What Do Defaults Tell Us About Discount Rates? Revealed Preference from 20 Years of Credit-Bureau Data

Joseph Hall
Georgia Institute of Technology

Job Talk

2026–27 Market

Roadmap

1. **The question and the data**
2. A sufficient statistic for the discount rate
3. Baseline estimates — and every reason not to believe them, addressed
4. Heterogeneity: who discounts, where, and when
5. Aggregation and what it means

The discount rate is the primitive of household finance

- Every intertemporal decision a household makes — saving, borrowing, defaulting, refinancing, retiring — is priced off one parameter
- Yet we mostly measure it by **asking hypothetical questions** (elicited rates: 30–100%+, unstable) or by structural estimation inside heavily parameterized models
- This paper: measure it from **what borrowers actually do at the highest-stakes margin they face** — default

Default is an intertemporal trade

- A borrower who defaults **keeps the remaining payments** (money today and tomorrow) and **accepts the consequences** (chargeoff, collections, credit exclusion — money later)
- Where the borrower stops paying therefore *reveals* the rate at which they trade the future for the present
- The paper turns this observation into a sufficient statistic that needs no structural model of the household

Preview of answers

- Revealed-preference annual discount rate: $\approx 27\%$ at the three-year horizon ($\widehat{\beta} = 0.787$, s.e. 0.003) — and the marginal-utility correction is *measured* to be nil
- An order of magnitude above risk-free rates; far below elicited rates; consistent with behavior in the credit-card literature
- The heterogeneity is the economics: rates decline monotonically in credit score ($26.7\% \to 23.2\%$) and in income ($25.8\% \to 21.8\%$); geography spreads β across $[0.73, 0.84]$
- And the headline twist: different loan horizons trace out a **steeply declining discount curve** — within person — a discount rate schedule

The data: the best you could ask for on this question

- **Equifax ACRO archive, 10% sample of US credit files**: every trade line, **57 snapshot months, 2005–2024** (quarterly to semiannual)
- Auto loans: **6.2 million** fixed-term loans, **277,026 defaults** usable for estimation, all 48 continental states + DC
- Same archive, same estimator, three more contract families: **58M first mortgages, 4.9M home-equity, 72M personal installment** — horizons from under two years to seventeen
- Terms, payments, balances, chargeoffs, timing — everything the estimator needs, at population scale, for 20 years

Contribution relative to three literatures

- **Measured discount rates** (elicitation, field experiments): we replace stated choices with consequential ones, at 6M-loan scale
- **Consumer default** (strategic default, liquidity default): we use default *timing* as information about preferences, not just a cost to be explained
- **Household finance measurement** (payment targeting, bureau-data methods): we contribute a practitioner's toolkit — rate imputation from quantity fields, power-preserving sample design — reusable beyond this paper

Roadmap

1. The question and the data
2. **A sufficient statistic for the discount rate**
3. Baseline estimates — and every reason not to believe them, addressed
4. Heterogeneity: who discounts, where, and when
5. Aggregation and what it means

Two periods, one identity

- Defaulting borrower keeps remaining payments A, loses chargeoff consequences B, with continuation default risk p_2
- Indifference at the observed stopping point delivers the **exact identity**

$$\frac{A}{(1-p_2)\,B+A} \;=\; \beta\,\lambda$$

- β: the discount factor. λ: relative marginal utility of the default state — the correction term the literature usually assumes away

From identity to estimator

- Implementation: two fixed-effects logits (default on log payment; default on log total obligation) within score-ventile $\times$ zip-by-month cells
- $\widehat{R} = \widehat{\gamma}_L / (\widehat{\gamma}_L + \widehat{\gamma}_m)$, annualized through mean defaulter horizon $\widehat{\Delta T}$:

$$\widehat{\beta} \; = \; \widehat{R}^{1/\widehat{\Delta T}}$$

- Identification: cross-sectional variation in payment size against obligation size, within cells that absorb local shocks and creditworthiness

The theory extends as far as anyone will push it

- T-period and continuous-time generalizations: the identity survives as the envelope condition of the optimal stopping problem (Appendix; available for the whiteboard)
- Heterogeneous agents: aggregation theory for what a population β distribution implies for measured aggregates — pooled estimation recovers a sensitivity-weighted average of type-level statistics (exact aggregation proposition)
- Marginal-utility correction: **measured directly in the panel** — the key input is the persistence of the marginal borrower's state, and 21.4M delinquency transitions put it at $\rho = 0.992 \Rightarrow \lambda = 1.00$, correction nil (the literature's lower-persistence calibrations move the rate to no less than 25%)

Roadmap

Baseline estimate

- $\widehat{\beta} = 0.787$ (s.e. 0.003) $\Rightarrow \approx 27\%$ **per year** at the three-year horizon
- The one approximation is measured away: at the panel's own distress persistence, $\lambda = 1.00$ — the correction is nil
- And the clock is not doing the work: the sensitivity ratio $R = 0.465$ would need $T = 18.8$ years on five-year loans to be consistent with $\beta = 0.96$
- Benchmarks: risk-free ∼2–4%; credit-card APRs ∼20–25%; elicited experimental rates 30–100%+

‣ measured correction ‣ is it just T? ‣ precision

Power here is a theorem

$$\mathrm{se}(\widehat{\beta}) = \frac{c}{\sqrt{n_{\text{defaults}}}}$$

- Rare-events information: every entry of the estimator's covariance scales with $1/n_d$ — **loans are free, defaults are power**
- Validated on 136 geographic clusters: log-log slope -0.615 (0.077); conditional on defaults, the loan count contributes **exactly nothing**: -0.016 (0.153)
- Consequence: every sample split in this paper is designed at the default margin (equal-default bands, default-floor clusters)

Threat 1: "you never observe the interest rate"

- Correct — no credit bureau records prices; the `rate` field is a payment-status code
- Answer: the contract's quantity side (L, m, T) **overidentifies** the rate — annuity inversion at origination, balance-path recovery from monthly snapshots
- Both validated on simulated bureau data with known rates; the two methods fail on **disjoint** reporting quirks, so their agreement is a data-quality diagnostic (Appendix: practitioner's guide)

▸ error taxonomy

Threat 2: "payment size is endogenous"

- We killed our own first instrument, in public: rate levels at origination fail three redesigns (vintage confounding — a purely cohort-level instrument cannot be rescued)
- The living design: **lender-menu leniency** — the staggered 60→72→84-month term expansion and lender size menus give two leave-one-out instruments in the *same* sample as the FE strategy; first stages $F = 221$ and $F = 47{,}000$
- Where the well-identified designs and the FE specifications meet: HAMP RD 0.76, bankruptcy RKD 0.78, and the FE family 0.79–0.80 across three specifications — the agreement table
- Lender-policy natural experiments are bundled treatments in this data (adoption events shift clientele too); the sharpness diagnostics and the graveyard are all reported

▸ the instrument graveyard

Threat 3: “defaulters are selected”

- Of course — the estimator *conditions* on the default margin; the question is what changes across comparisons
- Cross-group comparisons hold the margin fixed by design (equal-default bands); geographic correlations are re-estimated under default-composition controls
- Example below: the income–patience correlation **dies** under the selection control; the poverty and partisanship correlations **survive** — the paper reports both honestly

‣ full table

Threat 4: “they just can’t pay”

- **65% of 2.46M auto defaulters keep paying other creditors** six months *after* the chargeoff (10.7 other trades on average; only 1.6% default on everything) — the modal default is a portfolio choice
- $\widehat{\beta}$ is **flat across local labor shocks**: 0.774 / 0.786 / 0.789 by unemployment-change tercile
- The same borrower reveals 29% on personal loans, 12% on autos, 1% on the mortgage — **one budget cannot bind selectively**
- Distress is measured near-permanent ($\rho = 0.992$) — the transitory-shock story contradicts the data; the permanent-state story *is* the measured λ

Monte Carlo: why $\widehat{\beta}$ can’t acquit itself

Threat 5: "they just expect to be rich"

"Everyone was a temporarily embarrassed capitalist."

— John Steinbeck, "A Primer on the Thirties," *Esquire*, 1960

(the popular "temporarily embarrassed millionaires" is Ronald Wright's 2004 paraphrase)

- The objection: high borrowing rates reveal expected income growth, not impatience
- Three answers, in increasing strength: distress at the default margin is measured near-permanent ($\rho = 0.99$, 0.26% of spells cure per quarter); the growth calibrations large enough to matter are refuted by the data's own bound ($\lambda \geq 0.47$); and the SCF lower bounds hand every household its *most optimistic* lifecycle income path — the 16–20% floor survives
- Even a temporarily embarrassed capitalist should not pay 25% APR once the model has priced in the capital

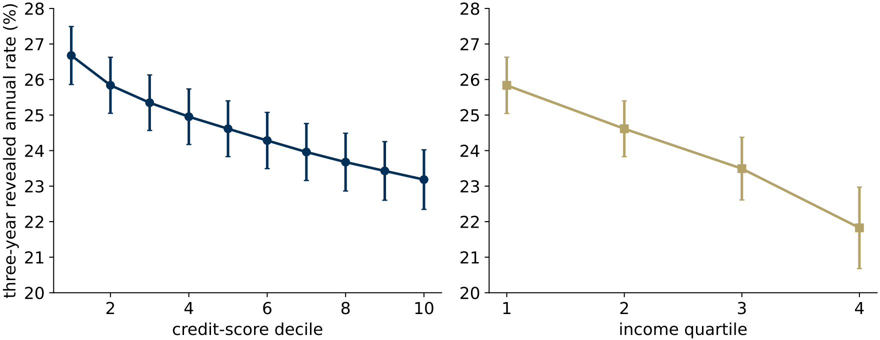
Roadmap

1. The question and the data
2. A sufficient statistic for the discount rate
3. Baseline estimates — and every reason not to believe them, addressed
4. **Heterogeneity: who discounts, where, and when**
5. Aggregation and what it means

The economics lives in the cross-section of β

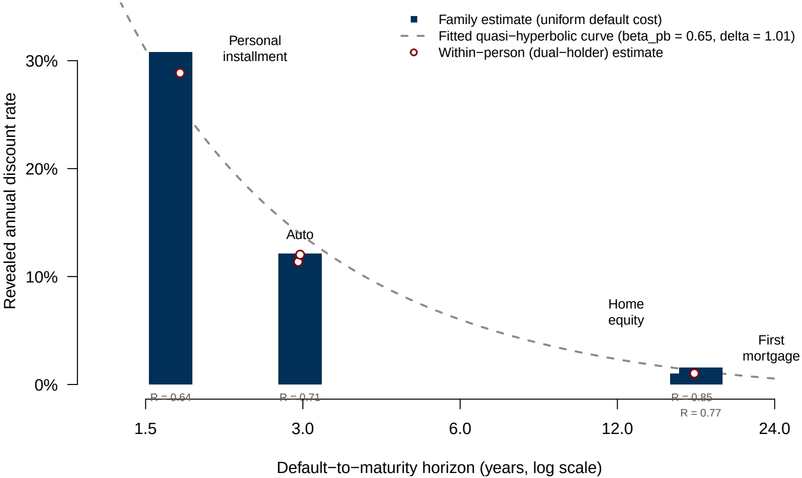
- A single aggregate discount rate answers no policy question — **who** is impatient governs the incidence of every intertemporal policy: moratoria, minimum payments, forbearance, retirement defaults
- This section: credit score, income, loan type, geography, political geography, and time
- Power-preserving design: every cell flagged by its default count, nothing silently pooled

By credit score and income: monotone, precise, modest



Varying-coefficient design: every default informs every cell, so the bands are **0.4pp** wide. Both gradients are monotone; both are modest against the 20–27% level. High default rates among low-score borrowers are mostly something other than impatience.

The revealed discount curve, traced loan by loan



Four families, one estimator, uniform default cost — the annual rate falls from **31% at a 1.7-year horizon to 1% at seventeen years**, and the dual-holder markers show it **within person**. The measurement is a schedule.

▸ (β_{pb}, δ) identification ▸ not liquidity

Geography: 154 power-designed clusters, 48 states

- Sub-state clusters built on zip3 adjacency to a default floor; estimated cluster-by-cluster with the full estimator
- Cross-cluster spread is **large**: β p10–p90 $[0.733, 0.835]$ — annual rates from $\sim$20% to $\sim$36%
- Decomposition: 84% of the cross-cluster variance in $\log \beta$ comes from the default sensitivities alone; the horizon contributes 2%

What geography correlates with

- Merged with IRS SOI zip3 covariates (loan-count weighted):
- **Poverty (EITC share):** $\rho_w = -0.37$, slope -0.42 (0.09) — poorer areas show *lower* measured discount rates; survives default-composition control
- Income level: raw positive, **dies under the selection control** — reported as a selection artifact
- Unemployment share, itemizer share: positive, state-FE fragile

Political geography: patience leans Republican

- County presidential returns (2008–2024) crosswalked to clusters
- 2020 GOP two-party share: $\rho_w = -0.19$ (Spearman -0.30); slope -0.088 (0.040), **strengthens** under the default-composition control (-0.101 (0.038))
- The 2016–2020 **swing is a precise zero** ($\rho_w = 0.01$) — the partisan *level* correlates with patience; the Trump-era movement is orthogonal to it
- Five-cycle panel (2008–2024) in estimation **[PLACEHOLDER: geography $\times$ cycle betas — within-cluster changes, lead-lag structure]**
- Companion analysis: enters the manuscript with the next estimation cycle; replication outputs in the project repository

Across time: patience is rising

- State $\times$ cycle cells, five presidential cycles: median β climbs **0.719** $\rightarrow$ **0.764** $\rightarrow$ **0.779** $\rightarrow$ **0.774** $\rightarrow$ **0.808** from 2005–08 to 2021–24
- Rates at the default margin fell from $\sim$39% to $\sim$24% over two decades — composition, credit access, or preference drift: the panel can ask
- Recession cohorts run more impatient (0.731 vs. 0.792) — a composition effect the paper flags and does not over-read

Roadmap

From micro β to macro objects

- Heterogeneous-agents aggregation: measured aggregates load on the *patient tail*, so population medians and macro calibrations answer different questions
- The 27% median coexists with standard macro calibrations, and the aggregation proposition states exactly how

What it means

- Incidence: debt relief and forbearance are transfers *to* the impatient margin, and the margin is now measured
- Pricing: lenders setting minimum payments face a distribution of β this paper measures directly
- Measurement: the toolkit (rate imputation, power-preserving splits, the exact identity) transfers to any bureau-data setting

Conclusion

- Defaults reveal discount rates: **≈27%/year** at the three-year horizon — and a **steeply declining schedule** across horizons, within person
- The best available data, attacked with an exact identity, a *measured* correction, designed power, and every objection diagnosed in the open
- The discount curve (four bars, one dashed line) is the picture to leave the room with the curve

Backup arsenal

Organized for the fight: theory, calibration, power, imputation, selection, data plumbing.

Backup: exact identity, full derivation

- Two periods: implicit differentiation of the period-1 threshold gives $\partial p_1/\partial L = f(y_1^*)\beta(1-p_2)\mathbb{E}[u'(C_2^P) \mid y_2 > y_2^*]/\Delta u_1'$ and $\partial p_1/\partial m$ as the current-payment analogue
- The ratio cancels the margin density and the utility scale, leaving $\beta\lambda$
- T periods: the same identity survives as the envelope condition of optimal stopping, provided the punishment has a flow component and good standing carries the option to default later

Backup: continuous-time statement

- Continuous time: default is a stopping time on the income process, and the identity appears as the smooth-pasting condition at the default boundary
- The response of stopping to a payment perturbation at horizon s is $D(s)$ times the survival-weighted marginal utility at s — ratios across horizons identify the discount function nonparametrically
- Two horizons point-identify quasi-hyperbolic (β_{pb}, δ); three make the exponential null testable

Backup: the marginal-utility correction, measured

- The key input is the persistence of the marginal borrower's state: 21,350,801 delinquency-spell transitions, **0.26% cure per quarter** $\Rightarrow \rho = 0.992$ annualized
- At measured ρ: $\lambda = 1.00$ at the Blundell–Pistaferri–Preston pass-through — **correction nil**; the headline stands as measured
- Literature-persistence sensitivity ($\rho = 0.95$, STY 2004): $\lambda = 0.95$, corrected rate 25% — the floor across every admissible calibration
- Internal-consistency bound: identification requires $\lambda \geq \widehat{\beta}^{\,T} = 0.47$, refuting exactly the calibrations that would generate large corrections

◂ back: baseline

Backup: the clock contributes little

- Decompose cross-group variance of $\log \beta$ into sensitivity (R) and horizon (T) contributions: geography **84%** ***R*-only** (rank correlation 0.97 at fixed T); state-cycle 89%
- Levels are T-proof: $R = 0.465$ requires $T = 18.8$ **years on five-year loans** to be consistent with $\beta = 0.96$; a 20% error in T moves the headline 27% $\rightarrow$ 22%
- Where T *does* carry magnitude — across loan families — that is the term-structure result, promoted to a finding

◂ back: baseline ◂ back: the curve

Backup: the estimator under liquidity-driven default

- Ground rule: in a simulated pure-liquidity economy, $\widehat{\beta}$ lands at **0.78, not 0** — obligations load mechanically on payments, so $\widehat{\beta}$ alone cannot referee
- Mixture path (true $\beta = 0.95$): $\widehat{\beta}$ falls 0.887 $\rightarrow$ 0.783 as the liquidity share goes 0 $\rightarrow$ 1
- Default-rate-matched surface, true $\beta \in [0.55, 0.99] \times$ shares up to 0.50: **no cell reaches the measured 0.787** — patient preferences plus admissible contamination cannot manufacture the estimate

◂ back: threat 4 ◂ back: the curve

Backup: two horizons identify (β_{pb}, δ)

- Quasi-hyperbolic $D(t) = \beta_{pb}\,\delta^t$: with horizon factors $b(T) = D(T)^{1/T}$,
  $$\ln \beta_{pb} = \frac{\ln b(T_1) - \ln b(T_2)}{1/T_1 - 1/T_2}$$
- Dual holders: auto $\times$ mortgage $\Rightarrow$ $\beta_{pb} = 0.71$, $\delta = 1.01$; auto $\times$ personal $\Rightarrow$ $\beta_{pb} = 0.55$, $\delta = 1.09$ — **large present bias, long-run rate $\approx$ 0** within measurement error
- Estimand discipline: every cross-person object is *horizon-labeled*; cross-horizon statements are statements about the curve; the schedule-proof heterogeneity object is the joint distribution of (β_{pb}, δ)

◂ back: the curve

Backup: power law — matrix statement

$$\mathrm{Avar}(\widehat{\beta}) = \nabla g' \, \Sigma \, \nabla g, \qquad \Sigma = V/n_d, \qquad \mathcal{I}(\gamma) \approx \mathbb{E}[n_d] \cdot M$$

- Rare-events logit information proportional to expected events (King–Zeng 2001); M = default-weighted design moment
- Horse race across 136 clusters: ln n_d coefficient -0.608 (0.102); ln n coefficient -0.016 (0.153)

◂ back: baseline

Backup: rate imputation — error taxonomy

- Method 1 exact except: odd first period (+18bp), balloons (adjusted inversion restores exactness)
- Method 2 exact except intermittent stale balances (-0.1bp median rule) and curtailments (-0.6bp median; mean version explodes)
- Financed fees: no bias; a $\sim$77bp note-rate/APR wedge at 2% fees — imputed rates lower-bound Reg-Z APRs

◂ back: threat 1

Backup: selection diagnostics, full table

- Poverty (EITC share): slope -0.42 (0.09) without the default-composition control, and the correlation survives with it
- Zip-level income: positive in the raw cross-section and gone under the control — reported as a selection artifact
- Partisanship level: -0.088 (0.040) raw, -0.101 (0.038) controlled; the 2016–2020 swing is a precise zero

◂ back: threat 3

Backup: the instrument graveyard, and the living design

- GS3-at-origination, three repairs, three failures: vintage controls ($\beta = 1.14$, CF $= 4.7$), AR(1) innovations (first-stage $F = 0.0$), imputed-rate target ($\beta = 1.07$) — a purely cohort-level instrument cannot be rescued
- Menu leniency: first stages $F_{dur} = 221$, $F_{size} = 47{,}288$ (lender-clustered); naive CF flips γ_m — long-term-menu lenders are subprime, menus proxy clientele
- Lender-FE iteration: within-lender instrument too weak for the control function — recorded; the agreement exhibit rests on the external designs and FE spec stability

◂ back: threat 2

Backup: data plumbing that survives review

- Single deterministic pipeline, fail-fast probes, restartable extracts; every exhibit regenerated from raw archives
- Known traps handled: zip zero-padding, integer overflow in population-weighted merges, CT's 2024 FIPS redistricting, left-truncated pre-2005 cohorts, censored 2021+ cohorts

Backup: what the archive contains

- 73 fields per trade-month; flags for every loan family; balloon, deferral, and narrative codes (fixed vs. variable rate)
- 57 snapshot months (283 archive files) $\times$ 10% of US credit files, 2005–2024