: Cluster means of z-scores; blue indicates values above and orange values below the cross-country mean.

The centroids in Figure 6 show that C4 lies more than one standard deviation below the mean on every digital, electrification and gender-inclusion indicator, whereas C3 departs from it mainly in inequality and education expenditure. Table 8 re-estimates the FE model of Section 4 within each cluster.

Table 8. Two-way fixed-effects estimates of the Gini index by cluster

Variable	C1 Digitally advanced	C2 Digital catching-up	C3 Unequal digital transition	C4 Rural digital periphery
Internet users (%)	-0.033	-0.037	-0.188***	-0.018
Fixed broadband (per 100)	0.037	0.075	0.409***	0.446
Mobile subscriptions (per 100)	-0.010	-0.028	0.043***	-0.066*
Rural population (%)	-0.067	-0.060	-0.151	0.034
ln GDP per capita	-1.455	3.608	-14.763***	7.413*
Education expenditure (% GDP)	-0.532	-0.710*	-2.052***	0.589
Observations / countries	271 / 41	219 / 40	47 / 10	114 / 32
R^2 (net of country and period effects)	0.091	0.104	0.476	0.171

Notes: Country and period effects; cluster-robust standard errors by country (reported in Supplementary Table B8). Estimates for C3 rely on 10 countries and should be interpreted with caution. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8 shows that the unequalising effect of broadband found in Section 4 is concentrated outside the advanced regime. Its coefficient is close to zero and insignificant in C1 (0.037) and C2 (0.075), but large in C3 (0.409, $p < 0.01$) and of similar size, though imprecise, in C4 (0.446). In C3, Internet use is strongly equalising (-0.188, $p < 0.01$). Broadband therefore widens inequality where it reaches a narrow, mostly urban elite, while broad-based Internet adoption compresses it.

5.3. Regime dynamics and external validation

To examine mobility between regimes, each country is assigned to the nearest 2015–2023 centroid in 2003–2008, 2009–2014 and 2015–2023. First-order Markov transitions (Quah, 1993) show that the regimes are highly persistent, with probabilities of remaining in the same regime between 0.82 and 0.98 (Supplementary Table B9). Only 6 of the 60 transitions observed for periphery countries are exits, and the countries that left started with much higher rural electrification (59.5% against 17.7%, $p = 0.001$) and expanded mobile access faster ($p = 0.004$), whereas initial Internet use did not differ (Supplementary Table B10).

5.4. Discussion

The clustering confirms that digital transformation produces distinct and persistent territorial regimes rather than a single gradient. The rural digital periphery is also a gender-excluded periphery (Mariscal et al., 2019), and bootstrap stability is high for C1 and C4 (Jaccard 0.89) but lower for C3 (0.53). Further indices, stability analyses and memberships are reported in Supplementary Appendix B.

6. Machine-Learning Regression as a Validation of the Econometric Model

6.1. Design

This section uses machine learning to validate, not to forecast: we ask whether a flexible algorithm, free from the linearity and additivity assumptions of Equation (1), recovers the panel relationships. The 712 observations of Section 4 are used with the Gini index as target and the regressors of Equation (1) as features; the interaction term is omitted, so any Internet × rurality interaction must be learned from the data.

Six algorithms are compared: Boosting (gradient-boosted trees), Decision Tree, K-Nearest Neighbours (KNN), Linear Regression, Random Forest (Breiman, 2001) and Linear Support Vector Machine. Performance is measured by nested cross-validation: the outer five-fold split is grouped by country, so that no country appears in both training and test folds, and an inner grouped search tunes the hyperparameters (Supplementary Table C1). Each model is evaluated on seven out-of-sample metrics (R^2, RMSE, MSE, MAE, MAPE, explained variance and maximum error), the ranks are averaged, and the gap between training and cross-validated R^2 is reported as an overfitting diagnostic. Table 9 reports the results.

Table 9. Out-of-sample performance of six regression algorithms (target: Gini index, levels)

Model	R^2 (CV)	RMSE	MSE	MAE	MAPE (%)	Expl. var.	Max error	R^2 (train)	Overfit gap	Mean rank	First places
Random Forest	0.277	6.888	47.44	5.260	14.48	0.278	25.47	0.699	0.422	1.14	6
KNN	0.252	7.005	49.07	5.303	14.63	0.255	24.68	0.780	0.528	1.86	1
Boosting (GBM)	0.244	7.042	49.60	5.380	14.79	0.245	26.93	0.565	0.321	4.00	0
Linear Regression	0.250	7.013	49.19	5.477	14.99	0.251	27.30	0.290	0.039	4.14	0
Linear SVM	0.230	7.108	50.53	5.477	14.65	0.234	28.59	0.264	0.034	4.86	0
Decision Tree	0.225	7.133	50.88	5.393	14.71	0.225	26.69	0.369	0.145	5.00	0

Notes: Nested cross-validation; outer 5-fold and inner 4-fold splits grouped by country (712 observations, 155 countries). Metrics refer to out-of-fold predictions; R^2 (train) is the mean in-sample R^2 of the outer-fold models. Mean rank and first places are computed over the seven out-of-sample metrics (gap excluded); ties are broken on unrounded values.

Table 9 shows that Random Forest is the best algorithm, ranking first on six of the seven metrics (mean rank 1.14), with the highest cross-validated R^2 (0.277) and the lowest RMSE and MAE. KNN ranks second but overfits more heavily, and Linear Regression (R^2 = 0.250) is close to the tree ensembles, which suggests that most of the predictable structure is linear. Figure 7 extends the comparison to the within-country dimension.

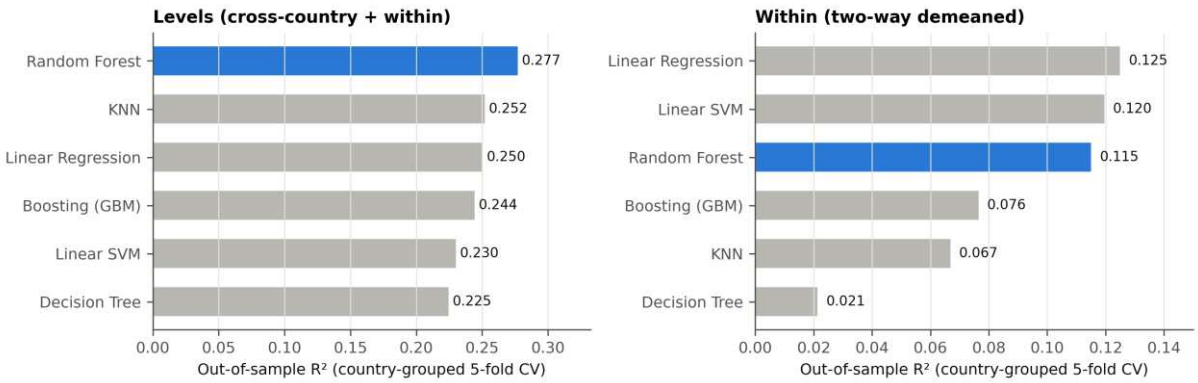


Figure 7. Cross-validated R^2 of the six algorithms in levels and on within-transformed data

Note: The within data are deviations from country and period means (two-way within transformation), mirroring the FE estimator. Country-grouped nested cross-validation in both panels.

The right panel of Figure 7 repeats the exercise after removing country and period means, which reproduces the identifying variation of the FE estimator. Linear Regression (R^2 = 0.125) and Linear SVM (0.120) then slightly outperform Random Forest (0.115) (Supplementary Table C2): within countries, a linear specification is adequate. Random Forest, the best overall algorithm, is retained and decomposed with SHAP values (Lundberg and Lee, 2017), shown in Figure 8.

6.2. Validation results

The validation proceeds in three steps: feature importance, comparison of effect sizes and signs, and the shape of the relationships.

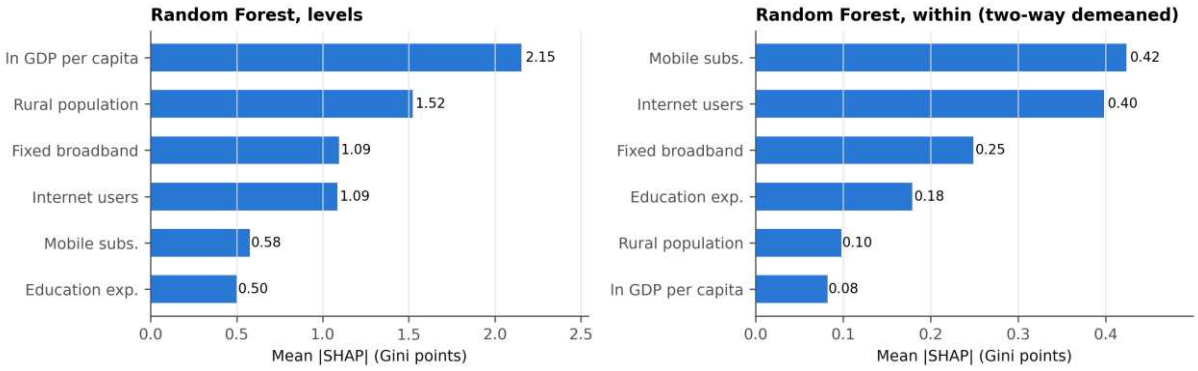


Figure 8. Feature importance of the Random Forest: mean absolute SHAP values

Note: Mean absolute contribution of each feature to the predicted Gini index, in Gini points. Permutation importances are reported in Supplementary Table C3.

Figure 8 shows that, in levels, the forest relies mainly on ln GDP per capita and rurality, reflecting persistent cross-country differences, whereas on within data mobile subscriptions, Internet use and broadband dominate and rurality and income become marginal, as in the FE model. Table 10 compares the linear slope of each SHAP profile with the corresponding econometric coefficient.

Table 10. Validation of the econometric estimates: panel coefficients versus Random-Forest SHAP slopes

Variable	Pooled OLS	RF-SHAP slope, levels [95% CI]	Same sign	Fixed effects	RF-SHAP slope, within [95% CI]	Same sign
Internet users (%)	-0.130***	-0.037 [-0.048; -0.020]	Yes	-0.063***	-0.040 [-0.055; -0.024]	Yes
Fixed broadband (per 100)	-0.182***	-0.082 [-0.121; -0.050]	Yes	0.095**	0.052 [0.015; 0.081]	Yes
Mobile subscriptions (per 100)	-0.014	-0.007 [-0.017; 0.003]	Yes	-0.028***	-0.024 [-0.033; -0.015]	Yes
Rural population (%)	-0.185***	-0.073 [-0.103; -0.036]	Yes	-0.070	0.003 [-0.082; 0.088]	No
ln GDP per capita	-0.359	-1.628 [-2.054; -0.814]	Yes	2.145	0.494 [-1.482; 1.893]	Yes
Education expenditure (% GDP)	-0.191	-0.131 [-0.639; 0.331]	Yes	-0.348	-0.312 [-0.575; -0.023]	Yes

Notes: The SHAP slope is the OLS slope of a feature's SHAP values on the feature, i.e. the average marginal effect implied by the Random Forest (Gini points per unit). 95% confidence intervals from 100 country-cluster bootstrap replications. Econometric coefficients from Table 3; *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 10 delivers the central validation result. In the within dimension, the Random Forest reproduces the sign of every significant FE coefficient, with bootstrap intervals that exclude zero: Internet use is equalising (−0.040; FE −0.063), broadband unequalising (0.052; FE 0.095) and mobile subscriptions equalising (−0.024; FE −0.028). In levels, broadband (−0.082) is negatively associated with inequality, so the forest independently reproduces the sign reversal of broadband between the two dimensions. Figure 9 shows the shape of the within relationships.

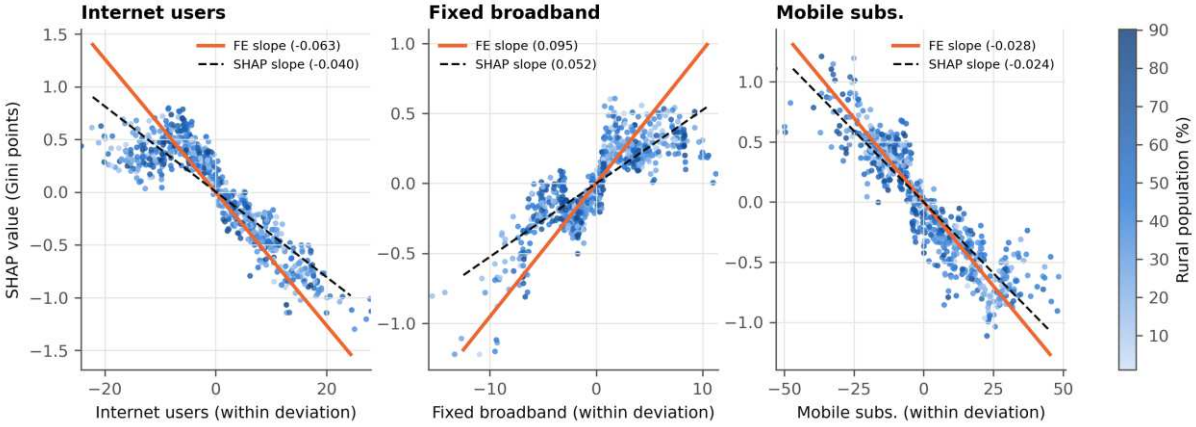


Figure 9. SHAP dependence of the digital variables in the within Random Forest, coloured by rurality

Note: Each point is a country-period observation. The solid line is the FE slope from Table 3; the dashed line is the SHAP slope. Colour intensity indicates the rural population share.

The profiles in Figure 9 are approximately linear and monotonic, and observations with high and low rurality overlap. SHAP interaction values (Lundberg et al., 2020) confirm that the Internet × rurality interaction accounts for only 5.0% of pairwise interaction strength, below the 6.7% expected under even spread (Supplementary Table C4).

6.3. Out-of-sample and causal machine-learning tests

Three further tests address the concern that this validation is circular (Supplementary Appendix C.3). In a temporal hold-out that trains on 2000–2017 and predicts 2018–2023, the fixed-effects model has the lowest error (RMSE 3.20), statistically indistinguishable from the within random forest (Diebold–Mariano p = 0.703), and removing the digital variables significantly worsens the forest's forecasts (p = 0.005). Double machine learning with country-grouped cross-fitting (Chernozhukov et al., 2018) yields −0.065 for Internet use and 0.081 for broadband; under region × period demeaning the Internet effect remains significant (−0.046, p = 0.001) and the broadband effect disappears (p = 0.46). A causal forest (Wager & Athey, 2018) estimates an average Internet effect of −0.060 (95% CI −0.116 to −0.004), similar across rurality terciles.

6.4. Discussion

The machine-learning evidence thus validates the econometric model on four counts. Linear models are as accurate as flexible ones within countries; the Random Forest independently recovers the signs of the significant FE coefficients and the sign reversal of broadband; it finds no meaningful Internet × rurality interaction; and out-of-sample and causal machine-learning tests reproduce the equalising effect of Internet use, including under region × period effects, but not a robust within-country effect of broadband. Tuning details and additional diagnostics are reported in Supplementary Appendix C.

7. Discussion

The three methods rely on different assumptions, and their results converge (Supplementary Table D1). Panel estimators control for heterogeneity and endogeneity but impose linearity, clustering is descriptive, and machine learning relaxes functional form without identifying causal effects, so a finding that survives all three is unlikely to be an artefact of one technique.

The most consistent result is the equalising effect of individual digital adoption, which survives eight panel estimators, two independent instruments, region × period effects, alternative inequality measures, spatial models, SHAP profiles and causal machine learning. This corroborates evidence that Internet and mobile diffusion compress income gaps (Canh et al., 2020; Lee et al., 2026), while the absence of robust thresholds qualifies claims of U-shaped trajectories (Baffour Gyau et al., 2025).

Broadband calls for a more cautious reading. All three methods reproduce its sign reversal, which would reconcile the contrasting findings of the literature (Houngbonon & Liang, 2021; Cao et al., 2025; He et al., 2025), but the within-country association disappears under region × period effects. The unequalising face of broadband is therefore a regional pattern, concentrated in the unequal digital transition regime, consistent with evidence that connectivity alone does not reduce inequality unless access is broadened and complemented by skills (Au, 2024; Chen et al., 2025).

The territorial hypothesis receives a nuanced answer. No method detects weaker returns to adoption in rural economies, but rurality defines a rural digital periphery where low connectivity, incomplete electrification and limited female financial inclusion overlap. Territorial divides are therefore better understood as cumulative exclusion than as weaker returns to adoption (Myovella et al., 2021; Rubiera-Morollón et al., 2026), and the persistence of the regimes implies that convergence will not occur spontaneously (Śledziewska & Łebkowska, 2026).

Read against the model of Section 2.8, the evidence supports H2 in full and H1 for adoption, but only partly H1 for infrastructure and not H3. The within-country broadband effect that the model attributes to skill complementarity disappears under region × period effects, and the human-capital interactions predicted by H3 are absent, whereas the Internet effect appears stronger in service-intensive economies. This suggests that the complementarity parameter β depends less on the national distribution of education than on where, and for whom, infrastructure is deployed, and on the structure of the labour market into which technologies diffuse. The model should therefore be extended by making the location of infrastructure deployment endogenous.

Methodologically, the triangulation shows that machine learning can strengthen econometric inference when it is used for validation with country-grouped cross-validation (Hoepner et al., 2021; Mulenga et al., 2026).

8. Policy Implications

The results carry three main implications. First, policies should prioritise broad-based adoption over infrastructure alone: infrastructure subsidies should be conditional on demand-side measures such as affordable tariffs, devices and digital skills (Chen et al., 2025; Au, 2024), and universal-service programmes should be evaluated on take-up among low-income, rural and female users, not only on coverage.

Second, policies should be regime-specific. In the rural digital periphery, electrification and mobile access should precede high-speed fixed networks (Myovella et al., 2021), whereas in the unequal digital transition regime digital investment should be paired with redistributive and labour-market policies (Cao et al., 2025).

Third, digital policies need an explicit gender lens, because mobile money can reduce women's financial exclusion (Kim, 2022), while digital credit may widen gender gaps without financial literacy (Johnen & Mußhoff, 2023).

To translate these results into magnitudes, Table 11 simulates scenarios in which each regime reaches the average digital profile of a more advanced regime, using the 2015–2023 cluster means and three sets of estimates: FE, FE with region × period effects and FE-2SLS with the telephony instrument (IV-C). Confidence intervals are obtained from 20,000 draws of the estimated coefficient vectors. Because the broadband coefficient is not robust to regional shocks, its contribution is reported separately and should be read as an upper bound.

Table 11. Simulated changes in the Gini index under regime catch-up scenarios

Scenario	Δ Internet / Δ broadband / Δ mobile	Baseline Gini	Adoption, FE	Adoption, FE + R×P	Adoption, IV-C	Broadband, FE
C4 → C2 levels	+42.2 / +11.9 / +42.2	38.7	−3.92 [−5.75; −2.11]	−3.78 [−5.56; −1.98]	−4.99 [−7.17; −2.87]	1.13 [0.16; 2.11]
C4 → C1 levels	+65.2 / +33.6 / +46.9	38.7	−5.54 [−8.34; −2.79]	−5.44 [−8.11; −2.77]	−7.39 [−10.72; −4.06]	3.20 [0.43; 5.91]
C2 → C1 levels	+23.1 / +21.7 / +4.6	36.5	−1.58 [−2.30; −0.89]	−1.06 [−1.71; −0.41]	−2.11 [−3.65; −0.57]	2.06 [0.28; 3.85]
C3 → C1 levels	+30.9 / +27.2 / +7.8	51.3	−2.17 [−3.10; −1.23]	−1.47 [−2.34; −0.62]	−2.86 [−4.87; −0.82]	2.59 [0.33; 4.82]

Notes: Each scenario raises Internet use, fixed broadband and mobile subscriptions from the 2015–2023 mean of the initial regime to that of the target regime (changes in percentage points or subscriptions per 100 people). Adoption = Internet use, including its interaction with the initial regime's rural share, plus mobile subscriptions; Broadband = broadband coefficient times the change in broadband. FE and FE + R×P from Table 5 (Panels A and B), IV-C from Table 5. Gini points; 95% intervals from 20,000 draws of the estimated coefficient vectors in brackets.

Closing the adoption gap between the rural digital periphery and the catching-up regime, which implies raising Internet use by 42 percentage points and mobile subscriptions by 42 per 100 people, would lower the Gini index by 3.9 points according to the FE model (95% CI 2.1 to 5.8), by 3.8 points under region × period effects and by 5.0 points with the external instrument, about one-tenth of the periphery's average Gini of 38.7. If, however, the accompanying broadband expansion had the within-country association estimated by the FE model, it would offset between about one-quarter and more than all of these gains. Figure 10 summarises the two channels across specifications.

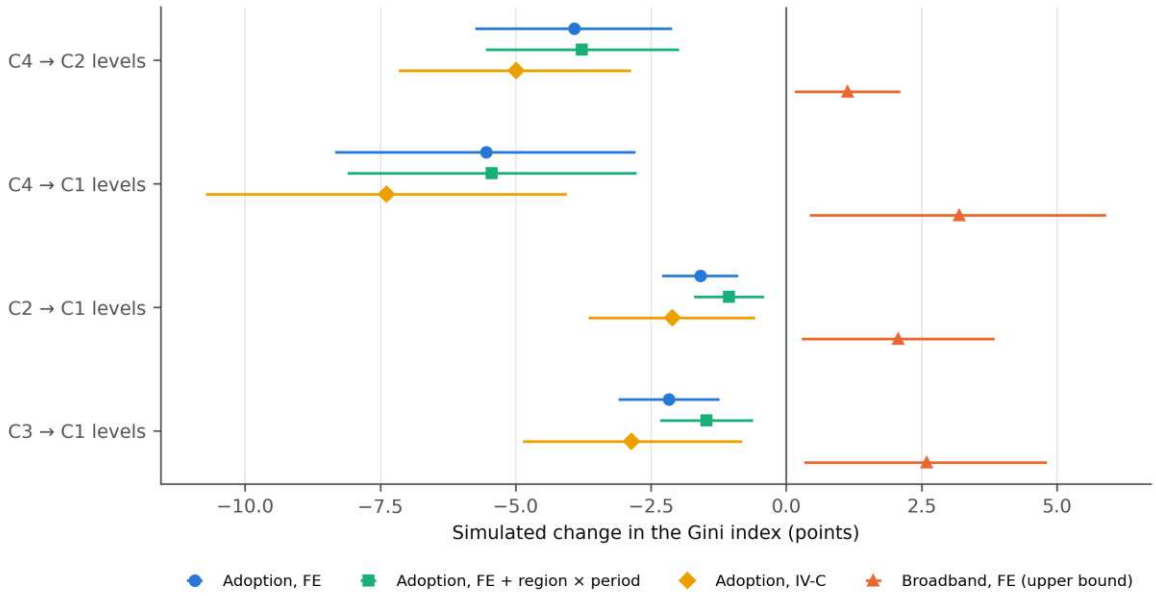


Figure 10. Simulated changes in the Gini index by scenario, channel and specification

Note: Point estimates with 95% intervals from Table 11. The broadband channel is shown for the FE estimate only and represents an upper bound, since the broadband coefficient is close to zero once region × period effects are included.

Figure 10 shows that the adoption gains are robust across specifications, whereas the broadband channel is sizeable in the FE estimates but vanishes under region × period effects. A complementary simulation of a broadband expansion without complementary adoption (Supplementary Table D2) shows that its distributional cost is concentrated in the unequal digital transition regime (4.1 Gini points, 95% CI 1.8 to 6.4). These magnitudes are partial-equilibrium extrapolations that indicate orders of magnitude rather than forecasts.

Finally, governments and international organisations should invest in subnational and gender-disaggregated digital indicators, without which territorial divides remain invisible to policy (Breen et al., 2025; Arora & Sapre, 2026).

9. Limitations and Future Research

The findings have several limitations. The Gini index combines income- and consumption-based surveys and is unevenly available, and the sample over-represents middle- and high-income countries; digital transformation is measured by access-oriented indicators; the analysis is national, and the European case study is imprecise; the shift-share instrument fails a placebo test for pre-existing trends, so geographic instruments and quasi-experimental evaluations are needed, particularly for broadband (Liu, 2017); and SHAP values and causal machine learning rely on selection on observables. An extended discussion is provided in Supplementary Appendix D.

10. Conclusions

This article examined whether digital transformation narrows or widens income inequality and whether its effects depend on territorial structure, combining a stylised model with panel econometrics, clustering and machine-learning validation for 155 countries over 2000–2023.

Three conclusions emerge. First, digital transformation has two distributional faces that differ in robustness. Internet adoption reduces income inequality, by about 0.6 Gini points per ten percentage points within

countries, and this result survives external instruments, regional shocks, alternative measures, spatial models and machine-learning validation. Fixed broadband is associated with lower inequality across countries and higher inequality within them, but the within-country effect disappears under region × period effects and should be read as a regional pattern. The evidence supports the opposite effects of adoption and the between–within reversal of broadband, but not skill-based amplification.

Second, territorial inequality operates through cumulative exclusion rather than through weaker returns to adoption. Four persistent digital–territorial regimes emerge, including a rural digital periphery where weak connectivity, incomplete electrification and low female financial inclusion overlap, and from which the rare exits follow electrification and mobile expansion. Closing its adoption gap could lower its Gini index by about 4 points.

Third, machine learning can strengthen econometric inference when it is used for validation rather than prediction. With country-grouped cross-validation, a random forest reproduces the fixed-effects signs and shows that within-country relationships are essentially linear, while temporal hold-out tests and causal machine learning confirm the equalising effect of Internet use. The design offers a replicable template in which estimators, clustering algorithms and learners are compared rather than chosen a priori.

Investing in networks is not sufficient: the distributional effect of digitalisation depends on who adopts, where and with which skills. Policies that combine infrastructure with affordability, digital literacy, electrification and gender-sensitive financial inclusion are more likely to make digital transformation a driver of cohesion rather than a new source of exclusion.

References

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter & D. Card (Eds.), Handbook of labor economics (Vol. 4B, pp. 1043–1171). Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5

Amber, H., & Chichaibelu, B. B. (2023). Narrowing the gender digital divide in Pakistan: Mobile phone ownership and female labor force participation. Review of Development Economics, 27(3), 1354–1382. https://doi.org/10.1111/rode.12994

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics, 11, 727–753. https://doi.org/10.1146/annurev-economics-080218-025643

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies, 58(2), 277–297. https://doi.org/10.2307/2297968

Arora, R., & Sapre, N. (2026). Rural–urban digital divide: Evidence from Indian states. International Journal of Finance and Economics, 31(2), 2912–2932. https://doi.org/10.1002/ijfe.70045

Arora, R., Garikipati, S., & Kaur, S. (2026). Community norms, peer influence, and women's digital financial inclusion: Evidence from India. World Development, 201, Article 107314. https://doi.org/10.1016/j.worlddev.2026.107314

Asongu, S. A., & Odhiambo, N. M. (2019). How enhancing information and communication technology has affected inequality in Africa for sustainable development: An empirical investigation. Sustainable Development, 27(4), 647–656. https://doi.org/10.1002/sd.1929

Asongu, S. A., & Odhiambo, N. M. (2020). Inequality and gender inclusion: Minimum ICT policy thresholds for promoting female employment in sub-Saharan Africa. Telecommunications Policy, 44(4), Article 101900. https://doi.org/10.1016/j.telpol.2019.101900
Au, A. (2024). How do different forms of digitalization affect income inequality? Technological and Economic Development of Economy, 30(3), 667–687. https://doi.org/10.3846/tede.2024.20562
Baffour Gyau, E., Li, Y., & Appiah, M. (2025). Global digital transformation: Discovering the impact of digitalization on income inequality in OECD countries, the moderating role of globalization. Economic Change and Restructuring, 58(1), Article 6. https://doi.org/10.1007/s10644-024-09843-2
Bauer, J. M. (2018). The internet and income inequality: Socio-economic challenges in a hyperconnected society. Telecommunications Policy, 42(4), 333–343. https://doi.org/10.1016/j.telpol.2017.05.009
Baylis, K., Heckelei, T., & Storm, H. (2021). Machine learning in agricultural economics. In Handbook of Agricultural Economics (Vol. 5, pp. 4551–4612). Elsevier. https://doi.org/10.1016/bs.hesagr.2021.10.007
Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press. https://doi.org/10.1007/978-1-4757-0450-1
Bhandari, A., & Shirazi, S. (2022). Economic dependency and rural exclusion: Explaining the global digital divide, 1995 to 2014. Sociology of Development, 8(3), 239–271. https://doi.org/10.1525/sod.2021.0011
Bilozubenko, V., Korneyev, M., Ivantsov, S., Yakovenko, V., Zhuravka, F., Shchytov, D., & Nahorianskyi, M. (2025). Clustering EU countries by the level of information and communication technologies development. Problems and Perspectives in Management, 23(2), 862–874. https://doi.org/10.21511/ppm.23(2).2025.63
Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073–1076. https://doi.org/10.1126/science.aac4420
Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1), 115–143. https://doi.org/10.1016/S0304-4076(98)00009-8
Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-experimental shift-share research designs. The Review of Economic Studies, 89(1), 181–213. https://doi.org/10.1093/restud/rdab030
Breen, C. F., Fatehkia, M., Yan, J., Zhao, X., Leasure, D. R., Weber, I., & Kashyap, R. (2025). Mapping subnational gender gaps in internet and mobile adoption using social media data. Proceedings of the National Academy of Sciences of the United States of America, 122(42), Article e2416624122. https://doi.org/10.1073/pnas.2416624122
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1–27. https://doi.org/10.1080/03610927408827101
Canh, N. P., Schinckus, C., Thanh, S. D., & Hui Ling, F. C. (2020). Effects of the internet, mobile, and land phones on income inequality and the Kuznets curve: Cross country analysis. Telecommunications Policy, 44(10), Article 102041. https://doi.org/10.1016/j.telpol.2020.102041
Cao, Q., Ni, P., & Guo, J. (2025). Broadband internet and income inequality among the floating population: Evidence from the “Broadband China” strategy in China. China Economic Review, 90, Article 102369. https://doi.org/10.1016/j.chieco.2025.102369
Chen, J. M. (2021). An introduction to machine learning for panel data. International Advances in Economic Research, 27(1). https://doi.org/10.1007/s11294-021-09815-6
Chen, M., Sun, Q., & Tang, L. (2025). Broadband infrastructure construction and consumption inequality. China and World Economy, 33(5), 49–77. https://doi.org/10.1111/cwe.12604
Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68. https://doi.org/10.1111/ectj.12097
Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. The Review of Economics and Statistics, 94(1), 260–272. https://doi.org/10.1162/REST_a_00139

Czernich, N., Falck, O., Kretschmer, T., & Woessmann, L. (2011). Broadband infrastructure and economic growth. The Economic Journal, 121(552), 505–532. https://doi.org/10.1111/j.1468-0297.2011.02420.x
Deng, X., Guo, M., & Liu, Y. (2023). Digital economy development and the urban-rural income gap: Evidence from Chinese cities. PLoS ONE, 18(2), Article e0280225. https://doi.org/10.1371/journal.pone.0280225
Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics, 4(1), 95–104. https://doi.org/10.1080/01969727408546059
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96) (pp. 226–231). AAAI Press.
Eurostat. (2026). Eurostat database: Income and living conditions, digital economy and society, and regional statistics. European Commission. https://ec.europa.eu/eurostat/data/database
Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. American Economic Review, 90(4), 869–887. https://doi.org/10.1257/aer.90.4.869
Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association, 97(458), 611–631. https://doi.org/10.1198/016214502760047131
Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586–2624. https://doi.org/10.1257/aer.20181047
Gomes, A., & Dias, J. G. (2025). Digital divide in the European Union: A typology of EU citizens. Social Indicators Research, 176(1), 149–172. https://doi.org/10.1007/s11205-024-03452-2
Halleck Vega, S., & Elhorst, J. P. (2015). The SLX model. Journal of Regional Science, 55(3), 339–363. https://doi.org/10.1111/jors.12188
Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics, 93(2), 345–368. https://doi.org/10.1016/S0304-4076(99)00025-1
Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251–1271. https://doi.org/10.2307/1913827
He, B., Nan, G., Xu, D., & Sun, J. (2025). Bridging or widening? The impact of the Broadband China policy on urban-rural income inequality. Humanities and Social Sciences Communications, 12(1), Article 555. https://doi.org/10.1057/s41599-025-04875-z
Hennig, C., & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. Journal of the Royal Statistical Society: Series C (Applied Statistics), 62(3), 309–369. https://doi.org/10.1111/j.1467-9876.2012.01066.x
Herrera, G. P., Constantino, M., Su, J.-J., & Naranpanawa, A. (2023). The use of ICTs and income distribution in Brazil: A machine learning explanation using SHAP values. Telecommunications Policy, 47(8), Article 102598. https://doi.org/10.1016/j.telpol.2023.102598
Hilbert, M. (2016). The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014. Telecommunications Policy, 40(6), 567–581. https://doi.org/10.1016/j.telpol.2016.01.006
Hoepner, A. G. F., McMillan, D., Vivian, A., & Wese Simen, C. (2021). Significance, relevance and explainability in the machine learning age: An econometrics and financial data science perspective. European Journal of Finance, 27(1-2), 1–7. https://doi.org/10.1080/1351847X.2020.1847725
Houngbonon, G. V., & Liang, J. (2021). Broadband internet and income inequality. Review of Network Economics, 20(2), 55–99. https://doi.org/10.1515/rne-2020-0042
Hua, Y., & Zhang, H. (2023). Internet penetration and income inequality: Evidence from the Chinese young labor market. Applied Economics, 55(54), 6444–6458. https://doi.org/10.1080/00036846.2022.2156471
Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15(4), 309–334. https://doi.org/10.1037/a0020761

Johnen, C., & Mußhoff, O. (2023). Digital credit and the gender gap in financial inclusion: Empirical evidence from Kenya. Journal of International Development, 35(2), 272–295. https://doi.org/10.1002/jid.3687
Kartiasih, F., Djalal Nachrowi, N., Wisana, I. D. G. K., & Handayani, D. (2023). Inequalities of Indonesia's regional digital development and its association with socioeconomic characteristics: A spatial and multivariate analysis. Information Technology for Development, 29(2-3), 299–328. https://doi.org/10.1080/02681102.2022.2110556
Kim, K. (2022). Assessing the impact of mobile money on improving the financial inclusion of Nairobi women. Journal of Gender Studies, 31(3), 306–322. https://doi.org/10.1080/09589236.2021.1884536
L'Abbate, S., Perchinunno, P., & Massari, A. (2026). Educational and digital territorial inequalities in Italy: A density-based clustering approach across degree of urbanization (DEGURBA) areas. Socio-Economic Planning Sciences, 107, Article 102581. https://doi.org/10.1016/j.seps.2026.102581
Lee, M.-H., Yu, C., & Kao, C.-H. (2026). ICT development and income inequality: The moderating role of the entrepreneurial environment. Telecommunications Policy, 50(1), Article 103084. https://doi.org/10.1016/j.telpol.2025.103084
LeSage, J., & Pace, R. K. (2009). Introduction to spatial econometrics. Chapman and Hall/CRC. https://doi.org/10.1201/9781420064254
Li, X., He, P., Liao, H., Liu, J., & Chen, L. (2024). Does network infrastructure construction reduce urban–rural income inequality? Based on the "Broadband China" policy. Technological Forecasting and Social Change, 205, Article 123486. https://doi.org/10.1016/j.techfore.2024.123486
Li, Z., Lunku, H. S., Yang, S., & Ngoma, J. B. (2025). Closing the vicious poverty gap: The role of digital technology and financial progress in mitigating income inequality in sub-Saharan Africa. International Journal of Economic Policy Studies, 19(2), 299–332. https://doi.org/10.1007/s42495-025-00153-7
Liang, D., Liu, Y., Zhou, M., Zhao, L., & Li, X. (2025). Does digital infrastructure exacerbate income inequality? Evidence from the Broadband China strategy. Structural Change and Economic Dynamics, 72, 360–373. https://doi.org/10.1016/j.strueco.2024.12.002
Liu, G., Long, W., & Luo, X. (2025). A random forest–based panel data approach for program evaluation. Journal of Applied Econometrics, 40(6), 591–607. https://doi.org/10.1002/jae.3123
Liu, S., & Yuan, J. (2025). Can government digitalization promote the urban–rural equalization of basic public services? Evidence from double machine learning. Applied Economics, 57(34), 5065–5080. https://doi.org/10.1080/00036846.2024.2364113
Liu, Y. (2017). Internet and income inequality: A research note. Economics Bulletin, 37(4), 2846–2853.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NIPS 2017) (pp. 4765–4774). Curran Associates.
Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9
Machado, J. A. F., & Santos Silva, J. M. C. (2019). Quantiles via moments. Journal of Econometrics, 213(1), 145–173. https://doi.org/10.1016/j.jeconom.2019.04.009
MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press.
Mariscal, J., Mayne, G., Aneja, U., & Sorgner, A. (2019). Bridging the gender digital gap. Economics: The Open-Access, Open-Assessment E-Journal, 13(2019-9). https://doi.org/10.5018/economics-ejournal.ja.2019-9
Morris, J., Morris, W., & Bowen, R. (2022). Implications of the digital divide on rural SME resilience. Journal of Rural Studies, 89, 369–377. https://doi.org/10.1016/j.jrurstud.2022.01.005

Mulenga, J., Simwanda, L., Pismennyi, S., Benešová, I., Heijman, W., & Tsenteradze, A. (2026). Income inequality in Southern Africa: Evidence from interpretable machine learning and panel econometric models. Socio-Economic Planning Sciences, 106, Article 102571. https://doi.org/10.1016/j.seps.2026.102571
Mundlak, Y. (1978). On the pooling of time series and cross section data. Econometrica, 46(1), 69–85. https://doi.org/10.2307/1913646
Myovella, G., Karacuka, M., & Haucap, J. (2021). Determinants of digitalization and digital divide in sub-Saharan African economies: A spatial Durbin analysis. Telecommunications Policy, 45(10), Article 102224. https://doi.org/10.1016/j.telpol.2021.102224
Nguyen, T. T., Nguyen, T.-T., & Grote, U. (2022). Internet use, natural resource extraction and poverty reduction in rural Thailand. Ecological Economics, 196, Article 107417. https://doi.org/10.1016/j.ecolecon.2022.107417
Nguyen, V. B. (2022). Does digitalization widen income inequality? A comparative assessment for advanced and developing economies. South East European Journal of Economics and Business, 17(2), 154–171. https://doi.org/10.2478/jeb-2022-0021
Olaoye, O. O., Shaddady, A., Tabash, M. I., & Ray, S. (2025). Does FinTech reduce gender asymmetry in access to finance in sub-Saharan Africa? Examining the role of digital inclusion. Journal of International Development, 37(3), 718–735. https://doi.org/10.1002/jid.3982
Quah, D. (1993). Empirical cross-section dynamics in economic growth. European Economic Review, 37(2–3), 426–434. https://doi.org/10.1016/0014-2921(93)90031-5
Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839
Pick, J., Sarkar, A., & Parrish, E. (2021). The Latin American and Caribbean digital divide: A geospatial and multivariate analysis. Information Technology for Development, 27(2), 235–262. https://doi.org/10.1080/02681102.2020.1805398
Richmond, K., & Triplett, R. E. (2018). ICT and income inequality: A cross-national perspective. International Review of Applied Economics, 32(2), 195–214. https://doi.org/10.1080/02692171.2017.1338677
Roodman, D. (2009). A note on the theme of too many instruments. Oxford Bulletin of Economics and Statistics, 71(1), 135–158. https://doi.org/10.1111/j.1468-0084.2008.00542.x
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
Rubiera-Morollón, F., Pérez-Gladish, B. M., & Fernández-García, T. (2026). Identifying the digital periphery of the European Union measuring the digitisation level of the EU NUTS 2 regions from a multiple criteria approach. Regional Science Policy and Practice, 18(6), Article 100311. https://doi.org/10.1016/j.rspp.2026.100311
Saba, C. S., & David, O. O. (2020). Convergence patterns in global ICT: Fresh insights from a club clustering algorithm. Telecommunications Policy, 44(10), Article 102010. https://doi.org/10.1016/j.telpol.2020.102010
Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. Journal of Econometrics, 190(2), 212–221. https://doi.org/10.1016/j.jeconom.2015.06.004
Shi, T., & Horvath, S. (2006). Unsupervised learning with random forest predictors. Journal of Computational and Graphical Statistics, 15(1), 118–138. https://doi.org/10.1198/106186006X94072
Simonian, J. (2020). Modular machine learning for model validation: An application to the fundamental law of active management. Journal of Financial Data Science, 2(2), 41–50. https://doi.org/10.3905/jfds.2020.1.027
Śledziewska, K., & Łebkowska, W. (2026). Digital readiness regimes and development dynamics: The role of governance and skills beyond connectivity. Technology in Society, 87, Article 103396. https://doi.org/10.1016/j.techsoc.2026.103396

Sohag, K., Badur, M. M., Ameer, W., & Vilamová, Š. (2024). Does ICT diffusion validate skill-biased technological change hypothesis? Evidence from the post-Soviet countries. Technology in Society, 76, Article 102450. https://doi.org/10.1016/j.techsoc.2023.102450
Tarafdar, P., & Swathysree, S. (2026). The intersectional impact of internet access on female income in India. Telecommunications Policy, 50(9), Article 103290. https://doi.org/10.1016/j.telpol.2026.103290
Tchamyou, V. S., Erreygers, G., & Cassimon, D. (2019). Inequality, ICT and financial access in Africa. Technological Forecasting and Social Change, 139, 169–184. https://doi.org/10.1016/j.techfore.2018.11.004
Tewathia, N., Kamath, A., & Ilavarasan, P. V. (2020). Social inequalities, fundamental inequities, and recurring of the digital divide: Insights from India. Technology in Society, 61, Article 101251. https://doi.org/10.1016/j.techsoc.2020.101251
Wang, H., Leng, H., & Yuan, M. (2025). From opportunity to inequality: How the rural digital economy shapes intra-rural income distribution. Humanities and Social Sciences Communications, 12(1), Article 534. https://doi.org/10.1057/s41599-025-04848-2
Wang, M., & Liu, J. (2025). Deciphering the digital divide: The heterogeneous and nonlinear influence of digital economy on urban-rural income inequality in China. Applied Economics, 57(33), 4861–4881. https://doi.org/10.1080/00036846.2024.2364101
Wang, S., Qu, C., & Yin, L. (2025). Digital literacy, labor migration and employment, and rural household income disparities. International Review of Economics and Finance, 99, Article 104040. https://doi.org/10.1016/j.iref.2025.104040
Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845
Wei, X., & Li, X. (2024). Internet use and rural income inequality: Evidence from China. Economic Research-Ekonomska Istrazivanja, 37(1), Article 2186914. https://doi.org/10.1080/1331677X.2023.2186914
Zhang, C., & Weng, X. (2024). Can broadband infrastructure construction promote equality of opportunity? Evidence from a quasi-natural experiment in China. Journal of Asian Economics, 93, Article 101759. https://doi.org/10.1016/j.asieco.2024.101759
Zhang, J., Gong, X., & Cheng, M. (2025). Broadband cities: Bridging urban-rural consumption gap with digital innovation. Cities, 167, Article 106315. https://doi.org/10.1016/j.cities.2025.106315
Zhang, J., Li, Y., & Fang, Z. (2025). Two digital divides and income inequality: A global perspective. Journal of Applied Economics, 28(1), Article 2568247. https://doi.org/10.1080/15140326.2025.2568247
Zhang, L., Liu, J.-K., Li, Z.-H., Yu, J.-Y., & Ding, C. J. (2025). Inclusive or fraudulent: Digital inclusive finance and urban–rural income gap. Asia-Pacific Financial Markets, 32(3), 821–854. https://doi.org/10.1007/s10690-024-09472-9

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx