International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

Data Analytics for the Effect Of Digital Inclusive Finance on Farmers' Entrepreneurship Decisions in Rural Areas Using the Supervised Learning-Based Regression in Rural China

Zheng Li[1], Ho-Jyun Wong[2], Shiwei Tan[3,*], Weikun Zhang[4]

[1]Innovation and Entrepreneurship Center, Applied Technology College of Soochow University, Suzhou, China

[2]School of Economics and Statistics, Guangzhou University, Guangzhou, China

[3]Business School, Lingnan Normal University, Zhanjiang, China

[4]School of Public and Social & Administration, Lingnan Normal University, Zhanjiang, China

*Corresponding Author.

Abstract

Farmers' entrepreneurship is essential in developing the countryside and improving farmers' income. Also, with the arrival of the digital economy, rural formal finance (FF), which was brought about by digital inclusion finance (DIF), promoted entrepreneurship among the farmers. Based on data analytics from 2021 to 2023 in China, this paper empirically performs the quantitative analysis of the effect of DIF on farmers' entrepreneurship decisions (FED). Further, it addresses the role of FF as an alternative for informal finance (IF) brought about by the development of DIF in the process of FED from a social network (SN) perspective using the supervised learning-based regression. The results show that DIF can significantly promote farmers' entrepreneurship (FE), especially in self-employed off-farm entrepreneurship. DIF has a positive effect on FE with low levels of education and weak social networks, suggesting that DIF has a truly "inclusive" role to play. Further mechanistic studies have found that DIF, by FF, compensates for the fact that farmers used to rely mainly on SN for financing from IF channels, increasing their financial accessibility, reducing their borrowing costs, and thus promoting their entrepreneurship. Our results from the estimation applied to the supervised learning algorithm considering the instrument variable provide scientific implications for promoting DIF matched with the rural credit system's perfection to improve farmers' production and operation.

Keywords: Digital Inclusive Finance (DIF), farmers' entrepreneurship decisions (FED), data analytics (DA), supervised learning-based regression (SLR), social networks (SN).

1. Introduction

Innovation and economic growth are driven by entrepreneurship. Farmers' entrepreneurship (FE) is rooted in agriculture and land-related businesses. Not all farmers can get entrepreneurship funding, especially in developing countries like China. Farmers struggle to get loans from banks due to a lack of collateral and reputation. Most of FE's capital comes from informal finance (IF), like family and friends. Society cannot solve farmers' financial problems if villages lend by acquaintance. Farmland mortgage loans and land capitalization have been implemented to encourage rural entrepreneurship, but farmers still have strict economic rules.

Respectively, established industry and commerce typically pave the way for entrepreneurial endeavors in urban centers. For a long time, rural areas have lacked many conditions that make entrepreneurship more successful, as in towns and cities, due to industrial characteristics, factor allocation, information, and credit constraints[1,2]. As more and more people from rural areas have left the agricultural workforce for urban centers, there has been a steady improvement and optimization of rural financial infrastructure and upgrades to the industrial structure, creating an environment where entrepreneurs can take root. Also, with the advancement of information technology such as the Internet, big data, and cloud computing, the traditional inclusive financial system has begun to play its inclusive function after digitization honestly. Digital inclusive finance (DIF) breaks through the problem of low financial accessibility and high financing costs in rural areas, making it possible to alleviate farmers' credit constraints when starting a business [3-5]. The annual report released by the Institute of DIF from Peking University shows that the impact of DIF is more significant where the rate of urbanization is lower. As a result, in the context of new technological changes, farmers' entrepreneurship decisions (FED) offer greater possibilities and increasingly need to be considered. Next, digital inclusion's formal financing would supersede the old-fashioned informal channels of reception through SN, like personal connections and favors. Does DIF help FED by making their finances more accessible and lowering their financing costs?

An extensive and in-depth study has been conducted in the existing literature on whether and how DIF affects resident FED. Numerous studies have demonstrated that DIF can significantly promote entrepreneurship decisions (ED) among residents [6-16]. Part of the studies has further investigated the mechanisms of DIF on residents' ED, with discussions from the perspectives of credit constraints, information constraints, financial literacy, and entrepreneurial opportunities, respectively [17-27]. It is also found that DIF can alleviate residents' credit Hull information constraints and promote entrepreneurship through channels such as improving residents' financial literacy and increasing entrepreneurial opportunities[28-33]. Besides, the existing literature has also examined the heterogeneity of DIF on resident entrepreneurship and found that DIF has a more notable impact on self-employment-based ED [34-37]. Some studies have focused on the effect of DIF on ED among transient populations and on entrepreneurial performance [38]. These researches are estimated by the ordinary linear regression, considering the condition distributions of the output features when the input features are given. As ED is an endogenous choice for families, this disturbance term is no longer constant, and every family faces different features. Currently, the ordinary linear regression estimation cannot remain consistent against the causal inference.

Below, the existing regression analysis only considers the influences of the feature-based factors, relying upon small sample sizes, and fails to solve the contradiction on ED, ranging from the simple descriptive statistical analysis to ordinary linear regression [39]. Toward the cause inference, the supervised learning-based regression with the flexible statistical technique is accessed to revisit the empirical evidence of ED using the survey microdata, outperforming the ordinary regression. Supervised learning is a machine learning technique that considers the effects of noise and outliers when running multiple regressions, extracting meaningful information hidden in the data to grasp the main characteristics of the correlations [40]. In fields of economy, finance, and agricultural production, this supervised learning is carried out to perform the cause inference that keeps the estimated analysis efficient, avoiding the impact on randomly processing the sample data. Suffice it to say that this supervised learning regression was appealed to frame the effect on ED by graphing a large set of the families' attributes.

Contrary to the ED of urban residents, FED is unique. Household wealth and credit constraints are key factors affecting FED through the expansion of agricultural production and operations [41-44]. Moreover, the average rural household has lower financial literacy, poorer financial accessibility, and higher financing costs. In that case, whether DIF also exerts its entrepreneurial effect in rural areas has attracted extensive interest. Some researchers revealed that DIF could significantly improve the operational efficiency of rural commercial banks [45-49]. For the relationship between DIF and FED, the existing literature has provided preliminary conclusions that DIF has made traditional credit constraints less critical for rural households and that financial literacy plays a vital role. For the relationship between DIF and FED, the existing literature has provided preliminary conclusions that DIF significantly increases farmers' entrepreneurial willingness and has attempted to explore the mechanisms behind this regarding credit constraints, among others. However, rural households' entrepreneurial choice decisions, processes, and characteristics differ from those of urban and transient families in that rural households face more substantial credit constraints and mostly rely on IF. IF channels such as friends and relatives and private lending account for a significant proportion of rural lending in rural societies based on networks of acquaintances, as opposed to FF [50]. Therefore, the characteristics of rural households and their models are considered when building on the supervised learning-based regression algorithm.

This study makes three contributions. First, unlike previous research, this paper exploits the data from the China Household Finance Survey, namely a nationwide family finance micro survey, to investigate the impact of FED by formal or informal finance in the context of DIF with supervised learning-based regression. Second, this algorithm was used to identify the relationship between DIF and FEB, which contributes to the fields of the families' research with ordinary statistical techniques, such as linear regression and logistic regression. Third, data analytics with survey dataset appealing to the supervised learning-based regression approach get the critical features on FED and determine the mechanisms of DIF with the coefficient values utilizing the instrumental variable, and are the models performing the unidentified recordings which we aim to explore in this paper. These findings show that farmers do not engage in agricultural entrepreneurship and prefer self-employment-based businesses in terms of the drive of DIF. This insight has important implications for which entrepreneurship the farmers prefer.

2. Methods

2.1 Data and variables

The farmer's microdata in this paper is drawn from the 2021 and 2023 individual samples of the China Household Finance Survey (CHFS) dataset published by the China Household Finance Survey and Research Centre, Southwest University of Finance and Economics. The CHFS data is a nationwide sample survey in rural China designed to collect information on the micro-level of household finance, including information on housing assets and financial wealth, debt and credit constraints, income and consumption, social security and insurance, intergenerational transfers, demographic characteristics and employment, and payment habits. The survey covers 29 provinces, municipalities, and autonomous regions, with over 40,000 households investigated in 2021 and over 35,000 in 2023. This data provides a comprehensive and detailed description of household economic and financial behaviors and offers much information for studying farmers' credit constraints, borrowing, and EB. In addition, the reason for using the data in 2021 and 2023 is that the data in these two years are highly time-sensitive. The surveys in these two periods provide rich information about the use of farmers' financial platforms and lending behavior, which is helpful for this study. Since the follow-up rate in 2023 was only about 50%, this paper did not use panel data but controlled the time-fixed effect and conducted a mixed cross-section analysis. After keeping the samples of farmer households and clearing the missing values of some variables, the total number of valid examples in two years is 36,960. In addition to the primary micro data, this paper uses the Peking University Digital Financial Inclusion Index and the regional characteristic indicators obtained from the provincial statistical yearbooks in China.

This paper comprehensively considers the multidimensional FED for the explained variable, referring to the existing literature [29-34]. First, the information by the CHFS, "Is your household currently engaged in a commercial or industrial business?" was used to identify the commercial or industrial ED. If the answer is "yes," the farmers are defined as entrepreneurial; if "no," then non-entrepreneurial. Secondly, based on the discussion of overall business entrepreneurship, this paper further classifies ED into self-employment and employer-based according to the number of employed people. It defines the farmers as self-employment entrepreneurship if it uses no additional employees and employer-based entrepreneurship if there is more than one employee [16]. Finally, considering that not all farmers' entrepreneurial fields are industrial and commercial and that someones still choose agricultural entrepreneurship, this paper refers to Huang et al. [51], who define agricultural entrepreneurship as farmers with agricultural income over $7370 or agricultural assets over $4442, as the CHFS dataset does not provide information on agricultural entrepreneurship. Generally speaking, smallholder-style farmers have low incomes from agricultural production. If the net agricultural income exceeds $14740, the household can have a specific scale of agricultural output and be regarded as a farming business.

For independent variables, this paper focuses on the impact of DIF on FED and analyzes the mechanisms behind the effect in terms of farmers' SN and lending behavior. Therefore, in the underlying analysis, the independent variable of this paper is the Peking University Digital Financial Inclusion Index in China (2014-2023), which measures the extent of DIF in different regions in terms of three dimensions consisting of coverage, usage, and digitization, respectively. The paper also examines the impact of three sub-indicators on FED: the coverage breadth of DIF, depth of use, and digitalization. To avoid the problem of reciprocal causality, the DIF indices selected for this paper are all indicators with a one-period lag. In a follow-up analysis, the article also examines the use of DIF by farmers and the impact of lending channels on entrepreneurship, further exploring how DIF affects FED. Referring to the existing literature, this paper measures the use of DIF by farmers based on whether they have used mobile payments and credit cards and calculates their availability by whether they successfully access loans from FF institutions. Access to loans is measured based on whether they have borrowed from banks, credit unions, private finance, and friends and relatives. In addition, this paper also examines the heterogeneous impact of DIF on FED and its influence mechanisms from the perspective of SN. This paper uses farmers' expenditure on favors and gifts in the past year to measure SN.

For control variables, to control interference from other factors as much as possible, the main household head and household information reflecting individual and farmers' characteristics and local ones reflecting the region level are selected as control variables in this paper. Personal and household characteristics include gender, age, education level, party membership, marital status, health status, economic and financial concerns, risk appetite, social security status, SN, etc. Regional characteristics include GDP per capita, fiscal revenue, the magnitude of marketization, and the region's education.

Table 1 Definitions and descriptive statistics of variable.

Variables	Definition	Mean value	Standard deviation
FED			
ED	Are the farmers engaged in industrial and commercial operations?(Yes=1,No= 0)	0.155	0.353
EE	=1 if the number of employees in business activities is more than 1,0; otherwise	0.034	0.188
SE	=1 if no employees in business activities,0 otherwise	0.12	0.323
AE	=1 if agricultural assets exceed $4442 or agricultural income exceeds $7370, 0; otherwise	0.055	0.226
DIF			
Total_dig	Total index of DIF	265.4	39.86
Coverage_dig	Coverage of DIF	243.7	41.77
Depth_dig	Usage depth of DIF	252.7	46.57
Digitization	The digitization of DIF	356.7	33.29
Use_fina	=1 if using mobile payment,0 if no	0.163	0.369
Use_finb	=1 if using a credit card,=0 otherwise	0.295	0.456
Availability_fina	Whether the applied loans from banks and other FF institutions were rejected (Yes=1, No=0)	0.295	0.465
Availability_finb	Whether the applied loans from banks and other FF institutions were approved (Yes=1, No=0)	0.175	0.389
Family Characteristics			
Age	The age of the head of household (HH)	54.51	11.92
Gender	The gender of HH (Male=1,Female=0)	0.853	0.357
Education	The education years of HH	7.758	3.418
Members	=1 if HH is a party member,0 otherwise	0.085	0.277
Marriage	=1 if HH is married,0 otherwise	0.875	0.317
Health	Self-assessment of the health status of HH (1 to 5 in integer level,=1 if very bad,=5 if very good)	2.734	1.061
Security	=1 if a household has endowment insurance,0 otherwise	0.746	0.437
Sensitivity	Self-assessment of the attention of HH to economic finance (1 to 5 in integer level,=1 if very low,=5 if very high)	2.681	2.243
Risk	Risk preference of HH (1 to 5 in integer level,=1 if very high,=5 if very low)	2.883	2.45
Network	Family social network characterized by the logarithm of favor gift expenses	2.426	3.766
Region characteristics			
Per_gdp	Logarithm of regional GDP per capita	10.88	0.365
Revenue	The logarithm of regional revenue income	8.656	0.411
Marketization	The magnitude of regional marketization	8.347	1.456
development	Number of colleges and universities in the region	99.78	37.35

The definitions of the critical variables and descriptive statistics are shown in Table 1, from which it is clear that the total proportion of farmers engaged in industrial and commercial business is 14%. However, of these, self-employment accounts for the vast majority, with only around 3% of entrepreneurs being employer-based businesses, indicating that most farmers are involved in entrepreneurial activities that are essentially family-based, self-employment-based businesses. The number of agricultural entrepreneurs is only 5%, meaning that only a tiny proportion of farmers are engaged in large-scale agricultural production and that the mainstream agricultural output in China is still small-scale family production. Regarding the Digital Inclusion Index, the mean value of the overall index is 264, while the standard deviation is 40, indicating a wide variation in the magnitude of DIF between regions. Additionally, only 16% of rural households used mobile payments, and 30% have used credit cards, indicating that despite the rapid development of DIF in recent years, few rural farmers in China have access to and use digital finance and enjoy the fruits of its growth. In the meantime, many loan applications were rejected compared to those approved, demonstrating that farmers face borrowing needs and credit constraints in their production and business activities.

2.2 Model

The main object of this paper is the effect of DIF on FED and their performance. This paper models ED as a binary choice and characterizes whether or not to start a business as a dummy variable as the explanatory variable $ED = f(DIF, family)$. Therefore, concerning existing research, the baseline model for this paper uses the supervised learning-based regression, which is as follows.

$$ED_{ipt} = \beta_0 + \beta_1 DIF_p + \beta_{1j} family_{jit} + \beta_{2j} region_{jpt} + \gamma_t + \varepsilon_{it} \quad (1)$$

Also,

$$P_{ED} = \frac{\mathrm{e}[\beta_0 + \beta_1 DIF_p + \beta_{1j} family_{jit} + \beta_{2j} region_{jpt} + \gamma_t + \varepsilon_{it}]}{1 + \mathrm{e}[\beta_0 + \beta_1 DIF_p + \beta_{1j} family_{jit} + \beta_{2j} region_{jpt} + \gamma_t + \varepsilon_{it}]} \quad (2)$$

Where ED_{ipt} denotes the dichotomous choice variable of whether household i from p region started a business in t time, P_{ED} refers to the probability of the predicted FED. The notion of this approach is to give the minimum error for the target output, i.e., $\min|ED - P|$ .This paper focuses on overall industrial and commercial entrepreneurship, while self-employment, employer-based, or agricultural entrepreneurship are also considered separately in the heterogeneity analysis section. DIF_p is the Digital Inclusion Index at the p region level. In robustness tests, this paper also uses household-level indicators of digital financial use. The $family_{jit}$ is a matrix of control variables at the individual and household levels in t time and $region_{jpt}$ a matrix of control variables at the district level in t time. Apart from the above, γ_t there is a time-fixed effect, ε_{it} a random error term, and β_1 the core coefficient of interest in Eq. (1).As DIF is not an exogenous variable, there are potential endogeneity issues. For example, despite controlling for critical characteristics at the individual, household, and regional levels, there may still be other factors that can affect both DIF and FED. Therefore, this paper exploits instrumental variables for the endogenous variable DIF and uses the supervised learning-based regression with the instrumental variable for robustness testing. Given that the instrument variable of DIF is z, $P(DIF) = P(DIF(z)\,|z = z*)$ and $z \perp P|DIF = DIF*$ .In equilibrium state, if $DIF = DIF*$, $ED_{ipt} = \beta_c + \beta_{c1} DIF_p + \beta_{c2} z$ and β_{c2} follows up the criterion of zero value. The β_1 can be presented as $E(\beta_1) + \frac{Cov(\beta_1, -\beta_{c2}/(\beta_{c1} - \beta_1))}{E[-\beta_{c2}/(\beta_{c1} - \beta_1)]}$ .So, β_1 is a coefficient value given the near-zero regularization term, of which the supervised learning algorithm is carried out below.

3. Results And Discussion

3.1 Basic regression results

This paper uses the supervised learning-based regression to empirically test the impact of DIF on the overall off-farm FED. While considering the total index of DIF, the effect of three sub-dimensions, consisting of the coverage breadth, the depth of use, and the magnitude of digitization, on FED is also analyzed separately, with the regression results shown in Table 2. Model (1) controls only for individual and household characteristics and time-fixed effects, to which model (2) adds regional economic characteristics. Additionally, models (3) to (5) report the estimation results for the three sub-dimensions, including coverage breadth, depth of use, and digitalization. The results of models (1) and (2) show that the total index of DIF positively affects FED with a 1% significance level, indicating that the higher the DIF, the higher the probability of farmers starting a business. Therefore, H1 holds. In addition, the results of the regressions on the various sub-dimensions show that the coverage breadth and depth of use have a significant positive effect on FED, while the magnitude of digitization does not significantly affect FED. This suggests that the impact of DIF on FED is seen more in its broad reach and depth of use. Farmers are more likely to benefit from the ease of borrowing from DIF outlets and digital technology usage scenarios such as mobile payments and internet lending. At the same time, the magnitude of digitization reflects the digital transformation of businesses more, and digitization's impact is likely to be more pronounced in more economically developed urban areas.

Table 2 Baseline regression results.

Variable	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Total_dig	0.406***	0.574***			
	(0.041)	(0.094)			
Coverage_dig			0.268***		
			(0.094)		
Depth_dig				0.477***	
				(0.053)	
Digitization					-0.01
					(0.046)
Age	-0.017***	-0.017***	-0.017***	-0.017***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gender	-0.042	-0.036	-0.038	-0.036	-0.037
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Education	0.031***	0.028***	0.028***	0.029***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Members	0.030	0.032	0.033	0.030	0.033
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
Marriage	0.411***	0.412***	0.41***	0.416***	0.41***
	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)
Health	-0.144***	-0.144***	-0.144***	-0.146***	-0.144***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Security	-0.136***	-0.126***	-0.126***	-0.127***	-0.124***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Sensitivity	0.024***	0.021***	0.022***	0.023***	0.023***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Risk	-0.027***	-0.028***	-0.028***	-0.029***	-0.029***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Network	0.025***	0.025***	0.025***	0.024***	0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Per_gdp		-0.311***	-0.198***	-0.286***	-0.057
		(0.059)	(0.065)	(0.048)	(0.041)
Revenue		0.128***	0.14***	0.15***	0.162***
		(0.042)	(0.042)	(0.042)	(0.042)
Marketization		0.042***	0.051***	0.028**	0.059***
		(0.012)	(0.012)	(0.012)	(0.012)
Development		-0.001	-0.001*	-0.001***	-0.001***
		(0.001)	(0.001)	(0.001)	(0.001)
Time-fixed	Yes	Yes	Yes	Yes	Yes
Constant	-1.264***	0.375	-0.28	0.345	-1.413***
	(0.109)	(0.558)	(0.624)	(0.515)	(0.481)
Observation	36960	36960	36960	36960	36960

Note: The standard errors are reported in parentheses. The average accuracy is 0.853.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

At the level of family characteristics, the results in Table 2 show that farmers' education level, economic sensitivity, and risk appetite significantly impact FED. A higher level of education indicates a higher level of human capital ownership. A higher level of human capital generally implies a higher level of financial literacy and entrepreneurial knowledge, thus increasing the motivation to start a business and improving the likelihood of entrepreneurship [38]. The higher the economic concern, sensitivity, and risk appetite, the more likely the business will be started. The findings of this paper are also consistent with existing research [41]. Age, on the other hand, is negatively associated with FED, suggesting that younger people are more adventurous and challenging, while older people have a relative preference for stability [16]. In addition, this paper finds that the family SN also has a significant positive effect on FED. Ma and Yang [32] showed that farmers with more SN would have more access to private borrowing and thus be more likely to start their businesses. Moreover, the less developed the FF is, the more significant the impact of private lending on the FED. It is precisely one of the aims of DIF to replace the absence of FF in the rural lending market.

3.2 Robustness

The index of DIF used above for the analysis is a macro-level indicator, while the overall may not affect every micro household. Regarding robustness, the paper first uses micro-household level indicators as a proxy for macro indicators. Models (1) - (3) in Table 3 report the impact of digital finance use at the household level on FED, respectively. These use_fina refers to whether a household uses mobile payments and use_finb credit cards. The total digital use is digital financial services if one of the two is present. The impact of DIF can be measured in another dimension by households' actual use of digital finance, as opposed to overall indicators at the macro level **Error! Reference source not found.Error! Reference source not found.**. The results show that those households using digital finance have greater chances of starting a business.

Table 3 Regression results from robustness.

	Alternatives for digital finance use to DIF			Lag DIF	IV Regression
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Use _ fina	0.514***				
	(0.02)				
Use _ finb		0.670***			
		(0.023)			
Total digital use			0.460***		
			(0.012)		
L2. DIF				0.674***	
				(0.116)	
DIF					0.421***
					(0.127)
Family	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes	Yes	Yes
Constant	-1.518***	-1.397***	-1.468***	0.534	-0.075
	(0.476)	(0.482)	(0.484)	(0.582)	(0.614)
Observation	36960	36960	36690	36960	36960

Note: The standard errors are reported in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The first stage regression with IV is positive at 1%, and the F-statistics from Stock-Yogo weak ID is 37752.01. The average accuracy is 0.789.

The above results suggest that the impact of DIF on FED is relatively robust. However, there may be endogeneity problems in the baseline model due to missing variables and reverse causality. On the one hand, though this paper controls for family and regional level characteristics as much as possible, there may still be some unobservable regional factors that influence DIF and FED. The omission of these variables would lead to biased results in the baseline model estimates. On the other hand, it may be that DIF is faster in areas with better formal entrepreneurial soil and higher entrepreneurial demand, and this reverse causality problem could also lead to skewed estimates. Seeking to circumvent the reverse causality problem, this paper reports the impact of the two-period lagged DIF on FED in a model (4), and the results show that the lagged DIF still has a positive and significant impact on rural entrepreneurship in the current period.

Furthermore, this paper finds IV for the endogenous variable DIF to overcome the omitted variable bias problem as much as possible, using the Surprised Learning with the IV. Referring to the existing literature [19,43], this paper chooses the difference between the two-lagged DIF and the current DIF multiplied by the two-lagged index as the instrumental variable. On the one hand, the product is highly correlated with the current period's index, satisfying the correlation condition; on the other hand, the product also does not directly affect FED, satisfying the exogeneity condition. A weak instrumental variables test is also conducted in this paper, and the first stage regression results indicate that the coefficient on the constructed instrumental variable is significantly positive and that the Stock-Yogo weak ID F statistics value of 37752.01 is well above the critical value level, indicating that there is no weak instrumental variables problem. In addition, model (5) reports the results considering the IV. The results from supervised learning-based regression considering the IV show that after accounting for the endogeneity problem caused by missing variables, DIF still has a significant positive impact on FED, and the findings of the benchmark regression are robust.

3.3 Heterogeneity

The analysis above focuses on the off-farm employment of farmers. At the same time, there are two main types of off-farm entrepreneurship among farmers (e.g., employer-based entrepreneurship and family-based self-employment entrepreneurship), with the former being more opportunity-based entrepreneurship and the latter being more survival-based entrepreneurship. So, does DIF have different impacts on different types of entrepreneurship? Models (1) and (2) in Table 4 report the effect of DIF on various types of entrepreneurship, respectively. The results of model (1) suggest that the effect of DIF on employer-based entrepreneurship is not significant, while model (2) suggests that DIF significantly promotes self-employment-based entrepreneurship. Possible reasons for this are as follows.

On the one hand, employer-based entrepreneurs face fewer financial constraints than self-employed ones, who have more access to finance, so the impact of DIF is not as significant. On the other hand, employer-based entrepreneurship often requires a higher loan amount, while limited access to DIF in rural areas affects self-employed entrepreneurship. The exact effect of heterogeneity needs to be followed up with further argumentation and research. After distinguishing employer-based and self-employed entrepreneurship, the paper also considers the impact of DIF on agricultural entrepreneurship. According to the data in this paper, the share of agricultural entrepreneurship is also 5%, higher than employer-based off-farm entrepreneurship. Model (3) shows that DIF significantly negatively impacts agricultural entrepreneurship, indicating the low likelihood of farmers doing agricultural entrepreneurship with DIF. The possible reason for this is that DIF in rural areas has provided the conditions for farmers to engage in off-farm entrepreneurship, and therefore, DIF has promoted off-farm entrepreneurship while squeezing out some of the agricultural entrepreneurship.

Table 4 Heterogeneity results on type of ED.

	EE	SE	AE
	Model (1)	Model (2)	Model (3)
DIF	0.170	0.491***	-1.462***
	(0.203)	(0.135)	(0.168)
Family	Yes	Yes	Yes
Region	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes
Observation	36960	36960	36960

Note: The standard errors are reported in parentheses. The average accuracy is 0.847.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

According to the theoretical analysis in this paper, the impact of DIF on FED may vary depending on their human capital status and the level of regional economic development. Depending on the education level, this paper classifies rural households with a high school education as a high human capital group, those with lower secondary education, and those below as having low human capital. Findings from models (1) and (2) in Table 5 indicate that the impact of DIF on FED with different types of human capital is positively significant, but the effect on FED with high human capital is as substantial as possible at the 10% level. This suggests that the impact of DIF on the FED of high human capital farmers is unstable, whereas that of low one is more pronounced. Moreover, based on regional GDP per capita, this paper classifies regions with GDP per capita above the average value as high economic development regions and those with GDP per capita below the average value as low economic development regions. Models (3) and (4) indicate that the impact of DIF on FED is more noticeable in less economically developed areas. To a certain extent, the above findings validate the "inclusive" H2 with DIF having a more significant impact on low economic development areas and farmers with low human capital.

Table 5 Heterogeneity results on human capital and economic development.

	High human capital	Low human capital	High economic development	Low economic development
	Model (1)	Model (2)	Model (3)	Model (4)
DIF	0.473*	0.406***	0.337	0.591**
	(0.282)	(0.144)	(0.223)	(0.240)
Family	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Time-foxed	Yes	Yes	Yes	Yes
Observation	5666	31294	15600	21360

Note: The standard errors are reported in parentheses. The above regression results are run using the Surprised Learning with IV. The average accuracy is 0.753. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4. Further Discussion

4.1 Mechanism I: social network

The Chinese countryside is a human relations and acquaintance society, and SN plays an essential role in the productive lives of rural households. Again, this paper's benchmark model regression results found that SN directly promotes FED. However, SN is more about providing IF, such as private lending to farmers, while DIF in rural areas is mainly in terms of FF channels. So, is DIF the alternative or complement to the traditional rural lending market? Is the impact of DIF on agricultural entrepreneurship related to farmers' SN? To this end, this paper further explores how DIF affects agricultural entrepreneurship from the perspective of SN. According to SN measured by their gift expenditure, this paper defines households with above-average gift expenditures as strong SN and those with below-average ones as weak SN. Models (1) and (2) in Table 6 report the impact of DIF on FED in different types of SN, respectively. According to the results, DIF only promotes FED in weak SN, while it has no impact on FED in strong SN. A comparison of SN coefficients also shows that for farmers with weak SN, the SN coefficients that impacted entrepreneurial choice became less significant with the entry of DIF. This also suggests that DIF has gone a long way to compensate for the more substantial credit constraints faced by households with weak SN in the past due to their small sizes. DIF has changed the past situation of some groups relying on SN for informal lending by bringing FF financing channels to rural areas. To obtain robust results, model (3) is subjected to interaction term analysis using the total sample, and the results show that SN has a negative moderating effect on the impact of DIF on FED.

Table 6 Mechanism identification of SN.

Variable	Strong SN	Weak SN	Interaction Term
	Model (1)	Model (2)	Model (3)
DIF	0.306	0.485***	0.458***
	(0.231)	(0.155)	(0.129)
SN	0.145***	0.000	0.086***
	(0.012)	(0.000)	(0.01)
DIF#SN			-0.19***
			(0.033)
Family	Yes	Yes	Yes
Region	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes
Observation	11052	25908	36960

Note: The standard errors are reported in parentheses. The above regression results are run using the Surprised Learning with IV. The average accuracy is 0.761. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To support these findings, the paper further investigates the impact of DIF on farmers' lending channels and analyzes the heterogeneity of this effect across SN sizes. Models (1) and (4) in Table 7 report the impact of DIF on formal and informal channels borrowed by rural households, respectively. The results show that DIF significantly facilitates formal channel borrowing from farmers without impacting informal channel borrowing. Meanwhile, models (2) and (3) analyze heterogeneity across SN and found that DIF only affected the borrowing behavior of farmers with weak SN, while it had no significant effect on the borrowing behavior of strong ones. This result confirms that DIF in rural areas has eased the farmers' financing constraints by facilitating formal borrowing. This effect is more pronounced in households with weak SN, suggesting that DIF complements informal borrowing, such as humane lending, in traditional rural China.

Table 7 Results with the consideration for sn and farmers' lending channel.

Variable	FF			IF
	Overall	Strong SN	Weak SN	Overall
	Model (1)	Model (2)	Model (3)	Model (4)
Dif	0.425***	0.32	0.486***	0.68
	(0.127)	(0.232)	(0.154)	(0.544)
Family	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes	No
Observation	36960	11052	25908	17645

Note: The standard errors are reported in parentheses. The above regression results are run using Surprised Learning with IV. Due to only the information provision in the CHFS 2023 dataset, the trained sample in Model (4) is from CHFS 2023, thus less than Model(1).* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Mechanism II: credit mitigation

The results in Tables 6 and 7 show how an essential mechanism through which DIF influences FED is providing credit to farmers facing strong credit constraints, thereby promoting entrepreneurship. The next step is to test this mechanism using a mediating effects model. In conjunction with the borrowing information provided by the CHFS data, this paper measures the farmers' credit needs and constraints regarding their experiences of being rejected and approved for borrowing. Table 8 reports the estimation results of the mediating effects model. Models (1) to (3) inform the mediating role of farmers' experience of loan rejection from formal sources in the impact of DIF on FED, and models (4) to (6) report that of loan approval from precious authorities. Findings show that farmers' previous experience of being rejected for loans in FF channels does not constitute a mediator of the impact of DIF on their entrepreneurship, while the approved one is a mediator. This suggests that DIF has increased the probability of farmers accessing loans in the FF sector, easing their credit constraints and promoting entrepreneurship. This finding validates H3, suggesting that DIF in rural areas has increased the farmers' financial accessibility and may also have reduced some of the finance costs, thereby increasing the probability of FED.

Table 8 Mechanism identification of credit mitigation.

Variable	Rejection	FED	FED	Acceptance	FED	FED
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
DIF	-0.045	0.17***	0.17***	0.084*		0.166***
	(0.028)	(0.037)	(0.038)	(0.045)		(0.04)
Availability_fina			0.04***			
			(0.01)			
Availability_finb					0.173***	0.028***
					(0.042)	(0.005)
Family	Yes	Yes	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes	No	No	No
Mediation			-1.1%			1.43%
Observation	36960	36960	36960	16158	16158	16158

Note: The standard errors are reported in parentheses. The above regression results are run using the Surprised Learning with IV. Due to only the information provision in CHFS 2023 dataset, the regression sample in Model (4) to (6) is from CHFS 2023, thus less than Model (1) to (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.3 Mechanism III: digital financial access

After analyzing the mechanisms of credit mitigation, this paper examines the mechanisms of the impact of DIF on FED from the perspective of the specific use of digital finance. DIF is an overall indicator at the macro level, and it is worth discussing further how farmers are affected by this trend. Therefore, based on the comprehensive analysis of the impact of DIF on FED, this paper further explores whether DIF has facilitated FED by increasing its utilization of digital finance. This paper draws on existing literature [41,42] to determine the actual use of digital finance in mobile payments and credit cards. DIF provides a scenario for farmers to use digital finance, and this use will, in turn, improve their financial literacy and financial awareness through information exchange and communication effects, thus promoting their entrepreneurship. The findings in Table 9 show that DIF in rural areas significantly encourages using digital finance, which further enables FED. The results of model (3) show a mediating effect of 6.71%.

Table 9 Mechanism identification of digital financial access.

	Model(1)	Model(2)	Model(3)
DIF	0.06***	0.11***	0.106***
Entrepreneurship	(0.03)	(0.02)	(0.018)
Digital Financial Access			0.124***
			(0.002)
family	Yes	Yes	Yes
region	Yes	Yes	Yes
Time-fixed	Yes	Yes	Yes
Mediation			6.71%
Observation	36960	36960	36960

Note: The standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5. Conclusion and Implication

In rural China, DIF offers more possibilities for the production and livelihood of farmers. Drawing on the big data analytics for CHFS, this paper uses the surprised learning-based regression with IV to examine the impact of DIF on FED and to explore the heterogeneity of this impact and the mechanisms of getting families' unidentified recordings. Compared to the conventional methods, the regression is utilized with the surprised learning-based algorithm following the consideration of IV, which can improve the accuracies of these estimations and quickly determine the influence features of DIF, simultaneously reducing the confounding factors in the decision-made process of FE. The study finds that DIF in rural areas significantly promoted FED, and the effect on self-employment-based entrepreneurship is more pronounced. Meanwhile, the impact of DIF on FED is more pronounced in households with low human capital and low economic development areas, which, to a certain extent, validates the "inclusive" nature of DIF. Additionally, from the perspective of SN, the paper explores how DIF promotes entrepreneurship by increasing the likelihood of farmers' access to FF channels and argues that DIF increases entrepreneurship rates by growing farmers' use of digital finance.

Based on the above findings, this paper makes the following policy recommendations. First, efforts can continue to actively promote the deepening of DIF in rural areas where the construction of the FF market is improved, increase the farmers' financial accessibility, and reduce their financing costs. Secondly, when comprehensively promoting and developing DIF in rural areas, there is a need for targeted publicity, explanation, and promotion based on the farmers' characteristics. For farmers with different levels of human capital and other SN, it is necessary to establish a good platform or mechanism for interaction and information exchange based on their actual borrowing needs and lending behavior to accelerate the digital transformation of rural financial institutions and use big data technology to track, assess and support the farmers' credit needs promptly to serve better their off-farm entrepreneurship and the expansion of agricultural production. Thirdly, FED should be actively encouraged, promoted, and supported to facilitate rural revitalization through local innovation close to home. Meanwhile, attention needs to be paid to the welfare effects of FED, focusing on the farmers' financial and mental stress in entrepreneurship. Enough technical support and assistance should be given to farmers regarding lending, management, and subsequent repayment using digital finance.

Acknowledgement

This research was funded by the Ministry of Education Humanities and Social Sciences Youth Project (Grant No. 23YJC630235), Guangdong Philosophy and Social Sciences Planning Youth Project (Grant No. GD22YGL19), and Guangdong Province Philosophy and Social Sciences Planning, Guangdong Provincial Ordinary University Characteristic Innovation Project (Grant No. 2023WCSCX060).

References

[1] Daming Zhou, and Tian Jie. Social Mobility among Entrepreneurial Migrants: A Case Study of the Dongguan Humen Zhisheng School. Chin. Sociol. Anthropol. 43 (2) (2010)25-50.
[2] Thomakis Zafeirios, and Irene Daskalopoulou. Entrepreneurial views and rural entrepreneurial potential: evidence from Greece. J. Knowl. Econ. 13(2022)11-34.
[3] Pan Guangwei. Promoting Inclusive Finance for Real Economy in China. China Econ. Transit. 2020, pp.113-119.
[4] Su Xin, and H. L. Chen. Research on digital finance in promoting the development of the agricultural industry. Acad. J. Bus. Manag. 37(2021)98-101.
[5] Zhang Ziqi. Research on Chinese Rural Digital Inclusive Finance Innovation Based on Blockchain Technology. 2022 International Conference on Urban Planning and Regional Economy (UPRE 2022). Atlantis Press, 2022.
[6] Jiang Lili, et al. The impact of the inclusive financial development index on farmer entrepreneurship. PloS One. 14(2019) e0216466.
[7] Aisaiti Gulizhaer, et al. An empirical analysis of rural farmers' financing intention of inclusive finance in China: The moderating role of digital finance and social enterprise embeddedness. Ind. Manage. Data Syst. 2019.
[8] Xie Wenwu, Tao Wang, and Xuan Zhao. Does digital inclusive finance promote coastal rural entrepreneurship? J. Coastal Res. 103(2020)240-245.
[9] Hasan Md Morshadul, Lu Yajuan, and Appel Mahmud. Regional development of China's inclusive finance through financial technology. Sage Open. 10(2020)2158244019901252.
[10] Yu Naishu, and Yanzhe Wang. Can digital inclusive finance narrow the Chinese urban–rural income gap? The perspective of the regional urban–rural income structure. Sustainability. 13(2021)6427.
[11] Hasan Morshadul, Thi Le, and Ariful Hoque. How does financial literacy impact on inclusive finance? Financ. Innov. 7(40) (2021)1-23.
[12] Liu Tian, Guangwen He, and Calum G. Turvey. Inclusive finance, farm households entrepreneurship, and inclusive rural transformation in rural poverty-stricken areas in China. Emerg. Mark. Financ. Tr. (57) (2021)1929-1958.
[13] Yu Chuanjiang, et al. Digital inclusive finance and rural consumption structure–evidence from Peking University digital inclusive financial index and China household finance survey. China Agr. Econ. Rev. 14(2022)165-183.
[14] Liu Ziqiang, Yihao Zhang, and Hongyi Li. Digital inclusive finance, multidimensional education, and farmers' entrepreneurial behavior. Math. Probl. Eng. 2021, pp. 1-13.
[15] Hasan Md Morshadul, Lu Yajuan, and Shajib Khan. Promoting China's inclusive finance through digital financial services. Glob. Bus. Rev. 23(2022)984-1006.
[16] Zhao Hongbo, Xiao Zheng, and Lin Yang. Does digital inclusive finance narrow the urban-rural income gap through primary distribution and redistribution? Sustainability. 14(2022) 2120.
[17] Khaleque Abdul. Performance of women entrepreneurs: Does access to finance really matter? Eurasian J. Bus. Econ. 11(2018)23-48.
[18] Xie Wenwu, Tao Wang, and Xuan Zhao. Does digital inclusive finance promote coastal rural entrepreneurship? J. Coastal Res. 103(2020)240-245.
[19] Lorenz Edward, and Sophie Pommet. Mobile money, inclusive finance and enterprise innovativeness: an analysis of East African nations. Ind. Innov. 28(2021)136-159.
[20] Garg A. K., and Neha Pandey. Making money work for the poor in India: Inclusive finance through bank-moneylender linkages. Online Working Paper. 2007.
[21] Chen Boou, and Chunkai Zhao. Poverty reduction in rural China: Does the digital finance matter? Plos One. 16(2021) e0261214.
[22] Yang Xiaolan, Yidong Huang, and Mei Gao. Can digital financial inclusion promote female entrepreneurship? Evidence and mechanisms. N. Am. J. Econ. Financ. 63(2022)101800.
[23] Noor Muhammad, Finnah Fourqoniah, and Muhammad Fikry Aransyah. The Investigation of financial inclusions, financial literation, and financial technology in Indonesia. Jurnal Perspektif Pembiayaan Dan Pembangunan Daerah. 8(2020)257-268.
[24] Liu Tian, Guangwen He, and Calum G. Turvey. Inclusive finance, farm households entrepreneurship, and inclusive rural transformation in rural poverty-stricken areas in China. Emerg. Mark. Financ. Tr. 57(2021)1929-1958.
[25] Ajide, Folorunsho M. Financial inclusion in Africa: does it promote entrepreneurship? J. Finan. Econ. Policy. 12(2020) 687-706.
[26] Xun Zhang, Guanghua Wan, Jiajia Zhang, et al. Digital economy, financial inclusion and inclusive growth. China Econ. 15(2020)92-105.

[27] Liu Yang, et al. Can digital financial inclusion promote China's economic growth? Int. Rev. Financ. Anal. 78(2021)101889.
[28] Xie Wenwu, Tao Wang, and Xuan Zhao. Does digital inclusive finance promote coastal rural entrepreneurship? J. Coastal Res. 103(2020)240-245.
[29] Liu Yang, Chunyu Liu, and Mi Zhou. Does digital inclusive finance promote agricultural production for rural households in China? Research based on the Chinese family database (CFD). China Agr. Econ. Rev. 13(2021) 475-494.
[30] Ji Xuanming, et al. Has digital financial inclusion narrowed the urban-rural income gap: The role of entrepreneurship in China. Sustainability. 13(2021)8292.
[31] Yanfang Niu, Limeng Ying, Jie Yang, et al. Organizational business intelligence and decision-making using big data analytics. Inform. Process. Manag. 58(6) (2021)102725.
[32] Yu Chuanjiang, et al. Digital inclusive finance and rural consumption structure–evidence from Peking University digital inclusive financial index and China household finance survey. China Agr. Econ. Rev. 14(2022)165-183.
[33] Vakkari Pertti. Searching as learning: A systematization based on literature. J. Inf. Sci. 42(1) (2016)7-18.
[34] Wang, Xue, and Guangwen He. Digital financial inclusion and farmers' vulnerability to poverty: Evidence from rural China. Sustainability. 12(2020)1668.
[35] Xie Wenwu, Tao Wang, and Xuan Zhao. Does digital inclusive finance promote coastal rural entrepreneurship? J. Coastal Res. 103(2020)240-245.
[36] Ye Yun, Yongjian Pu, and Ailun Xiong. The impact of digital finance on household participation in risky financial markets: Evidence-based study from China. Plos One. 2022, 17, e0265606.
[37] Li Yuanyuan, et al. Does digital finance promote entrepreneurship of migrant? Evidence from China. Appl. Econ. Lett. 29(2022)1829-1832.
[38] Liu Jiamin, et al. Can New-Type Urbanization Construction Narrow the Urban–Rural Income Gap? Evidence from China. Sustainability. 14(2022)14725.
[39] Gnecco Giorgio, Nutarelli Federico, Selvi Daniela. Optimal data collection design in machine learning: the case of the fixed effects generalized least squares panel data model. Machine Learning. 7(2021), 1549-1584.
[40] Gan Jiangzhang, Wen Guoqiu, Yu Hao, Zheng Wei, Lei Cong. Supervised feature selection by self-paced learning regression. Pattern Recognition Letters. 132(2020), 30-37.
[41] Han Guyuan, Youfei Hu, and Yuhan Zhang. Digital inclusive finance, innovation and entrepreneurship. Sci. J. Econ. Manag. Res.2019, 1.
[42] Li Jie, Yu Wu, and Jing Jian Xiao. The impact of digital finance on household consumption: Evidence from China. Econ. Model. 86(2020)317-326.
[43] Ye Yun, Yongjian Pu, and Ailun Xiong. The impact of digital finance on household participation in risky financial markets: Evidence-based study from China. Plos One. 17(2022) e0265606.
[44] Liu Ying, Haoran Zhao, Jieguang Sun, et al. Digital Inclusive Finance and Family Wealth: Evidence from LightGBM Approach. Sustainability. 14(2022)15363.
[45] Zhang Miao, Juanjuan Yang. Research on financial technology and inclusive finance development. 2018 6th International Education, Economics, Social Science, Arts, Sports and Management Engineering Conference (IEESASM 2018). Atlantis Press, 2019.
[46] Heping Ge, Bowen Li, Decai Tang, et al. Research on digital inclusive finance promoting the integration of rural three-industry. Int. J. Env. Res. Pub. He. 19(2022)3363.
[47] Zitian Fu, Yujiao Zhou, Weifeng Li, et al. Impact of digital finance on energy efficiency: Empirical findings from China. Environ. Sci. Pollut. Res. 30(2022)13-35.
[48] Fei Li, Yufei Wu, Jinli Liu, et al. Does digital inclusive finance promote industrial transformation? New evidence from 115 resource-based cities in China. Plos One. 17(2022) e0273680.
[49] Jingrong Li, Bowen Li. Digital inclusive finance and urban innovation: Evidence from China. Rev. Dev. Econ. 26(2022)1010-1034.
[50] Liu Yang, Chunyu Liu, and Mi Zhou. Does digital inclusive finance promote agricultural production for rural households in China? Research based on the Chinese family database (CFD). China Agr. Econ. Rev. 13(2021)475-494.
[51] Qian Huang, Hongzhi Zhu, Huiyuan Su. Research on the impact of digital inclusive finance on farmers' entrepreneurial choice. Fujian Forum (Humanit. Soc. Sci. Ed.).08(2021)85-102.