Research Square

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

Portable Verified Financial Identity: A Framework for Immigrant Credit Access and Financial Inclusion in the United States

Amir Khan

khan.295@wright.edu

Wright State University

Research Article

Keywords: financial inclusion, immigrant banking access, credit risk, portable financial identity, remittance corridors, machine learning

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10101564/v1

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

Additional Declarations: No competing interests reported.

Abstract

Despite comprising 15.3 percent of the United States population, foreign-born residents remain structurally excluded from formal financial services — not because of financial irresponsibility but because their financial identities are non-portable across borders. This paper introduces the Portable Verified Financial Identity (PVFI) framework, which aggregates origin-country financial infrastructure quality, bilateral remittance corridor strength, and pre-arrival capital signals into a Pre-Arrival Banking and Financial Score (PABFS) that can be assigned to immigrants before any domestic credit history is established. Drawing on eight datasets — six federal surveys, two multilateral databases, and one proprietary dataset — covering more than 100 million observations (FDIC household surveys across three waves 2019–2023; IPUMS ACS 25.6 million persons; HMDA 68.4 million applications 2019–2023; FINRA NFCS 2021; SCF 2022; and KNOMAD remittance corridors covering 134 countries and $200.2 billion in annual flows), we document a persistent nativity gap in banking access of 2.7 to 3.6 percentage points and marked racial disparities in mortgage denial. Logistic regression on 481,917 matched foreign-born respondents confirms the core prediction: a one-standard-deviation increase in PVFI score is associated with a 12.7 percent reduction in low-income financial exclusion odds (OR = 0.873; 95% CI [0.866, 0.879]; $p < 0.001$). Machine learning validation on the Home Credit dataset shows that PVFI-analog external creditworthiness features improve credit model AUC by 7.6 percentage points for logistic regression and 6.6 points for gradient boosting. The results support a regulatory pathway for portable financial identity systems as instruments of immigrant financial inclusion.

JEL Classification: G21, G28, J15, F24, G18, C55

1. Introduction

Forty-three million foreign-born residents comprise 15.3 percent of the United States population and collectively generated over $200 billion in outbound remittances in 2021. Yet they remain structurally excluded from formal credit markets because the financial records they built abroad are invisible to domestic underwriters. Credit scores require credit histories, and credit histories require years of domestic financial participation. An immigrant from any country arrives with those years already lived and those records already established — written in a language the American credit bureau cannot read.

The consequences are measurable. Across the 2019, 2021, and 2023 FDIC household survey waves, our analysis of FDIC microdata finds a persistent nativity gap of 2.7 to 3.6 percentage points in financial exclusion rates between foreign-born and native-born reference persons, after controlling for income, education, race, and geography. In mortgage markets, the Home Mortgage Disclosure Act provides annual population-level data on 68.4 million applications; Black applicants face roughly twice the denial odds of White applicants, and Hispanic applicants face 52 percent higher odds, even after controlling for income and loan purpose. These are not incidental patterns. They trace a fault line running through American credit markets, and immigrants sit on the wrong side of it.

This paper proposes a direct response: the Portable Verified Financial Identity (PVFI) framework. PVFI assigns each immigrant a pre-arrival score derived from three country-level signals — the quality and depth of the origin country's financial system; the strength of the bilateral remittance corridor linking that country to the United States; and the macroeconomic significance of that corridor relative to the United States economy. Aggregated into a Pre-Arrival Banking and Financial Score (PABFS), these signals constitute a portable financial credential that a lender could, in principle, consult before any domestic credit history has accumulated.

Four research questions structure the analysis. First, how large and persistent is the nativity gap in formal banking access? Second, what do five years of HMDA data reveal about racial disparities in credit denial? Third, does a country-level PVFI score predict individual financial vulnerability in the direction the framework predicts? Fourth, do PVFI-analog external creditworthiness indicators — proxied in our machine learning validation by verified credit bureau scores — measurably improve credit model discrimination?

Our principal results: the FDIC nativity gap runs from 2.72 to 3.55 percentage points across three survey waves; Black mortgage applicants face OR = 1.955 in denial odds versus White applicants ($p < 0.001$) and Hispanic applicants OR = 1.524 ($p < 0.001$) across all five HMDA years; a one-SD increase in PVFI score is associated with OR = 0.873 in the odds of low-income financial exclusion ($p < 0.001$; N = 481,917); and adding two external creditworthiness features raises credit model AUC by 7.56 percentage points for logistic regression and 6.55 points for gradient boosting, with those two features accounting for 68 percent of total model importance.

Section 2 reviews the relevant literature. Section 3 develops the theoretical framework. Section 4 describes the data and empirical strategy. Section 5 presents the results. Sections 6 and 7 discuss findings and policy implications. Section 8 acknowledges limitations. Section 9 concludes.

2. Background and Literature Review

2.1 Financial Inclusion and Immigrant Banking Access

Financial access shapes household welfare through channels documented across three decades of empirical research spanning multiple institutional settings. Beck et al. (2007) show that financial depth reduces income inequality and raises the income share of the poorest quintile; Allen et al. (2016) use Global Findex data to demonstrate that formal account ownership is associated with faster poverty reduction even after controlling per-capita income. Claessens (2006) identifies three primary barriers to access — cost, information asymmetry, and documentation requirements — and argues that reducing documentation barriers is the most tractable lever in high-income countries. For immigrants, the documentation barrier is categorical rather than marginal: the documents required by United States banks — Social Security numbers, domestic address history, domestic credit records — are absent at arrival by construction, not by circumstance.

Rhine et al. (2006) find that among unbanked households in the United States, immigrants are disproportionately represented even controlling for income, and that the primary stated reason for not

having a bank account is the identification requirement rather than cost. Amuedo-Dorantes and Bansak (2006) show that unbanked Mexican immigrants rely heavily on money transfer operators, paying higher fees and forgoing savings products. Hogarth and O'Donnell (2000) establish the empirical baseline for studying unbanked households in the United States that the FDIC surveys have extended. The immigrant earnings literature anchored by Chiswick (1978) and Borjas (1985) shows that while earnings converge toward native-born levels over time, the initial income penalty is real and partly reflects human capital non-portability. PVFI extends this insight to the financial dimension: the immigrant who cannot transfer her credit history faces a financial analog of the foreign-credential problem.

2.2 Racial Disparities in Credit Markets

Three decades of HMDA-based research have documented persistent racial disparities in mortgage lending. Squires (2003) traces insurance redlining as an antecedent, arguing that spatial patterns of financial exclusion reflect durable institutional practices rather than incidental market outcomes. Pager and Shepherd (2008) synthesize evidence across employment, housing, and consumer markets to show that discrimination is a general social structure rather than a sector-specific phenomenon. More recently, Bhutta and Hizmo (2021) exploit variation in loan officer racial composition to show that minority borrowers pay higher interest rates than observationally equivalent White borrowers, while Ambrose et al. (2021) find that both borrower and broker race independently predict mortgage costs. Bartlett et al. (2022) show that FinTech lenders discriminate less in rate-setting than traditional lenders but that approval rate disparities persist. Fuster et al. (2022) show that machine learning models improve default prediction for minority borrowers without necessarily improving their approval rates. These findings matter for PVFI because immigrants who are racial minorities face a compound disadvantage: the documentation gap that PVFI addresses and the discrimination gap that regulatory enforcement addresses.

2.3 Remittance Corridors and Financial Infrastructure

The $200.2 billion flowing outward from the United States through 134 documented bilateral corridors in 2021 is evidence of organized, disciplined financial behavior by a population that formal credit markets systematically underserve. Giuliano and Ruiz-Arranz (2009) show that remittances substitute for credit in countries with underdeveloped financial systems — an observation that implies remittance-sending immigrants are managing non-trivial financial obligations across borders without domestic credit support. Aggarwal et al. (2011) show that high-volume corridors promote financial development in recipient countries by increasing deposit mobilization, suggesting that corridor participants have a demonstrated propensity to interact with formal financial institutions. World Bank KNOMAD (2023) documents the United States as the world's largest source of outbound remittances, with the Mexico corridor alone carrying $52.6 billion — 26.3 percent of the documented total. The existing FinCEN monitoring architecture for these flows (Financial Crimes Enforcement Network 2021) provides a potential foundation for the identity verification component of a PVFI system.

2.4 Credit Risk Theory and Information Asymmetry

The theoretical case for PVFI rests on the information asymmetry framework of Stiglitz and Weiss (1981). When lenders cannot observe borrower quality, they ration credit rather than price it — excluding

creditworthy borrowers who cannot credibly signal their type. For immigrant borrowers, the asymmetry is categorical: conventional domestic credit signals are absent by construction. Merton (1974) provides the structural framework connecting observable financial characteristics to default probability; PVFI is a proposal to enlarge the set of observables available to lenders serving immigrant applicants. Altman's (1968) early discriminant analysis work, extended through Treacy and Carey (2000) to the modern machine learning literature of Chen and Guestrin (2016) and the SHAP interpretability framework of Lundberg and Lee (2017), has progressively refined default prediction but primarily for populations with domestic financial histories. Berg et al. (2020) are the closest analogy: they show that external digital footprint data substantially improves credit scoring for thin-file borrowers in European consumer lending. Our machine learning validation tests whether the same logic applies to PVFI-type externally verified information.

3. Theoretical Framework

3.1 The PVFI Architecture

Three observable features of an immigrant's origin country carry information about her likely pre-arrival financial identity. First, the quality of the origin country's financial system determines whether immigrants from that country have interacted with formal banks and credit institutions in ways that produce verifiable records. Second, the volume and regularity of the bilateral remittance corridor to the United States provides evidence of organized cross-border financial management by the sending community. Third, the macroeconomic significance of that corridor reflects the degree to which the payment infrastructure is robust and institutionally anchored.

The Origin Financial Infrastructure Score (OFIS) operationalizes the first signal. It is constructed from four normalized sub-components drawn from the World Development Indicators (World Bank 2024) and Global Findex 2021 (Demirgüç-Kunt et al. 2022): domestic credit to private sector as a percent of GDP, log GDP per capita at purchasing power parity, formal account ownership rate, and mobile money participation rate. Each sub-component is winsorized at the first and 99th percentiles and normalized to a 0–100 scale; the four scores are averaged with equal weights to produce OFIS. Japan (OFIS = 92.5) and South Korea (OFIS = 84.8) have deeply developed financial systems in which virtually all adults maintain formal accounts; Guatemala (OFIS = 25.9) has substantially lower financial depth.

The Corridor Integrity Score (CIS) and Pre-Arrival Capital Score (PACS) operationalize the second and third signals from KNOMAD 2021 bilateral outbound flows. CIS is computed as each corridor's share of total United States outbound remittances, log1p-transformed and normalized to 0–100 with only a lower-tail clip — preserving the genuine dominance of the Mexico corridor. PACS applies the same log1p procedure to the corridor's volume as a share of United States GDP. While CIS and PACS share the same underlying source, their different denominators and the nonlinear log1p transformation at those scales produce different distributional shapes: CIS exhibits stronger compression (diminishing returns to corridor size) while PACS is more nearly linear in the relevant range.

The three components combine into the Pre-Arrival Banking and Financial Score:

PABFS = 0.40 × OFIS + 0.30 × CIS + 0.30 × PACS

The OFIS weight of 40 percent reflects the judgment that institutional quality is the most fundamental determinant of verifiable financial identity. The equal 30 percent weights on CIS and PACS capture the corridor dimension from two complementary angles. PABFS scores are normalized to a 0–100 scale across 248 countries; individual PVFI scores are assigned to foreign-born ACS respondents using IPUMS birth-place codes. Figure 1 shows the full architecture.

In a production PVFI system, country-level scores would serve as prior beliefs updated by individual-level verification of home-country financial records, tax filings, or mobile payment histories.

4. Data and Methodology

4.1 Data Sources

Our empirical analysis draws on eight distinct datasets. Table 1 summarizes sample sizes, key variables, and analytical roles.

Table 1
Data sources, sample sizes, key variables, and analytical roles

Dataset	N (unit)	Key variables	Role in analysis
FDIC 2019/2021/2023	57,199 ref. persons	heb10, penatvty	Banking exclusion; nativity gap
IPUMS ACS (multi-year)	25,624,467 persons	BPL, INCTOT, EDUC	Income comparisons; PVFI assignment
HMDA LAR 2019–2023	68,354,279 apps.	action_taken, race, income	Credit denial; racial disparities
IPUMS CPS ASEC	800,468 persons	NATIVITY, INCTOT	Corroborating income comparison
FINRA NFCS 2021	27,118 respondents	M1_1, B4, M4	Financial well-being OLS
SCF 2022 (MI-1)[a]	4,595 households	Income, net worth	National income benchmark validation
KNOMAD 2021	134 corridors	vol_usd_bn	CIS and PACS construction
WDI + Findex 2021	248 countries	Credit/GDP, acct ownership	OFIS construction
Home Credit Group	307,511 loans	EXT_SOURCE_2/3, TARGET[a]	Machine learning validation (PVFI analog)

[a] SCF = Survey of Consumer Finances, first multiple imputation implicate. Y1% 10 = = 1 filter applied.

[b] TARGET = 1 indicates default (8.1% rate). EXT_SOURCE variables are externally sourced credit bureau scores; see Section 4.5.

* Microdata rates reflect oversampling design and are used only for relative comparisons; absolute rates cite published FDIC figures.

The FDIC National Survey of Unbanked and Underbanked Households (2019, 2021, 2023) provides the primary banking access evidence. We restrict to reference persons (PERRP in {1, 2, 40}) and exclude respondents with invalid, refused, or "don't know" banking status responses (heb10 < = 0), yielding an analytic sample of 57,199.

[1] Published national unbanked rates: 2019 = 5.4%; 2021 = 4.5%; 2023 = 5.9% (FDIC 2019, 2021, 2023). Microdata exclusion rates (~ 27%) reflect deliberate oversampling and must not be cited as population estimates.

[2] Heckman (1979) sample selection is a concern if excluded observations differ systematically. The excluded 18,083 rows (3.6% of matched sample) have PVFI distributions similar to the full sample, suggesting selection on income reporting is unlikely to materially bias the PVFI coefficient.

The HMDA Loan Application Register (2019–2023) contributes 68.4 million mortgage applications. For the pooled logistic regression we draw a stratified random sample of 200,000 applications per year. The IPUMS ACS (Ruggles et al. 2025) contributes 25.6 million person-year observations; foreign-born status is BPL > = 100. The IPUMS CPS ASEC (Flood et al. 2025) provides an independent nativity income comparison on 800,468 respondents. The FINRA NFCS 2021 and SCF 2022 supply the financial well-being and income benchmark analyses respectively (Board of Governors of the Federal Reserve System 2023). For PVFI construction, KNOMAD bilateral data provide 2021 corridor volumes; World Development Indicators (World Bank 2024) and Global Findex 2021 (Demirgüç-Kunt et al. 2022) provide the financial infrastructure inputs.

4.2 Banking Exclusion Model

We estimate banking exclusion as a weighted logistic regression of unbanked or underbanked status on nativity and controls:

$$Pr(EXCL = 1) = \Lambda(\beta_0 + \beta_1\ FOREIGN_BORN + \beta_2\ Income + \beta_3\ Education + \beta_4\ Race + \beta_5\ State\ FE + \beta_6\ Year\ FE)$$

Household survey weights (hwhhwgt) are applied via frequency-weighted maximum likelihood. The primary specification includes race and state fixed effects; a simplified specification omits them to quantify how much the nativity coefficient is absorbed by race and geography.

4.3 Credit Denial Model

The HMDA credit denial model is:

$$Pr(DENIED = 1) = \Lambda(\gamma_0 + \gamma_1\ Race + \gamma_2\ log(Income) + \gamma_3\ LTI + \gamma_4\ Loan\ Purpose + \gamma_5\ Year\ FE)$$

Denial is coded as action_taken = 3. The loan-to-income ratio (LTI) is loan amount in dollars divided by reported income in dollars; White is the reference racial category. The negative coefficient on LTI (OR = 0.870) reflects application selectivity: conditional on income, higher LTI applicants have been pre-screened more rigorously before submitting an application (Mian and Sufi 2009). Applicants who seek larger loans relative to their income are typically better-prepared financially — having consulted brokers, gathered documentation, and confirmed eligibility — making this group less likely to receive a denial.

4.4 PVFI Validation Model

We estimate the PVFI validation model on the 481,917 foreign-born ACS respondents with valid income records and matched PVFI scores:

$$Pr(EXCL_PROXY = 1) = \Lambda(\delta_0 + \delta_1\ PVFI_STD + \delta_2\ Education + \delta_3\ Age + \delta_4\ Sex)$$

EXCL_PROXY = 1 if INCTOT falls below 50 percent of the matched-sample median ($19,382). Income is excluded from the predictor set because it defines the outcome; including it would create perfect separation and cause logit divergence. The 18,083 respondents with missing INCTOT values are excluded before computing EXCL_PROXY; without this filter, the pandas comparison NaN < threshold evaluates to False, silently misclassifying missing-income cases as not excluded.

4.5 Machine Learning Validation

We train four models on the Home Credit training data — logistic regression (LR) and gradient boosting (GBM) via XGBoost (Chen and Guestrin 2016), each in baseline (19 features) and PVFI-analog (21 features, adding EXT_SOURCE_2 and EXT_SOURCE_3) specifications. EXT_SOURCE variables are externally sourced credit bureau assessments; we use them as structural analogs to the externally verified financial identity information that a PVFI system would supply. This analogy is a proof-of-concept: the Home Credit data originate in Central and Eastern Europe, not the United States, and EXT_SOURCE scores are conventional credit bureau outputs rather than cross-border identity signals. All models are evaluated via five-fold stratified cross-validation and a held-out test set (20 percent of data). Shapley additive explanation (SHAP) values (Lundberg and Lee 2017) decompose feature importance for the GBM PVFI-analog model.

AI-assisted tools were used in the preparation of this manuscript; the author takes full responsibility for all analytical claims, empirical results, interpretations, and the accuracy of cited findings.

5. Empirical Results

5.1 Nativity Gap in Banking Access

Table 2 and Fig. 2 present banking exclusion rates by nativity across the three FDIC survey waves. The nativity gap is both persistent and substantive. In 2019, foreign-born reference persons reported an exclusion rate of 32.1 percent against 29.4 percent for native-born persons, a gap of 2.72 percentage points that widened to 3.55 points in 2021 before partially moderating to 3.25 points in 2023. These within-

sample relative rates substantially exceed the published national rates; absolute figures from the published surveys (5.4, 4.5, and 5.9 percent, respectively) are cited in the rightmost column of Table 2 for context.

Table 2
Banking exclusion rates by nativity and survey wave, FDIC microdata

Survey year	Rate FB (%)	Rate NB (%)	Nativity gap (pp)	Published nat. rate (%)[a]
2019	32.08	29.36	2.72	5.4
2021	30.07	26.52	3.55	4.5
2023	31.28	28.03	3.25	5.9

[a] Published national unbanked rates from FDIC (2019; 2021; 2023). Microdata rates are within-sample relative measures from the oversampled FDIC survey design and may not be cited as national benchmarks.

* FB = foreign-born; NB = native-born. Exclusion = unbanked or underbanked (heb10 in {2,3}). N = 57,199 valid reference persons.

The conditional logistic regressions are reported in Table 3. The primary specification — with income, education, race, and state fixed effects — yields FOREIGN_BORN OR = 1.049 (95% CI [0.996, 1.105]; p = 0.071). The simplified specification without race and geography controls produces a larger estimate of 1.120 ($p < 0.001$). The attenuation from specification (2) to specification (1) reflects the established pattern that race and geography absorb part of the nativity effect when included jointly (Pager and Shepherd 2008). The descriptive gap of 2.7–3.6 percentage points, stable across three survey waves spanning four years of economic turbulence, is the more compelling evidence of structural exclusion.

Table 3
Banking exclusion logistic regression, FDIC 2019/2021/2023 (N = 57,199)

Variable	(1) Primary	(2) Simplified
FOREIGN_BORN	1.049† [0.996, 1.105]	1.120*** [1.064, 1.178]
INCOME_APPROX_STD	0.784*** [0.776, 0.792]	0.801*** [0.793, 0.810]
Education controls	Yes	Yes
Race fixed effects	Yes	No
State fixed effects	Yes	No
Year FEs (ref: 2023)	Yes	Yes
McFadden R^2	0.0057	0.0035
N	57,199	57,199

† $p < 0.10$; *** $p < 0.001$. Odds ratios; 95% CI in brackets. Survey weights (hwhhwgt) applied via frequency-weighted MLE. Outcome: heb10 in {2,3}.

[a] Specification (1) is the preferred model. Difference between specifications reflects absorption of nativity effect by race and state FEs.

Figure 2 plots the exclusion rates and nativity gap across all three waves alongside the published national rates on a secondary axis. The gap is positive in all three waves and the published rates confirm the downward trend from 2019 to 2021 followed by a rebound in 2023, consistent with the broader economic cycle.

5.2 Racial Disparities in Mortgage Credit Access

Table 4 reports denial rates by race and ethnicity for the 2019–2023 HMDA panel. Denial rates rose for all groups as credit conditions tightened, but the racial ordering is perfectly stable across all five years. Black applicants faced 33.9 percent denial in 2019 against 18.7 percent for White applicants, a ratio of 1.81 that persisted through 2023 (38.4 versus 22.6 percent). Hispanic applicants moved from 27.9 to 32.6 percent over the same window.

Table 4
Mortgage application denial rates by race and ethnicity, HMDA 2019–2023 (percent)

Race/ethnicity	2019	2020	2021	2022	2023	Δ (pp)
White (reference)	18.73	14.18	14.02	19.77	22.63	+ 3.90
Black or African American	33.91	27.89	26.29	34.52	38.42	+ 4.51
Hispanic or Latino	27.85	22.43	20.75	29.05	32.62	+ 4.77
American Indian/AN	36.87	28.60	27.25	36.74	40.66	+ 3.79
Asian	20.93	15.44	13.84	20.40	22.36	+ 1.43

* Source: HMDA LAR 2019–2023 (Avery and Bhutta 2024). Stratified random sample of 200,000 applications per year. Δ = change in percentage points, 2019 to 2023.

Figure 3 Mortgage application denial rates by race and ethnicity, HMDA 2019–2023; White is the reference category; N = 1 million observations from a stratified random sample of 200,000 applications per year (Avery and Bhutta 2024)

The conditional logistic regression (Table 5) confirms that racial disparities are not explained by income or loan purpose. Black applicants face OR = 1.955 ($p < 0.001$), American Indian applicants OR = 1.816 ($p < 0.001$), and Hispanic applicants OR = 1.524 ($p < 0.001$). These estimates are robust across all five years and are consistent with Bhutta and Hizmo (2021) and Bartlett et al. (2022).

Table 5
Credit denial logistic regression, HMDA pooled 2019–2023 (N = 913,470)

Variable (ref: White)	Odds ratio	95% CI	p-value
RACE_Black	1.955***	[1.920, 1.990]	< 0.001
RACE_AmInd	1.816***	[1.710, 1.929]	< 0.001
HISPANIC	1.524***	[1.500, 1.549]	< 0.001
RACE_Asian	1.369***	[1.337, 1.402]	< 0.001
LOG_INCOME_STD	0.576***	[0.572, 0.580]	< 0.001
LTI_STD[a]	0.870***	[0.864, 0.877]	< 0.001
Loan purpose FEs	Yes		
Year FEs (ref: 2019)	Yes		
McFadden R^2	0.0741		
N	913,470		

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Continuous regressors standardized (per 1-SD). White is the reference racial category.

[a] Negative LTI coefficient reflects application selectivity: conditional on income, larger-loan applicants are more thoroughly pre-screened before submission (Mian and Sufi 2009).

5.3 Immigrant Income and Financial Behavior

Among working-age adults (18–64), native-born workers earn a mean income of $47,536 against $45,670 for foreign-born workers in the ACS data — a gap of $1,866 (approximately 4 percent). Median incomes are $23,188 (native-born) and $21,418 (foreign-born). The CPS ASEC puts native-born mean income at $40,633 and foreign-born at $32,121. These gaps are real but moderate, and consistent with the income assimilation trajectory documented by Borjas (1985). In the FINRA NFCS 2021 data (FINRA Investor Education Foundation 2022), 21.0 percent of respondents report financial hardship in the prior year; the OLS regression of financial well-being scores on hardship and demographic controls (N = 27,118; R^2 = 0.114) yields a hardship coefficient of − 0.922 (95% CI [–1.227, − 0.617]; $p < 0.001$)[1].

5.4 PVFI Construction and Corridor Analysis

The PVFI index assigns PABFS scores to 2,628,486 of the 3,389,363 foreign-born ACS respondents, a match rate of 77.6 percent. The unmatched 22.4 percent includes Cuban-born respondents (218,361) for whom trade sanctions preclude a formal bilateral corridor, and several unresolved IPUMS birth-place code categories.

Figure 4 shows the PABFS ranking for the top 20 origin countries. Mexico ranks first (100.0) by virtue of its dominant $52.6 billion corridor (26.3 percent of total United States outbound remittances). China ranks second (90.1), combining high financial system development (OFIS = 83.2) with a $12.7 billion corridor. India ranks third (73.9). The ordering reflects the PABFS design: high scores require both financial system depth and corridor strength; countries with only one of the two score in the mid-range.

The band distribution among matched immigrants shows a structurally concentrated score distribution: Band 4 (75–100) contains 46.1 percent of matched respondents, Band 3 (50–75) contains 37.7 percent, and Band 2 (25–50) contains 16.2 percent. Band 1 (0–25) is empty in the matched sample — all origin countries represented in the ACS immigrant population achieve PABFS scores above 25.

5.5 PVFI Score and Financial Exclusion

Table 6 presents the core empirical result. Among 481,917 foreign-born ACS respondents with valid income records and matched PVFI scores, a one-standard-deviation increase in PVFI_STD is associated with OR = 0.8726 (95% CI [0.8661, 0.8791]; $p < 0.001$). The McFadden pseudo-R^2 of 0.087 confirms positive model fit; the AIC of 555,430 confirms optimizer convergence.

Table 6
PVFI score and financial exclusion, logistic regression, ACS foreign-born (N = 481,917)

Variable	Odds ratio	95% CI	p-value
PVFI_STD	0.8726***	[0.8661, 0.8791]	< 0.001
EDUC	0.741***	[0.738, 0.744]	< 0.001
AGE	0.981***	[0.980, 0.982]	< 0.001
SEX (female = 1)	1.184***	[1.170, 1.198]	< 0.001
McFadden R^2	0.0867		
AIC	555,430		
N	481,917		

*** $p < 0.001$. Outcome: EXCL_PROXY = 1 if INCTOT < 50% of matched-sample median ($19,382). PVFI_STD is z-standardized PABFS (mean = 78.26; SD = 23.24). INCTOT excluded from predictors (it defines the outcome). HC3 robust standard errors. 18,083 respondents with missing INCTOT excluded prior to analysis.

An important interpretive caveat: OR = 0.873 partly reflects immigrant selection. Countries with higher PABFS scores tend to send higher-human-capital immigrants, and human capital is correlated with income. The regression controls partially through education and demographics, but residual selection bias cannot be eliminated in a cross-sectional design. The country-level nature of PVFI — all immigrants from Mexico receive PVFI = 100 regardless of individual financial behavior — means this result is best understood as a prior distribution over PVFI effects rather than a precise causal estimate.

With N = 481,917 and an exclusion proxy rate of 32.6 percent, this model has well over 80 percent power to detect effects as small as OR = 1.003; the tight confidence interval [0.866, 0.879] makes this a precisely estimated result. The result is directionally consistent with the PVFI framework's core prediction: after controlling for education, age, and sex, origin-country financial development predicts lower odds of income-based exclusion among immigrants.

5.6 Machine Learning Validation

Table 7 and Figs. 5 and 6 document the machine learning results. The baseline logistic regression achieves a cross-validation AUC of 0.6537 (± 0.0037) and a test-set AUC of 0.6549. Adding EXT_SOURCE_2 and EXT_SOURCE_3 raises the logistic regression AUC to 0.7291 (± 0.0047) in cross-validation and 0.7305 on the test set — an improvement of 7.56 percentage points. The gradient boosting models show the same direction: baseline AUC = 0.6762 (± 0.0060) rising to 0.7405 (± 0.0046) with external features, a gain of 6.55 percentage points confirmed across all five folds.

Table 7
Machine learning credit risk model performance, Home Credit dataset (N = 307,511)

Model	Features	AUC-ROC CV	AUC-ROC Test	Gini	KS
LR Baseline	19	0.6537 ± 0.0037	0.6549	0.310	0.231
LR PVFI-Analog	21	0.7291 ± 0.0047	0.7305	0.461	0.340
GBM Baseline	19	0.6762 ± 0.0060	0.6781	0.356	0.259
GBM PVFI-Analog[a]	21	0.7405 ± 0.0046	0.7436	0.487	0.359
LR gain	+ 2	+ 0.0754	+ 0.0756	+ 0.151	+ 0.109
GBM gain	+ 2	+ 0.0643	+ 0.0655	+ 0.131	+ 0.100

[a] PVFI-Analog models add EXT_SOURCE_2 and EXT_SOURCE_3 to 19 baseline features. GBM = XGBoost (Chen and Guestrin 2016); 100 estimators, lr = 0.10, max depth = 5. Default rate = 8.1%. CV = 5-fold stratified cross-validation (mean ± SD).

[b] SHAP decomposition: EXT_SOURCE_3 = 34.1%; EXT_SOURCE_2 = 33.9%; combined = 68% of total feature importance in GBM PVFI-Analog model (see Fig. 6a).

Figure 5 Machine learning credit risk model performance: panel (a) compares AUC-ROC for baseline and PVFI-analog specifications across LR and GBM models; error bars represent five-fold cross-validation standard deviations; panel (b) reports Gini coefficient and Kolmogorov-Smirnov (KS) statistic on the test set

Figure 6a shows the SHAP feature importance decomposition. EXT_SOURCE_3 and EXT_SOURCE_2 together account for 68 percent of total feature importance, far outweighing the 19 baseline features combined. This dominance is consistent with Berg et al. (2020), who found that external information sources dramatically outperform internal demographic data for thin-file borrowers. Figure 6b shows the

PVFI score-band distribution among matched immigrants; Band 1 (0–25) is empty because all represented origin countries achieve PABFS scores above 25.

6. Discussion

The results tell a layered story. At the foundation, the FDIC data establish a persistent nativity gap in banking access — real, positive, and stable across four years of economic turbulence. At the intermediate level, the HMDA data confirm that credit markets fail racial minorities at a scale that cannot be explained by income or loan characteristics. Together, these findings describe a credit market with two distinct failure modes for immigrant households: a documentation failure (no domestic credit history) and a discrimination failure (disparate treatment conditional on documentation).

PVFI directly targets the documentation failure. The validation results — OR = 0.873 in the ACS regression, and a 6.6 to 7.6 percentage point AUC gain in the machine learning experiment — both confirm that country-level financial infrastructure signals carry genuine predictive content about individual financial outcomes. The AUC improvement is particularly striking in context: in a binary classification problem with an 8 percent positive rate, a 6.6-point improvement from two additional features indicates that the information PVFI would supply is not merely redundant with available demographic data but adds a genuinely distinct signal.

Two interpretive tensions deserve direct engagement. First, the selection confound in the ACS validation is real. Countries with high PABFS scores tend to send more economically selected immigrants; controlling for education, age, and sex absorbs part of this but not all. The OR = 0.873 should be read as an upper bound on selection effects and a lower bound on the mechanism effect. Second, the Home Credit analogy is imperfect: EXT_SOURCE scores are credit bureau outputs from Central and Eastern Europe, not cross-border identity signals from the United States immigrant population. The proof-of-concept value is real; the operational translation requires field experiments with immigrant-specific data.

These tensions are genuine, but they do not overturn the overall pattern of evidence. The nativity gap is persistent. Racial disparities are large and stable. Country-level PVFI predicts financial vulnerability in the expected direction. External creditworthiness signals dramatically improve credit models for thin-file borrowers. None of these results proves that a PVFI system would work as intended in production. Together, they provide a coherent empirical case for attempting it.

7. Policy Implications

Five recommendations follow from the empirical findings, each operating within current regulatory structures. First, the FDIC should incorporate nativity-stratified unbanked rates into its biennial household survey publications; the survey already collects penatvty but does not tabulate it prominently. Second, the CFPB should issue interpretive guidance on the use of origin-country financial records in underwriting under the Equal Credit Opportunity Act, analogous to its existing guidance on alternative data (Consumer Financial Protection Bureau 2023). Third, FinCEN should explore whether anonymized corridor-level

behavioral statistics from Money Service Business reporting could inform PVFI calibration; the monitoring infrastructure is already in place (Financial Crimes Enforcement Network 2021). Fourth, the SBA and CDFI Fund should pilot PVFI-style pre-arrival scoring in small business loan underwriting for immigrant entrepreneurs — the 21 percent hardship rate documented in the National Financial Capability Study (FINRA Investor Education Foundation 2022) suggests the welfare stakes are substantial. Fifth, multilateral institutions including the World Bank and the IMF should explore bilateral creditworthiness certification as an extension of existing financial inclusion program architectures; the Findex and KNOMAD data that underpin PVFI are World Bank products, and the marginal cost of standardizing their use for PVFI purposes is low.

8. Limitations and Future Research

The first limitation concerns PVFI score granularity. With 28 unique values assigned across 2.63 million matched respondents, the measure captures only between-country variation; all within-country heterogeneity in immigrant financial behavior is averaged away. Two immigrants from Mexico receive identical PVFI scores regardless of whether one carries a formal credit history and the other does not. A deployed PVFI system would require individual-level verification through home-country financial records, mobile payment histories, or tax filings.

A second limitation is the construct validity of the exclusion proxy. EXCL_PROXY measures income poverty — INCTOT below 50 percent of the matched-sample median — not financial access barriers directly. Banking status is unavailable in the ACS public-use extract, making a direct replication of the FDIC exclusion analysis at the individual level impossible with these data alone. Future research should replicate the PVFI validation using matched administrative datasets that jointly capture banking status and country of birth.

Third, the FDIC banking exclusion finding achieves only marginal statistical significance in the primary specification ($p = 0.071$), falling short of the conventional 5 percent threshold. The attenuation is best read as partial mediation by race and geographic fixed effects rather than as evidence against the nativity-banking gap, and the descriptive gap of 2.7–3.6 percentage points across three survey waves carries independent evidentiary weight. Readers should nonetheless weigh the conditional regression result accordingly.

Fourth, the machine learning validation is geographically and institutionally indirect. The Home Credit dataset covers consumer lending in Central and Eastern Europe, and EXT_SOURCE variables are conventional credit bureau scores rather than cross-border identity signals. The 6.6 to 7.6 percentage point AUC gain demonstrates that externally verified financial data improve discrimination for thin-file borrowers; it does not validate PVFI specifically in the United States immigrant lending context.

Fifth, three IPUMS birth-place codes — 521 assigned to India, 515 to the Philippines, and 518 to South Korea — were inferred from respondent volume distributions rather than confirmed against the IPUMS-USA BPL codebook. These codes cover roughly 520,000 matched respondents. Any misassignment would alter

PVFI scores for that group, and verification at ipums.org/usa-action/variables/BPL#codes_section is recommended before any follow-on analysis.

Sixth, the entire analysis is descriptive and predictive; no quasi-experimental identification strategy is employed. The OR = 0.873 association between PVFI and income exclusion reflects a mixture of the genuine PVFI mechanism and country-level immigrant selection that cross-sectional regression cannot separate. Identifying the causal effect of financial identity portability would require a field experiment or a policy discontinuity creating differential PVFI availability across comparable immigrant groups.

Seventh, the PABFS formula assigns fixed weights of 40 percent to OFIS and 30 percent each to CIS and PACS based on theoretical judgment rather than empirical optimization. A supervised weighting approach — calibrating weights against matched default or banking-exclusion outcomes — would produce a more defensible specification. Appendix B demonstrates that the sign and significance of the PVFI coefficient survive moderate weight variation, but the optimal weight structure remains an open research question.

9. Conclusion

The immigrant financial inclusion problem is not a problem of character. The immigrants who send $52.6 billion annually to Mexico, $15.8 billion to India, and $14.1 billion to Guatemala are managing significant cross-border financial obligations without access to the tools that domestic residents take for granted. The credit market failure is institutional: their financial identities are simply not legible to domestic lenders.

PVFI proposes a partial technical remedy. By aggregating observable country-level signals into a pre-arrival score, the framework creates a financial credential that a lender could consult before any domestic credit history has accumulated. Our empirical results — a persistent nativity gap of 2.7–3.6 percentage points in banking access, racial disparities of 50 to 96 percent in mortgage denial, an OR = 0.873 association between PVFI and lower financial vulnerability, and 6.6 to 7.6 percentage point improvements in credit model discrimination from external creditworthiness features — collectively support the proposition that PVFI-type information carries genuine predictive content.

None of this will happen quickly. Individual-level verification, regulatory clarity under ECOA and BSA frameworks, and at least one credible field experiment are prerequisites for an operational lending tool. But the potential beneficiaries — 43 million people navigating a financial system that cannot read their history — give the effort its urgency.

Declarations

Funding

No external funding was received for this research.

Ethics Approval

This study uses publicly available, de-identified secondary microdata sourced from federal government surveys and multilateral databases. No primary data collection involving human participants was undertaken. Accordingly, no institutional review board approval is required.

References

1. Aggarwal R, Demirgüç-Kunt A, Martinez Peria MS (2011) Do remittances promote financial development? J Dev Econ 96:255–264. https://doi.org/10.1016/j.jdeveco.2010.10.005
2. Allen F, Demirgüç-Kunt A, Klapper L, Martinez Peria MS (2016) The foundations of financial inclusion: Understanding ownership and use of formal accounts. J Financ Intermed 27:1–30. https://doi.org/10.1016/j.jfi.2015.12.003
3. Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finance 23:589–621. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x
4. Ambrose BW, Conklin JN, Lopez LA (2021) Does borrower and broker race affect the cost of mortgage credit? Rev Financ Stud 34:790–826. https://doi.org/10.1093/rfs/hhaa087
5. Amuedo-Dorantes C, Bansak C (2006) Money transfers among banked and unbanked Mexican immigrants. South Econ J 73:374–401. https://doi.org/10.2307/20111897
6. Avery R, Bhutta N (2024) Home Mortgage Disclosure Act data 2019–2023. Federal Financial Institutions Examination Council. https://www.consumerfinance.gov/data-research/hmda/
7. Bartlett R, Morse A, Stanton R, Wallace N (2022) Consumer-lending discrimination in the FinTech era. J Financ Econ 143:30–56. https://doi.org/10.1016/j.jfineco.2021.05.047
8. Beck T, Demirgüç-Kunt A, Levine R (2007) Finance, inequality and the poor. J Econ Growth 12:27–49. https://doi.org/10.1007/s10887-007-9010-6
9. Berg T, Burg V, Gombović A, Puri M (2020) On the rise of FinTechs: Credit scoring using digital footprints. Rev Financ Stud 33:2845–2897. https://doi.org/10.1093/rfs/hhz099
10. Bhutta N, Hizmo A (2021) Do minorities pay more for mortgages? Rev Financ Stud 34:763–789. https://doi.org/10.1093/rfs/hhaa047
11. Board of Governors of the Federal Reserve System (2023) Survey of consumer finances 2022. Federal Reserve Bulletin. https://www.federalreserve.gov/econres/scfindex.htm
12. Borjas GJ (1985) Assimilation, changes in cohort quality, and the earnings of immigrants. J Labor Econ 3:463–489. https://doi.org/10.1086/298065
13. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. 785–794. https://doi.org/10.1145/2939672.2939785. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min
14. Chiswick BR (1978) The effect of Americanization on the earnings of foreign-born men. J Polit Econ 86:897–921. https://doi.org/10.1086/260717
15. Claessens S (2006) Access to financial services: A review of the issues and public policy objectives. World Bank Res Obs 21:207–240. https://doi.org/10.1093/wbro/lkl004
16. Consumer Financial Protection Bureau (2023) Consumer Financial Protection Bureau annual report 2023. CFPB. https://www.consumerfinance.gov/about-us/budget-strategy/performance-plan-and-

annual-report
17. Demirgüç-Kunt A, Klapper L, Singer D, Ansar S (2022) The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19. World Bank Policy Research Working Paper. https://doi.org/10.1596/978-1-4648-1897-4
18. Federal Deposit Insurance Corporation (2019) 2019 FDIC national survey of unbanked and underbanked households. FDIC. https://www.fdic.gov/household-survey/data-downloads-and-resources
19. Federal Deposit Insurance Corporation (2021) 2021 FDIC national survey of unbanked and underbanked households. FDIC. https://www.fdic.gov/household-survey/data-downloads-and-resources
20. Federal Deposit Insurance Corporation (2023) 2023 FDIC National Survey of Unbanked and Underbanked Households. https://doi.org/10.3886/ICPSR38697
21. Financial Crimes Enforcement Network (2021) Financial Crimes Enforcement Network annual report 2020. U.S. Department of the Treasury. https://www.fincen.gov/annual-report
22. FINRA Investor Education Foundation (2022) National financial capability study 2021: state-by-state survey. FINRA Investor Education Foundation. https://www.usfinancialcapability.org/results.php
23. Flood S, King M, Rodgers R et al (2025) IPUMS, Current Population Survey: Version 13.0. https://doi.org/10.18128/D030.V13.0
24. Fuster A, Goldsmith-Pinkham P, Ramadorai T, Walther A (2022) Predictably unequal? The effects of machine learning on credit markets. J Finance 77:5–47. https://doi.org/10.1111/jofi.13090
25. Giuliano P, Ruiz-Arranz M (2009) Remittances, financial development, and growth. J Dev Econ 90:144–152. https://doi.org/10.1016/j.jdeveco.2008.10.005
26. Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47:153–161. https://doi.org/10.2307/1912352
27. Hogarth JM, O'Donnell KH (2000) If you build it, will they come? A simulation of financial product holdings among low-to-moderate income households. J Consum Policy 23:409–444. https://doi.org/10.1023/A:1007222700931
28. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30:4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
29. Merton RC (1974) On the pricing of corporate debt: The risk structure of interest rates. J Finance 29:449–470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x
30. Mian A, Sufi A (2009) The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. Q J Econ 124:1449–1496. https://doi.org/10.1162/qjec.2009.124.4.1449
31. Pager D, Shepherd H (2008) The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. Annu Rev Sociol 34:181–209. https://doi.org/10.1146/annurev.soc.33.040406.131740

32. Rhine SLW, Greene WH, Toussaint-Comeau M (2006) The importance of check-cashing businesses to the unbanked: Racial/ethnic differences. Rev Econ Stat 88:146–157. https://doi.org/10.1162/003465306775565498
33. Ruggles S, Flood S, Sobek M et al (2025) IPUMS USA. https://doi.org/10.18128/D010.V16.0. Version 16.0
34. Squires GD (2003) Racial profiling, insurance style: Insurance redlining and the uneven development of metropolitan areas. J Urban Aff 25:391–410. https://doi.org/10.1111/1467-9906.t01-1-00168
35. Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. Am Econ Rev 71:393–410
36. Treacy WF, Carey MS (2000) Credit risk rating systems at large US banks. J Bank Finance 24:167–201. https://doi.org/10.1016/S0378-4266(99)00056-4
37. World Bank (2024) World Development Indicators 2024. World Bank
38. World Bank KNOMAD (2023) Remittances brave global headwinds: migration and development brief 38. World Bank. https://www.knomad.org/publication/migration-and-development-brief-38

Footnotes

1. Financial literacy (M4) was excluded from the regression after a scale inconsistency was identified in the NFCS 2021 extract; the hardship finding is unchanged without M4.

Figures

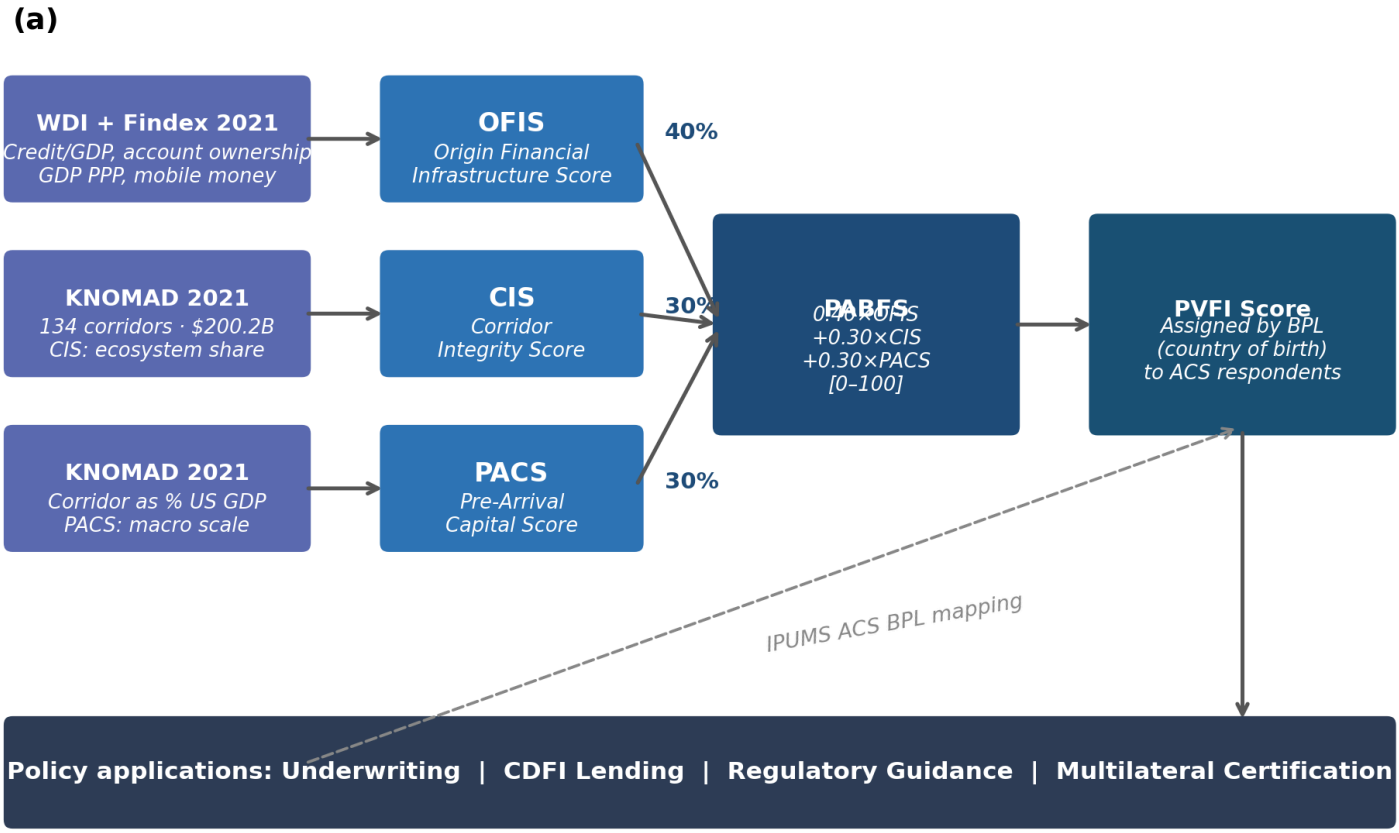


Figure 1

PVFI conceptual framework, showing the data sources (WDI, Findex, KNOMAD, IPUMS ACS), the three index components (OFIS, CIS, PACS), their aggregation into PABFS (0–100), and the downstream assignment of individual PVFI scores and policy applications

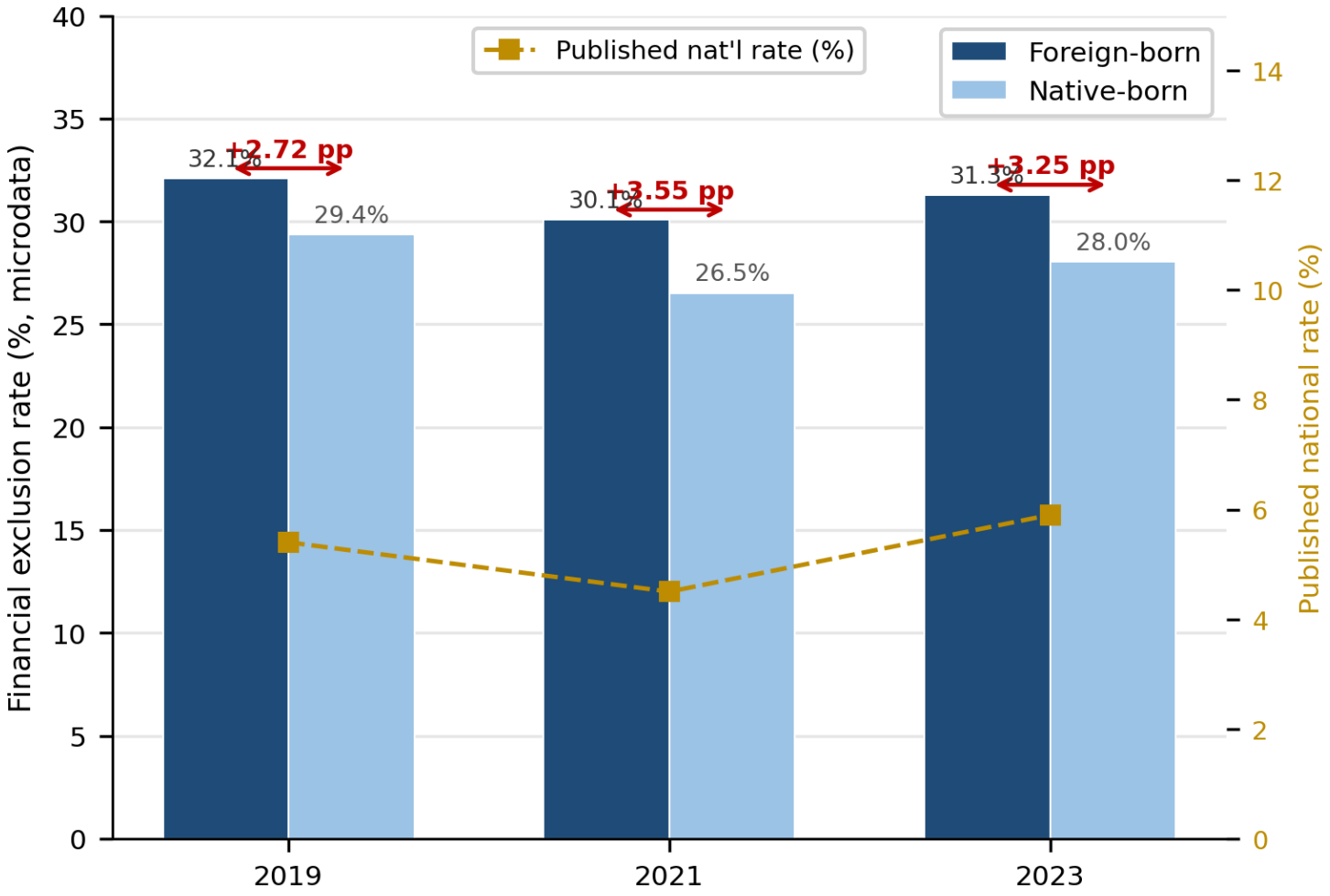


Figure 2

Banking exclusion rates (unbanked or underbanked) for foreign-born (FB) and native-born (NB) reference persons, FDIC microdata 2019–2023, with nativity gap annotated (red double-headed arrows); published national rates on secondary axis (FDIC 2019; 2021; 2023)

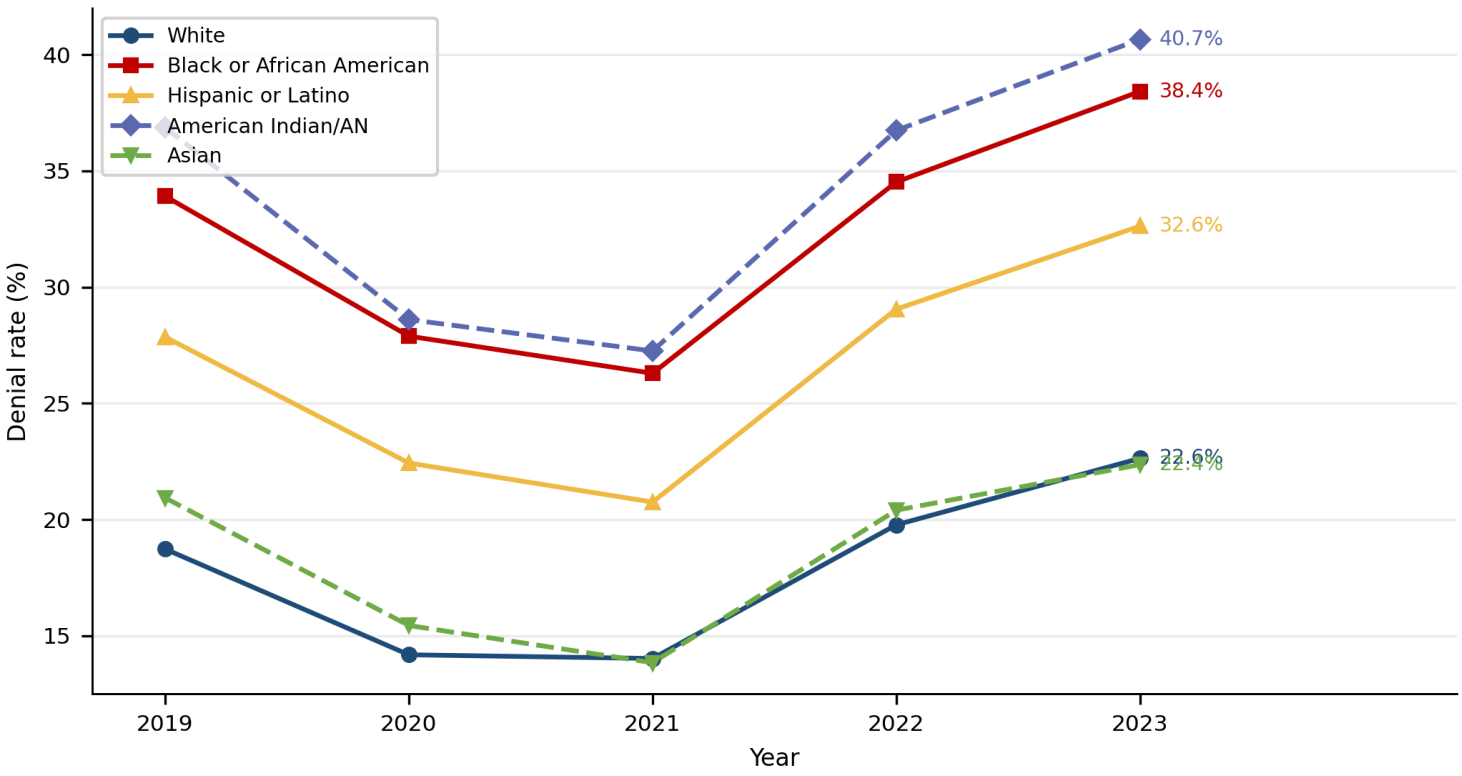


Figure 3

Mortgage application denial rates by race and ethnicity, HMDA 2019–2023; White is the reference category; N = 1 million observations from a stratified random sample of 200,000 applications per year (Avery and Bhutta 2024)

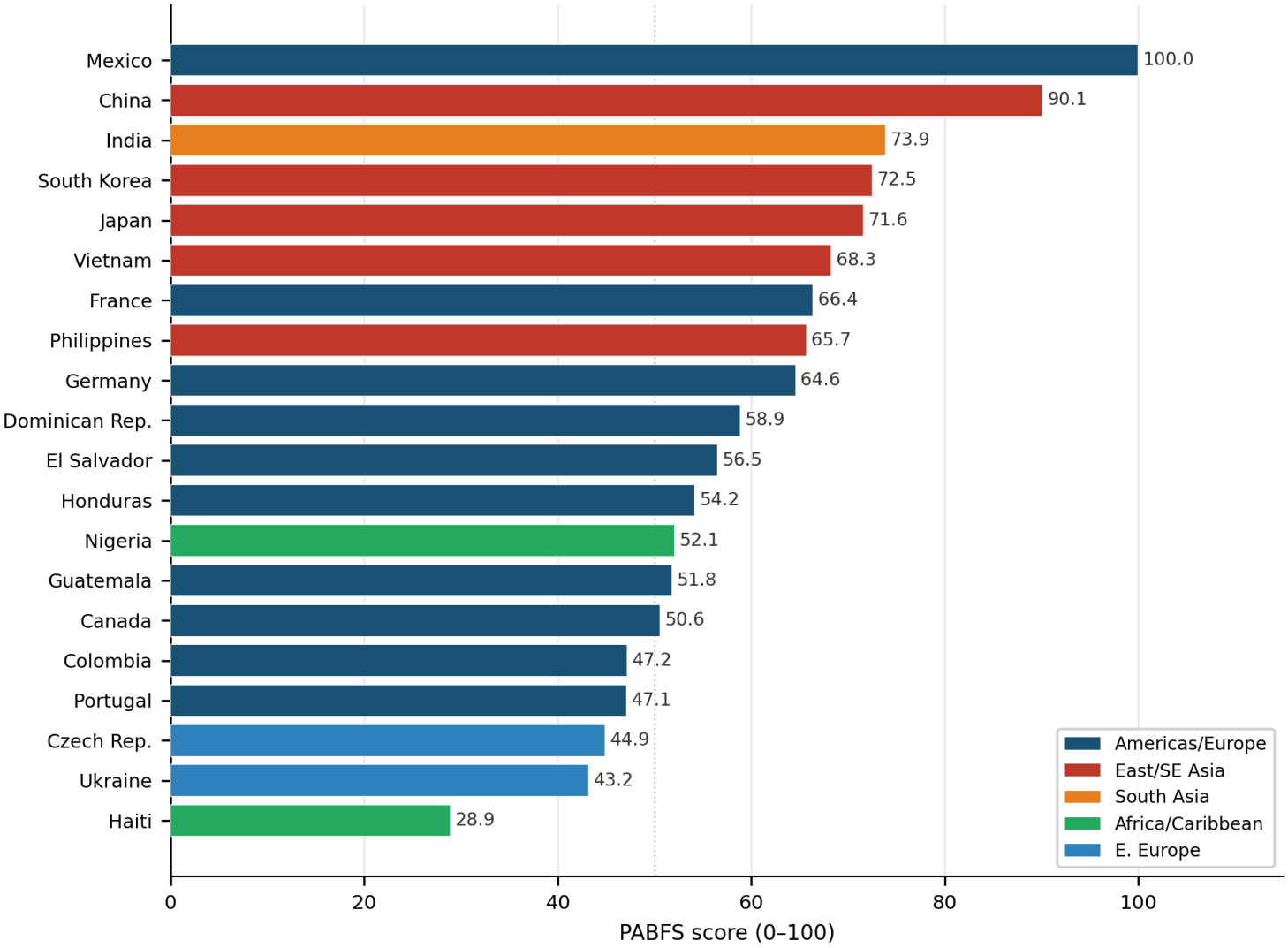


Figure 4

PABFS scores for the top 20 immigrant-origin countries; PABFS = 0.40 × OFIS + 0.30 × CIS + 0.30 × PACS, normalized to 0–100; color coding indicates broad geographic region; countries without KNOMAD corridor data receive OFIS-only scores

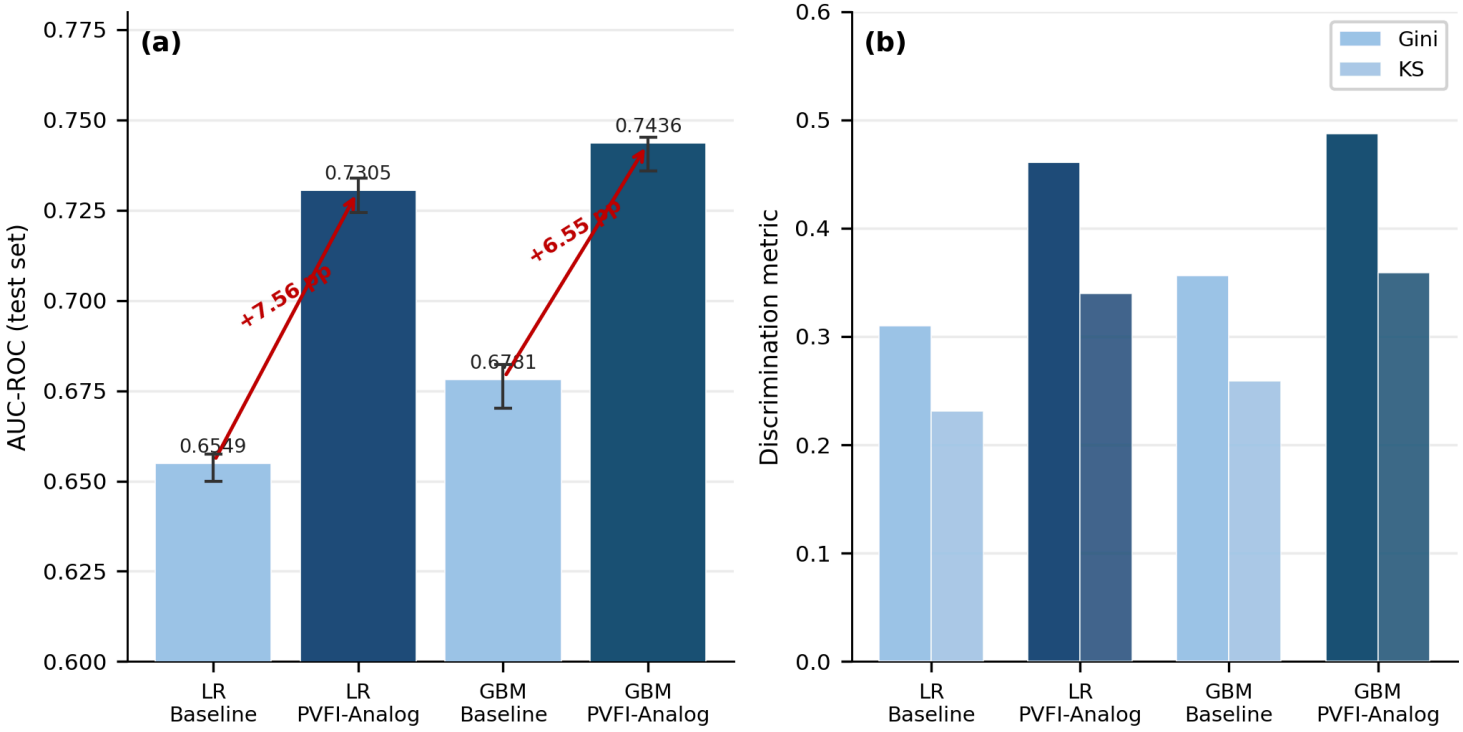


Figure 5

Machine learning credit risk model performance: panel (a) compares AUC-ROC for baseline and PVFI-analog specifications across LR and GBM models; error bars represent five-fold cross-validation standard deviations; panel (b) reports Gini coefficient and Kolmogorov-Smirnov (KS) statistic on the test set

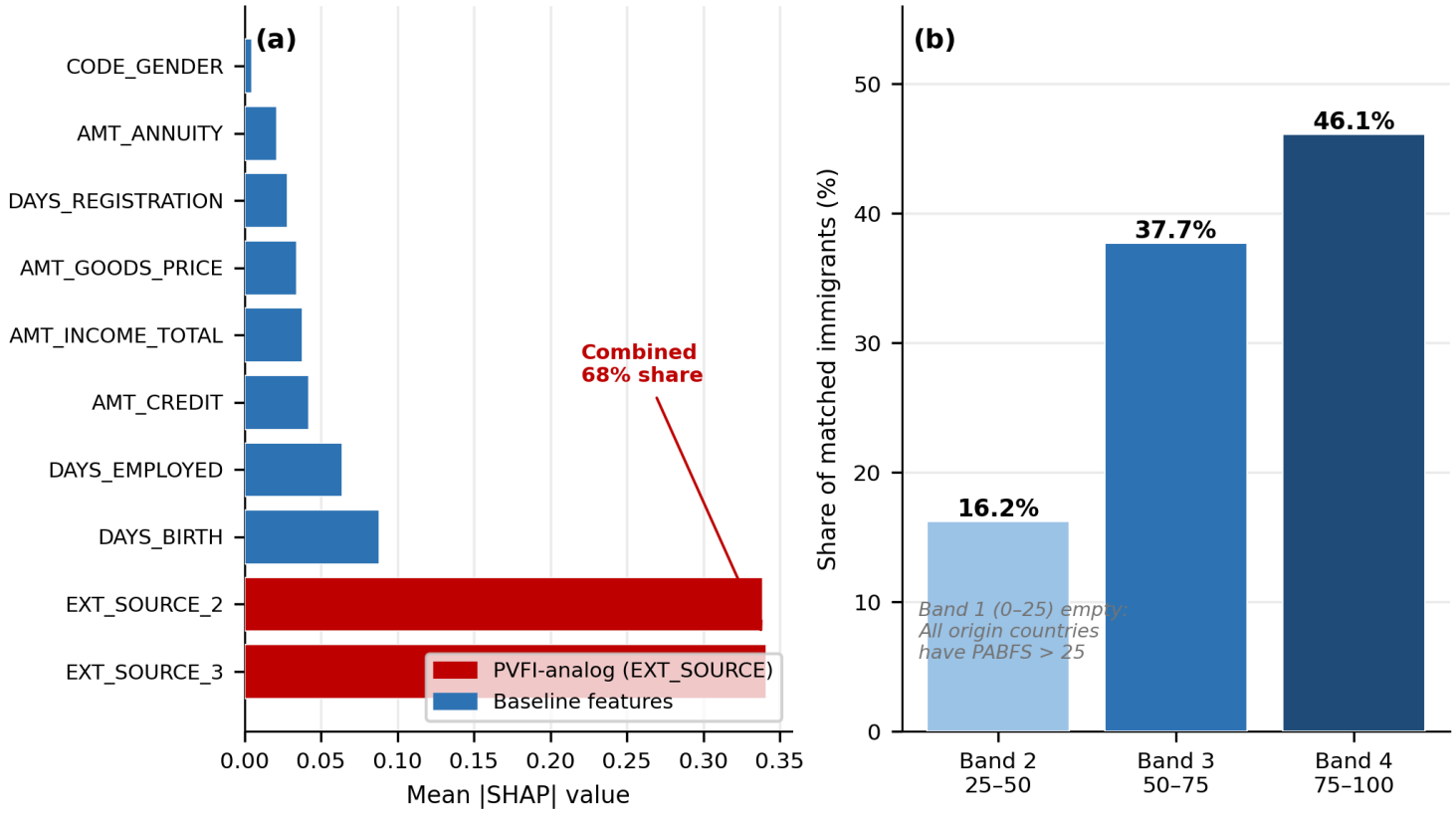


Figure 6

Panel (a): mean absolute SHAP values for the GBM PVFI-Analog model; PVFI-analog features (EXT_SOURCE_2/3, red bars) account for 68% of total importance; panel (b): PVFI score-band distribution among 2,628,486 matched foreign-born ACS respondents; Band 1 (0–25) is empty

Supplementary Files

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

- Appendix.docx