Approval odds, discouraged borrowers and experiment checks for a credit card marketplace

Rodrigo Hermont Ozon

2026-11-28

1. Question

A card marketplace must show approval odds that are true, recommend cards by value and the chance of approval, and read its experiments without fooling itself. This note reports what public data says on each point. Code and results: github.com/rhozon/credit-card-marketplace-analytics.

2. Data

The New York Fed's SCE Credit Access Survey (39,025 interviews of 17,904 people, October 2013 to October 2025) records card applications in the past 12 months, their outcome, discouragement (not applying for fear of a denial), a self-reported score band, balances, late payments and the stated chance of approval. It has no income, age or education. The Criteo Uplift v2.1 experiment (13,979,592 users, 85% exposed to advertising) and the Hillstrom e-mail test (64,000 customers, three arms) are the experiments. FRED and the CFPB complaint API give context and a guardrail.

3. Methods

Approval among applicants is modelled by a logit with seven visitor-facing inputs, a spline logit and monotone gradient boosting, trained to October 2022, calibrated on 2023 and tested on 2024 to 2025. Selection into applying is handled by a probit with sample selection fitted by maximum likelihood (Van de Ven and Van Praag, 1981), with the chance of needing US$ 2,000 as the excluded variable. Stated beliefs are matched to the outcome the same respondent reports in a later wave. Experiments are read after a sample ratio test, an A/A split, standardised differences and a classifier test of balance; effects use the delta method and regression adjustment (Deng et al., 2013; Lin, 2013), Holm across metrics, the mixture SPRT for continuous monitoring (Johari et al., 2022), and X- and DR-learners judged by Qini curves on held-out users.

4. Results

Approval model (test 2024-2025, n = 1,525)	AUC	Brier	Calibration slope
Logit, 7 inputs, isotonic	0.868	0.098	0.95
Spline logit	0.868	0.100	0.89
Monotone gradient boosting	0.858	0.101	0.78
Score band rate only	0.843	0.105	0.92
Base rate	0.500	0.142	-

Applications shown as "high" odds were approved 97.1% of the time in the test, "medium" 84.2% and "low" 58.1%. Approval among applicants fell from 83.6% (2019) to 78.1% (2025).

The selection model estimates a correlation of 0.86 between the unobservables of applying and of approval. Applied naively, the applicants' model gives discouraged borrowers 54% approval odds; corrected, 23%, and 6.6% of them exceed 50%. The two-step estimator gave an inverse Mills coefficient of 2.5, outside the admissible range, and was discarded.

People who stated a chance below 50% (mean 14%) were approved 56% of the time when they later applied. The model's Brier score on 3,774 belief-outcome pairs is 0.088 against 0.135 for the stated beliefs; the most pessimistic fifth relative to the model is discouraged 11.4% of the time against 0.5% in the middle fifth.

In Criteo the sample ratio is exact (p = 0.999) but balance fails: seven of twelve features differ by more than 0.02 standard deviations and a classifier predicts the arm with AUC 0.508 against a shuffled-label maximum of 0.502. The visit effect is 1.03 points raw and 0.70 adjusted, so the adjustment corrects bias as well as variance. Repeated looks at an A/A test inflate false positives to 20% (10 looks) and 40% (100 looks); the mixture SPRT stays at or below 1.7%. The top 10% by the X-learner gain 67 visits per 1,000 treated against 13 at random. In Hillstrom all six arm-metric effects survive Holm and covariate adjustment cuts variance by about 3%.

5. Limits

Score bands and outcomes are self-reported; the card requested is unknown, so odds are generic. Card terms in the calculator are examples: the CFPB plan files could not be downloaded automatically. Selection rests on one exclusion variable. The Criteo effect size is uncertain until the assignment log is checked.

6. References

Deng, A., Xu, Y., Kohavi, R. and Walker, T. (2013). Improving the sensitivity of online controlled experiments by utilizing pre-experiment data. WSDM, 123-132. doi:10.1145/2433396.2433413

Diemert, E., Betlei, A., Renaudin, C., Amini, M. R., Gregoir, T. and Rahier, T. (2021). A Large Scale Benchmark for Individual Treatment Effect Prediction and Uplift Modeling. arXiv. doi:10.48550/arXiv.2111.10106

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47(1), 153. doi:10.2307/1912352

Johari, R., Koomen, P., Pekelis, L. and Walsh, D. (2022). Always Valid Inference: Continuous Monitoring of A/B Tests. Operations Research, 70(3), 1806-1821. doi:10.1287/opre.2021.2135

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. The Annals of Applied Statistics, 7(1), 295-318. doi:10.1214/12-AOAS583

Van de Ven, W. P. M. M. and Van Praag, B. M. S. (1981). The demand for deductibles in private health insurance. Journal of Econometrics, 17(2), 229-252. doi:10.1016/0304-4076(81)90028-2