Proceedings of Business and Economic Studies, 2026, Volume 9, Issue 8
http://ojs.bbwpublisher.com/index.php/PBES
Online ISSN: 2209-265X
Print ISSN: 2209-2641

Research on the Impact of Digital Inclusive Finance on Urban and Rural Residents' Income: Causal Effect Analysis Based on the Dual Machine Learning Model

Yamin Sun

Guan County Statistics Bureau, Guan County, Liaocheng 252500, Shandong, China

Abstract: Narrowing the urban-rural income gap and promoting the sustained growth of urban and rural residents' income are core issues in advancing common prosperity. Digital inclusive finance, relying on mobile internet and big data credit scoring, provides a new path to overcome the geographical constraints and entry barriers of traditional finance. Based on panel data of prefecture-level cities in China from 2011 to 2024, this paper uses a dual machine learning (DML) partially linear model to identify its causal effect on the per capita disposable income of urban and rural residents. The study found that for every 1-unit increase in the digital inclusive finance index, urban and rural residents' income increased by approximately RMB 0.274 per person after controlling for the double fixed effects (using the 2011 average as the base period). This conclusion remained robust even after changing the machine learning algorithm, adjusting the cross-fitting split ratio, introducing the time × city and province high-dimensional fixed effects, shortening the sample window, using the city clustering standard error, and adding control variables. Mechanism testing showed that entrepreneurial activity, industrial structure upgrading, and education expenditure levels constituted three significant mediating paths. In terms of heterogeneity, the income-increasing effect was stronger in the eastern, western, and southern regions, as well as in non-old industrial cities, large cities, central cities, and non-resource-based cities, while it was insignificant or weak in the northeast and old industrial cities. Further analysis showed that digital inclusive finance significantly suppressed the Theil index and the urban-rural income ratio. This study provides reusable methods and empirical support for differentiated digital finance policies.

Keywords: Digital inclusive finance; Urban and rural residents' income; Urban-rural income gap; Dual machine learning; Causal identification

Online publication: August 31, 2026

1. Introduction

Solidly advancing common prosperity requires the urban and rural income pattern to shift from "total expansion" to "structural convergence." The "14th Five-Year Plan" emphasizes urban-rural integration and inclusive finance. The Central Financial Work Conference in 2024 listed inclusive finance and digital finance as the "five major articles." Existing literature confirms that digital inclusive finance can promote entrepreneurship, improve agricultural total factor productivity, and drive industrial upgrading [1]. Inter-provincial evidence also shows that it promotes common prosperity through entrepreneurial activity [2], but its net effect on urban and rural residents' income is still affected by high-dimensional mixing and misconfiguration of the functional form. Traditional two-way fixed effects are difficult to handle the nonlinear synergistic relationship between "digital index–income" and are prone to bias due to the omission of variables such as regional digital infrastructure and financial ecosystem.

This paper adopts a dual machine learning (DML) partial linear model, with random forest as the main component and Lasso/GBM as the auxiliary component to fit the redundant parameter (nuisance) functions $(\hat{m}\cdot\hat{l}\cdot)$ and $(\hat{m}\cdot\hat{l}\cdot)$. By cross-fitting and Neyman orthogonalization, the regularization bias is weakened, and a consistent estimate of the average treatment effect (ATE) is obtained [3]. The contribution of this paper is: to use DML to identify the causal relationship of the income increase effect of digital inclusive finance in prefecture-level cities, to characterize heterogeneity from the two-layer perspective of "geographical location + city attributes," and to incorporate mechanism variables and gap variables into a unified framework.

2. Theoretical analysis and research hypotheses

Based on the theories of financial development, endogenous growth, and dual economic structure, digital inclusive finance affects income through three channels. First, it alleviates financial exclusion: mobile payment and big data credit scoring replace physical outlets and collateral, enabling long-tail groups to obtain working capital and smoothing productive investment. Second, it optimizes capital allocation: it reduces information asymmetry, incentivizes entrepreneurship among rural and urban marginal groups, and facilitates intertemporal allocation of education expenditures by families [4]. The driving effect of digital finance on urban entrepreneurial activity is already backed by city-level evidence [5]. Third, it reallocates urban and rural factors: e-commerce and supply chain finance promote the transfer of labor to high-productivity sectors and extend the agricultural industry chain [6]. Therefore, it is proposed that: H1: digital inclusive finance significantly improves the income level of urban and rural residents. H2a: digital inclusive finance promotes the growth of urban and rural residents' income by increasing entrepreneurial activity. H2b: digital inclusive finance promotes the growth of urban and rural residents' income by promoting the upgrading of industrial structure. H2c: digital inclusive finance promotes the growth of urban and rural residents' income by increasing the level of education expenditures.

3. Research design

3.1. Model setting

This paper employs DML to evaluate the income effect of digital inclusive finance, addressing high-dimensional heterogeneity and nonlinear bias. Let *i* represent a prefecture-level city and t represent the year. The partially linear DML is defined as follows:

$$\mathrm{Incom}_{it} = \beta_1 \mathrm{Dif}_{it} + \theta(X_{it}) + U_{it}\,, \tag{1}$$

$$E(U_{it}|X_{it},Dif_{it})=0 \tag{2}$$

$Incom_{it}Dif_{it}X_{it}\theta$ (·) $\beta_1\hat{m}(X)=\hat{E}(Incom|X)\hat{l}(X)=\hat{E}(Dif|X)Incom-\hat{m}(X)]Dif-\hat{l}(X)]\beta_1$ Among them, $Incom_{it}$ represents the income level of urban and rural residents in prefecture-level city i in year t; the core explanatory variable represents the digital inclusive finance index of prefecture-level city, which is a set of high-dimensional control variables and is an unknown function. represents the ATE. In the estimation, let and be fitted by random forest (main model) and Lasso and GBM (auxiliary verification), respectively. After 5-fold cross-fitting, the orthogonalized residuals [pairs] are regressed to obtain a consistent estimate. The Neyman orthogonal score function can eliminate the regularization convergence order bias.

3.2. Variables and data

The core explanatory variable is the Digital Inclusive Finance Index (DIF), taken from the sixth edition of the Peking University Digital Inclusive Finance Index (PKU-DFIIC) for prefecture-level cities, jointly compiled by the Peking University Digital Finance Research Center and the Ant Group Research Institute. The publicly available sixth edition covers 31 provinces, 337 prefecture-level cities, and approximately 2800 counties from 2011 to 2023. The 2024 prefecture-level city observations in this paper were obtained by the research team from the Peking University Digital Finance Research Center using pre-release supplementary data of the same caliber (obtained in 2025). The index is weighted by three dimensions: coverage breadth, usage depth, and digitalization level, and the overall index has been standardized. The explained variable is per capita disposable income of urban and rural residents; control variables include the logarithm of per capita GDP, population density, fiscal revenue and expenditure pressure, government intervention, and the proportion of science and technology expenditure; mechanism variables are entrepreneurial activity (number of newly registered enterprises/resident population), industrial structure upgrading (value added of tertiary industry/secondary industry), and education expenditure; the gap variables are the Theil index and the urban-rural income ratio. The final sample matched 259 prefecture-level cities, spanning 14 years from 2011 to 2024, with N = 3,626 after balancing [7]. Descriptive statistics are shown in **Table 1**.

Table 1. Descriptive statistics

(1) Variable type	(2) Variable name	(3) Variable symbol	(4) *N*	(5) Mean	(6) SD	(7) Min	(8) Max
Core explanatory variables	Digital Inclusive Finance	Dif	3,626	210.6	82.56	17.02	403.1
Explanatory variable	Urban and rural residents' income	Incom	3,626	160.0	79.76	26.21	1,238
Mechanism variables	Entrepreneurial activity	this	3,626	137.2	134.5	17.56	5,080
	Industrial structure upgrading	they go	3,626	1.163	0.619	0.175	5.715
	Education expenditure	edu	3,626	174.7	39.19	43.63	396.1
Control variables	Economic development level	lngdp	3,626	10.81	0.545	8.807	12.56
	population density	hum	3,626	5.603	1.325	2.354	9.106
	scientific and technological level	sci	3,626	0.0180	0.0186	7.02e-05	0.193
	Fiscal revenue and expenditure pressure	fan	3,626	0.131	0.107	-0.0104	0.806
	Degree of government intervention	gov	3,626	0.208	0.105	0.0560	0.872

4. Empirical analysis

4.1. Benchmark regression

This paper uses a dual machine learning model to empirically test whether digital inclusive finance in prefecture-level cities has a positive effect on the income of urban and rural residents. Based on Equation (1), **Table 2** reports the regression results of the level of digital inclusive finance on the income of urban and rural residents. In **Table 2**, column (1) only controls for feature variables, and the Dif coefficient is 0.4918 (1% significant); in column (2), after adding time and individual double fixed effects, it drops to 0.2739 (1% significant). The economic meaning is that for every 1-point increase in the index, the average income of urban and rural residents in prefecture-level cities increases by about 0.274 yuan/person, and H1 is true. Compared with the traditional two-way fixed effects model, DML effectively reduces the estimation variance and improves the inference accuracy by screening high-dimensional variables and cross-fitting.

Table 2. Benchmark regression results

	(1)	(2)
	Incom	Incom
Dif	0.4918***	0.2739***
	(0.0152)	(0.0276)
_cons	-0.2302	0.9282*
	(0.5527)	(0.5229)
Control variables	yes	yes
Time fixed effect	no	yes
Individual fixed effects	no	yes
N	3626	3626

4.2. Robustness test

To ensure the reliability of the baseline conclusions, this paper conducts robustness checks from three dimensions: algorithm selection, parameter setting, and model setting.

Replacing the random forest in the baseline regression with five algorithms—lasso regression, elastic net, support vector machine, gradient boosting, and neural network—the Dif coefficients were all significantly positive at the 1% level (0.9859, 0.6181, 0.6481, 0.3777, and 0.3661, respectively), confirming that the baseline conclusions do not depend on a specific learner. Further adjusting the cross-fit split ratio from 1:4 to 1:2 and 1:7, the Dif coefficients remained significantly positive at the 1% level (0.3381 and 0.2551, respectively), indicating that the conclusions are unaffected by changes in the split ratio.

Further modifications were made to the model settings and sample conditions (**Table 3**): In column (1), the sample period was shortened to 2016–2024, and the core coefficients remained significant; in column (2), robust standard errors were adopted for city-level clustering, and the statistical inference results remained unchanged; in column (3), population mobility rate and openness to the outside world were added to the control variables, and although the Dif coefficient decreased slightly, it was still highly significant, indicating that the model did not have serious omission bias. Due to space limitations, the results of introducing the time × city interaction fixed effect and the province fixed effect are not shown, but their coefficients are also significantly positive at the 1% level.

Table 3. Robustness checks

	Shrink window	Clustering SE	Increase control
Dif	0.4298***	0.3018***	0.2653***
	(0.0520)	(0.0536)	(0.0279)
Control variables	yes	yes	yes
Time/Individual FE	yes	yes	yes
N	2331	3626	3626

Based on the above tests, the positive effect of digital inclusive finance on the income of urban and rural residents is highly robust.

4.3. Mechanism testing

To examine the mechanism path, based on Jiang Ting's [6] definition of the "two-step method"—only the first step, Dif→Med, needs to be identified cleanly, and the second step, Med→Y, needs to be supported by theory without performing a separate causal regression—we construct the first-step equation:

$$Med_{it} = \gamma_2 Dif_t + \varphi_2(X_t) + \varepsilon_t \quad (3)$$

$$Incom_{it} = \beta_1^{'} Dif_t + \lambda Med_t + \psi(X_t) + \mu_t \quad (4)$$

Among them, Med_{it} represents entrepreneurial activity, industrial structure upgrading, and education expenditure, in that order. This paper focuses on testing the significance and does not substitute Med in equation (4) back into Incom to perform a Sobel-style mediation coefficient product test.

Table 4 shows that the estimated coefficients of the Digital Inclusive Finance Index (Dif) for entrepreneurial activity, industrial structure upgrading, and education expenditure are 0.2586, 0.0023, and 0.1352, respectively, all significantly positive at the 1% level. For every 1-unit increase in the Digital Inclusive Finance Index, the Industrial Structure Upgrading Index increases by approximately 0.0023 units. Given that the mean for Industrial Structure Upgrading is only 1.163, this increase is equivalent to 0.2% of its mean, indicating that while the driving effect of digital inclusive finance on industrial structure is statistically significant, its economic scale is relatively moderate. This demonstrates that digital inclusive finance significantly promotes urban and rural residents' income growth through the aforementioned three paths, validating hypotheses H2a, H2b, and H2c.

Table 4. Mechanism test results

	(1)	(2)	(3)
	Entrepreneurial activity	Industrial structure upgrading	Education expenditure
Dif	0.2586***	0.0023***	0.1352***
	(0.0871)	(0.0002)	(0.0175)
_cons	0.9996	0.0063	-0.1828
	(1.5633)	(0.0039)	(0.3115)
Control variables	yes	yes	yes
Time fixed effect	yes	yes	yes
Individual fixed effects	yes	yes	yes
N	3626	3626	3626

4.4. Heterogeneity analysis

(1) Geographical heterogeneity. Considering the uneven regional development in my country, this paper further examines the regional heterogeneity of the income effect of digital inclusive finance. The sample is divided into four major regions: East, Central, West, and Northeast. Table 5 shows the geographical grouping: East (0.4928, 1% significant) > West (0.3796, 1% significant) > Central (0.3208, 1% significant), Northeast -0.0581 (not significant). The East has well-developed infrastructure, the Central and Western regions have a higher marginal effect of "filling the gaps" in infrastructure, while the Northeast is constrained by population outflow and industrial transformation. Further dividing by North and South, the South (0.4303, 1% significant) is higher than the North (0.1763, 1% significant), but is not listed due to space limitations. This reflects the active private economy and high acceptance of digital finance in the South, while the North has a higher proportion of heavy industry and a lower degree of marketization, which limits the effectiveness of digital finance.

Table 5. Results of heterogeneity test

Dimension	Grouping	Dif coefficient	Standard error	*N*
Geographical location	East	0.4928**	(0.0293)	1092
	Central	0.3208***	(0.0282)	1050
	West	03796***	(0.0224)	1050
	Northeast	-0.0581	(0.0996)	434
City attributes	Old industrial	0.0977*	(0.0551)	1260
	Non-old industry	0.4411***	(0.0207)	2366
	Big cities	0.4060***	(0.0257)	1302
	Small and medium-sized cities	0.2153***	(0.0334)	2324
	Central cities	0.3526***	(0.0326)	448
	Non-central cities	0.1899***	(0.0332)	3178
	Resource-based	0.2258***	(0.0302)	1456
	Non-resource type	0.3076***	(0.0301)	2170
Control variables		yes		
Time/Individual FE		yes		

(2) Urban heterogeneity. Considering the significant differences among cities in terms of industrial structure, administrative level, and population size, this paper further divides the sample into old industrial/non-old industrial, large/medium-sized, central/non-central, and resource-based/non-resource-based cities for regression analysis. The results are shown in Table 5. Table 5 shows that the non-old industrial city group (0.4411) is significantly stronger than the old industrial city group (0.0977); large cities (0.4060) > medium-sized cities (0.2153); central cities (0.3526) > non-central cities (0.1899); and non-resource-based cities (0.3076) > resource-based cities (0.2258*). The above grouping results indicate that a sound digital infrastructure, a vibrant private sector, and a diversified industrial structure are important prerequisites for the full release of the income-generating effect of digital inclusive finance.

4.5. Further analysis of urban-rural income gap

Based on the above empirical results, this paper uses the Theil index and the urban-rural income ratio to measure the urban-rural income gap, respectively. The regression results are shown in **Table 6**. **Table 6** shows that after controlling for the linear and quadratic terms of the variables, the estimated coefficients of the Theil index and the urban-rural income ratio by Dif are significantly negative at the 1% level (−0.0010 and −0.0008), respectively, indicating that digital inclusive finance, It is also inclusive in bridging the urban–rural divide, aligning with Yang *et al.* [8], who argue that digital inclusive finance is conducive to narrowing the urban–rural gap.

Table 6. Results of the urban-rural income gap test

	Theil index	Urban-rural income ratio
Dif	-0.0010***	-0.0008***
	（0.0002）	（0.0002）
_cons	0.0010	0.0016
	（0.0040）	（0.0040）
Control variables	yes	yes
Control variable quadratic term	yes	yes
Time/Individual FE	yes	yes
N	3626	3626

Note: theil1 is the Theil index, and theil2 is the urban-rural income ratio. Columns (1) and (3) contain only linear terms of the control variables, while columns (2) and (4) contain both linear and quadratic terms.

5. Research conclusion and policy recommendations

Empirical evidence shows that digital inclusive finance significantly increases income and narrows the urban-rural gap; entrepreneurship, industrial upgrading, and education spending are key mechanisms; the effects show a gradient of “stronger in the east/west/south than in the northeast/north” and “stronger in non-old industrial/large cities/central/non-resource-based areas.” The policy implications are as follows:

First, a tiered regional supply model. Eastern regions prioritize technology-driven micro and small enterprise loans, central and western regions supplement first-time loans in county-level areas, and the Northeast promotes synergy between finance and technological upgrading. Second, targeted mechanism activation. Online entrepreneurship guarantees are embedded in government services, digital supply chain finance is promoted in industries, and installment payments for education are offered to human resources. Third, eliminating rural credit blind spots. Data on contracted land transfer, agricultural machinery operation, and agricultural insurance are collected and incorporated into credit reporting systems, while retaining age-appropriate offline channels. Fourth, heterogeneous assessment. A joint “finance-industry” KPI is established for the Northeast, old industrial cities, and resource-based cities, providing positive incentives in the Macroprudential Assessment.

Disclosure statement

The author declares no conflict of interest.

References

[1] Yang W, Su L, Wang M, 2020, Digital Inclusive Finance and Urban and Rural Residents' Income: A Mediating Effect Analysis Based on Economic Growth and entrepreneurial behavior. Journal of Shanghai University of Finance and Economics, 22(4): 83–94.

[2] Han L, Peng Y, Meng Q, 2023, Digital Inclusive Finance, Entrepreneurial Activity and Common Prosperity: An Empirical Study Based on Inter-Provincial Panel Data in My Country. Soft Science, 37(3): 18–24.

[3] Chernozhukov V, Chetverikov D, Demirer M, et al., 2018, Double/Debiased Machine Learning for Treatment and Structural Parameters. The Econometrics Journal, 21(1): 1–68.

[4] Zhang B, Li N, 2022, Research on the Impact of Digital Inclusive Finance on the Non-Agricultural Transfer of Rural Labor Force - An Empirical Analysis Based on Mlogit and Threshold Models. World Agriculture, (9): 65–75.

[5] Zhao T, Zhang Z, Liang S, 2020, Digital Economy, Entrepreneurial Activity and High-Quality Development: Empirical Evidence from Chinese Cities. Management World, 36(10): 65–76.

[6] Jiang T, 2022, Mediating and Moderating Effects in Empirical Studies of Causal Inference. China Industrial Economics, (5): 100–120.

[7] Guo F, Wang J, Wang F, et al., 2020, Measuring the Development of Digital Inclusive Finance in China: Index Compilation and Spatial Characteristics. Economics (Quarterly), 19(4): 1401–1418.

[8] Yang D, Dai H, Huang F, 2022, A Study on the Threshold Effect of Digital Inclusive Finance on the Income Gap Between Urban and Rural Residents—An Empirical Analysis Based on Different Development Dimensions. Economic and Management Review, 38(3): 89–101.