Sustainable Development

WILEY

RESEARCH ARTICLE OPEN ACCESS

Unlocking Financial Inclusion: The Dynamics of Bank Account Ownership in Urban Slums

Davide Moro[1] | Sivia Trentin[2] | Lucia Dalla Pellegrina[3,4,5,6] | Paolo Landoni[1]

[1]Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy | [2]School of Economics and Statistics, University of Milano-Bicocca, Milano, Italy | [3]Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Milano, Italy | [4]Center for Interdisciplinary Studies in Economics, Psychology and Social Sciences (CISEPS), University of Milano-Bicocca, Milano, Italy | [5]Solvay Brussels School of Economics and Management, Université Libre de Bruxelles, Bruxelles, Belgium | [6]Center for European Studies (CefES), University of Milano-Bicocca, Milano, Italy

Correspondence: Lucia Dalla Pellegrina (lucia.dallapellegrina@unimib.it)

Received: 22 June 2025 | **Revised:** 20 December 2025 | **Accepted:** 6 January 2026

Keywords: financial inclusion | financial literacy | gender | sustainable development | urban slums

ABSTRACT

Financial inclusion is a key driver of sustainable development, contributing to poverty reduction (SDG 1), gender equality (SDG 5), and reduced inequalities (SDG 10). Despite extensive financial-inclusion policies in India, residents of urban slums remain largely excluded from formal banking systems. This study examines the determinants of formal bank account use in Seelampur, one of Delhi's largest informal settlements, interpreting active account ownership as a core indicator of effective financial inclusion. Using original survey data from 221 adults collected in 2019, the analysis combines logistic regression, Partial Least Squares–Structural Equation Modeling (PLS-SEM), and hierarchical cluster analysis. The results show that education, income, and institutional engagement—proxied by tax compliance and prior borrowing—significantly increase the likelihood of using a formal bank account. PLS-SEM reveals that institutional engagement mediates the effect of socio-economic status on financial inclusion, highlighting the central role of formalization pathways. At the same time, a persistent gender gap emerges: women are significantly less likely than men to be financially included, even after controlling for socio-economic characteristics. Cluster analysis uncovers substantial heterogeneity within the slum population, identifying distinct behavioral profiles and showing that financial exclusion is concentrated among low-income, female-dominated groups with weak links to formal institutions. Overall, the findings suggest that sustainable financial inclusion requires policies that go beyond physical access to banking infrastructure and address gender-specific barriers, institutional participation, and localized outreach in informal urban contexts.

JEL Classification: G21, G23, G53, I24

1 | Introduction

Financial inclusion is a key driver of equitable development and a cornerstone of the Sustainable Development Goals (SDGs), particularly SDG 1 (No Poverty), SDG 5 (Gender Equality), and SDG 10 (Reduced Inequalities). Access to formal financial services enables households to save securely, receive transfers, and invest in education or business activities, thereby reducing vulnerability to shocks (Beck et al. 2007; Demirgüç-Kunt and Klapper 2013). Despite the global expansion of inclusive-finance policies, a large share of the urban poor remains outside the formal financial system. In urban slums, the combination of irregular income, low literacy, weak institutional trust, and limited financial literacy creates persistent barriers to full participation.

India provides a particularly relevant context to investigate financial exclusion. Government initiatives such as the Pradhan Mantri Jan Dhan Yojana, the Aadhaar digital-identity program, and the diffusion of Unified Payments Interface (UPI) have substantially increased bank account ownership nationwide (Reserve Bank of India 2018; Demirgüç-Kunt et al. 2022). Yet these achievements often fail to translate into effective, sustained use among informal workers, particularly among slum residents (Banerjee et al. 2015). In these communities, proximity to bank branches does not guarantee inclusion. Documentation frictions, high transaction costs, and sociocultural constraints—especially those affecting women—continue to limit access to and utilization of formal accounts (Allen et al. 2016).

While numerous studies have examined rural financial perspectives and national-level inclusion patterns, evidence at the urban-slum level remains scarce. Research rarely integrates gender, religion, and institutional engagement within the same analytical framework or explores the heterogeneity of inclusion pathways beyond average effects. Few studies use original microdata to capture the specific mechanisms shaping inclusion in informal urban settlements. Addressing this gap, our study focuses on Seelampur, one of Delhi's largest and most densely populated slums, where formal banking initiatives coexist with enduring structural and behavioral barriers.

Using a unique dataset of 221 adult residents, we investigate the determinants of formal bank account ownership, emphasizing both structural variables (income, education) and institutional factors (tax compliance, prior borrowing). To capture different dimensions of inclusion, we employ a triangulated empirical strategy combining (i) logistic regression, providing baseline insights; (ii) Partial Least Squares–Structural Equation Modeling (PLS-SEM), which uncovers latent relationships suitable for small samples; and (iii) hierarchical clustering, which identifies behavioral profiles of inclusion, limiting endogeneity within the variables at stake. This mixed design provides a more nuanced understanding of how formalization, gender, and institutional engagement interact in shaping financial inclusion in a slum context.

Our findings contribute to both financial-inclusion and sustainable-development literatures. They highlight how structural capabilities (education, income) and institutional pathways (tax participation, documented transactions) jointly enhance access to formal accounts. Moreover, the results reveal a persistent gender gap that disadvantages women even after controlling for socio-economic characteristics. By focusing on a marginalized urban community, the study offers insights into how national policies can be localized and made more inclusive.

The remainder of the paper proceeds as follows: Section 2 reviews the literature and develops empirical hypotheses. Section 3 outlines the research design and methods. Section 4 presents the institutional context, while Section 5 describes the data and variables. Section 6 reports the econometric results, while Section 7 focuses on PLS-SEM analysis. Section 8 illustrates the cluster analysis, and Section 9 concludes with policy implications and directions for future research.

2 | Literature Review and Hypotheses

2.1 | Financial Inclusion, Sustainable Development, and Structural Determinants

Financial inclusion—the ability of individuals, particularly marginalized groups, to access and effectively use formal financial services—has become a cornerstone of sustainable development, directly contributing to poverty reduction (SDG 1), gender equality (SDG 5) and reduced inequalities (SDG 10) (Ledgerwood et al. 2013; Demirgüç-Kunt et al. 2018; Oanh and Dinh 2024; World Bank 2024; Nefla et al. 2025). Inclusive financial systems enhance resilience by allowing households to save, borrow, and insure against shocks, thus promoting equitable growth and urban stability (Allen et al. 2016; Swamy 2014).

Despite these benefits, financial access remains uneven across populations. Empirical research highlights the role of income, education, and literacy in enabling individuals to meet account requirements and navigate formal systems (Sarma and Pais 2011; Grohmann et al. 2018; Lusardi and Mitchell 2011). However, beyond these structural drivers, institutional engagement—such as tax payment, possession of identity documents, or prior borrowing—emerges as a crucial mechanism linking individuals to the formal economy (Sengupta 2019; Galiani et al. 2022). Given these premises, we formulate the following hypothesis:

H1. *Higher levels of income, education, and institutional engagement are positively associated with bank account ownership among slum residents.*

2.2 | Social and Cultural Dimensions: Gender, Religion, and Intersectional Barriers

A robust literature identifies gender as one of the most persistent axes of financial exclusion. Women often face restricted mobility, limited control over resources, and sociocultural norms that constrain independent financial decision-making (Kabeer 2001; Swamy 2014). While several microfinance initiatives have sought to empower women, evidence of transformative effects remains mixed due to household-level bargaining dynamics and male appropriation of benefits (Holvoet 2005; Karim 2011).

Religious affiliation further shapes financial behavior through both constraints and community support. In India, Muslim households frequently experience exclusion from formal banking owing to socio-economic marginalization and preferences for Sharia-compliant instruments (Us-Saqib and Arif 2012; Nolan 2015). Yet Islamic social-finance networks can also facilitate entry into formal channels when institutional trust is built (Kim et al. 2020). Based on literature, we formulate the following hypotheses:

H2. *Women are significantly less likely than men to own and use bank accounts, reflecting persistent gendered barriers to financial inclusion.*

H3. *Religious affiliation (Hindu* vs. *Muslim) influences financial inclusion in gender-specific ways, due to differences in sociocultural norms and community networks.*

Together, these findings illustrate that financial inclusion is not only an economic issue but also a reflection of intersectional inequalities rooted in social norms, religion, and household power structures.

2.3 | Urban-Slum Contexts and Behavioral Heterogeneity

Although India has achieved considerable national progress in financial inclusion, urban informal settlements remain underserved. Slum dwellers, despite their physical proximity to banks, face barriers of documentation, irregular income, and institutional mistrust (Nolan 2015; Swamy 2014). Standard products rarely suit their irregular earnings or caregiving responsibilities, leading to dormant accounts and persistent exclusion (Barkema et al. 2024; Hansen et al. 2019).

Empirical research based on primary slum-level data is extremely limited, leaving important knowledge gaps on how gender, religion, and institutional engagement jointly shape inclusion. Moreover, traditional regression approaches capture only average effects, masking the heterogeneity of financial behaviors within slum populations. Building on recent works emphasizing heterogeneity assessed by mixed and latent-variable methods (Galiani et al. 2022; Colamartino et al. 2025), this study combines logistic regression, PLS-SEM, and hierarchical cluster analysis to reveal both observable determinants and unobserved behavioral typologies. The following hypothesis is therefore formulated:

H4. *Distinct latent financial-behavior profiles exist within the slum population, implying that the determinants of financial inclusion vary across subgroups beyond average regression effects.*

By addressing theoretical and methodological gaps, this paper contributes to the sustainable development literature in three ways. First, it provides original slum-level evidence from Seelampur, illuminating the micro-mechanisms of inclusion in marginalized urban contexts. Second, it offers a gender-disaggregated and institution-focused perspective that connects financial inclusion to equality and empowerment. Third, it introduces a triangulated analytical framework—Logit + PLS-SEM + Cluster—to capture the complex, multidimensional nature of financial inclusion as a pathway toward sustainable development.

3 | Research Design

The research design is anchored in the four hypotheses developed in Section 2. H1–H3 concern observable relationships between individual characteristics and bank account ownership and are tested using logistic regression, while H4 focuses on behavioral segmentation explored through cluster analysis. A PLS-SEM extension triangulates the findings by modeling indirect and latent effects among structural and institutional constructs. This multi-method framework enhances explanatory power and robustness: the logistic model identifies statistically significant predictors; the PLS-SEM validates whether those effects persist once interrelations and mediating variables are considered; and the cluster analysis uncovers unobserved heterogeneity beyond average effects (Galiani et al. 2022).

3.1 | Empirical Strategy

3.1.1 | Step 1: Logistic Regression

The first stage estimates a binary logit model where the dependent variable *Usebankaccount* equals 1 if the respondent owns and uses a formal bank account and 0 otherwise.[1]

Explanatory variables capture demographic traits (gender, age, religion), household structure (family size, children's education status), socio-economic position (income, education, homeownership), and institutional engagement (tax payment, loan history). Formally:

$$P_i = \frac{e^{\beta_0+\beta X_i}}{1+e^{\beta_0+\beta X_i}} \tag{1}$$

where $Y_i$indicates account ownership/use and $X_i$is the vector of explanatory variables described above. This stage tests H1–H3 by quantifying the marginal effects of income, education, gender, religion, and institutional engagement on financial inclusion.

3.1.2 | Step 2: PLS-SEM Extension

Recognizing the small sample size ($N = 221$) and the interdependence among socio-economic and institutional variables, we complement the regression with PLS-SEM analysis. This technique is appropriate for small, non-normal samples and formative indicators typical of primary data. This addresses the methodological critique that a single reduced-form logit cannot capture latent pathways (See Colamartino et al. 2025).

3.1.3 | Step 3: Cluster Analysis

To capture unobserved heterogeneity (H4), we apply hierarchical cluster analysis to standardized demographic, socio-economic, and institutional variables. We use Ward's linkage and an appropriate distance metric for mixed data, following the procedure detailed later in the paper. The clusters represent behavioral profiles of financial inclusion (e.g., "structural access" vs. "institutional gateway"), showing how determinants co-occur within subgroups beyond average regression effects.

3.2 | Design Logic and Rationale

Combining regression, PLS-SEM, and clustering serves three purposes: triangulation and validity (multiple techniques mitigate method-specific bias), theory–method alignment (each stage maps to specific hypotheses and constructs), and context sensitivity (the approach accommodates small, heterogeneous, informal-settlement data that challenge single-method inference). This integrated design tests directional mechanisms while recognizing that inclusion is multidimensional and context-specific, shaped by structural inequalities and institutional engagement.

All variables, measurement details, and diagnostic tests are presented in the following section. This combination of econometric and multivariate techniques allows a rigorous yet context-sensitive examination of financial inclusion among Seelampur residents.

4 | The Institutional Context

India has witnessed a remarkable transformation in financial inclusion over the past decade, driven by ambitious policy reforms and digital innovation. The share of adults owning a bank account rose from 35% in 2011 to nearly 80% in 2017, positioning India among the world's fastest improvers (Demirgüç-Kunt et al. 2018). Central to this progress has been the Pradhan Mantri Jan Dhan Yojana, launched in 2014, which encouraged banks to open zero-balance accounts for the unbanked and promoted gender inclusion—more than half of all PMJDY accounts are now held by women.

This expansion reflects a strong commitment to SDG 1 (No Poverty) and SDG 5 (Gender Equality), yet progress in access has not always translated into effective use. Many accounts remain dormant, and inclusion gaps persist among marginalized groups—especially urban slum residents, informal workers, women, and migrants (Swamy 2014; Ghosh and Vinod 2017). These groups face both structural and institutional constraints: low literacy, irregular earnings, lack of identification documents, and weak trust in formal institutions.

A closer look at national patterns underscores these inequalities. Account ownership among the less educated lags significantly behind that of better-educated groups, and women's access, though improved, still trails men's usage rates (Demirgüç-Kunt et al. 2018). Financial participation also mirrors the employment structure: individuals outside the labor force or in the informal sector exhibit greater exclusion. The urban poor, often engaged in cash-based or undocumented work, remain particularly vulnerable to exclusion despite their geographical proximity to banks and financial intermediaries.

At the institutional level, India's financial-inclusion architecture relies on an expanding ecosystem of public and private banks, fintechs, and non-governmental organizations (NGOs). Policy instruments such as the Aadhaar digital ID system, the Unified Payments Interface (UPI), and networks of business correspondents have extended the reach of financial services to millions (Swamy 2014; Grohmann et al. 2018). However, these technologies do not automatically overcome socio-cultural or behavioral barriers. For many low-income women, for instance, digital tools remain inaccessible due to limited device ownership or digital literacy, creating a "second-generation divide" in financial inclusion (Galiani et al. 2022).

Urban informal settlements like Seelampur, Delhi's largest slum, exemplify this paradox. Residents live within a few kilometers of bank branches but face exclusion rooted in the absence of documentation, precarious livelihoods, and fragmented institutional trust. These communities thus represent a critical frontier for sustainable-development policy: improving inclusion here means addressing not only physical access but also behavioral, institutional, and gendered dimensions of financial participation.

By situating our micro-level investigation in Seelampur, this study provides an empirical lens through which to assess the real-world effectiveness of India's financial-inclusion initiatives. It highlights the persistent gap between formal policy achievements and the lived realities of marginalized urban populations—an issue central to both national development priorities and the global SDG agenda.

5 | Data Description

This study relies on original primary data collected through a face-to-face household survey conducted in Seelampur, Delhi's largest urban slum, during November–December 2019.[2] The survey was implemented in partnership with the Maverick Foundation, an NGO with long-standing experience in community-based research and financial-education programs.[3]

5.1 | Sampling Design and Fieldwork

Given the absence of an official sampling frame for informal settlements, we adopted a stratified random sampling strategy to ensure heterogeneity in gender, religion, education, and household composition. Enumerators—trained in ethical data-collection practices—used structured questionnaires[4] translated into Hindi and Urdu to minimize linguistic bias and nonresponse.[5] A total of 221 individuals were interviewed across Seelampur's main clusters.

Seelampur is one of Delhi's most densely populated and socioeconomically disadvantaged areas, with an estimated population exceeding 400,000 residents, though official figures vary due to rapid migration and informal settlement expansion.[6] The sample size, though modest, provides rare micro-level insight into slum residents' financial behaviors, a population typically underrepresented in national surveys such as the Global Findex or the Indian National Sample Survey (Demirgüç-Kunt et al. 2018; Nolan 2015).

5.2 | Key Variables

The analysis centers on formal bank account usage. The outcome binary variable (*Usebankaccount*) is equal to 1 if the respondent owns and actively uses a formal bank account and 0 otherwise. This operationalization follows standard definitions of financial inclusion in the development-finance literature (Allen et al. 2016; Sarma and Pais 2011). Despite policy efforts such as the Pradhan Mantri Jan Dhan Yojana, only 33.5% of respondents reported using a bank account, confirming the persistence of exclusion even in urban areas with physical access to banks.

Explanatory variables reflect the multidimensional framework outlined in Section 2:

TABLE 1 | Summary statistics of the variables.

Label	Description	Mean	Std. error	Min	Max
Usebankaccount	Use of a bank account (yes = 1)	33%			
Gender	Male = 1	27%			
Age	Age of the interviewee	36.17	146.7	14	90
Religion	Hindu or Muslim (Muslim = 1)	38%			
Peopleinthefamily	Number of family members	4.4	1.7	1	12
Ownership	Own house: (yes = 1)	80%			
Qualification	Illiterate: base category; Class 5: fifth year of education; Class 8: eighth year of education; Class 10: tenth year of education; Class 12: twelfth year of education; Graduate: first level degree; Master: master's degree				
Class 5		12%			
Class 8		18%			
Class 10		9%			
Class 12		12%			
Graduate		7%			
Master		0.5%			
Childrened	No children: base category; Too Young: children have not yet reached school age; School Going: child with a higher level of education attends school; College Going: child with a higher level of education attends college; Drop Out from the school: children withdrawn from the school system				
No children	27.15%				
Too young	7.24%				
School going	42.08%				
College going	3.17%				
Dropout	5.43%				
Percapitaincomerupees	Per capita monthly family income; Rupees	6896 (about 85 USD)	9770	4000	6896
Payingtaxes	Payment of taxes (yes = 1)	18%			
Haveyoureceivedaloan	Receipt of a previous loan; yes = 1	11%			

- Demographic factors: *Gender* (male = 1), *Age* (years), and *Religion* (Hindu = 0, Muslim = 1). These variables relate directly to H2 and H3, testing gender and cultural differentials in financial access.
- Household structure: *Family size* and *children's education status*, capturing dependency ratios and intergenerational educational aspirations, which influence saving behavior and liquidity needs.
- Socio-economic position: *Per-capita monthly income*, *educational attainment*, and *home ownership*. These variables operationalize the structural dimensions of H1, linking human and financial capital to inclusion.
- Institutional engagement: *Tax payment status* and *previous loan history*, proxies for formalization and interaction with state or financial institutions—core to H1 and the latent construct "Institutional Engagement" later modeled in the PLS-SEM.

The survey instrument was adapted from the Global Findex and OECD/INFE financial literacy questionnaires to ensure cross-national comparability and internal consistency of measures. All variables were cleaned for outliers and coded consistently across models. Categorical predictors were converted to binary or ordinal scales depending on their analytical role.

5.3 | Descriptive Statistics

Table 1 presents descriptive statistics for all variables. The data reveal substantial heterogeneity across respondents. The average age is 36 years (range 14–90); 27% of respondents are male; and 38% identify as Muslim. The mean household size is 4 members, reflecting dense family structures typical of informal settlements.

Socio-economic conditions are precarious: average per-capita income is approximately 6900 rupees (≈ 85 USD) per month,[7] and only 18% report paying taxes. Educational attainment is low but varied—45% of respondents completed primary education, while fewer than 10% hold tertiary degrees. Eighty percent own their dwellings, though tenure security is often informal. Only 11% have previously received a loan.[8]

A variance-inflation-factor (VIF) analysis confirms that multicollinearity is not a concern: none of the independent variables exceed the critical threshold of 5, validating their joint inclusion in the regression models.[9]

5.4 | Contribution of the Dataset

This dataset provides a rare, high-resolution perspective on financial inclusion among India's urban poor. By integrating both structural attributes (income, education) and institutional dimensions (tax compliance, loan history), it allows for a comprehensive test of H1–H4. Moreover, the inclusion of gender and religion identifiers enables intersectional analysis, while institutional proxies support the construction of latent factors in the PLS-SEM and the segmentation of respondents in the cluster analysis.

Thus, the Seelampur dataset serves not only as the empirical foundation for the econometric models but also as a conceptual bridge linking the micro-level realities of slum residents to the broader sustainable-development agenda.

6 | Regression Analysis

6.1 | Baseline Regression Model

This section presents the empirical results from the regression models testing the observable determinants of financial inclusion, as formulated in H1–H3. The dependent variable, *Usebankaccount*, is a binary indicator coded as 1 if the respondent owns and uses a formal bank account and 0 otherwise. Given the dichotomous nature of the dependent variable, we estimate a logistic regression model to identify the probability of bank account ownership as a function of individual and household characteristics. The baseline specification includes explanatory variables capturing demographic traits (gender, age, religion), household structure (family size, children's education status), socio-economic position (income, home ownership, education), and indicators of institutional engagement (tax compliance, loan history).

Formally, the probability of account usage is modeled as:

$$logit\left(\theta_i\right) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \boldsymbol{\beta_1 X_i} + \varepsilon_i \quad (2)$$

where β_0 represents the intercept, the vector β_1 summarizes the parameters associated with the explanatory variables included in X_i (detailed in the previous section), while ε_i is a zero-mean error term. The *logit*(.) function converts the probability of using a bank account into a linear value. For categorical variables, the baseline category corresponds to the most frequently observed category in the sample.

The model, estimated on cross-sectional data, is designed to uncover theoretically grounded pathways shaping financial inclusion. Gender, religion, and education are treated as exogenous demographic factors that precede financial behavior, while income, tax compliance, and loan history capture mechanisms through which inclusion operates. Although causality cannot be fully established, the analysis identifies reasonable directional relationships consistent with prior evidence. The comprehensive specification and inclusion of key controls highlight the structural and institutional drivers of financial inclusion.

6.2 | Results

The results of the baseline regression analysis are presented in Table 2, with additional insights provided through the marginal effects of statistically significant variables, as determined by the Wald test in Table 3.

6.2.1 | Socio-Economic Set-Up, Education, and Demographic Variables (H1)

Consistent with H1, both socio-economic and institutional variables in Table 2 emerge as strong and statistically significant predictors of financial inclusion. The coefficient on per-capita income is positive and significant ($\beta \approx 0.0002$, $p = 0.015$), and its marginal effect of roughly +0.09 (Table 3) implies that even modest income increases translate into sizeable gains in formal participation. Education shows an even stronger pattern: compared to illiterate respondents, those completing Class 10 are 17 percentage points more likely to hold an account, and graduates are 56 points more likely. These large effects suggest that human-capital accumulation not only raises awareness and financial capability but also facilitates the procedural requirements of interacting with formal institutions, consistent with prior evidence from cross-country data (Allen et al. 2016; Grohmann et al. 2018; Sarma and Pais 2011).

Beyond these structural factors, the results also underscore the critical role of institutional engagement. Individuals who pay taxes are 37 percentage points more likely to possess an account ($p < 0.001$), while those with prior borrowing experience are 21 points more likely ($p < 0.01$). These large marginal effects indicate that contact with formal systems, whether fiscal or credit, creates a virtuous cycle of trust, documentation, and familiarity that lowers transaction costs and fosters sustained participation in the financial sector (Galiani et al. 2022; Sengupta 2019).

Demographic controls behave as expected: age displays a mild inverted-U shape, with middle-aged respondents showing higher inclusion, while family size and homeownership remain statistically insignificant. Taken together, these baseline findings confirm that financial inclusion in marginalized urban settings is simultaneously a function of capability and connectivity, that is, the material and cognitive capacity to engage with formal finance and the relational exposure that builds confidence in it. The outsized marginal effect of tax payment highlights that formalization processes are not merely fiscal instruments but effective vehicles of inclusion, reinforcing the idea that participation in public and financial institutions is mutually reinforcing. These results provide strong support for H1 and fit within the

TABLE 2 | Logit regression results: Estimated coefficients.

Coefficient	Variable	Estimated coefficient	Std error	*p*
β_0	Intercept	−1.6344	1.5623	0.2955
β_1	*Percapitaincomerupees*	0.0002	0.0001	0.0151**
β_2	*GenderM = 1 = 1*	1.9905	0.5506	0.0003***
β_3	*Age*	−0.0203	0.0297	0.4938
β_4	*Religion Muslim = 1*	0.8858	0.4706	0.0598*
β_5	*PayingtaxesYes = 1*	2.5009	0.7126	0.00004***
β_6	*Haveyoureceivedaloan = 1*	1.6865	0.6249	0.0070***
c	*QualificationClass10 = 1*	1.3990	0.7820	0.0736*
β_8	*QualificationClass12 = 1*	1.2304	0.8030	0.1255
β_9	*QualificationClass8 = 1*	0.7111	0.6951	0.3063
β_{10}	*QualificationClass5 = 1*	0.2664	0.7430	0.7200
β_{11}	*QualificationGraduate = 1*	4.1333	1.1367	0.0003***
β_{12}	*QualificationMaster = 1*	14.6886	1455.39	0.9919
β_{13}	*Ownership = 1*	0.1784	0.5659	0.7526
β_{14}	*Ppeopleinthefamily*	−0.1828	0.1721	0.2882
β_{15}	*ChildrenedCollegeGoing = 1*	−1.3042	1.2490	0.2964
β_{16}	*ChildrenedDropOut = 1*	−0.9525	1.0383	0.3590
β_{17}	*ChildrenedNoChildren = 1*	−2.5709	0.8843	0.0036***
β_{18}	*ChildrenedSchoolFinished = 1*	−0.5339	0.8973	0.5519
β_{19}	*ChildrenedTooYoung = 1*	−0.6896	1.0115	0.4954
Pseudo-R^2: 0.46				

Note: Significant level: *10%, **5% e *** 1%.

TABLE 3 | Logit regression results: Marginal effects (significant variables).

Coefficient	Variable	Marginal effect	Std error	*p*
β_1	*Percapitaincomerupees*	0.00002	0.0001	0.0151**
β_2	*GenderM = 1*	0.2717	0.5506	0.0003***
β_4	*Religion Muslim = 1*	0.0973	0.4706	0.0598*
β_5	*PayingtaxesYes = 1*	0.3694	0.7126	0.00004***
β_6	*HaveyoureceivedaloanYes = 1*	0.2092	0.6249	0.0070***
β_7	*QualificationClass10 = 1*	0.1698	0.7820	0.0736*
β_{11}	*QualificationGraduate = 1*	0.5637	1.1367	0.0003***
β_{17}	*ChildrenedNoChildren = 1*	−0.2230	0.8843	0.0036***

Note: Significant level: * 10%, ** 5% e *** 1%.

sustainable-development perspective that links education (SDG 4), income opportunities (SDG 8), and institutional access (SDG 10) as complementary levers for inclusive financial systems.

6.2.2 | Gender and Religion Effects (H2 and H3)

The baseline results in Tables 2 and 3 provide clear evidence that both gender and religion shape financial inclusion outcomes, validating hypotheses H2 and H3. The coefficient for gender is large, negative, and highly significant across all baseline models ($\beta \approx -0.64$, $p < 0.01$), corresponding to an average marginal effect of about −0.18. This implies that, holding other factors constant, women are nearly one-fifth less likely than men to own and actively use a bank account. Such a pronounced gap persists even after accounting for education, income, and institutional engagement, indicating that gender disparities in financial inclusion are not simply a by-product of socio-economic disadvantage but reflect entrenched social norms, differential mobility, and constrained decision-making power within households. These findings are consistent with long-standing evidence that patriarchal structures and gendered financial behaviors restrict women's economic agency in South Asia (Kabeer 2001; Swamy 2014) and with more recent analyses linking women's empowerment and inclusive finance to progress toward SDG 5 (Gender Equality) (Kim et al. 2020).

The religion variable, which distinguishes Muslim from Hindu respondents, also exerts a statistically significant influence in the baseline models. The positive coefficient ($\beta \approx 0.34$, $p \approx 0.05$) and marginal effect of roughly +0.10 indicate that Muslim respondents are about 10 percentage points more likely to hold a bank account than their Hindu counterparts, after controlling for other covariates. This result diverges from national-level patterns that typically report lower financial inclusion among Muslims due to income disparities, perceived discrimination, or preference for informal or Sharia-compliant instruments (Demirgüç-Kunt et al. 2018; Kim et al. 2020). The Seelampur context—where community-based organizations and NGOs play an active role in facilitating formal access—may mitigate these constraints and instead translate religious identity into stronger local institutional linkages. Hence, while gender operates as a persistent axis of exclusion, religion in this localized setting appears to function as a context-dependent inclusion channel, reflecting the effectiveness of community-embedded outreach initiatives.

Overall, the baseline regressions in Tables 2 and 3 confirm that gender and religion both matter for financial inclusion, albeit in opposite directions: women remain systematically disadvantaged, while Muslim affiliation is associated with higher inclusion probabilities in this specific urban slum environment. These patterns underscore the multidimensional nature of exclusion and inclusion, reinforcing the need for gender-sensitive and culturally grounded interventions that recognize how social identity interacts with structural and institutional variables to determine access to finance.

6.3 | Discussion on Regression Analysis and Sustainable-Development Implications

The regression results offer a preliminary but coherent empirical validation of the theoretical framework and literature presented in Section 2. Structural and institutional variables (education, income, tax compliance) exert the strongest influence on financial inclusion, while gender and religion shape the social accessibility of finance.

From a sustainable development perspective, these findings provide three key insights:

1. Education as empowerment: Expanding education access among slum residents, especially women, has immediate spillover effects on financial participation—directly advancing SDGs 4 (Quality Education) and 5 (Gender Equality).
2. Institutional pathways: Encouraging formal interactions such as tax registration or microcredit participation fosters trust in financial institutions and accelerates progress toward SDG 10 (Reduced Inequalities).
3. Intersectionality in inclusion policy: Financial inclusion strategies must consider how gender and religion jointly condition access to formal finance, ensuring that interventions such as the Jan Dhan Yojana reach the most excluded subgroups.

Together, these findings reaffirm that financial inclusion is a multidimensional process—one that extends beyond physical access to banking infrastructure to encompass the social, institutional, and behavioral dimensions essential for achieving sustainable development.

6.4 | Robustness Checks

Having established the baseline and gender-specific patterns, we next verify the stability of our findings through a set of robustness checks. First, we evaluate the goodness of fit of the logistic model using a likelihood ratio test based on deviance, comparing the full specification with a null model containing only the intercept. The test statistics allow us to reject the null hypothesis that the reduced model is sufficient, indicating that the estimated model explains the data significantly better than the parsimonious alternative.[10]

Second, we re-estimate the regressions separately for male and female respondents to explore potential heterogeneity in the determinants of financial inclusion. The results of these gender-stratified regressions[11] confirm that the direction and magnitude of the coefficients remain broadly consistent with the baseline estimates. Among men, income and tax payment retain strong positive effects, whereas for women, education and prior borrowing experience are the key drivers of inclusion. This stability across sub-samples supports the robustness of the main findings and suggests that the observed gender gap reflects genuine behavioral and institutional asymmetries rather than sample-composition bias.

6.5 | Endogeneity and Limitations

Although the regression models provide useful insights into the correlations between bank account usage, it is important to acknowledge the limitations inherent in drawing causal conclusions from cross-sectional observational data. Several explanatory variables included in the analysis—particularly per capita income, tax payment, and previous loan history—may be endogenous to financial inclusion. For example, the decision to open a bank account may facilitate access to formal employment and credit (Allen et al. 2016), thereby increasing income and the likelihood of paying taxes. Similarly, prior borrowing experience may reflect a reverse channel in which financial inclusion enables or incentivizes credit-taking rather than the reverse.

These forms of reverse causality and omitted variable bias are well-documented challenges in the financial inclusion literature. Unobserved factors such as trust in institutions, risk preferences, or financial literacy (Galiani et al. 2022; Cole et al. 2011) may confound the estimated relationships if they simultaneously influence both the independent variables and the likelihood of holding a bank account. While we control for observable demographic traits (e.g., age, gender, education, religion, household size), we cannot rule out residual confounding due to latent behavioral or psychological variables.

In light of these concerns, the results presented in the regression analysis should be interpreted as identifying structural associations consistent with causal mechanisms, rather than estimating treatment effects in a strict econometric sense. Nevertheless, the evidence supports theoretical predictions from earlier work suggesting that institutional engagement—such as paying taxes or accumulating credit history—can promote financial inclusion through formalization feedback loops (Sengupta 2019; Allen et al. 2016; Karlan and Valdivia 2011). Similarly, the consistent significance of educational attainment aligns with existing findings that link human capital to financial capability (Lusardi and Mitchell 2011; Addai 2017; Samer et al. 2015).

The limitations of our identification strategy also point to valuable directions for future research. While this study offers exploratory evidence of potential mechanisms, more rigorous causal inference in the regression ambit would require either panel data that tracks individuals over time or exogenous variation introduced through natural experiments, instrumental variables, or randomized interventions. For instance, future work could exploit discontinuities in policy access, such as eligibility thresholds for financial inclusion schemes (Demirgüç-Kunt et al. 2018), or draw on random assignment of financial literacy programs (Karlan and Valdivia 2011) to better isolate treatment effects.

Finally, we note that despite the relatively small sample size, the use of a carefully designed and context-rich dataset enhances both the internal validity and the interpretive depth of our findings. While external generalizability remains limited, this mixed-method approach responds to recent calls for more localized and granular studies of financial behavior in informal urban settings (Barkema et al. 2024; Nolan 2015), and it sets the stage for the cluster analysis presented in the following section. Building on the regression estimates, which identify the main observable determinants of formal bank account ownership, we next turn to a PLS-SEM specification. This approach will allow us to model latent constructs and simultaneous relationships among variables, thereby clarifying the pathways through which socio-economic capability and institutional engagement shape financial inclusion.

7 | PLS-SEM Analysis

The PLS-SEM method is particularly appropriate for small-sample and exploratory studies (Colamartino et al. 2025), as it allows for the simultaneous estimation of relationships between latent constructs—Socio-Economic Status (SES) and Institutional Engagement (IE)—and their observed indicator, that is, formal account ownership. In contrast to logistic regression, which examines the marginal effect of individual predictors on the binary outcome of account ownership, PLS-SEM is designed to model causal paths that incorporate measurement error and capture indirect effects among variables.

In our specification, financial inclusion is treated as a latent construct measured by bank account ownership, while two additional latent constructs capture the underlying determinants of inclusion. The first, socio-economic status, is reflected by income, home ownership, and educational attainment. The second, institutional engagement, is measured through tax compliance and borrowing history. Demographic characteristics such as gender, age, religion, and household size are introduced as observed exogenous variables, shaping both socio-economic status and financial inclusion. Formally, the structural model links financial inclusion to socio-economic status, institutional engagement, and demographic traits, while the measurement model connects the latent constructs to their observable indicators.

The results from the PLS-SEM analysis[12] suggest that the main paths identified are consistent with those found in the baseline regressions. The path from socio-economic status to financial inclusion is positive and significant, indicating that higher income and education levels increase the likelihood of account ownership. The link between institutional engagement and financial inclusion is also positive, showing that tax compliance and loan history are key drivers of inclusion, especially for women. The gender effect remains strongly positive in favor of men, confirming the persistence of gender disparities. Religious affiliation shows a weaker but still positive effect for Muslims, aligning with the marginal significance observed in the regression analysis.[13]

Taken together, the PLS-SEM results corroborate the baseline regression findings while offering additional conceptual insights. They emphasize that financial inclusion in slums is not simply the outcome of isolated demographic variables but rather the product of broader latent dimensions such as socio-economic status and institutional engagement. By incorporating PLS-SEM into our analysis, we not only confirm the robustness of our conclusions but also address methodological critiques by demonstrating that our results hold under an alternative, theoretically richer framework.

7.1 | Discussion on PLS-SEM

The PLS-SEM results provide additional insight into the mechanisms through which individual characteristics shape formal bank account ownership. The model identifies two formative latent constructs—Socio-Economic Status (SES) and Institutional Engagement (IE)—whose relationships with bank account ownership help clarify the pathways suggested by the baseline logistic regression.

First, SES exhibits a strong and positive effect on IE, indicating that individuals with higher education and income are more likely to be connected to formal institutions (through tax payment or documented borrowing). This mediating structure confirms that capability factors do not influence financial inclusion solely through direct channels but operate partly by increasing an individual's degree of economic formalization.

Second, IE shows a significant direct effect on formal account ownership, reinforcing the idea that institutional contact—particularly tax payment and prior engagement with regulated credit—constitutes an important pathway into formal finance. This aligns with findings in development economics emphasizing that regular interactions with the state or regulated intermediaries reduce informational frictions and lower barriers to account uptake.

Third, the indirect effect of SES on financial inclusion, transmitted through IE, is statistically meaningful, suggesting that the role of institutional formalization is as relevant as socioeconomic capability. This complements the regression results, where income and education exhibit direct significance: the SEM shows how these variables exert influence by revealing latent structure beneath the observable indicators.

Regarding endogeneity, PLS-SEM does not provide causal identification in the econometric sense, but it mitigates several endogeneity concerns highlighted in the logistic regression (Section 4) in two ways: (1) Measurement error is explicitly modeled, reducing attenuation bias in SES and IE—an important advantage in small-N, survey-based settings; (2) Simultaneous estimation of multiple relationships helps address potential omitted-variable bias arising from correlated latent constructs. By modeling SES → IE → Account Ownership jointly, PLS-SEM captures interdependencies that a single-equation logit cannot.

While PLS-SEM cannot resolve reverse causality or fully eliminate omitted-variable bias, it provides a more structurally grounded representation of the determinants of inclusion and strengthens confidence in the direction and plausibility of the mechanisms identified. In this sense, it complements rather than substitutes the regression analysis, offering a small-sample-appropriate framework that clarifies the channels linking capability, engagement, and formal inclusion.

8 | Cluster Analysis

While the logistic regression presented in Section 6 identifies the average marginal effects of key socio-economic and institutional factors on bank account ownership, it implicitly assumes that these relationships are homogeneous across the population. However, as highlighted in our hypotheses (H4), the determinants of financial inclusion are likely to operate differently across subgroups within the slum population.

To capture this heterogeneity, we complement the regression with a cluster analysis. The purpose of combining the two methods is therefore not redundancy but complementarity: regression and PLS-SEM provide a theory-driven test of associations consistent with our hypotheses (H1–H3), while cluster analysis adopts a data-driven approach to reveal latent behavioral profiles that may not align with the regression's linear assumptions. This dual-method design is consistent with methodological recommendations emphasizing that regression models capture average associations, whereas clustering techniques are particularly suited to uncovering unobserved heterogeneity and latent subgroups (Everitt et al. 2011; Wedel and Kamakura 2000). In the context of financial inclusion, recent work has similarly highlighted the importance of combining econometric analysis with segmentation methods to better understand heterogeneous financial behaviors (Galiani et al. 2022). In other words, the regression analysis addresses *whether* certain factors matter on average, while the cluster analysis explores *how* combinations of these factors co-occur in practice, offering a more nuanced understanding of financial behavior in Seelampur.

Based on measures of similarity between elements, the objective is to create groups that are homogeneous internally and heterogeneous between, measuring the similarity between statistical units through distance measures. The smaller the distance between two elements, the more they will belong to the same group. The concept of distance can take on different meanings; specifically, its general definition is the following:[14]

$$L_{jk} = \sum_{k=1}^{n} \left(\left| x_i(k) - x_j(k) \right|^N \right)^{\frac{1}{N}} \tag{3}$$

where $x_i(k)$ are the k explanatory variables pertaining to the i-th individual and $x_j(k)$ are the k explanatory variables about all other j-th individuals, k as defined by the variables illustrated in Section 3. The algorithm that characterizes clustering methodology based on the concept of distance consists of two steps. In the first step, we assign each observation to the closest representative group (chosen as the cluster element that minimizes the distances from all the data in a specific cluster) according to a specific distance metric. The second step consists of updating: once the k clusters have been determined, the representative groups' composition is updated, and the previous step is repeated until convergence is reached.

We opted for an agglomerative hierarchical clustering, using Ward's method and Gower's distance matrix (Gower 1971), as the dataset contains both numerical and categorical variables.[15] The steps of the implemented algorithm are as follows: (i) analysis of each point as a separate cluster; (ii) calculation of the Gower distance matrix between clusters; (iii) union of the clusters closest to each other according to Ward's method; (iv) repeat steps (ii) and (iii) until all the points join together in a single group; (v) the obtained hierarchy is returned.

The Gower distance matrix contains all pairwise distances between the observations. The procedure is the following: for each category $k = 1, \ldots, n$ a score s_{ijk} included in the interval [0,1] is defined. If two points x_i and x_j are close with respect to the characteristic (variable) k, then s_{ijk} is close to 1; on the contrary, it will be close to 0. The s_{ijk} score is calculated in different ways depending on the type of characteristic being described: for quantitative variables $s_{ijk} = 1 - |x_{ik} - x_{jk}| / R_k$, with R_k is the range of the variable k, while for qualitative and dichotomous variables $s_{ijk} = 1 \; if \left\{ x_{ik} = x_{jk} \right\}$.

Once having calculated s_{ijk}, we evaluated the score δ_{ijk}, which is equal to 1 if the two points x_i and x_j can be compared with respect to characteristic k, and 0 otherwise (e.g., due to missing values). When δ_{ijk} is null, by convention s_{ijk} is also set to zero.

For dichotomous variables, the following applies:

$$\delta_{ijk} = \{ 1 \; if \, i, j + 1 \; if \, i+, j - 1 \; if \, i-, j + 0 \; if \, i, j- \tag{4}$$

Once these quantities are defined, the Gower distance is computed as a simple average of the scores, that is,

$$S_{ij} = \frac{\sum_{k=1}^{p} s_{ijk}\delta_{ijk}}{\sum_{k=1}^{p} \delta_{ijk}} \quad (5)$$

The *daisy* function of the R software was used.[16] The numerical variables were rescaled in the interval [0,1].

The clusters are then joined based on the Ward link, which maximizes the metric of the average silhouette (i.e., the internal criterion that is used to evaluate the goodness of the clustering), obtained as follows:

$$S_i = \frac{D_{min_i}{}^{bet} - \underline{D_i}^{with}}{max\left\{ D_{min_i}{}^{bet}; \underline{D_i}^{with} \right\}} \quad (6)$$

where $D_{min_i}{}^{bet}$ is the average distance of point i from the points in the nearest cluster, and $\underline{D_i}^{with}$ is the average distance of point i from all other points in the same cluster. $\underline{D_i}^{with}$ ranges in the interval [−1, 1], where values close to 1 indicate a well-clustered point, values close to −1 indicate bad clustering, while values close to 0 indicate that a reclassification may be appropriate.

In the hierarchical clustering process with Ward's method and Gower distance matrix, the clusters are merged in such a way as to minimize the sum of the squares of the differences between the observations and the centroids of the clusters to which they belong. At each step of the algorithm, more similar clusters are merged, considering the distance matrix as a similarity measure. The clustering obtained accounts for the mixed nature of the data.

8.1 | Results

Using the clustering metric, we obtained 2–9 clusters. Table 4 reports the average silhouette values obtained for each clustering. The optimal number of 4 clusters, corresponding to the maximum silhouette score, is chosen. The same conclusion is obtained by analyzing the dendrogram obtained from the application of the algorithm (Figure 1), showing that the increase in the length of the object fusions seems to slow down to around 4 or 5 clusters. Notice that the variable relating to the use of a bank account was not considered for the implementation of the clustering algorithm to create groups that were homogeneous within themselves and heterogeneous among themselves, which could then be associated with the use of a bank account.

TABLE 4 | Cluster analysis: Average silhouette values.

Number of clusters	Average silhouette
2	0.2111
3	0.198
4	0.2165
5	0.1968
6	0.1819
7	0.1957
8	0.1914
9	0.2111

The number of statistical units detected in each of the four clusters is: 38 (cluster 1); 118 (cluster 2); 30 (cluster 3); 35 (cluster 4). The most populated cluster is the second, followed by the first. The fact that an optimal number of clusters of four was detected does not preclude the number of real classes from being two, as in the *Usebankaccount* variable. This difference between the number of clusters and the number of real classes of the reference variable could mean that each class is characterized by more than one behavior, which did not emerge from the regression analysis.

From Table 5, it can be observed that the composition of clusters 2 and 4 is rather unbalanced: in cluster 2, for the most part (87%), the subjects do not use a bank account, while in cluster 4, the most part (around 89%) are bank account users. Regarding cluster 1, most subjects (68%) are non-users of a bank account, although the proportion is not as unbalanced as for the other clusters. The composition, however, is perfectly balanced in cluster 3, with exactly half of the subjects using a bank account—a pattern that may indicate the presence of heterogeneous individuals whose financial behaviors do not align clearly with the other clusters, potentially capturing residual or unstructured variation. Note, in addition, that cluster 3 is also the least populated. It can, therefore, reasonably be thought that the subjects included in clusters 1 and 2 better identify the characteristics of subjects who do not use a bank account, while the subjects in cluster 4 better identify those who use it. Regarding cluster 3, further investigations will be carried out to provide an interpretation of the characteristics of the subjects contained therein.

To profile users and non-users of bank accounts, descriptive statistics for the four clusters were analyzed (Table 6). The patterns reveal notable socio-economic and demographic contrasts that help explain financial-inclusion behaviors.

Income levels vary substantially across clusters. Cluster 1 shows the lowest income range, while Cluster 4 displays the highest and most dispersed values, confirming that higher income is positively associated with account ownership. Age differences are also evident: Cluster 3 includes older respondents (average 43 years), whereas Cluster 1 is younger and predominantly composed of non-users, suggesting that younger individuals are less likely to engage with formal banking. Household size appears similar across groups, providing little evidence of any relationship between family composition and inclusion.

Gender composition differs sharply. Clusters 1 and 2 are overwhelmingly female (82% and 99%, respectively) and largely made up of non-users, whereas Clusters 3 and 4 are predominantly male. This pattern reinforces the gender gap highlighted in the regression analysis, indicating that women remain structurally and institutionally disadvantaged in accessing formal finance.

Religion is relatively balanced across clusters, with Hindus slightly prevailing. Unlike the regression analysis, however, clustering does not reveal strong religious differentiation. Tax compliance, by contrast, shows a clear divide: over 97% of Cluster

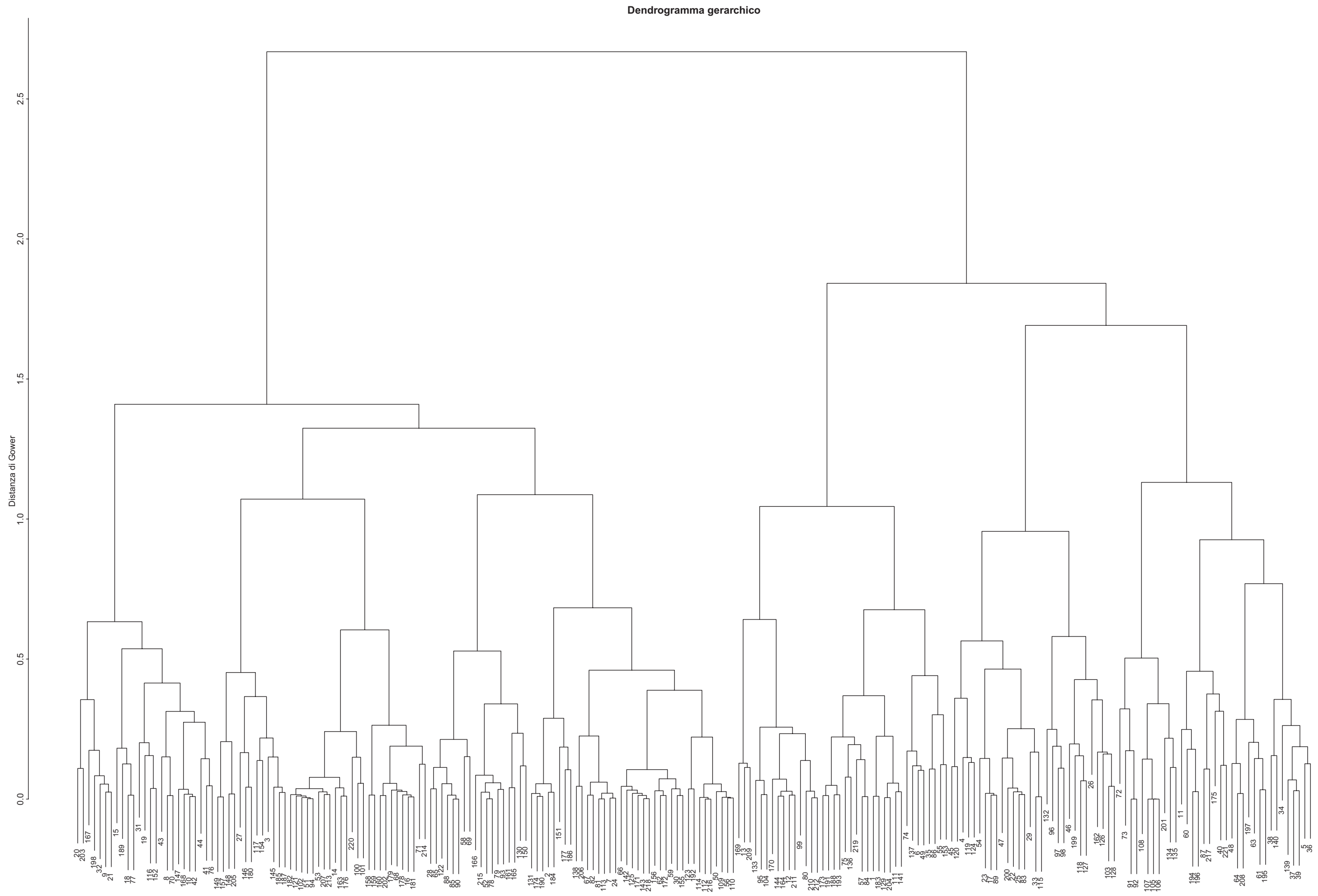

FIGURE 1 | Cluster analysis: Dendrogram.

TABLE 5 | Cluster composition—*Usebankaccount*.

Cluster	No	Yes
1	68.42%	31.58%
2	86.44%	13.56%
3	50.00%	50.00%
4	11.43%	88.57%

4 members—those with active bank accounts—pay taxes, while non-payers dominate Clusters 1–3. This finding confirms the regression result that formalization and tax engagement strongly correlate with financial inclusion.

Loan experience and education appear less discriminating. Across all clusters, most respondents have never received a loan, and education levels do not clearly separate users from non-users. Yet, Cluster 1, mainly female non-users, includes many respondents whose children have completed school, suggesting an indirect link between family education and future inclusion. Similarly, home ownership does not vary systematically with account use, confirming the regression finding that property ownership is not a decisive factor.

Overall, the cluster analysis complements the regression results by reaffirming the central roles of income and institutional engagement while illustrating the persistent gender disparities that shape financial inclusion in Seelampur.

8.2 | Discussion on Cluster Analysis and Causality Issues

This study represents a significant departure from prior research on financial inclusion by applying cluster analysis—a method largely absent from existing development economics literature. While cluster analysis is not a causal tool in the econometric sense, it improves causal interpretation by revealing heterogeneous behavioral profiles that are obscured in linear models. In settings like Seelampur—characterized by small samples, diverse socio-economic conditions, and non-linear behavioral patterns—this capacity to detect heterogeneity provides valuable information on where and for whom causal mechanisms may differ.

The four clusters identified display distinct socio-economic and institutional configurations. Clusters 1 and 2 consist largely of low-income women with weak institutional engagement and represent the structurally excluded; Cluster 4, composed predominantly of higher-income, tax-compliant men, reflects the fully included, and Cluster 3 forms a transitional group between these extremes. This segmentation shows that income, gender, and institutional engagement operate differently across subgroups, offering a deeper and more

TABLE 6 | Outcome from cluster analysis.

Per-capita income					
Min	**Max**	**Median**	**Mean**	**Variance**	**Coefficient of variation**
Cluster 1					
0	20,000	4833	5916	18,132,699	0.7198
Cluster 2					
3333	21,000	3000	3778	9,818,205	0.8294
Cluster 3					
1250	20,000	4000	5003	15,836,187	0.7954
Cluster 4					
2000	66,667	12,000	20,096	33,325,764	0.9084

Age					
Min	**Max**	**Median**	**Mean**	**Variance**	**Coefficient of variation**
Cluster 1					
14	35	21	21.47	19.82	0.2073
Cluster 2					
17	90	35	39.04	186.82	0.35
Cluster 3					
20	68	43	43.5	186.53	0.314
Cluster 4					
17	67	32	36.17	252.68	0.4395

Family members					
Min	**Max**	**Median**	**Mean**	**Variance**	**Coefficient of variation**
Cluster 1					
1	9	4	4.211	3.3599	1.0574
Cluster 2					
1	12	4	4.483	3.4826	0.4163
Cluster 3					
1	7	4	4.5	1.6379	0.2844
Cluster 4					
1	7	4	4.286	1.7983	0.3129

Gender		
Cluster	**Woman**	**Man**
1	81.58%	18.42%
2	99.15%	0.85%
3	6.67%	93.33%
4	34.29%	65.71%

Religion		
Cluster	**Hindu**	**Muslim**
1	60.53%	39.47%
2	65.25%	34.75%

(Continues)

TABLE 6 | (Continued)

Religion		
Cluster	**Hindu**	**Muslim**
3	53.33%	46.67%
4	62.82%	37.14%

Tax payment		
Cluster	**No**	**Yes**
1	92.11%	7.89%
2	99.15%	0.85%
3	96.67%	3.33%
4	2.86%	97.14%

Previous loans		
Cluster	**No**	**Yes**
1	79.00%	21.00%
2	88.00%	12.00%
3	93.00%	7.00%
4	97.00%	3.00%

Education							
Cluster	**Class10**	**Class12**	**Class5**	**Class8**	**Graduate**	**Illiterate**	**Master**
1	5.26%	10.53%	18.42%	39.47%	7.89%	15.79%	2.63%
2	65.25%	5.93%	11.86%	3.39%	12.71%	0.85%	0.00%
3	20.00%	6.67%	50.00%	3.33%	16.67%	3.33%	0.00%
4	17.14%	20.00%	11.43%	20.00%	8.57%	22.86%	0.00%

House ownership		
Cluster	**Own house**	**Rented house**
1	100.00%	0.00%
2	66.10%	33.90%
3	90.00%	10.00%
4	97.17%	2.86%

Children education						
Cluster	**College going**	**Drop out from school**	**No children**	**School finished**	**School going**	**Too young**
1	0.00%	0.00%	0.00%	94.74%	0.00%	5.26%
2	58.47%	1.69%	5.93%	5.93%	17.80%	10.17%
3	50.00%	10.00%	10.00%	13.33%	16.67%	0.00%
4	25.71%	5.71%	5.71%	37.14%	20.00%	5.71%

context-sensitive interpretation of the pathways suggested by regression analysis.

From a causal-reasoning standpoint, cluster analysis contributes by uncovering treatment heterogeneity—variation in behavioral responses that standard regressions suppress. The concentration of bank account users with higher income and tax compliance in Cluster 4 reinforces the regression-based interpretation that economic formalization is a key driver of inclusion (Sengupta 2019; Allen et al. 2016). Meanwhile, the mixed outcomes in Cluster 3, despite similar observable traits, point to possible unobserved moderators such as trust, information asymmetries, and intra-household decision-making processes (Galiani et al. 2022). Identifying such subgroups strengthens inference by showing where unobserved heterogeneity—not model misspecification—likely drives residual variance in the logit model.

Methodologically, clustering imposes fewer parametric restrictions than logistic regression and incorporates both categorical and continuous variables through Gower distance, making it suitable for informal and non-linear urban contexts where socio-economic traits and institutional norms interact in complex ways (Barkema et al. 2024). The alignment between the characteristics of inclusive clusters (e.g., high income, tax compliance) and the key regression predictors provides a non-parametric cross-validation of the hypothesized mechanisms.

Variables such as religion, loan history, and homeownership do not define the clusters, suggesting that their effects—statistically significant in some regressions—are either indirect or context-dependent. Distinguishing between variables that form stable behavioral segments and those that do not helps clarify which relationships are robust enough to have plausible causal relevance.

In sum, although cluster analysis cannot eliminate endogeneity or omitted-variable problems, it strengthens causal interpretation by mapping how combinations of traits co-occur with inclusion and by identifying heterogeneity in the mechanisms suggested by the logit and PLS-SEM results. This triangulated approach aligns with recent calls in development economics for mixed, context-sensitive methodologies to explore causal pathways (Karlan and Valdivia 2011; Galiani et al. 2022), enhancing the internal validity and policy relevance of the study's conclusions.

9 | Conclusions, Future Research, and Policy Implications

This study examined the determinants of formal bank account ownership among residents of Seelampur, Delhi, one of the largest urban slums in India. Using original survey data, it combined logistic regression, PLS-SEM, and hierarchical cluster analysis to identify the structural and institutional drivers of financial inclusion. The results from the regression analysis highlight the importance of education, income, and institutional engagement—through tax payments and prior borrowing experience—as significant predictors of formal bank account ownership. In contrast, a persistent gender gap continues to disadvantage women, while Muslim respondents show slightly higher inclusion rates, likely reflecting the role of community-based outreach and social networks.

The findings from the PLS-SEM analysis confirm and extend the regression results. The PLS-SEM model identifies two latent constructs—Socio-Economic Status (SES) and Institutional Engagement (IE)—both of which exert significant direct effects on financial inclusion. In addition, the results indicate that higher education and income levels increase the likelihood of engaging with formal institutions such as tax systems and regulated lenders. This mediating pathway underscores how capability and institutional participation interact, offering a deeper structural explanation for the mechanisms observed in the regression results.

The cluster analysis further enriches this interpretation by uncovering the heterogeneous behavioral profiles that underlie the average relationships captured in the regression and PLS-SEM models. Four distinct clusters emerge: two dominated by low-income women with weak institutional engagement, one transitional cluster with mixed inclusion outcomes, and one cluster composed largely of higher-income, tax-compliant men with strong engagement in formal institutions. These findings show that formal inclusion is not only a function of individual characteristics but also of how socio-economic traits and institutional behaviors combine within specific subgroups. Cluster analysis therefore adds a valuable non-parametric validation of the mechanisms identified through regression and PLS-SEM and highlights multiple, context-specific pathways to financial inclusion.

In addition, in the context of this study, it emerges that most respondents who hold a formal account report doing so with a public-sector commercial bank (approximately two-thirds),[17] followed by private-sector commercial banks and a smaller number of cooperative or regional-rural banks. This distribution reflects both the predominance of state-owned banks in low-income urban areas and the limited reach of private institutions in informal settlements. It also reinforces the centrality of public financial infrastructure in extending formal inclusion to marginalized populations.

The study's integrated approach provides an empirically grounded understanding of how formalization and capability interact in shaping inclusion outcomes in marginalized urban settings. Policy strategies should therefore combine capability enhancement—through education and financial literacy—with formalization pathways, such as simplified identification and tax procedures, to promote sustained inclusion. Finally, community-based organizations and gender-sensitive programs remain essential for transforming access into meaningful financial participation, contributing to the achievement of SDGs 1, 5, and 10.

Finally, the empirical insights from this study support a multi-level set of implications. For policymakers, simplifying documentation and strengthening local financial infrastructure—particularly through public-sector banks—can reduce administrative frictions and expand formalization. Financial institutions should prioritize gender-sensitive onboarding and neighborhood-level outreach to better engage low-income women. Community organizations can play a vital role in rebuilding trust and improving financial literacy, especially among the structurally excluded clusters identified in this study. Future research should build on these findings by using panel data or quasi-experimental approaches to better address causality, expanding the analysis to multiple slum areas, and incorporating richer measures of financial behavior and institutional trust. This would further enhance understanding of how financial inclusion evolves and how policy interventions can be tailored to heterogeneous urban populations.

Acknowledgements

The authors would like to thank Chiara Fornaciari, Mattia Messaggio, and Camilla Mondani for their valuable support in the data collection process. WE are indebted to Buyan Arvijikh Boldbaatar for his insightful feedback on earlier drafts of this paper. Lucia Dalla Pellegrina acknowledges support by the Baffi Center at Bocconi University. Open access publishing facilitated by Universita degli Studi di Milano-Bicocca, as part of the Wiley - CRUI-CARE agreement.

Endnotes

[1] For conceptual consistency, formal bank account ownership refers exclusively to accounts held in regulated financial institutions—namely, commercial banks, cooperative banks, or postal banks. Accounts opened with non-regulated microfinance organizations, self-help groups, or informal savings schemes were not classified as formal accounts. This operationalization follows the World Bank's Global Findex definition of financial inclusion.

[2] A map of Seelampur can be found in Appendix F (Figure A4).

[3] Data collection was approved by the Maverick Foundation's internal ethics committee and followed voluntary informed-consent procedures consistent with the World Bank's ethical research guidelines.

[4] See Appendix A for details on the questionnaire.

[5] Enumerators were selected from the local community to increase trust and accuracy of responses, reducing social-desirability bias and nonresponse rates.

[6] The Seelampur population exceeds 400,000 according to Delhi's Urban Development Department (2019), though exact figures vary due to rapid migration and informal settlement expansion.

[7] Currency conversions (INR to USD) are based on the 2019 average exchange rate of 81 INR = 1 USD (World Bank, 2020).

[8] See Appendix B for a breakdown of owners and non-owners of a bank account.

[9] See Table A2 in Appendix C for details. Figure A1 in Appendix C reports the correlation heatmap.

[10] See Appendix D (Goodness of Fit and Model Validation).

[11] See Appendix D (Gender Stratified Regressions).

[12] See Appendix E for technical details on the PLS_SEM model. Specifically, Figure A2 a graphical representation of the measurement model, while a graphical representation of the structural model is reported in Figure A3.

[13] The measurement model shows satisfactory reliability and convergent validity (all standardized loadings >0.70; Cronbach's $\alpha = 0.79$–0.84; AVE >0.5), and discriminant validity is confirmed by the Fornell–Larcker criterion. The structural model explains 63% of the variance in financial inclusion ($R^2 = 0.63$), with socio-economic status ($\beta = 0.38$, $p < 0.01$) and institutional engagement ($\beta = 0.41$, $p < 0.01$) exerting significant direct effects. The overall goodness-of-fit index (SRMR = 0.07) indicates a well-specified model (Colamartino et al. 2025). See Tables A7–A10 in Appendix E for details. Together, these results confirm that the structural and institutional pathways identified in the regression analysis remain robust when modeled through latent constructs and causal paths. Full PLS-SEM measurement and structural model statistics, including factor loadings, reliability coefficients, and bootstrapped path estimates, are available from the authors upon request.

[14] Where with $N = 1$ we obtain the Manhattan distance, with $N = 2$ the Euclidean distance and with $N = \infty$ the Chebychev distance.

[15] Ward's method, also known as the minimum variance method, is based on the choice of clusters to merge at each iteration, based on an optimal value of some objective function. Specifically, it aims to minimize the sum of squared errors within clusters. In the hierarchical clustering process with Ward's method and Gower distance matrix, the clusters are merged in such a way as to minimize the sum of the squares of the differences between the observations and the centroids of the clusters to which they belong. At each step of the algorithm, more similar clusters are merged, considering the aforementioned distance matrix as a similarity measure.

[16] Cluster Analysis Basics and Extensions. R package version 2.1.4. Maechler M. and Rousseeuw P. and Struyf, A. and Hubert M. and Hornik K., 2022.

[17] Consistent with this distribution, respondents generally hold accounts with government banks, while a smaller share also uses formal-sector institutions such as Kotak Mahindra Bank, HDFC Bank, Axis Bank, and Yes Bank.

[18] The regression output of the null model is available upon request.

[19] The presence of a large number of zero values in the loadings matrix is an inherent feature of the adopted PLS-SEM model specification. Within this framework, each indicator loads exclusively on the latent construct to which it is theoretically assigned and does not contribute to the measurement of other constructs.Since all constructs in our model are specified as reflective and unidimensional, and cross-loadings are not allowed, the software reports a non-zero loading only in the column corresponding to the relevant construct, while assigning zeros to all remaining columns. As a result, the loadings matrix appears structurally "sparse," that is, characterized by a high proportion of zero entries, for the following reasons:

(i) the three socio-economic status indicators load exclusively on SES;(ii) the institutional engagement indicators load exclusively on CI;(iii) *usebankaccount* loads only on FI;(iv) the exogenous variables (gender, age, religion, and household size), being single-item measures rather than latent constructs, exhibit a unit loading on their own dimension and zero loadings on all others.

References

Addai, B. 2017. "Women Empowerment Through Microfinance: Empirical Evidence From Ghana." *Journal of Finance and Accounting* 5, no. 1: 1–11. https://doi.org/10.11648/j.jfa.20170501.11.

Allen, F., A. Demirgüç-Kunt, L. Klapper, and M. S. M. Peria. 2016. "The Foundations of Financial Inclusion: Understanding Ownership and Use of Formal Accounts." *Journal of Financial Intermediation* 27: 1–30. https://doi.org/10.1016/j.jfi.2015.12.003.

Banerjee, A., D. Karlan, and J. Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7, no. 1: 1–21. https://doi.org/10.1257/app.20140287.

Barkema, H. G., J. A. Coyle-Shapiro, and E. M. le Grand. 2024. "How Do Fieldworkers in Poverty Craft Meaningful Roles to Achieve Social Impact? Female Teachers in Slums in India." *Academy of Management Journal* 67, no. 1: 232–261. https://doi.org/10.5465/amj.2021.1244.

Beck, T., A. Demirgüç-Kunt, and M. S. M. Peria. 2007. "Reaching Out: Access to and Use of Banking Services Across Countries." *Journal of Financial Economics* 85, no. 1: 234–266. https://doi.org/10.1016/j.jfineco.2006.07.002.

Colamartino, C., F. Manta, and P. Toma. 2025. "NFTs for Certified Products: A Heritage to Protect on the "Table" of the Metaverse." *Applied Economics* 57, no. 32: 4785–4800.

Cole, S., T. Sampson, and B. Zia. 2011. "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?" *Journal of Finance* 66, no. 6: 1933–1967. https://doi.org/10.1111/j.1540-6261.2011.01680.x.

Demirgüç-Kunt, A., and L. Klapper. 2013. "Measuring Financial Inclusion: Explaining Variation in Use of Financial Services Across and Within Countries." *Brookings Papers on Economic Activity*: 279–321. http://www.jstor.org/stable/23594869.

Demirgüç-Kunt, A., L. Klapper, D. Singer, and S. Ansar. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. World Bank Publications.

Demirgüç-Kunt, A., L. Klapper, D. Singer, S. Ansar, and J. Hess. 2022. "The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19." *World Bank*. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099818107072234182.

Everitt, B. S., S. Landau, M. Leese, and D. Stahl. 2011. *Cluster Analysis*. 5th ed. Wiley. https://doi.org/10.1002/9780470977811.

Galiani, S., P. Gertler, and C. Navajas-Ahumada. 2022. “Trust and Saving in Financial Institutions by the Poor.” *Journal of Development Economics* 159: 102981. https://doi.org/10.1016/j.jdeveco.2022.102981.

Ghosh, S., and D. Vinod. 2017. “What Constrains Financial Inclusion for Women? Evidence From Indian Micro Data.” *World Development* 92: 60–81. https://doi.org/10.1016/j.worlddev.2016.11.011.

Gower, J. C. 1971. “A General Coefficient of Similarity and Some of Its Properties.” *Biometrics* 27: 857–871. https://doi.org/10.2307/2528823.

Grohmann, A., T. Klühs, and L. Menkhoff. 2018. “Does Financial Literacy Improve Financial Inclusion? Cross-Country Evidence.” *World Development* 111: 84–96. https://doi.org/10.1016/j.worlddev.2018.06.020.

Hansen, N., M. A. Huis, and R. Lensink. 2019. “Microfinance Services and Women's Empowerment.” In *Handbook on Ethics in Finance*, 1–22. Springer International Publishing. https://doi.org/10.1007/978-3-030-29371-0_4.

Holvoet, N. 2005. “The Impact of Microfinance on Decision-Making Agency: Evidence From South India.” *Development and Change* 36, no. 1: 75–102. https://doi.org/10.1111/j.0012-155X.2005.00403.x.

Kabeer, N. 2001. “Conflicts Over Credit: Re-Evaluating the Empowerment Potential of Loans to Women in Rural Bangladesh.” *World Development* 29, no. 1: 63–84. https://doi.org/10.1016/S0305-750X(00)00081-4.

Karim, L. 2011. “Microfinance and Its Discontents: Women in Debt in Bangladesh.” U of Minnesota Press. https://doi.org/10.5749/minnesota/9780816670949.001.0001.

Karlan, D., and M. Valdivia. 2011. “Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions.” *Review of Economics and Statistics* 93, no. 2: 510–527. https://doi.org/10.1162/REST_a_00074.

Kim, D. W., J. S. Yu, and M. K. Hassan. 2020. “The Influence of Religion and Social Inequality on Financial Inclusion.” *Singapore Economic Review* 65, no. 1: 193–216. https://doi.org/10.1142/S0217590817460031.

Ledgerwood, J., J. Earne, and C. Nelson. 2013. *The New Microfinance Handbook: A Financial Market System Perspective*. World Bank Publications. https://doi.org/10.1596/978-0-8213-8927-0.

Lusardi, A., and O. S. Mitchell. 2011. “Financial Literacy Around the World: An Overview.” *Journal of Pension Economics and Finance* 10, no. 4: 497–508. https://doi.org/10.1017/S1474747211000448.

Nefla, D., R. Hassan, S. Jellouli, and S. Ben Thabet. 2025. “Financial Inclusion as a Sustainable Development Imperative in Africa: Trends, Gaps, and Future Directions.” *Sustainable Development*. https://doi.org/10.1002/sd.70491.

Nolan, L. B. 2015. “Slum Definitions in Urban India: Implications for the Measurement of Health Inequalities.” *Population and Development Review* 41, no. 1: 59–84. https://doi.org/10.1111/j.1728-4457.2015.00026.x.

Oanh, T. T. K., and L. Q. Dinh. 2024. “Digital Financial Inclusion, Financial Stability, and Sustainable Development: Evidence From a Quantile-On-Quantile Regression and Wavelet Coherence.” *Sustainable Development* 32, no. 6: 6324–6338. https://doi.org/10.1002/sd.3021.

Pandey, A., R. Kiran, and R. K. Sharma. 2025. “Examining Financial Inclusion Success in North India via Financial Inclusion Drivers, Financial Literacy and Financial Initiatives: A Variance Based-SEM Approach.” *Journal of Open Innovation: Technology, Market, and Complexity* 11: 100542. https://doi.org/10.1016/j.joitmc.2025.100542.

Reserve Bank of India. 2018. “Annual Report 2017–2018.” Reserve Bank of India. https://www.rbi.org.in.

Samer, S., I. Majid, S. Rizal, M. R. Muhamad, and N. Rashid. 2015. “The Impact of Microfinance on Poverty Reduction: Empirical Evidence From Malaysian Perspective.” *Procedia - Social and Behavioral Sciences* 195: 721–728. https://doi.org/10.1016/j.sbspro.2015.06.343.

Sarma, M., and J. Pais. 2011. “Financial Inclusion and Development.” *Journal of International Development* 23, no. 5: 613–628. https://doi.org/10.1002/jid.1698.

Sengupta, U. 2019. “State-Led Housing Development in Brazil and India: A Machinery for Enabling Strategy?” *International Journal of Housing Policy* 19, no. 4: 509–535. https://doi.org/10.1080/19491247.2018.1510076.

Swamy, V. 2014. “Financial Inclusion, Gender Dimension, and Economic Impact on Poor Households.” *World Development* 56: 1–15. https://doi.org/10.1016/j.worlddev.2013.10.019.

Us-Saqib, N., and G. M. Arif. 2012. “Time Poverty, Work Status and Gender: The Case of Pakistan.” *Pakistan Development Review* 51, no. 1: 23–46. https://doi.org/10.30541/v51i1pp.23-46.

Wedel, M., and W. A. Kamakura. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. 2nd ed. Springer. https://doi.org/10.1007/978-1-4615-4651-1.

World Bank. 2024. “World Bank Country Classifications by Income Level for 2024–2025.” https:// blogs. world bank. org/ en/ opend ata/ world-bank-country-classifications -by-income-level-for-2024-2025.

Appendix A

Questionnaire

Maverick Foundation.

Microenterprise and Microeconomics.

[Area:_____________________________].

Section A: Personal Information

1. Name of Respondent: _________________________
2. Gender:

M

F

T

3. Age in years: _____________________
4. Business Relationship:
   i. Owner
   ii. Worker
   iii. Support Services
   iv. Unemployed
   v. Pension
5. Qualification:
   i. Illiterate
   ii. Class 5
   iii. Class 8
   iv. Class 10
   v. Class 12
   vi. Graduate
   vii. Masters
6. Religion:
   i. Hindu
   ii. Muslim
   iii. Sikh
   iv. Other
7. Cast:
   i. General
   ii. OBC
   iii. SC
   iv. ST
8. Marital Status:
   i. Married
   ii. Unmarried
   iii. Divorced
   iv. Widow
9. Spouse Occupation:
   i. Same Business
   ii. Job
   iii. Labor Work
   iv. Unemployed
   v. NA
10. Children:
    i. 1
    ii. 2
    iii. More than 3
    iv. NA
11. Small children (0–14):
    i. Yes
    ii. No
12. Newborn (0–6):
    i. Yes
    ii. No
13. Children Education:
    i. No children
    ii. Too young
    iii. School Going
    iv. College Going
    v. Drop Out from school
    vi. School finished
14. Children Support:
    i. No children
    ii. Supporting
    iii. No support
15. Family members: ________________________
16. Years in the area:
    i. Less than 1
    ii. 1–5 years
    iii. More than 5 years

Section B: Enterprise

1. Type of business:
   i. Skilled
   ii. Labor
   iii. Service
   iv. NA
2. Business Trait:
   i. Recycling
   ii. Manufacturing
   iii. Artisan
   iv. Handicraft
   v. Eatery
   vi. Tailoring
   vii. Dairy
   viii. Roadside Vendor
   ix. Garments
   x. Servant
   xi. NA
   xii. Others
3. Experience in the business area: _________________________
4. How did you get into this business?
   i. By Chance
   ii. Family
   iii. By Choice
   iv. NA
5. Place of Business?
   i. Own Shop
   ii. Rented Shop
   iii. Roadside Shade
   iv. Movable Cart
   v. Outside Seelampur
   vi. House
   vii. NA
6. How many workers directly earn from your business?
   i. Self
   ii. 1–2

iii. 3–5
iv. Others

Section C: Skills for Enterprise

1. Does your business need specific skills?
   i. Yes
   ii. No
   iii. NA
2. What type of skills [__________________________________]
3. How did you learn the skills needed for your business?
   i. Family
   ii. By practice
   iii. Govt.
   iv. By school
   v. Others
4. Do you feel that training can improve your business?
   i. Yes
   ii. No
   iii. NA
5. How? [____]

Section D: Economic Status

1. How much do you earn? [_______________________________]
2. Total family income [______________________________]
3. Growth
   i. Yes
   ii. No
   iii. NA
4. Do you use a bank account?
   i. Yes
   ii. No
5. Do you pay any type of tax?
   i. Yes
   ii. No
6. Have you ever taken a loan from the Bank or local lender?
   i. Yes
   ii. No
   iii. Not needed
7. Purpose of loan:
   i. Business
   ii. Health
   iii. Debt
   iv. Education
   v. Personal
   vi. House
8. Tot. Loan [____]
9. Do you save money?
   i. Yes
   ii. No
10. Ownership
   i. Own House
   ii. Rented House
11. Have Scooter or Bike?
   i. Yes
   ii. No
12. Have own loading vehicle?
   i. Yes
   ii. No
13. Have TV?
   i. Yes
   ii. No

Section E: Happiness Index and Future

1. Are you happy with your business?
   i. Yes
   ii. No
2. Will you put your children into the same business?
   i. Yes
   ii. No
3. Did you ever make a growth plan for your business?
   i. Yes
   ii. No
4. Do you feel that you will ever change your business?
   i. Yes
   ii. No

Appendix B

Mean Values—User/Non-User of Bank Account

Table A1 provides a comparison of the variables dividing individuals who use a bank account (*Usebankaccount*: YES) and those who do not (*Usebankaccount*: NO). The evidence emerging from this analysis suggests that interviewees who use a bank account have a significantly higher average income (12,886 rupees) compared to non-users (3,881 rupees). Regarding gender, females constitute 47.30% of bank account users, but a higher percentage (86.39%) of non-users. Contrarily, males constitute 52.70% of bank account users but a lower percentage (13.61%) of non-users. The average age of bank account users (35.24 years) is slightly lower than that of non-users (36.64 years). Among bank account users, 56.76% identify as Hindu and 43.24% as Muslim. Among non-users, 65.31% identify as Hindu and 34.69% as Muslim. A higher percentage (47.30%) of bank account users pay taxes than non-users (2.72%). Furthermore, bank account users (9.52%) are less likely to have received a loan compared to non-users (14.86%). Bank account users

TABLE A1 | Mean values for users and non-users of a bank account.

Variable	***Usebankaccount*: YES**	***Usebankaccount*: NO**
Percapitaincomerupees	12,886	3,881
Gender: Female	47.30%	86.39%
Gender: Male	52.70%	13.61%
Age	35.24	36.64
Religion: Hindu	56.76%	65.31%
Religion: Muslim	43.24%	34.69%
Payingtaxes	47.30%	2.72%
Haveyoureceivedaloan	9.52%	14.86%
Qualification: Graduate	1.36%	51.02%
Qualification: Illiterate	51.02%	0.00%
Qualification: Master	0.00%	1.35%
Ownership	89.19%	75.51%
Peopleinthefamily	4176	4524

have a lower percentage of graduates (1.36%) but a higher percentage of illiterates (51.02%) compared to non-users, who have a higher percentage of graduates (51.02%) and a very low percentage of illiterates. Moreover, a higher percentage (89.19%) of bank account users own their house than non-users (75.51%), whereas the average family size (4.176) among bank account users is slightly lower than that among non-users (4.524). In summary, individuals who use a bank account tend to have higher incomes, are more likely to be male, younger on average, have a diverse religious distribution with a slightly higher proportion of Hindus, are more likely to pay taxes, have lower rates of loan reception, lower rates of education (especially graduates), higher rates of home ownership, and slightly smaller family sizes compared to those who do not use a bank account. The differences observed across these variables highlight socio-demographic disparities between bank account users and non-users in this context.

Appendix C

Correlation Analysis

The following table reports the Cramer's V correlation coefficients between the response variable *Usebankaccount* and the covariates.

Appendix D

Robustness Analysis

This appendix complements the empirical analysis by providing the full set of econometric diagnostics, estimation procedures, and robustness checks. It includes model-significance statistics (Wald χ^2 and Likelihood-Ratio tests), the log-likelihood function, and parameter-stability diagnostics. It also details the estimation of marginal effects, the probit re-specification, the instrumental-variable (IV-logit) approach used to test potential endogeneity of income, and the bootstrap resampling procedure applied to verify standard-error robustness. All results confirm the stability and reliability of the estimates discussed in the main text.

TABLE A2 | Correlations between covariates and *Usebankaccount*.

Variable	Correlation with *Usebankaccount*
Percapitaincomerupees	0.7675
Gender	0.4171
Age	0.4809
Religion	0.0833
Payingtaxes	0.5518
Haveyoureceivedaloan	0.0796
Qualification (average years of school)	0.4265
Ownership	0.1617
Peopleinthefamily	0.2702
Childrened (average no. Children)	0.2230

Goodness of Fit and Model Validation

After the estimation and interpretation of the relevant parameters, we evaluate the goodness of the implemented model. We use the concept of deviance as a measure of the goodness of fit to the data through a likelihood ratio test. The latter compares the likelihood of the complete model with the likelihood of the null model containing only the intercept (null model).

We define two models, M_1 and M_2, which respectively represent the null (without covariates) model and the complete model (with covariates). The null hypothesis tests that the reduced model M_1 is sufficient to explain the data.

The test statistic compares the log-likelihood of the two models based on the following specification:

$$\Lambda = -2\left(l_{M_1} - l_{M_2}\right) \sim \chi^2_{(k_2-k_1)} \text{s. t. H0} \tag{A1}$$

where $k_2 > k_1$. We calculate the p-value to evaluate the outcome of the test (acceptance or rejection of the null hypothesis), obtained as $P\left(\chi^2_{(k_2-k_1)} > \lambda_{oss}\right)$, where λ_{oss} represents the value of the test statistic calculated on the available data.

The results of the statistical analysis are the following:

$$\lambda_{oss} = -2(-140.491 + 76.043) = 128.9 \tag{A2}$$

$$p\text{-}value = P\left(\chi^2_{(19)} > 128.9\right) \approx 0 \tag{A3}$$

This leads to the conclusion that the LRT applied to the parsimonious model leads to a rejection of the null hypothesis at any admissible significance threshold. It is therefore possible to assess that the complete model M_2 explains the data significantly better than the parsimonious model.[18]

Gender-Stratified Regressions

Several recent empirical studies in India demonstrate that gender differences in digital access, technology use, and institutional engagement

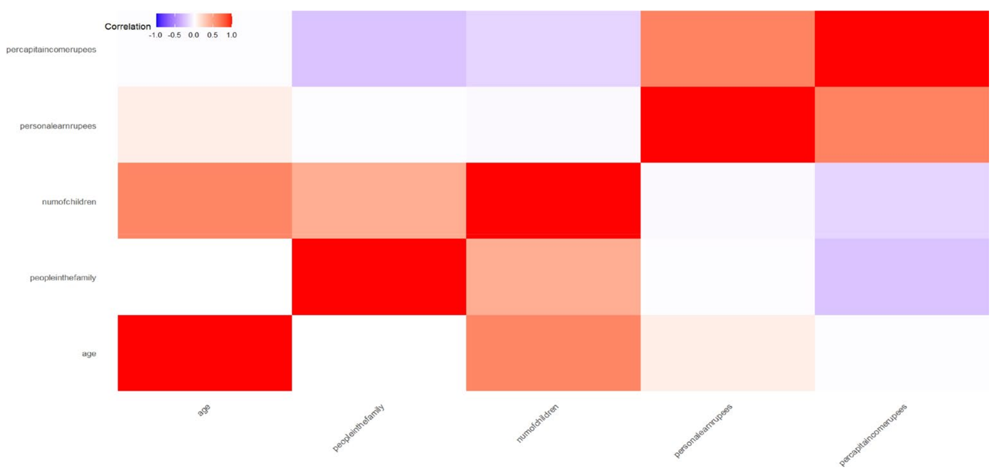

FIGURE A1 | Correlations between covariates and Usebankaccount.

financial services (Pandey et al. 2025). These findings reinforce our hypothesis (H2) and, together with the strong gender coefficient observed in the pooled model, justify stratifying the analysis by gender. Prior research consistently shows that men and women follow distinct financial pathways, shaped by differences in autonomy, mobility, bargaining power, and institutional engagement (Kabeer 2001; Hansen et al. 2019). Testing for gender heterogeneity, therefore, enables us to assess whether the determinants of financial inclusion operate differently across male and female respondents.

Results (Tables A3–A6) confirm this expectation. Among men, income, education, and religion are the primary determinants of inclusion: higher earnings, higher schooling, and Muslim affiliation all significantly increase the likelihood of owning a bank account. For women, however, religion plays no significant role. Instead, institutional engagement emerges as decisive: women who pay taxes or have previously borrowed are substantially more likely to use bank accounts, highlighting that financial formalization operates through distinct channels across genders.

The gender-stratified results reveal important contrasts in the pathways to financial inclusion. For men, financial inclusion is strongly associated with economic status and community background, as higher income, higher education, and Muslim affiliation significantly increase the likelihood of account ownership. This pattern suggests that male inclusion is embedded in structural inequalities, where economic standing and religious networks shape opportunities to access formal financial services. Such findings are consistent with cross-country evidence that men's banking behavior often reflects broader socio-economic positioning (Allen et al. 2016; Demirgüç-Kunt et al. 2018).

For women, by contrast, institutional engagement emerges as the critical channel. Tax compliance and loan history are the main predictors of financial inclusion, while income and religious identity are not significant. This indicates that women enter the financial system less through their economic or social status and more through direct contact with formal institutions. This mechanism resonates with studies showing that women's financial empowerment in India is often mediated by administrative processes—such as documentation, credit access, or tax filing—rather than by labor market earnings alone (Ghosh and Vinod 2017; Pandey et al. 2025).

The divergence between men and women underscores that financial inclusion is not only unequal in outcomes but also gendered in processes.

TABLE A3 | Estimated coefficients: Males.

Coefficient	Variable	Estimated coefficient	Std error	p
β_0	Intercept	−9.7213	4.5133	0.0312**
β_1	*Percapitaincomerupees*	0.0010	0.0005	0.0332**
β_2	*Religion Muslim = 1*	3.3590	1.6919	0.0471**
β_3	*PayingtaxesYes = 1*	1.5208	2.0024	0.4476
β_4	*HaveyoureceivedaloanYes = 1*	1.3875	1.9449	0.4756
β_5	*QualificationClass10 = 1*	6.8776	3.5314	0.0515*
β_6	*QualificationClass12 = 1*	9.4553	3.8861	0.0150**
β_7	*QualificationClass8 = 1*	5.9093	3.2159	0.0661*
β_8	*QualificationClass5 = 1*	5.8213	3.5901	0.1049
β_9	*QualificationGraduate = 1*	6.0591	3.5006	0.0835*
β_{10}	*ChildrenedCollegeGoing = 1*	0.4932	15.097	0.9739
β_{11}	*ChildrenedDropOut = 1*	3.4607	4.6654	0.4582
β_{12}	*ChildrenedNoChildren = 1*	−6.1946	2.1586	0.0041***
β_{13}	*ChildrenedSchoolFinished = 1*	−1.5072	1.7789	0.3968
β_{14}	*ChildrenedTooYoung = 1*	11.7250	2512.3425	0.9963
Pseudo-R^2: 0.6448				

Note: Significant level: * 10%, ** 5% e *** 1%.

TABLE A4 | Marginal effects (significant variables): Males.

Coefficient	Variable	Marginal effect	Std error	p
β_1	*Percapitaincomerupees*	0.0001	0.0005	0.0332**
β_2	*Religion Muslim = 1*	0.2217	1.6919	0.0471**
β_5	*QualificationClass10 = 1*	0.3928	3.5314	0.0515*
β_6	*QualificationClass12 = 1*	0.5311	3.8861	0.0150**
β_7	*QualificationClass8 = 1*	0.2905	3.2159	0.0661*
β_9	*QualificationGraduate = 1*	0.3012	3.5006	0.0835*
β_{12}	*ChildrenedNoChildren = 1*	−0.3563	2.1586	0.0041***

Note: Significant level: * 10%, ** 5% e *** 1%.

TABLE A5 | Estimated coefficients: Females.

Coefficient	Variable	Estimated coefficient	Std error	*p*
β0	Intercept	−2.3127	0.5884	0.0001***
β1	*Percapitaincomerupees*	0.0002	0.0001	0.0225**
β2	*Religion Muslim = 1*	−0.1702	0.6057	0.7788
β3	*PayingtaxesYes = 1*	2.8476	1.0364	0.0060***
β4	*HaveyoureceivedaloanYes = 1*	1.4069	0.7007	0.0447**
β5	*QualificationClass10 = 1*	1.8045	0.8811	0.0406**
β6	*QualificationClass12 = 1*	−0.3957	1.1307	0.7264
β7	*QualificationClass8 = 1*	0.0194	0.9668	0.9840
β8	*QualificationClass5 = 1*	−3.3150	2.6608	0.2128
β9	*QualificationGraduate = 1*	4.2242	1.4103	0.0027***
β10	*QualificationMaster = 1*	15.5713	3956.1806	0.9969
β11	*ChildrenedCollegeGoing = 1*	−0.9359	1.5828	0.5542
β12	*ChildrenedDropOut = 1*	−16.386	1242.5272	0.9895
β13	*ChildrenedNoChildren = 1*	−1.3854	0.8973	0.1226
β14	*ChildrenedSchoolFinished = 1*	−1.3517	0.9813	0.1684
β15	*ChildrenedTooYoung = 1*	−1.1048	1.1745	0.3469
Pseudo-R^2: 0.4396				

Note: Significant level: * 10%, ** 5% e *** 1%.

TABLE A6 | Marginal effects (significant variables): Females.

Coefficient	Variable	Marginal effect	Std error	*p*
β1	*Percapitaincomerupees*	0.0001	0.0001	0.0225**
β3	*PayingtaxesYes = 1*	0.4031	1.0364	0.0060***
β4	*HaveyoureceivedaloanYes = 1*	0.1521	0.7007	0.0447**
β5	*QualificationClass10 = 1*	0.2502	0.8811	0.0406**
β9	*QualificationGraduate = 1*	0.6329	1.4103	0.0027***

Note: Significant level: * 10%, ** 5% e *** 1%.

Men's pathways reflect structural inequalities tied to resources and networks, whereas women's pathways highlight the importance of institutional gateways that can either enable or constrain access. This expands on earlier work by Kabeer (2001), which showed that financial inclusion policies targeting women often fail if they do not address institutional barriers and intra-household power asymmetries.

From a policy perspective, these insights suggest that gender-sensitive interventions are essential for meeting SDG 5 (Gender Equality) and SDG 10 (Reduced Inequalities). For men, policies that stabilize income and reduce community-based barriers may be most effective. For women, interventions that simplify tax documentation, expand credit history opportunities, and enhance trust in financial institutions could have disproportionate benefits. Importantly, our results demonstrate that strategies cannot be one-size-fits-all: promoting inclusive and sustainable development in urban slums requires recognizing and addressing the distinct mechanisms by which men and women engage with financial systems.

Appendix E

PLS-SEM Models: Technical Details

PLS-SEM—Model 1

A first specification of the model was structured as follows:

- Financial Inclusion (FI): latent variable, measured by the ownership of a bank account. The construct is treated as reflective: financial inclusion, not directly observable, manifests through the indicator "bank account ownership," which reflects its level.
- Socio-economic Status (SES): reflective latent variable manifested by income, home ownership, and education level, which reflects the level of SES.
- Institutional Involvement (CI): reflective latent variable, measured by tax compliance and past loan history, which reflects the degree of an individual's involvement with institutions.

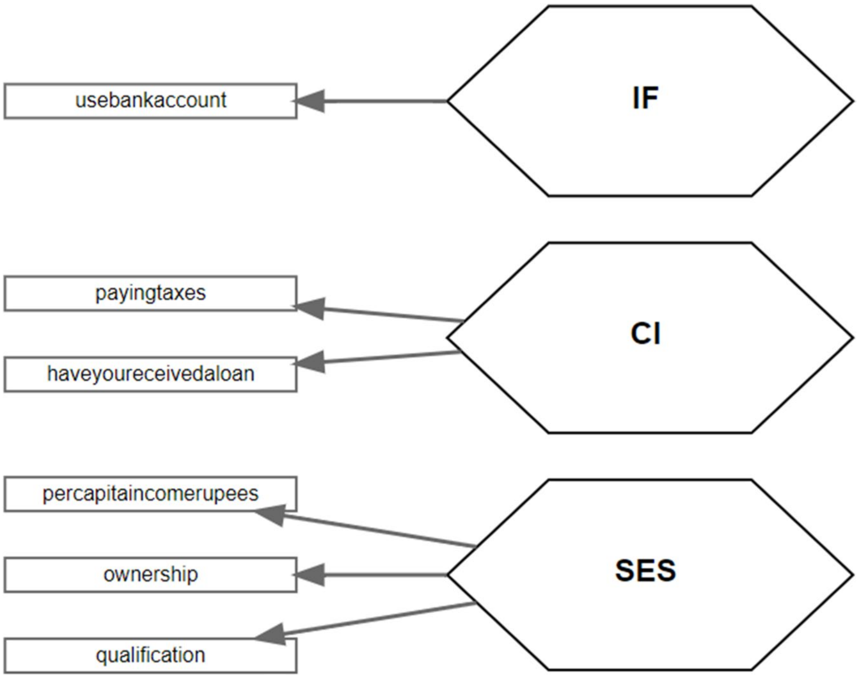

FIGURE A2 | Measurement model.

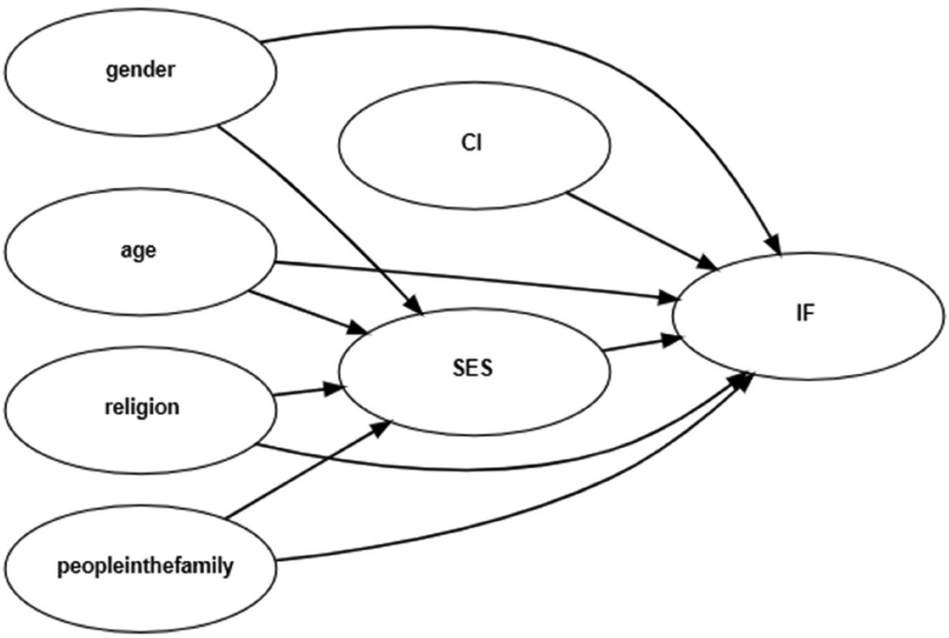

FIGURE A3 | Structural model.

- Gender, age, religion, and household size: exogenous variables affecting both SES and FI.

In this framework, the measurement model specifies how each latent construct is linked to its observed indicators, ensuring that the measures reliably reflect the underlying concepts. In contrast, the structural model captures the relationships among the latent variables and exogenous predictors. Specifically, it models the effects of socio-economic status, institutional involvement, and demographic traits on financial inclusion, thereby testing the theoretical dependencies among the constructs within the PLS-SEM framework.

Financial Inclusion Versus Socio-Economic Status

- R^2 and AdjR^2:
    - FI (Financial Inclusion): $R^2 = 0.402$, Adj$R^2 = 0.385$ indicate that about 40% of the variance in financial inclusion is explained by the model. This is a moderate value, implying acceptable predictive capacity.
    - SES (Socio-economic Status): $R^2 = 0.138$, Adj$R^2 = 0.122$ indicate that roughly 12%–14% of the variance in SES is explained by the model variables. This relatively low value suggests that SES is only weakly influenced by the included variables (age, gender, religion, number of household members).
- Path Coefficients (direct effects of exogenous on endogenous variables):
    - Financial Inclusion (FI):
        - SES → FI = 0.144. Positive and moderate: increases in SES lead to higher financial inclusion, though the effect is not strong.

- CI → FI = 0.390. The largest effect: institutional involvement (tax compliance, loan history) is the primary driver of financial inclusion.
- Gender → FI = 0.252. Positive in favor of males, in line with prior literature.
- Age → FI = −0.029. Negligible and slightly negative.
- Religion → FI = 0.054. Small positive value, indicating a very weak positive effect; religious affiliation has marginal influence on FI.
- Household size → FI = −0.065. Suggests that larger households are slightly less financially included.

◦ Socio-economic Status (SES):

- Gender → SES = 0.268. Indicates that gender affects SES; males exhibit, on average, higher SES than females.
- Age → SES = −0.128. Moderately negative: SES tends to decline slightly with age.
- Religion → SES = 0.107. Suggests a moderate positive effect of religion on SES.
- Household size → SES = −0.197. A comparatively stronger negative effect than age or religion: larger families tend to have lower SES.

TABLE A7 | FI vs. SES.

	FI	SES
R^2	0.402	0.138
AdjR^2	0.385	0.122
SES	0.144	NA
CI	0.390	NA
gender	0.252	0.268
age	−0.029	−0.128
religion	0.054	0.107
peopleinthefamily	−0.065	−0.197

Overall Interpretation

Financial inclusion is most influenced by institutional involvement, followed by gender and then socio-economic status. SES itself is only partially explained by the model's exogenous variables; household size and gender are the most relevant among them, while religion and age exhibit rather marginal effects. These relationships confirm that financial inclusion depends not only on socio-economic status but also on institutional and demographic factors.

TABLE A8 | Reliability.

	alpha	rhoC	AVE	rhoA
SES	0.401	0.686	0.439	0.498
CI	−0.031	0.554	0.498	−0.110
gender	1.000	1.000	1.000	1.000
age	1.000	1.000	1.000	1.000
religion	1.000	1.000	1.000	1.000
peopleinthefamily	1.000	1.000	1.000	1.000
FI	1.000	1.000	1.000	1.000

Reliability

- Socio-economic Status (SES):
  - alpha = 0.401. Very low, indicating poor internal consistency among indicators.
  - rhoC = 0.686. Just below the 0.7 threshold: fair but not optimal composite reliability.
  - rhoA = 0.498. Low, confirming weak internal consistency.
  - AVE = 0.439. Below the 0.5 threshold: insufficient convergent validity.
  - Interpretation: as a latent construct in this model, SES is poorly reliable; the current indicators are not sufficiently suited to represent it.
- Institutional Involvement (CI):
  - alpha = −0.031. A negative value signals severe inconsistency among indicators; some may even be inversely correlated.
  - rhoC = 0.554. Below 0.7: low composite reliability.
  - rhoA = −0.110. Negative and thus highly problematic.
  - AVE = 0.498. Just below 0.5.
  - Interpretation: CI is unreliable in the current specification; the indicators do not consistently capture institutional involvement.

Loadings

Outer loadings indicate how strongly each observed indicator relates to its latent construct (standardized correlation). Interpretation guidelines: ≥0.70 = strong; 0.40–0.70 = acceptable in exploratory studies (especially if construct AVE > 0.50); < 0.40 = weak and typically reconsidered.[19]

TABLE A9 | Loadings.

Indicator	SES	CI	gender	age	religion	peopleinthefamily	FI
percapitaincomerupees	0.872	0.000	0.000	0.000	0.000	0.000	0.000
ownership	0.454	0.000	0.000	0.000	0.000	0.000	0.000
qualification	0.592	0.000	0.000	0.000	0.000	0.000	0.000
payingtaxes	0.000	0.990	0.000	0.000	0.000	0.000	0.000
haveyoureceivedaloan	0.000	0.128	0.000	0.000	0.000	0.000	0.000
usebankaccount	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Gender	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Age	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Religion	0.000	0.000	0.000	0.000	1.000	0.000	0.000
peopleinthefamily	0.000	0.000	0.000	0.000	0.000	1.000	0.000

TABLE A10 | Bootstrap.

Path	Original Est.	Bootstrap mean	Bootstrap SD	T Stat.	2.5% CI	97.5% CI
SES → FI	0.144	0.150	0.066	2.180	0.023	0.282
CI → FI	0.390	0.389	0.080	4.888	0.226	0.540
gender → SES	0.268	0.271	0.064	4.179	0.138	0.390
gender → FI	0.252	0.252	0.066	3.842	0.120	0.381
age → SES	−0.128	−0.121	0.088	−1.447	−0.275	0.073
age → FI	−0.029	−0.029	0.047	−0.603	−0.119	0.066
religion → SES	0.107	0.103	0.065	1.651	−0.026	0.227
religion → FI	0.054	0.054	0.055	0.990	−0.053	0.164
peopleinthefamily → SES	−0.197	−0.199	0.054	−3.650	−0.301	−0.090
peopleinthefamily → FI	−0.065	−0.065	0.044	−1.478	−0.149	0.023

Bootstrap Method

Following the PLS-SEM estimation, the path coefficients represent the estimated effects among model variables. However, being sample-based, they do not by themselves indicate whether a path is statistically significant. To address this, bootstrap resampling assesses the stability and reliability of the coefficients by generating thousands of “virtual” samples from the original data, re-estimating the PLS-SEM on each, and recording the coefficients. This yields an empirical distribution for each path, from which one can compute the mean, standard deviation, and confidence intervals (CIs). If the CI does not include zero, the path is statistically significant. In this way, bootstrap turns deterministic PLS-SEM estimates into stronger inferential evidence, distinguishing reliable relationships from potentially spurious ones.

- Significant paths:
    - SES → FI: 0.144, $T = 2.18$, 95% CI [0.023, 0.282] → positive and significant: SES increases FI.
    - CI → FI: 0.390, $T = 4.888$, 95% CI [0.226, 0.540] → strongly positive and significant: CI is the main driver of FI.
    - Gender → SES: 0.268, $T = 4.179$, 95% CI [0.138, 0.390] → positive and significant: gender affects SES.
    - Gender → FI: 0.252, $T = 3.842$, 95% CI [0.120, 0.381] → positive and significant: confirms a direct gender effect on FI.
    - Household size → SES: −0.197, $T = -3.65$, 95% CI [−0.301, −0.090] → negative and significant: larger families have, on average, lower SES.
- Non-significant paths:
    - Age → SES ($T = -1.447$)
    - Age → FI ($T = -0.603$)
    - Religion → SES ($T = 1.651$)
    - Religion → FI ($T = 0.990$)
    - Household size → FI ($T = -1.478$)

Synthesis

The key determinants of financial inclusion are institutional involvement, gender, and socio-economic status—with CI as the strongest driver, followed by gender differences and SES. SES, in turn, is significantly influenced by gender and household size, suggesting that men and smaller families tend to enjoy better socio-economic positioning. By contrast, age, religion, and household size (as a direct predictor of FI) have no significant direct effects on financial inclusion, indicating a negligible role in this model. Finally, the bootstrap confidence intervals corroborate the robustness of the results: only significant paths have intervals that exclude zero, further supporting the stability of the observed relationships.

Appendix F

Map of Seelampur

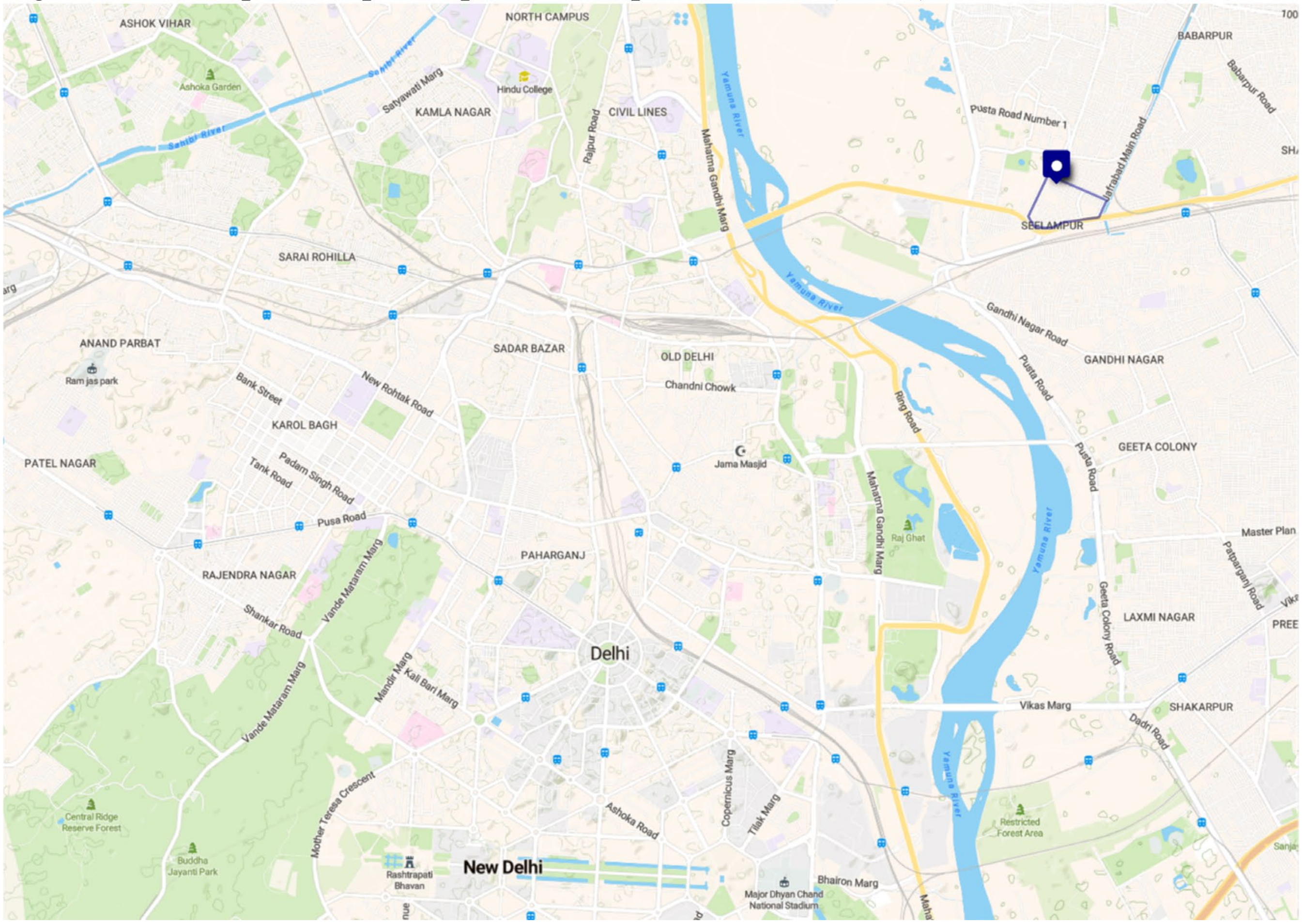

FIGURE A4 | Seelampur's Map—OpenStreetMap contributors (ODbL).