Which Signals Predict Lasting Engagement?

Parameter weights for ranking where a partner bank should do outreach

Outreach Discovery team, LEAD ASPIRE 2026

October 12, 2026

Abstract

When people have no credit score, we measure trust as behavior: do they stay engaged? We fit a logistic regression of 90-day program retention on nine standardized signals. **All data are simulated (1,200 synthetic participants) to validate the analysis pipeline; no real participant data or empirical findings are reported.** Relationship signals (consented visits, opt-in response, relationship duration) dominate retention, but they exist only after contact. Among signals available before contact, income regularity (+), on-time rent (+) and balance volatility (−) carry the usable information. We use those weights to rank areas for outreach, and the engagement signals to build the day-45 follow-up list. The real pilot will supply real data.

Keywords: financial inclusion; outreach prioritization; logistic regression; retention; synthetic data.

1 Hypothesis

Retention is a function of the nine signals:

$$\operatorname{logit} P(\text{retained at 90 days}) = \beta_0 + \sum_{k=1}^{9} \beta_k x_k. \qquad (1)$$

We test $H_0 : \beta_k = 0$ for each parameter. We expect (i) engagement signals to be strongest, (ii) pre-contact financial signals to have smaller but real effects, and (iii) some signals to add nothing.

2 Methods

Data (synthetic). 1,200 simulated participants (median-area retention near 50%; observed 51%) with nine signals grouped as payment reliability, cash-flow stability and relationship engagement. Effects were set from our design assumptions and are not estimates from real people.

Model. Logistic regression by maximum likelihood on z-scored inputs, so each odds ratio (OR) is per 1 SD. Wald z, p-values and 95% CIs are from the fitted covariance matrix.

Validation. Fit on 600, evaluate on a separate 600 (AUC, calibration, Brier score). PCA checks that the signals form three groups.

2.1 Financial-inclusion context

The FDIC household survey documents banking access and use of financial services [1]. CFPB research examines consumers with limited or unscored credit histories [2]. These publications motivate the outreach problem; they are not sources for the simulated coefficients.

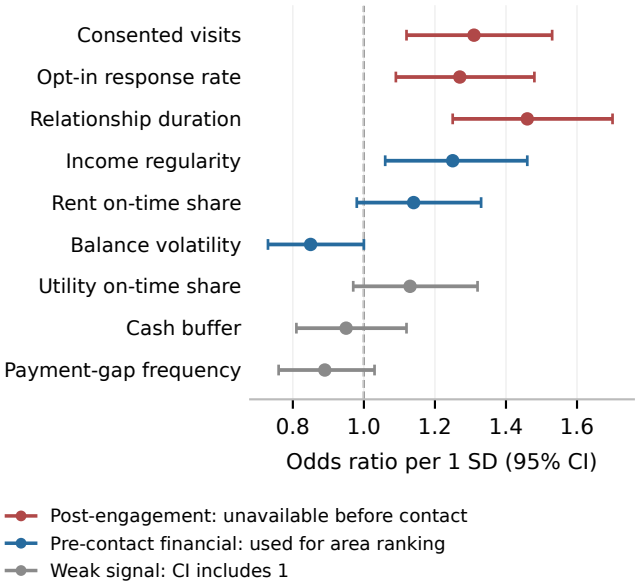


Figure 1: Parameter associations with 90-day retention in synthetic data. Points are ORs per 1 SD; bars are 95% CIs. Engagement signals are excluded from pre-contact area ranking. Rent is borderline ($p = 0.086$).

3 Results

Consented visits, opt-in response and relationship duration are the largest and clearly significant. Income regularity and balance volatility are significant but smaller. On-time rent is borderline ($p = 0.09$). Utility on-time share, cash buffer and payment-gap frequency are not distinguishable from zero.

Holdout AUC is 0.66 (modest) and calibration is close to the diagonal (Brier 0.234), so predicted probabilities are reasonable even though separation is moderate. PCA: the first three components each explain roughly 24%.

Parameter	β	SE	z	p	OR	95% CI (OR)
β_0 (intercept)	+0.04	0.06	+0.62	0.535	–	–
Consented visits	+0.27	0.08	+3.44	< 0.001	1.31	[1.12, 1.53]
Opt-in response rate	+0.24	0.08	+3.08	0.002	1.27	[1.09, 1.48]
Relationship duration	+0.38	0.08	+4.76	< 0.001	1.46	[1.25, 1.70]
Income regularity	+0.22	0.08	+2.71	0.007	1.25	[1.06, 1.46]
Rent on-time share	+0.13	0.08	+1.72	0.086	1.14	[0.98, 1.33]
Balance volatility	−0.16	0.08	−1.98	0.047	0.85	[0.73, 1.00]
Utility on-time share	+0.12	0.08	+1.55	0.121	1.13	[0.97, 1.32]
Cash buffer	−0.05	0.08	−0.60	0.550	0.95	[0.81, 1.12]
Payment-gap frequency	−0.12	0.08	−1.53	0.127	0.89	[0.76, 1.03]

Table 1: Logistic regression, $n = 1{,}200$, synthetic data from the pipeline. Intercept and effect sizes are illustrative. Values are reproduced from the source paper; validation uses a separate 600-participant holdout.

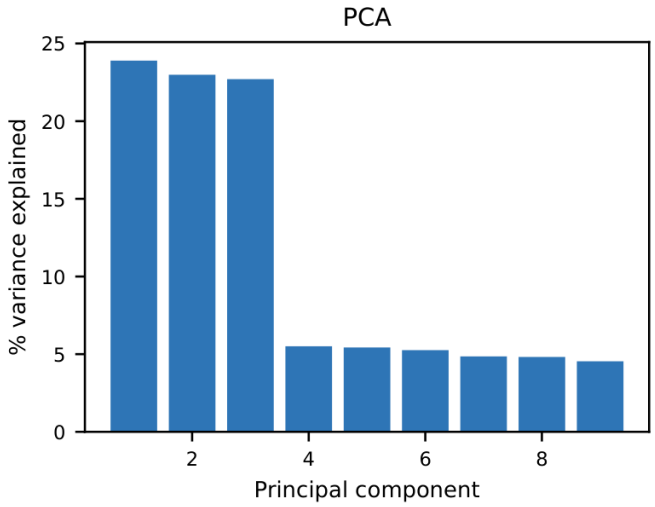


Figure 2: PCA variance explained. The first three components each explain roughly 24%, consistent with three signal groups. Original synthetic-data plot retained.

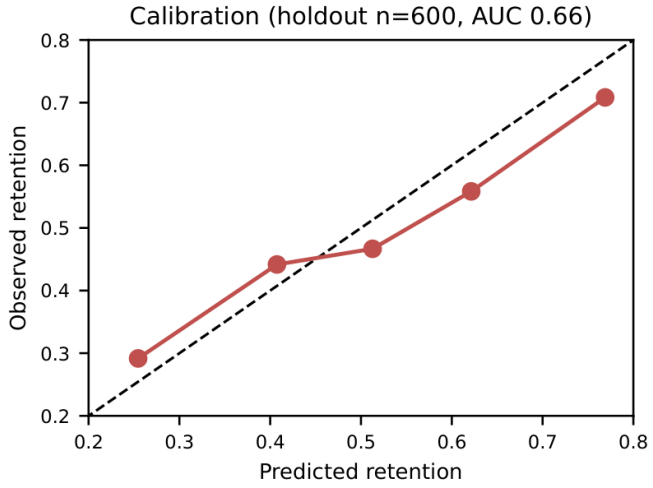


Figure 3: Holdout calibration ($n = 600$, AUC 0.66, Brier score 0.234). The dashed diagonal represents perfect calibration. Original synthetic-data plot retained.

4 Conclusion

Relationship signals dominate retention once a person is engaged. They drive the day-45 follow-up list. **Before contact, the usable parameters are income regularity, on-time rent and balance volatility.** We aggregate these to area level, apply their fitted β weights, and rank where to go.

Engagement signals are excluded from area ranking because they do not exist for communities not yet reached, and using them would be circular.

Limits: synthetic data, modest AUC; weights must be re-estimated on real pilot data before any decision relies on them.

Interpretation boundary

The coefficients describe an intentionally simulated demonstration. Neither statistical significance here nor the external references establish real-world retention effects or justify deployment without pilot validation.

References

[1] Federal Deposit Insurance Corporation, *2023 FDIC National Survey of Unbanked and Underbanked Households*, Nov. 2024. Available: fdic.gov/household-survey.

[2] Consumer Financial Protection Bureau, *Technical Correction and Update to the CFPB's Credit Invisibles Estimate*, June 2025. Available: consumerfinance.gov, Research Reports.