Research Article

An Empirical Analysis of Loan Repayment Behavior and Default Rates on Digital Lending Platforms: Evidence from an Emerging Market

Edmund Mallinguh[1], Christopher Wasike[2]

1. Department of Accounting and Finance, Kiriri Women's University of Science and Technology, Kenya; 2. Department of Business Administration, Kiriri Women's University of Science and Technology, Kenya

The rapid growth of digital lending platforms has revolutionized access to credit in emerging economies, providing financial opportunities to underserved individuals. Despite these advancements, high loan default rates remain a substantial challenge, threatening the sustainability of these lending platforms and the broader financial ecosystem. This study examines the determinants of loan default among users of mobile-based digital lending platforms in Kenya, a leading adopter of digital financial services. Employing a sample of 161 responses from borrowers, the analysis uses a probit regression model to identify primary predictors of loan default. The study focuses on borrower demographics like age and place of residence and loan characteristics, such as repayment term (period) and interest rates. The results reveal that older customers and those in rural areas are more likely to default on their loans, whereas longer repayment periods reduce the probability of default. Interestingly, the interest rate does not significantly affect default behavior, implying that borrowers prioritize access to credit over its cost. These results underscore the essence of tailoring digital lending practices to the diverse borrower needs and circumstances. By addressing the factors that drive loan defaults, digital platform lenders can improve risk assessment frameworks, enhance financial inclusion, and enhance the sustainability of digital lending platforms. The findings provide actionable insights for policymakers, lenders, and other industry players seeking to mitigate default risks and foster a more robust digital financial ecosystem.

Corresponding author: Edmund Mallinguh, eddie.mallie@gmail.com

1. Introduction

The rapid growth of digital lending platforms, particularly mobile phone-based platforms, has revolutionized access to credit in emerging economies[1]. These platforms have bridged the gap for underbanked populations, offering quick and convenient financial solutions in emerging economies like Kenya. With the advent of mobile-based digital lending platforms, individuals previously excluded from formal financial systems now have unprecedented access to credit. These digital platforms leverage technology to provide quick and convenient loans, bypassing traditional requirements like collateral or extensive credit history. Despite their transformative potential, digital lending platforms face challenges, notably high default rates, threatening their sustainability and impact. Understanding the determinants of loan default in this context is crucial for enhancing the efficacy of these lending platforms and promoting financial inclusion, especially in markets with weak regulation[2].

Previous research has extensively explored loan default prediction using traditional banking datasets. However, there is still limited focus on digital lending platforms, especially in developing markets, thus requiring further research[3]. These digital lending platforms face unique challenges, including limited borrower credit histories, high transaction costs, and operational risks. Besides, demographic and economic factors like age, repayment term, interest rate, and place of residence remain understudied in this context. Addressing these gaps is critical for tailoring credit assessment models to the local market and ensuring financial inclusion without compromising profitability.

Kenya presents a unique case for analyzing digital lending. The country has one of the highest mobile penetration rates in Africa, and mobile money platforms (like M-Pesa, Artel Money, and Orange Money) have laid the groundwork for the widespread adoption of digital financial services. Nonetheless, the ease of access to credit raises concerns about over-indebtedness and default rates, particularly among vulnerable borrowers. Loan default affects lenders' profitability and undermines trust in digital financial systems, posing broader economic implications. Identifying the factors influencing default behavior is essential for designing effective risk management strategies and fostering sustainable growth in the digital lending sector[4].

Loan default remains a pressing challenge for digital lending platforms in growing economies like Kenya. Whereas these platforms have democratized access to credit, their over-reliance on unconventional data sources and algorithms for credit risk assessment has yielded mixed results. Higher default rates compromise the financial stability of digital lending platforms and deter potential investors, limiting their capacity to serve a broader population. Investors in digital loan markets

doi.org/10.32388/DKJLUJ

demonstrate the ability to differentiate between risk and profit[5]. The hurdle is further compounded by the lack of detailed empirical studies exploring the specific factors that drive loan default in the context of digital lending in Kenya.

This study will employ econometric modeling to examine the predictors of loan default among users of digital lending platforms in the Kenyan context. By focusing on key variables like age, repayment term, interest rate, and place of residence, the study aims to uncover actionable insights that can inform policy and operational strategies. Our findings contribute to the growing literature on digital lending and offer practical recommendations for enhancing credit risk management in this rapidly evolving sector.

The paper is organized as follows: Section 2 is divided into two main parts: materials and methods. The materials part explores empirical literature on digital lending, a brief analysis of the Kenyan context, and hypotheses formulation. The methods part has different subsections; the first is about data collection, ethics, sampling, and variable description. The other subsection focuses on developing the econometric model employed in the study by developing equations determining loan default by digital borrowers. In Section 3, we present empirical findings outlining significant loan default predictors on digital platforms, presenting specific contributions of each factor through marginal effect analysis. Section 4 discusses the empirical results, links these findings to current literature, and offers explanations where none exist. In the final Section 5, we present our conclusions.

2. Literature Review

2.1. Existing Literature on Digital Lending

Digital lending has revolutionized the financial sector, leveraging technology to streamline loan processes and improve access to credit. The innovation has gained traction globally, with various regions adopting different approaches, such as online platforms and mobile money services. Primary drivers include filling credit gaps, data analytics, and regulatory support, contributing to financial inclusion and economic growth. In emerging markets, digital lending platforms are crucial for funding MSMEs, offering faster and more efficient credit evaluation than traditional methods[1]. The digitalization of lending processes has increased speed, accuracy, and efficiency in origination and renewal. While digital lending offers numerous advantages for both banks and consumers, challenges persist, particularly in developing economies, where solutions are needed to address obstacles faced by clients[6][7].

doi.org/10.32388/DKJLUJ

Factors influencing loan repayment differ across studies. For instance, age is irrelevant in predicting defaults in the Ghanaian market but significant in Tanzania, where younger borrowers between 18 and 24 years are considered more reliable[8][9]. Inasmuch as digital credit enhances borrowing opportunities for borrowers excluded from conventional markets, it accounts for 90 percent of blacklistings in Kenya due to high default rates and stricter consequences for defaults than other credit markets[10]. These findings underpin the complex dynamics of digital lending and the need for careful consideration of borrower characteristics and loan terms.

Research on rural and urban borrowers and digital loan default reveals crucial insights. Rural borrowers are less likely to default on small business loans than urban borrowers, with social capital significantly reducing default rates[11]. Interest rates, loan size, loan diversion, and collateral significantly determine default rates, particularly for low-income urban borrowers[12]. Credit rationing criteria can influence loan access and repayment performance, with the technique effectively identifying a third of creditworthy borrowers in one study. Tightening loan contract terms, like reducing grace periods and rejecting applications with long processing times, may boost the pool of creditworthy borrowers[13].

Recent evidence highlights the complexity of loan default prediction and its determinants. Novel multiperiod default prediction techniques that outperform traditional methods in accurately forecasting time-to-default are necessary[14]. The loan repayment period is a significant factor in default risk, with longer-term loans showing substantially higher default rates than shorter-term loans. Despite substantial loan demand, many digital borrowers are unaware of loan terms and incur late fees. While brief financial literacy intervention improves knowledge, it may not significantly increase timely repayments and may modestly increase loan demand and default likelihood[15]; these findings underscore the importance of considering multiple variables and time horizons when assessing loan default risk.

Moreover, existing studies on digital loans and interest rates reveal that loan and borrower-specific factors significantly impact interest rates and default status across online peer-to-peer lending platforms. Formal regulations increase interest rates for small microloans and supply shortages[12][16]. Interest rates play a crucial role in default probability, with higher interest rates increasing default risk and lower base rates potentially leading banks to lend to riskier borrowers. Besides, unclear regulations on interest rate limits in online lending create uncertainty in determining rates[17][18].

That notwithstanding, higher default rates are a concern, with empirical evidence suggesting that increasing loan delivery time may decrease default likelihood. High default rates and potential predatory behaviors raise consumer protection concerns[19]. Further studies are necessary to explore digital loan default, factoring in market differences and dynamics.

2.2. Contextual Analysis Specific to the Kenyan Market

Digital lending has become increasingly prevalent in Kenya's financial sector. While it offers easier access to loans, it also raises concerns about regulation and consumer protection. Commercial banks have embraced digital lending, positively impacting technical efficiency. The correlation between digital lending duration and non-secured loans is positive and significant, while digital lending costs negatively correlate with non-secured loans[20][21]. Despite their popularity, most digital lenders operate in an unregulated environment, potentially exploiting customers through unfavorable interest rates, unfettered access to personal data, and unethical debt collection techniques.

Digital lending is expanding access to credit for underserved communities, particularly in rural areas[22]. There is evidence of regional disparities in loan accessibility and default behavior between urban and rural-based borrowers. Digital credit costs are comparable to traditional bank lending rates. Factors influencing digital loan uptake include employment status and loan application processes, with age and loan considerations having less significant effects[23]. Evidence demonstrates a high demand for digital loans, particularly among low-income earners and micro-entrepreneurs, owing to their accessibility and speed compared to traditional banking systems.

Unfortunately, default rates in the country's digital lending space have been a growing concern. High interest rates, limited financial literacy, and unemployment have been linked to increased default probabilities. Besides, concerns about predatory lending practices by digital lenders persist due to weak consumer protection and low financial literacy among borrowers[24].

2.3. Hypotheses Formulation

The formulation of hypotheses provides a structured framework for investigating the relationships between borrower characteristics and the likelihood of loan default. The hypotheses are grounded in theoretical insights and contextual factors particular to the Kenyan digital lending market. The factors considered, namely, age, repayment term, interest rate, and place of residence (rural or urban), are important in understanding borrower behavior and risk assessment. The study seeks to test the absence of significant relationships, providing a rigorous foundation for analysis and interpretation by focusing on null hypotheses.

- H_1: *The age of a borrower on digital platforms does not significantly affect the likelihood of loan default.*

Research on age and financial decision-making reveals complex relationships. Whereas some studies demonstrate that financial sophistication peaks in the early 50s, others suggest that basic financial skills remain stable or improve with age. However, more complex financial decisions are challenging for older adults. Age-related cognitive decline is associated with decreased financial literacy and numeracy[25]. Be that as it may, older adults often benefit from greater experience-based knowledge and lower negative emotions about financial decisions, which can positively impact their decision-making.

However, within the context of this study, the null hypothesis assumes no significant difference in default likelihood across age groups.

- H_2: *The repayment term (period) of digital loans has no meaningful effect on the likelihood of loan default.*

Digital loan repayment periods significantly impact borrower behavior and loan outcomes. The findings of a Mexican study suggest that shorter repayment periods can lead to higher default rates[12]. Likewise, in Kenya, small businesses prefer longer repayment periods, allowing for better cash flow management and increased loan uptake[26]. This hypothesis assumes no impact from repayment term duration, serving as a baseline for testing this variable. The null hypothesis assumes no impact from repayment term duration, serving as a baseline for testing this variable.

- H_3: *The cost of digital credit (Interest rate) substantially impacts the likelihood of loan default by the borrower.*

Research suggests a positive relationship between loan interest rates and default probability. Higher interest rates increase loan default probability, particularly for business borrowers with low profit margins. Likewise, lower base interest rates can lead banks to lend to riskier borrowers, potentially increasing default rates[17]. Although higher interest rates may exacerbate default risks, the null hypothesis posits no significant relationship between interest rates and default behavior, ensuring an unbiased examination of this variable.

- H_4: *Place of residence, whether rural or urban, does not significantly influence the likelihood of loan default.*

Place of residence is a proxy for access to financial resources and economic opportunities. While urban borrowers might enjoy greater financial inclusion, rural borrowers could face unique hurdles. The null hypothesis assumes no significant differences in default rates between rural and urban residents.

3. Methodology

3.1. Data Collection

3.1.1. Description of the Questionnaire and Data Collection Process

The data for this study were collected through an online survey targeting individuals who have used digital lending platforms, specifically mobile phone-based platforms, in Kenya. The questionnaire was structured to capture quantitative and qualitative information concerning the borrowing behavior, demographic characteristics, and perceptions of digital lending users. The survey instrument included closed-ended questions focusing on key aspects of digital lending, including demographic information, economic background, loan details, credit scoring, repayment behavior, and user satisfaction.

To ensure data quality, the questionnaire underwent pre-testing with a small group of respondents (approximately 10) to pinpoint and handle ambiguities in question phrasing or structure. Based on the feedback, minor revisions were made to improve the tool's clarity. Data collection was conducted over one month using a secure online platform, ensuring the anonymity and confidentiality of responses. The tool had a screening question; respondents who had never used digital lending platforms were terminated from the survey.

3.1.2. Sample Techniques and Size

The study employed a purposive sampling technique to target respondents with prior experience using digital lending platforms. This approach was necessitated by the study's focus on understanding loan default determinants in the context of digital borrowing. A total of 283 responses were initially collected. However, after applying a filtering question to exclude individuals who had not used digital lending services, the dataset was reduced to 161 valid responses.

The sample size of 161 respondents is sufficient to provide meaningful insights into the study's research questions while accounting for the statistical power required for the econometric analysis. Respondents were distributed across different demographic groups, including urban and rural residents, to capture diverse perspectives.

3.1.3. Ethical Considerations in Data Collection

Ethical considerations were integral to the data collection process. Respondents were provided with an information sheet detailing the study's purpose, rights, and measures taken to ensure data confidentiality. Informed consent was obtained before respondents participated in the survey, with assurance that their data would only be used for academic research purposes and stored securely to prevent unauthorized access. The study also adhered to the principles outlined in the General Data Protection Regulation (GDPR) and similar local laws governing data protection and privacy, as suggested by other researchers[27].

3.2. Variable Description

The study focused on variables derived from the literature and contextual understanding of digital lending platforms in Kenya. The dependent variable is Loan Default (Binary: 0 = No Default, 1 = Default). This variable captures whether a respondent had defaulted on a loan taken via digital lending platforms. This binary outcome is critical in modeling the factors influencing loan repayment behavior. There are four independent factors in this study. Age represents the respondent's age in years.

Age was hypothesized to influence loan repayment ability, as younger individuals might have different financial behaviors than older ones. Repayment Term indicates the duration of the loan repayment period in months. This variable measures the flexibility of repayment schedules and their association with default likelihood. Place of Residence (Categorical: Rural = 0, Urban = 1) shows whether the respondent resides in a rural or urban area. This variable reflects access to financial resources and economic opportunities, which may affect repayment capacity. The last predictor is the interest rate charged on loans as a percentage. Higher interest rates were hypothesized to increase the probability of default.

3.3. Econometric Modeling Approach

3.3.1. Model Specification

The primary objective of the econometric analysis was to identify factors influencing loan default among users of digital lending platforms. The study adopted the probit model to predict digital loan default. The independent variable (Loan default) is dichotomous, making the Probit model ideal for probability estimation. The Probit model provides a way to compute marginal effects, allowing for intuitive interpretations of the practical significance of predictor variables. The technique aligns with prior research on binary financial outcomes such as default behavior and credit risk modeling. The model has been adopted in finance and economics[28].

The Probit model estimates the probability (Y_i) of loan default as:

doi.org/10.32388/DKJLUJ

$$Y_i = \int \frac{1}{0} \frac{\textit{if the borrower defaults on the loan,}}{\textit{Otherwise.}} \tag{1}$$

The latent variable model underlying the probit regression is given by:

$$Y_i^* = (\beta_o + \beta_1 Age + \beta_2 Repayment\ Term + \beta_3 Place\ of\ Residence + \beta_4 Interes\ Rate + \epsilon_i) \tag{2}$$

Where:

a. Y_i*: the unobserved continuous latent variable that determines loan default

b. β_o: the intercept,

c. $\beta_o,\ \beta_1,\ \beta_2,\ \beta_{3,},\ \beta_4$: the coefficients associated with explanatory variables.

d. ϵ_i: The error term is assumed to follow a standard normal distribution.

The observed outcome, P_i is related to the latent variable as follows:

$$(Y_i) = \int \frac{1}{0} \frac{\textit{if the borrower defaults on the loan,}}{\textit{Otherwise}} \tag{3}$$

$$Y_i = \int \frac{1}{0} \frac{if\ P_i^* > 0,}{if\ P_i^* < 0.} \tag{4}$$

The probability of loan default is modeled as follows:

$$P\left(Y_i\ = 1|X_i\right) = \phi\left(\beta_o + \beta_1 Age + \beta_2 Repayment\ Rate + \beta_3 Place\ of\ Residence + \beta_4 Interest\ rate\right) \tag{5}$$

Where $\Phi(\cdot)$ represents the cumulative distribution function (CDF) of the standard normal distribution.

3.3.2. Marginal Effects

The marginal effects, derived from the probit model, quantify the impact of a one-unit change in an independent variable on the probability of loan default. For continuous variables, the marginal effect is:

$$\frac{\partial P\left(Y_i\ = 1|X_i\right)}{\partial P_{ij}} = \phi\left(X_i'\beta\right).\beta_j \tag{6}$$

Where $\phi(\cdot)$ is the probability density function (PDF) of the standard normal distribution and β_j is the coefficient of the j^{th} independent variable. For binary independent variables, the marginal effect is given as:

$$P = (Y_i\ = 1|X_{iJ} = 1, X_{i-J}\ ) - (Y_i\ = 1|X_{iJ} = 0, X_{i-J}\ ) \tag{7}$$

Where X_{i-J} denotes all other covariates held constant.

doi.org/10.32388/DKJLUJ

3.3.3. Model Estimation

The study estimated the model using maximum likelihood estimation (MLE), which optimizes the log-likelihood function:

$$\ln L = \sum_{i=1}^{N} [Y_i \ln(\phi \left(X_i' \beta\right)) + (1 - Y_I) \ln(1 - \phi \left(X_i' \beta\right))], \tag{8}$$

Where N is the total number of observations.

3.3.4. Robustness Checks

Robustness checks were conducted to ensure the model's reliability. Specifically, the variance inflation factor (VIF) was calculated to confirm the absence of multicollinearity issues among independent variables. Moreover, robust standard errors were estimated using the Huber-White sandwich estimator.

4. Results

4.1. Descriptive Statistics

The dataset comprises 161 valid responses from an online survey targeting users of digital lending services. The survey initially attracted 280 responses. However, responses from individuals who indicated they had never used digital lending services were terminated at an initial screening question to focus exclusively on relevant participants. This filtering ensured the study's findings directly apply to users with firsthand experience of such platforms.

Table 1 presents key descriptive statistics for selected variables used in the analysis. Continuous variables were summarized using central tendency (mean, median), dispersion (standard deviation, minimum, maximum), and correlation. Frequency distribution and proportion are used for categorical ones. The average respondent age is approximately 35 years, with most respondents between 21 and 65. The repayment period ranges from 1 to 12 months, with an average duration of 6.8 months. Interest rates range between 10 percent and 30 percent, with a mean of 18.5 percent, suggesting considerable variation in borrowing terms across respondents.

Additionally, the table summarizes the distribution of participants by their place of residence, revealing that a significant majority (73.91%) reside in urban areas, with only 26.09% in rural regions. These statistics provide an overview of the sample characteristics and lay the groundwork for understanding the subsequent analytical results. This distribution highlights the dominance of urban respondents in the sample, suggesting the higher adoption of digital lending platforms in urban areas.

Variable, N = 161	**Mean**	**Std. Dev.**	**Min**	**Median**	**Max**	**(1)**	**(2)**	**(3)**
Age (Years) (1)	35.4	10.2	21	34	65	1		
Term (Months) (2)	12.12	9.54	1	6	12	-0.155	1	
Interest (Monthly %) (3)	18.5	5.2	10	18	30	0.051	-0.012	1
Place of Residence	*Proportion (%)*							
a) Rural	42							
b) Urban	119							

Table 1. Descriptive Statistics

(1)—Age, (2)—Repayment Term, (3)—Interest Rate

Other variables, such as monthly income, loan amount, and awareness of credit scores, were initially considered but excluded from the final model due to multicollinearity or statistical insignificance. Nevertheless, these descriptive insights from the tables serve as a foundation for the econometric analysis, helping to contextualize the relationships examined in the subsequent sections.

4.2. Probit Model Estimation

The Probit regression results suggest that Age, Repayment Term, and Place of Residence are significant predictors of loan default behavior. At the same time, Interest Rate has no statistically significant relationship with loan default. The detailed results are presented in Table 2 below, showing the estimated coefficients, standard errors, z-values, p-values, and significance levels.

Variable	Estimate	Std. Error	z-value	*p*-value
Intercept	-0.313	0.564	-0.555	0.578
Age	0.028*	0.012	2.226	0.026
Interest Rate	0.010	0.012	0.820	0.412
Repayment Term	-0.254*	0.101	-2.504	0.012
Place of Residence (Urban)	-0.650*	0.320	-2.032	0.042

Table 2. Probit Regression Output

**Significance levels:* **p* < 0.05.*

The coefficient for Age ($\beta_1 = 0.028,\ p = 0.026$) is statistically significant—the positive relationship suggests that as the borrower's age increases, the likelihood of loan default slightly increases. This could be attributed to older individuals facing higher financial pressures, or it could reflect a cohort effect where older borrowers may have different repayment behavior than younger borrowers. The coefficient for Place of Residence ($\beta_2 = -0.650,\ p = 0.042$) indicates statistical significance at 5%. The negative coefficient suggests that borrowers in urban areas are less likely to default than in rural settings. This could reflect greater financial stability in urban areas due to better access to income-generating opportunities or financial services found in such places than in rural areas.

The coefficient for Interest Rate or the cost of digital loans ($\beta_3 = 0.010,\ p = 0.412$) suggests that the variable is not statistically significant in predicting loan default. This implies that, within this model, the interest rate has no strong effect on the likelihood of loan default. This finding could be surprising, as one might expect higher interest rates to increase default risk; however, this relationship may not hold in this dataset or be masked by other variables in the current model. The coefficient for Place of Residence ($\beta_4 = -0.650,\ p = 0.042$) indicates statistical significance at 5%. The negative coefficient suggests that borrowers in urban areas are less likely to default than in rural settings. This could reflect greater financial stability in urban areas due to better access to income-generating opportunities or financial services found in such places than in rural areas.

The Probit model provides valuable insights into the factors influencing loan default behavior. Among the variables tested, Age, Repayment Term, and Place of Residence show significant relationships with the likelihood of loan default, while Interest Rate does not. These findings contribute to a better understanding of the dynamics behind loan default and can inform lending policies to reduce default risks.

4.3. Marginal Effects Analysis

While the Probit model estimates the likelihood of loan default based on various explanatory variables, the marginal effects provide a more intuitive understanding of how changes in these variables impact the probability of loan default. The marginal effects reflect the change in the probability of the dependent variable (loan default) for a one-unit change in the independent variable, holding all other variables constant. This allows for a clearer interpretation of the practical significance of each predictor.

4.3.1. Model Results and Interpretation

The marginal effects were calculated to assess how the predictors influence the probability of loan default. The results are presented in Table 3, which includes the average marginal effects (AME), standard errors, z-values, p-values, and significance levels.

Variable	AME	Std. Error	z-value	*p*-value
Age	0.0091	0.0039	2.343	0.019
Interest Rate	0.0032	0.0038	0.825	0.409
Repayment Term	-0.0834	0.0316	-2.638	0.008
Place of Residence (Urban)	-0.2136	0.1008	-2.119	0.034

Table 3. Marginal Effects Output

AME denotes Average Marginal Effect; **Significance levels:* $^{**}p < 0.01$, $p < 0.05$.

The marginal effect of Age (0.009) implies that for each additional year of age, the probability of loan default increases by 0.91%. This positive effect indicates that older individuals are more likely to default on loans, holding other factors constant. While the increase in default probability per year is small, it is statistically significant (*p = 0.019*), suggesting a meaningful relationship between age and loan default behavior. The marginal effect for Interest Rate (*0.003)* illustrates that for each one percentage point increase in the interest rate, the probability of loan default increases by only *0.32%.* However, the (*p = 0.409)* suggests that this effect is not statistically significant, implying that Interest Rate has no meaningful impact on loan default probability when other variables are considered. The finding is consistent with the Probit model results, where Interest Rate did not emerge as a significant predictor.

The marginal effect for the Repayment Term (*-0.083)* indicates that for each additional unit increase in the repayment term, the probability of loan default decreases by *8.34%.* This negative relationship suggests that longer repayment periods reduce the likelihood of default because they allow borrowers more time to manage repayment and avoid defaulting on their loans. The effect is statistically significant *(p = 0.008),* reinforcing the Repayment Term's importance in reducing the likelihood of loan default. The marginal effect for Place of Residence (*-0.214*) suggests that urban residency lowers the probability of loan default by *21.36%* compared to the rural one. This significant negative effect *(p = 0.034*) indicates that urban borrowers are substantially less likely to default on their loans than rural borrowers, potentially due to better access to financial resources, job opportunities, and other forms of economic stability available in urban settings than in rural ones.

The marginal effects analysis provides a deeper understanding of the practical implications of the Probit model results. The findings show that Age, Repayment Term, and Place of Residence are significant determinants of loan default. Age enhances the likelihood of default, while longer repayment terms and urban residence significantly reduce the probability of default. The Interest Rate's non-significance in the marginal effects analysis supports the earlier Probit model results, indicating that Interest Rates have no substantial impact on default behavior.

These marginal effects are crucial for understanding how each factor directly influences the probability of loan default in practical terms. They assist policymakers and lenders in developing targeted interventions more effectively to reduce default levels.

4.4. Robustness Checks

Table 4 represents the results of checks to confirm that the findings are not sensitive to model specifications, estimation methods, or potential data irregularities.

doi.org/10.32388/DKJLUJ

4.4.1. Likelihood Ratio Test

The likelihood ratio test compared the full probit model (with all key variables) to a nested model that excluded secondary variables. The Chi-Square statistic = *0.00* and p = 1.00 suggest no improvement in model fit between the two specifications. Thus, additional terms do not significantly enhance the model's explanatory power. As a result, the refined model remains the preferred specification for analysis due to its parsimony and interpretability.

Robustness Check	**Metric**	**Value**	**95% CI**	**Interpretation**
Likelihood Ratio Test	Chi-Square Statistic	0.00	-	No improvement between the full and nested models (*p = 1.00)*.
Cross-Validation (5-fold)	Accuracy	65.8%	[61.2%, 70.4%]	Reasonable predictive performance, highlighting out-of-sample stability.
	Kappa	0.21	[0.14, 0.24]	Moderate agreement between predicted and actual classifications.
Heteroskedasticity-Robust Standard Errors	Z-Statistic	Age: 2.455, Rep Term: -2.561 Residence: -2.068	-	Robust to potential heteroskedasticity, confirming the reliability of results.
Alternative Model Specification (Nested Model)	Log-Likelihood	-93.039	-	Consistent coefficients and significance levels with the refined model.

Table 4. Robustness Checks Results

CI—Confidence Interval; Rep Term—Repayment Term; Residence—Place of Residence

4.4.2. Cross-Validation

A 5-fold cross-validation procedure was employed to evaluate the probit model's predictive performance. The results yielded an average accuracy of 65.8% (*at 95% CI = 61.2%—70.4%)*. The findings suggest reasonable predictive ability for out-of-sample data. The moderate Kappa statistic *(0.21)* reflects the model's ability to distinguish between loan default and non-default cases beyond random chance. The cross-validation results highlight the model's practical utility in classification tasks and ensure that the findings are generalizable to other samples beyond the current dataset.

4.4.3. Heteroskedasticity Checks

Robust standard errors were calculated using the heteroskedasticity-consistent covariance matrix estimator (HC0) method to account for potential heteroskedasticity in the data. These robust errors confirmed the statistical significance of key variables: Age positively affects loan default probability *($z = 2.455$, $p < 0.05$)* while Repayment Term (*$z = -2.561$, $p < 0.05$*) and Place of Residence—Urban *($z = -2.068$, $p < 0.05$)* are negatively associated with default. The consistency of results with robust standard errors underscores the reliability of the findings, even in the presence of potential heteroskedasticity.

4.4.4. Alternative Model Specification

A nested model was fitted to assess whether excluding specific variables would alter the results. The log-likelihood value for the nested model *(-93.039)* is identical to the refined model, suggesting no loss in model fit. Coefficients for the primary predictors remained consistent in magnitude and statistical significance, reinforcing the robustness of the refined model specification.

Therefore, the robustness checks demonstrate that the probit model is well-specified and produces reliable estimates. The likelihood ratio test confirms the sufficiency of the refined model, while cross-validation indicates its practical predictive utility. Adjusting for heteroskedasticity ensures the validity of significance tests, and alternative specifications highlight the stability of the results.

4.4.5. A note about the Hypotheses

The first hypothesis postulated that age does not significantly affect the likelihood of loan default. The findings indicate that age substantially influences the likelihood of loan default, with older borrowers more likely to default than younger borrowers. Therefore, the study rejects the null hypothesis. This suggests that age is an important variable for digital lending platforms when assessing default risk. The second hypothesis stated that the repayment term does not significantly affect the likelihood of loan default. The analysis shows that the repayment term is associated substantially with loan default. Borrowers with longer repayment terms were less likely to default than those with shorter terms. Hence, the null hypothesis is rejected.

The third hypothesis states that interest rate substantially affects the likelihood of loan default. The results confirm that interest has no impact on default likelihood. Consequently, the study fails to reject the null hypothesis. The last one hypothesized that the place of residence does not significantly affect the likelihood of loan default. The analysis demonstrates that place of residence substantially affects loan default. Borrowers in rural areas have a higher default probability than urban residents, possibly due to limited access to financial resources and economic opportunities. Thus, the null hypothesis fails to hold.

5. Discussion

5.1. Implications of Findings

The probit regression model and marginal effects findings offer critical insights into the factors influencing loan default in digital lending platforms.

The positive correlation between age and loan default suggests that older individuals may be more likely to default. This may be attributed to decreased income stability, health-related issues, or a preference for less formal financial arrangements. For digital lending platforms, the finding demonstrates the significance of designing more tailored loan products for older borrowers, such as lower amounts or extended repayment terms, to minimize default risk. Digital lenders might also consider offering financial literacy programs targeted at older borrowers, ensuring they comprehend the implications of digital borrowing. These findings mirror similar results in the Tanzanian market[8].

The negative association between urban residence and loan default is significant, indicating that borrowers in urban areas have a lower probability of defaulting on loans. This is consistent with the idea that urban residents may have better access to steady income sources, higher financial literacy, and more diversified economic opportunities than their rural counterparts. This could imply the need to develop differentiated strategies for rural and urban clients for digital lending platforms. Rural borrowers may face more hurdles related to income stability and access to credit, leading to higher default rates. Thus, platforms targeting rural areas should offer more flexible repayment schedules or risk-adjusted loan conditions. The findings contrast those by Burlando et al.[12] on a contextual basis.

The significant negative impact of repayment terms on loan default suggests that longer repayment periods are associated with lower default probabilities. The finding aligns with financial theory, which posits that extended periods provide borrowers with more manageable repayment schedules, lowering the likelihood of default. Digital lending platforms may adopt this insight to offer more customized repayment terms based on the borrower’s financial ability. Shorter repayment terms might be more suitable for borrowers with stable incomes, whereas longer terms could be ideal for those in precarious economic situations. The findings are supported by existing literature in the Kenyan Market[26].

5.2. Insights into the Significance of Variables

The significant effect of age on loan default suggests that financial institutions should carefully consider the demographics of their borrowers. Older customers may face specific hurdles that increase their likelihood of default, including fixed incomes, health issues, and less familiarity with digital platforms. With digital lending becoming more prevalent, platforms may need to implement strategies to address the needs of older borrowers, such as financial programs and tailored loan products. Understanding age-related risk factors will also assist digital lenders in designing risk-adjusted interest rates.

Besides, the effect of place of residence further underscores the importance of geographic and socio-economic factors in predicting loan default. A lower probability of loan default for urban residents may indicate their access to well-paying jobs and more formal financial services. Contrarily, rural borrowers often have limited access to financial products and more irregular incomes, which face a higher risk of default. This highlights the need for digital lending platforms to adjust their loan offerings based on the residence location of their clients. For instance, digital lenders targeting rural-based customers might benefit from offering smaller loan amounts with flexible repayment schedules and personalized support.

Be that as it may, the significant negative relationship between repayment period and loan default reinforces the importance of providing flexible repayment plans that accommodate the financial capabilities of digital borrowers. A longer repayment term allows customers to spread out their payments, reducing the strain on their cash flow and lowering the chances of default. Digital lending platforms may consider providing more flexible lending options, with varying repayment schedules based on the individual borrower’s ability to repay. Tailored repayment periods can also reduce the risk of borrowers defaulting and enhance their satisfaction with the platform’s services.

Finally, there was no strong relationship between interest rates and digital loan default, suggesting that credit access is prioritized over cost. These results lend confidence to the model's utilization in informing policy decisions, such as tailoring digital lending strategies to demographic and socioeconomic characteristics (e.g., age and place of residence). The robustness of the results supports their application in real-world decision-making and further academic research.

6. Conclusion

The study explored the determinants of loan default among users of digital lending platforms in Kenya, focusing on key factors like age, repayment term, interest rate, and place of residence. The findings demonstrated that age significantly influences digital loan default behavior, with older borrowers exhibiting a higher likelihood of default. Similarly, repayment terms (period) emerged as a significant determinant, where longer terms reduced the probability of default. The place of residence, specifically rural areas, was also found to increase default risk, underpinning the impact of geographic disparities on credit performance. Conversely, interest rates had no meaningful effect on loan default, suggesting that borrowers prioritize access to credit over cost considerations in the digital lending context.

The results have important implications for various stakeholders in the digital lending ecosystem. Digital lenders should consider demographic and geographic factors in credit risk assessment. Tailored strategies, like flexible repayment terms and localized risk assessments for rural borrowers, help mitigate default risks. Given the increasing penetration of digital lending platforms, regulatory structures should ensure that credit products are designed to balance financial inclusion with responsible lending practices. Policies promoting financial literacy among older borrowers and rural populations could help reduce default rates. As an investor, grasping the determinants of loan default can help assess the risk profile of digital lending portfolios and inform investment decisions.

While this study provides valuable insights, certain limitations should be noted. The reliance on self-reported survey data introduces the potential for response bias, especially regarding sensitive financial information. The data were collected at a specific time, and evolving market dynamics or economic conditions may affect borrowing behavior over time. Specific factors, such as income stability, education level, and behavioral traits, were excluded from the analysis and could have contributed to default behavior. Future research could explore other areas like diverse regional contexts, which could validate the findings and provide a more comprehensive understanding of default determinants. Also, behavioral and psychometric variables in credit risk models should be considered for deeper insights into borrower decision-making.

In conclusion, this study underscores the significance of demographic and geographic factors in shaping loan default behavior within the digital lending ecosystem. By addressing the limitations and leveraging the findings, stakeholders can promote responsible lending practices, boost financial inclusion, and strengthen the sustainability of digital lending platforms.

Notes

JEL Codes: G1, G2, O1, 03

References

1. [a], [b] *Arun T, Markose SM, Murinde V, Kostov P, Khan A, Arı N, et al. Digital Lending in Fintech India - UK Comparison. SSRN Electronic Journal. 2023;*
2. [^] *Tampuri Jnr MY. Digital Lending in Emerging Economies: the Nexus between Financial Innovation, Inclusive Digital Finance, Risk and Consumer Protection. SSRN Electronic Journal. 2023;*
3. [^] *Ebong J, Babu G. Demand for credit in high-density markets in Kampala: Application of digital lending and implication for product innovation. Journal of International Studies. 2020;13(4).*
4. [^] *Mishra AK, Lence SH. Risk management by farmers, agribusinesses, and lenders. Agricultural Finance Review. 2005;65(2).*
5. [^] *Setiawan B, Andrianto Y, Safira F. Investment Decisions and Investor Experience in Choosing Financial Technology (Fintech) Services: Peer-To-Peer Lending. In 2020.*
6. [^] *Mohammed M, Basha J, Vidyalaya R, Gupta A, Raphael D, Rajeev V. The Role of Digital Platforms in Revolutionizing the Lending Sector. Int J Innov Res Sci Eng Technol. 2021;10(6).*
7. [^] *Rasulbek Boltaboevich S, Malikakhon Valijon Kizi K, Durdona Shukhrat Kizi U. Digital Lending: Its Advantages And Problems. International Journal of Research. 2020;07(04).*
8. [a], [b] *Nyahende VR. The Influence of Students' Loans Borrowers' Characteristics on Default Rate in Tanzania. Higher Education Studies. 2013;3(4).*
9. [^] *Adanu K, Boateng E. Predicting loan repayment default among second-tier borrowers in Ghana. International Journal of Entrepreneurship and Small Business. 2015;24(3).*
10. [^] *Johnen C, Parlasca M, Mußhoff O. Promises and pitfalls of digital credit: Empirical evidence from Kenya. PLoS One. 2021;16(7 July).*
11. [^] *Deyoung R, Glennon D, Nigro PJ, Spong K. Small Business Lending and Social Capital: Are Rural Relationships Different? The Journal of Entrepreneurial Finance. 2019;21(2).*
12. [a], [b], [c], [d] *Burlando A, Kuhn M, Prina S. Too Fast, Too Furious? Digital Credit Delivery Speed and Repayment Rates. SSRN Electronic Journal. 2023;*
13. [^] *Hunte CK. Controlling loan default and improving the lending technology in credit institutions. Savings and Development. 1996;20(1 DEC.).*

14. ^Wang Z, Jiang C, Zhao H. DEPICTING RISK PROFILE OVER TIME: A NOVEL MULTIPERIOD LOAN DEFAUL T PREDICTION APPROACH. MIS Q. 2023;47(7).
15. ^Brailovskaya V, Dupas P, Robinson J. Is Digital Credit Filling a Hole or Digging a Hole? Evidence from Mal awi. Economic Journal. 2024;134(658).
16. ^Santoso W, Trinugroho I, Risfandy T. What Determine Loan Rate and Default Status in Financial Technol ogy Online Direct Lending? Evidence from Indonesia. Emerging Markets Finance and Trade. 2020;56(2).
17. a, bDivino JA, Lima ES, Orrillo J. Interest rates and default in unsecured loan markets. Quant Finance. 2013; 13(12).
18. ^PELITA T, SIHABUDIN S, DJUMIKASIH D. Legal Analysis of The Problem of The Imposition of Interest Rat e On Online Loans. International Journal of Environmental, Sustainability, and Social Science. 2021;2(2).
19. ^Kaur K, Ilavarasan V. Digital Loan Sharks in India & Regulatory Framework: An Assessment. In: ACM Inte rnational Conference Proceeding Series. 2021.
20. ^Ochieng OF, Musau SM. Digital Lending and Loan Portfolio of the Listed Commercial Banks in Kenya. Th e International Journal of Business & Management. 2022;10(4).
21. ^Getugi JC, Osoro C, Kihara A. Digital Lending and Technical Efficiency of Commercial Banks in Kenya. Int ernational Journal of Finance. 2023;8(3).
22. ^Tetteh GK. Local digital lending development and the incidence of deprivation in Kenya. Financial Innov ation. 2023;9(1).
23. ^Kamau CG. Digital Credit in Kenya: A Survey of Costs, Uses and Borrowers Considerations in Relation to L oan Uptake. East African Journal of Business and Economics. 2021;3(1).
24. ^Chen A, Even-Tov O, Kang JK, Wittenberg Moerman R. Digital Lending and Financial Well-Being: Throu gh the Lens of Mobile Phone Data. SSRN Electronic Journal. 2023;
25. ^Gamble KJ, Boyle PA, Yu L, Bennett DA. Aging and financial decision making. Manage Sci. 2015;61(11).
26. a, bMurage M. Mobile Loan Frequency of Repayment as a Predictor of the Financial Performance of SMEs i n Urban Informal Settlements in Kenya. African Journal of Empirical Research. 2021;2(1 & 2).
27. ^Mostert M, Bredenoord AL, Van Der Slootb B, Van Delden JJM. From privacy to data protection in the EU: I mplications for big data health research. Vol. 25, European Journal of Health Law. 2017.
28. ^Lin TH. A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant an alysis, logit, probit, and neural networks models. Neurocomputing. 2009;72(16–18).

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.