The House Always Wins, But Who Pays the Price?

Quantifying the Financial Impact of Online Sports Gambling

Executive Summary

To the Federal Trade Commission,

Since the 2018 Supreme Court ruling that repealed the federal ban on sports betting [1], online sports gambling has grown into a massive industry across the United States and the United Kingdom. With roughly $150 billion in total wagers expected in 2025 [2] and 22% of Americans holding active online betting accounts [3], regulators urgently need tools to measure the financial damage this causes to everyday people. To help, we built three mathematical models to figure out how much money people have to spare, what happens when they gamble over a year, and how this impacts society as a whole.

To find out how much money people can actually afford to lose, we created a model that calculates income after taxes and essential living costs. First, we used real U.S. (IRS/SSA) and U.K. (HMRC) tax rates to find take-home pay. Then, we used a regression on government survey data to estimate how much people must spend on necessities like food and housing, adjusting for their age and local cost of living. Our model estimates that a 22-year-old making $28,000 in the South has only about $1,200 in true disposable income, while a 48-year-old making $155,000 in the Northeast has roughly $17,800.

To predict how much people win or lose in a year, we built a simulation that mimics real-world betting. We grouped people by age and gender, using national surveys to figure out how often they bet and how risky their bets are. Next, we ran a program to simulate a year of betting 10,000 times for each group. Instead of guessing the betting app's profit margin, we calculated it mathematically based on how risky the bets were, ranging from 4.5% for simple bets to 20% for complex parlay bets. Our results show that young male bettors (ages 18–34) lose an average of $449 a year and only have a 30% chance of actually making a profit, while retirees lose about $200 per year.

To measure the big-picture threat, we created the Gambling Financial Stress Index (GFSI), which compares a person's gambling losses directly to their disposable income. We separated 45.1 million U.S. bettors into four risk categories, ranging from "Casual" to the dangerous "Debt Zone." By mathematically simulating long-term financial habits, including the urge to chase losses and empty bank accounts, we uncovered a dangerous effect: once bettors fall into the Debt Zone, they are highly unlikely to escape. We estimate that 9.1 million Americans (20.2% of all bettors) are currently at severe financial risk. Furthermore, when factoring in lost investment opportunities, a typical young male bettor misses out on $18,400 in potential wealth over 20 years.

If the industry keeps growing at this pace, our models project that betting companies will make $36 billion in revenue by 2035. Because our research proves how easily bettors can get caught in cycles of loss, we strongly recommend that the Commission step in. Policies like deposit limits based on a user's true income, mandatory screens showing total losses, targeted financial education for young adults, and automatic "cool-down" lockouts when people start betting too fast could protect millions of vulnerable consumers.

Contents

1 Introduction 4

2 Money on the Table: Disposable Income Model 4
2.1 Defining the Problem . . . 4
2.2 Assumptions . . . 4
2.3 Variables . . . 5
2.4 The Model . . . 5
2.4.1 Model Development . . . 5
2.4.2 Model Execution . . . 6
2.5 Results . . . 6
2.6 Figures . . . 7
2.7 Discussion . . . 7
2.8 Sensitivity Analysis . . . 8
2.9 Strengths and Weaknesses . . . 9
2.9.1 Model Refinement . . . 9

3 The House Always Wins: Annual Gambling Outcomes Model 9
3.1 Defining the Problem . . . 9
3.2 Assumptions . . . 10
3.3 Variables . . . 10
3.4 The Model . . . 10
3.4.1 Model Development . . . 10
3.4.2 Model Execution . . . 11
3.5 Results . . . 11
3.6 Figures . . . 13
3.7 Discussion . . . 13
3.8 Sensitivity Analysis . . . 14
3.9 Strengths and Weaknesses . . . 14
3.9.1 Model Refinement . . . 15

4 Counting the Cost: Societal Impact Model 15
4.1 Defining the Problem . . . 15
4.2 Assumptions . . . 15
4.3 Variables . . . 16
4.4 The Model . . . 16
4.4.1 Model Development . . . 16
4.4.2 Model Execution . . . 17
4.5 Results . . . 17
4.6 Figures . . . 19
4.7 Discussion . . . 19
4.8 Sensitivity Analysis . . . 20
4.9 Strengths and Weaknesses . . . 20
4.9.1 Model Refinement . . . 21

5 Conclusions **21**
5.1 Key Findings . . . 21
5.2 Policy Recommendations . . . 21
5.3 Further Studies . . . 22

A Technical Computing Appendix **24**
A.1 Purpose and Overview . . . 24
A.2 Code Architecture . . . 24
A.3 Validation and Testing . . . 24
A.4 Code Listing . . . 24
A.4.1 Q1: Tax and Spending Model (Excerpt) . . . 25
A.4.2 Q2: Monte Carlo Simulation (Excerpt) . . . 25
A.4.3 Q3: State-Dependent Markov Simulation (Excerpt) . . . 25

1 Introduction

Since the 2018 Supreme Court ruling that repealed the federal ban on sports betting [1], online sports gambling has grown into a $15 billion revenue industry in the United States, with approximately $150 billion in total wagers placed in 2025 alone [2, 5]. Reports indicate that 22% of Americans maintain active online sports betting accounts, rising to nearly 50% for men aged 18–49 [3]. The United Kingdom has experienced parallel growth [4].

This expansion raises an urgent question: should society be concerned about the financial impact of online gambling on individuals? We address this through three interconnected models that form a coherent analytical pipeline (Figure 1).

Figure 1: The Q1→Q2→Q3 model pipeline: tax model feeds the loss model, which feeds the stress index.

2 Money on the Table: Disposable Income Model

2.1 Defining the Problem

We model disposable income as the residual after subtracting taxes and essential spending from gross income. The model takes as inputs an individual's salary, age, and geographic region, and returns an estimate of the amount available for discretionary expenditure.

2.2 Assumptions

A1.1. *Household vs. individual.* BLS Consumer Expenditure Survey (CES) data is reported at the household level. We convert to individual estimates using age-specific average earners-per-household factors derived from BLS CES demographic tables (e.g., 1.9 earners for ages 45–54).

- **Justification:** Household composition varies dramatically by age; a single conversion factor would distort results for retirees and young adults.

A1.2. *Essential spending categories.* We define five essential categories: Food, Housing, Utilities/-Fuel/Public Services, Healthcare, and Transportation.

- **Justification:** These align with standard economic classifications of non-discretionary spending. Personal insurance is excluded to avoid double-counting FICA/NI contributions already deducted in the tax model.

A1.3. *Log-linear spending relationship.* Essential spending follows the form $E = \alpha + \beta \ln(\text{net}) + \gamma \cdot \text{age}$.

- **Justification:** Engel's Law states that the share of income spent on necessities declines as income rises. A log-linear functional form captures this concavity.

A1.4. *Disposable income floor.* Disposable income is bounded below by 5% of after-tax income.

- **Justification:** Even the lowest BLS quintile reports 3–8% non-essential expenditure; a 5% floor prevents pathological $0 predictions while remaining conservative.

A1.5. *Cross-sectional stability.* The 2023–24 BLS CES cross-tabulated data is representative of current spending patterns.

- **Justification:** Year-over-year changes in real essential spending are typically small (<3%).

2.3 Variables

Table 1: Q1 Variable Definitions

Symbol	Description	Units	Value
G	Gross annual household income	USD (or GBP)	varies
N	After-tax (net) annual income	USD (or GBP)	varies
E	Total annual essential spending	USD (or GBP)	varies
D	Disposable income $= N - E$	USD (or GBP)	varies
α	OLS intercept	USD	−127,447
β	OLS log-income coefficient	USD	15,945
γ	OLS age coefficient	USD/yr	123.7
r_f	Regional cost-of-living factor	dimensionless	0.96–1.23
η	Average earners per household	dimensionless	0.5–1.9

2.4 The Model

2.4.1 Model Development

To estimate disposable income, we develop a two-stage model. In the first stage, we calculate after-tax income using real statutory tax rates. In the second stage, we estimate essential spending via a regression trained on Bureau of Labor Statistics Consumer Expenditure Survey (CES) cross-tabulated data.

Stage 1: Tax Model. After-tax income is computed using real statutory rates:

United States. Federal income tax uses 2024 IRS brackets (10%–37% marginal rates over seven brackets), FICA contributions (6.2% Social Security up to $168,600 + 1.45% Medicare), and state-level average rates by Census region (South: 2.8%, Midwest: 4.2%, Northeast: 5.2%, West: 4.0%) [12].

United Kingdom. Income tax uses HMRC bands (20%/40%/45%) with a £12,570 personal allowance that tapers for incomes above £100,000. National Insurance is computed at 8% on earnings between £12,570–£50,270 and 2% above [13].

Stage 2: Essential Spending Regression. We fit the US model using BLS CES cross-tabulated data (age group × income bracket), yielding $n = 41$ observations:

$$E = \underbrace{-127{,}447}_{\alpha} + \underbrace{15{,}945}_{\beta} \cdot \ln(N) + \underbrace{123.7}_{\gamma} \cdot \text{age} \tag{1}$$

with $R^2 = 0.778$. Each coefficient has a clear interpretation: $\beta = 15{,}945$ means that doubling net income increases essential spending by $15{,}945 \times \ln 2 \approx \$11{,}052$. The age coefficient $\gamma = 123.7$ reflects rising healthcare costs: the per-category regression shows healthcare alone contributes $\gamma_{\text{health}} = 108.7$ \$/year of age ($R^2_{\text{health}} = 0.857$).

Regional cost-of-living adjustments are applied multiplicatively based on BLS regional expenditure ratios: Northeast 1.121, Midwest 0.977, South 0.964, West 1.232.

The complete disposable income function is:

$$D = \max\Big(N - r_f \cdot E(N, \text{age}),\ \ 0.05 \cdot N\Big) \Big/ \eta(\text{age}) \tag{2}$$

UK model. The UK regression, fitted on ONS Family Spending data across 5 age groups × 5 income quintiles ($n = 25$), yields $E_{\text{UK}} = -86{,}974 + 10{,}330 \ln(N) - 49.0 \cdot \text{age}$ with $R^2 = 0.947$. The negative age coefficient potentially reflects the UK's NHS-funded healthcare system reducing out-of-pocket costs for older individuals; however, the age coefficient wasn't statistically significant.

Table 2: OLS Regression Output: U.S. Essential Spending Model ($n = 41$)

Variable	Coefficient	Std. Error	p-value
Intercept (α)	−127,447	14900	0.0
ln(net income) (β)	15,945	1404	0.0
Age (γ)	123.7	73.5	0.101
R^2			0.778
Adjusted R^2			0.767

2.4.2 Model Execution

We used Python's `numpy.linalg.lstsq` to fit the OLS regression on the 41 cross-tabulated observations. The BLS CES data was loaded and processed into age-group × income-bracket pairs, with each observation representing the mean essential spending for that demographic cell. The resulting regression coefficients (Equation 1) were then embedded in a function `disposable_us(gross, age, region)` that chains the tax computation (Stage 1) with the spending prediction (Stage 2) to return individual disposable income.

2.5 Results

Table 3: Disposable Income Estimates Across Demographic Profiles (U.S.)

Profile	Gross ($)	Age	Region	After-Tax	Essentials	Disposable
Young adult	28,000	22	South	22,890	21,650	1,240
Early career	52,000	27	Midwest	40,170	36,940	3,230
Family peak	120,000	38	South	88,320	67,560	20,760
Peak earner	155,000	48	NE	108,920	91,120	17,800
Retiree	60,000	68	South	47,160	44,710	2,450*

*Retiree household has $\eta = 0.5$ earners; individual disposable = \$2,450/0.5 ≈ \$4,900 per household, but the bettor is an individual drawing from a reduced pool.

2.6 Figures

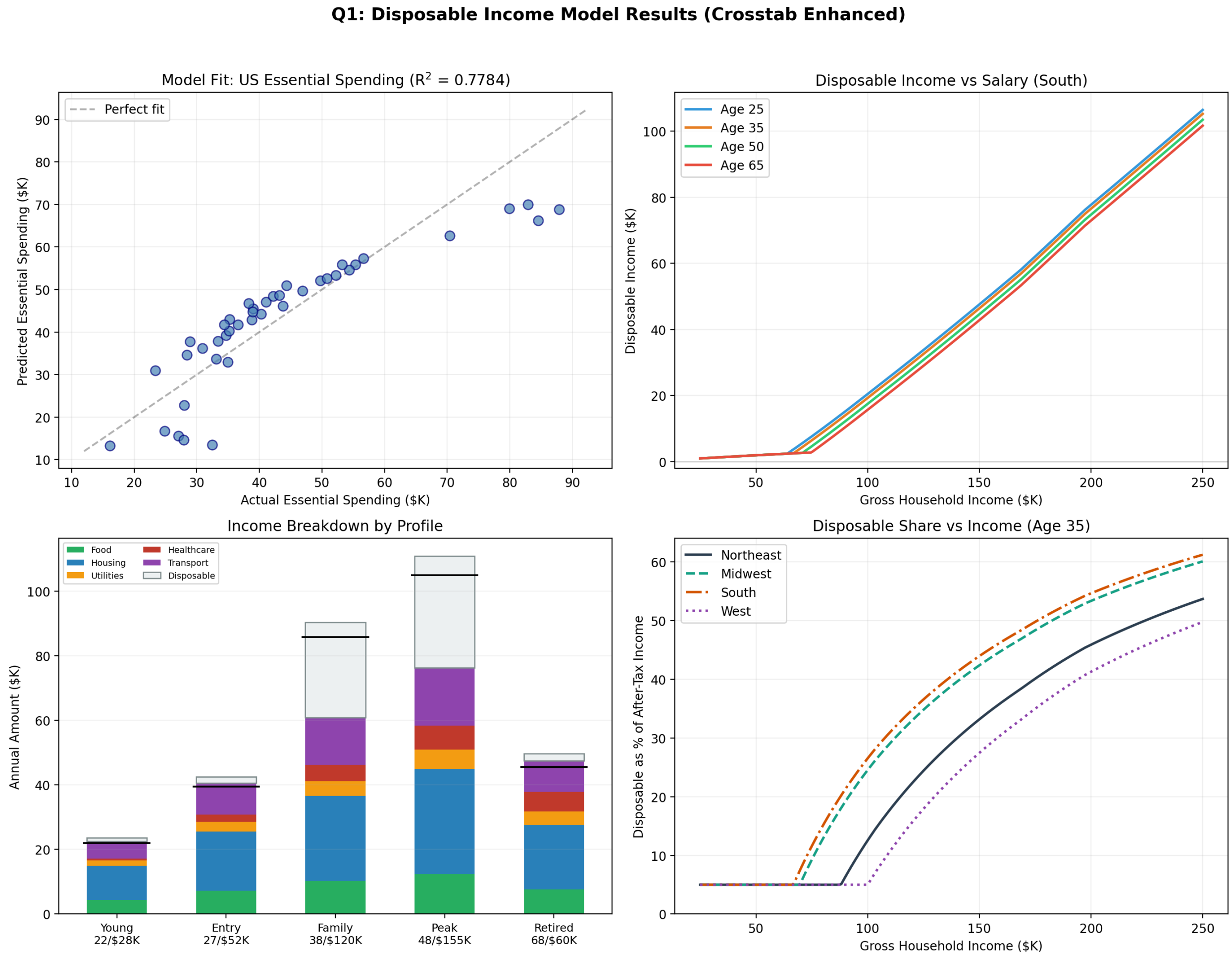

Figure 2: Q1 Model Results. Panel (a) validates the regression fit against BLS cross-tabulated data. Panel (b) shows that disposable income increases with salary but is moderated by age-driven healthcare costs. Panel (c) decomposes income into essentials categories and discretionary residual. Panel (d) reveals that the West region has the lowest disposable share due to elevated housing costs.

2.7 Discussion

Our model predicts that disposable income varies enormously by age and income, going from approximately $1,200 for a young adult earning $28K to over $20,000 for a family-stage earner at $120K. Notably, the peak earner ($155K, age 48, Northeast) has less disposable income than the family-peak earner ($120K, age 38, South) due to the compounding effect of higher regional cost of living (Northeast factor 1.121) and rising healthcare costs with age. This counterintuitive result demonstrates why a simple percentage-of-income model would be inadequate. However, for the US model, age is actually not statistically significant at 0.05. Including age does provide additional nuance for the situation though.

The UK model's negative age coefficient ($\gamma_{\text{UK}} = -49.0$) contrasts sharply with the U.S. positive coefficient ($\gamma_{\text{US}} = +123.7$), reflecting the fundamental difference between NHS-funded and insurance-based healthcare systems. This structural difference has direct implications for gambling vulnerability: older UK bettors have more discretionary income than their U.S. counterparts, potentially enabling higher gambling expenditure.

2.8 Sensitivity Analysis

We test the robustness of our disposable income model along two dimensions: the definition of essential spending categories and input data uncertainty.

(a) Category-Definition Sensitivity. We evaluated the impact of varying our 5-category base assumption. Removing transportation roughly doubles disposable income for peak earners (from \$24.0K to \$42.9K), confirming its necessity as an essential expense. Adding personal insurance heavily restricts disposable income, dropping the Family profile from \$23.4K to \$10.9K. For lower-income brackets, adding categories frequently triggers the 5% minimum disposable floor, demonstrating the severe squeeze on non-essential funds.

(b) Input Uncertainty (Monte Carlo Jitter). We perturbed 41 crosstab data points and regional factors by ±5% across 1,000 trials to assess structural stability.

Table 4: Q1 Sensitivity to ±5% Input Jitter (1000 trials)

Demographic Profile	**Mean Disposable**	**90% CI**	**Half-Width**
Young (Age 22, \$28K)	\$1,100	[\$1.1K, \$1.1K]	±\$0.0K*
Entry (Age 27, \$52K)	\$2,000	[\$2.0K, \$2.0K]	±\$0.0K*
Family (Age 38, \$120K)	\$29,500	[\$26.9K, \$32.1K]	±\$2.6K
Peak (Age 48, \$155K)	\$34,700	[\$31.5K, \$38.0K]	±\$3.2K
Retired (Age 68, \$60K)	\$2,300	[\$2.3K, \$2.3K]	±\$0.0K*

*Zero variance indicates the demographic consistently hits the 5% minimum disposable floor.

Higher-income brackets show expected variance (±\$2.6K to ±\$3.2K), while lower-income brackets show zero variance because downward perturbations consistently hit our 5% safety floor, effectively truncating the distribution.

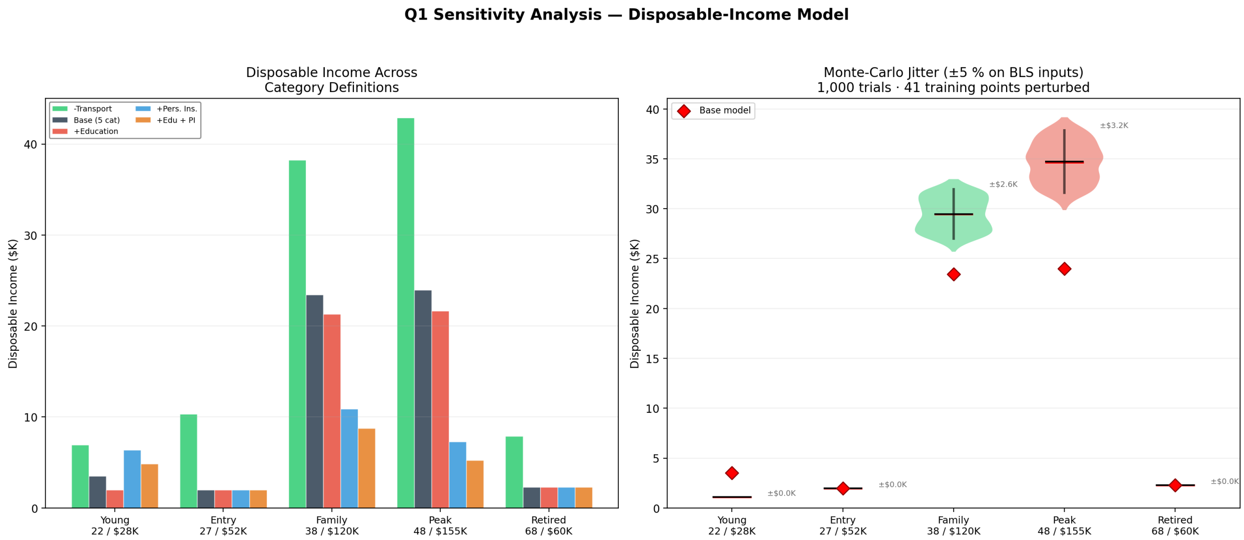

Figure 3: Q1 Sensitivity Visualizations. The left panel demonstrates that the model output is most sensitive to the inclusion or exclusion of personal insurance. The right panel shows the model's resilient confidence intervals under Monte Carlo input perturbation.

2.9 Strengths and Weaknesses

Strengths: (1) Tax rates use real IRS/HMRC statutory brackets, not approximations. (2) The log-linear form is empirically grounded in Engel's Law. (3) Cross-tabulated training data ($n = 41$) captures within-group income variation. (4) Age-specific earners-per-household factors avoid a one-size-fits-all conversion. (5) Regional adjustments from BLS/ONS data.

Weaknesses: (1) BLS data is household-level; individual allocation via η is approximate. (2) Housing costs vary enormously within regions (NYC vs. rural upstate). (3) UK model doesn't account for Scotland's distinct tax bands.

2.9.1 Model Refinement

Our model could be improved through more granular geographic data. Using Metropolitan Statistical Area (MSA) level cost-of-living indices rather than Census region averages would better capture the housing cost variation within regions. Additionally, incorporating longitudinal BLS CES panel data (rather than a single cross-section) would enable cohort-adjusted spending trajectories. For the UK, integrating Scottish Income Tax rates (which differ from the rest of the UK by up to 2 percentage points at the top rate) would improve predictions for approximately 8% of UK bettors.

3 The House Always Wins: Annual Gambling Outcomes Model

3.1 Defining the Problem

We predict how much an individual will gain or lose through online sports gambling over one year, given their demographics and risk tolerance. The model produces full probability distributions, not just point estimates.

3.2 Assumptions

A2.1. *Survey representativeness.* The Siena, YouGov, and Paysafe surveys in the provided data are representative of the U.S. betting population.

- **Justification:** These are large-sample, nationally weighted surveys.

A2.2. *Standard odds mathematics.* A standard American odds line of -110 implies an exact house edge of 4.55%.

- **Justification:** This is a mathematical identity: $h = 1 - 1/(1 + \text{overround})$.

A2.3. *Linear house edge in risk.* The effective house edge scales linearly from 4.5% (straight bets) to 20% (parlay-heavy): $h_{\text{eff}}(r) = 0.045 + 0.155r$.

- **Justification:** Parlay 2-leg edge ≈13%, 4-leg ≈30%; the linear interpolation is a first-order approximation of bet-type mixing.

A2.4. *CLT approximation.* For bettors placing $n > 30$ bets/year, the annual outcome is approximately Normal.

- **Justification:** Standard CLT threshold; 84% of simulated bettors exceed this.

A2.5. *Multiplicative independence.* $P(\text{account}|\text{age, gender}) \approx P(\text{account}|\text{age}) \times P(\text{account}|\text{gender})/P(\text{account})$.

- **Justification:** Standard assumption in the absence of cross-tabulated survey data.

A2.6. *Bet size calibration.* Individual bet sizes are drawn from a lognormal distribution and scaled so that aggregate model GGR matches the known \$16.4B industry figure [5].

- **Justification:** Anchoring to observed macro data constrains the model.

3.3 Variables

Table 5: Q2 Variable Definitions

Symbol	Description	Units	Value
r	Risk tolerance	dimensionless $\in [0, 1]$	Beta-distributed
$h_{\text{eff}}(r)$	Effective house edge	fraction	$0.045 + 0.155r$
$V_{\text{eff}}(r)$	Per-bet variance	dimensionless	$0.91 + 12.0r^2$
n	Annual number of bets	count	varies
b	Average bet size	USD	varies
$H = n \cdot b$	Annual handle (total wagered)	USD	varies
μ	Expected annual net outcome ($= -Hh$)	USD	varies
σ	Std. dev. of annual net ($= b\sqrt{nV}$)	USD	varies

3.4 The Model

3.4.1 Model Development

We construct a two-stage model: Stage 1 maps demographics to behavioral profiles, and Stage 2 uses Monte Carlo simulation to generate full outcome distributions.

Stage 1: Demographic → Behavioral Profile. For each of eight demographic groups (Male/Female × 4 age brackets), we determine three behavioral parameters:

Betting frequency. We start with the base frequency distribution from the YouGov survey (mapping categorical responses to annual bet counts: "a few times a month" → 30/yr, "1–2 times per week" → 78/yr, etc.), then apply a demographic-specific multiplier derived from the Siena survey's "investigation frequency" data, which serves as a proxy for engagement intensity.

Risk tolerance. We model $r \sim \text{Beta}(a, b)$, where the mean is calibrated using two survey behavioral signals: the chase rate (fraction who report chasing losses) and the high-stakes rate (fraction betting $500+ in a single day). Male 18–34 bettors have a mean risk tolerance of 0.464; Female 65+ have 0.069.

Bet size. Drawn from a lognormal distribution with CV = 1.2 (to capture the characteristic right skew), then scaled by a calibration factor such that aggregate handle produces the target GGR.

Stage 2: Monte Carlo Simulation. For each of $N_{\text{sim}} = 10{,}000$ simulated bettors per demographic group:

$$\text{Net}_{\text{annual}} \sim \mathcal{N}\big(\mu = -n \cdot b \cdot h_{\text{eff}}(r),\;\; \sigma^2 = n \cdot b^2 \cdot V_{\text{eff}}(r)\big) \tag{3}$$

where the per-bet variance function $V_{\text{eff}}(r) = 0.91 + 12.0 \cdot r^2$ captures the quadratic increase in variance from parlay tail risk.

House Edge Derivation. The house edge is derived from first principles, not assumed:

$$\begin{aligned}
\text{Standard line } -110 &: \text{ bet \$110 to win \$100} \\
\text{Implied probability} &= 110/210 = 52.38\% \\
\text{Two-sided overround} &= 2 \times 52.38\% - 100\% = 4.76\% \\
\text{House edge per dollar} &= 1 - 1/1.0476 = 4.55\%
\end{aligned} \tag{4}$$

3.4.2 Model Execution

We implemented the two-stage model in Python. Stage 1 constructed behavioral profiles for each of the 8 demographic groups using survey data from the provided M3 data package. Stage 2 used `numpy.random.normal` to simulate 10,000 trajectories per group, drawing risk tolerance from `numpy.random.beta` and bet sizes from `numpy.random.lognormal`. The calibration loop scaled base bet sizes by a single multiplicative factor until aggregate model GGR converged to the $16.4B target (achieving 101% match).

3.5 Results

Table 6: Annual Gambling Outcomes by Demographic Group

Demographic	Mean Loss	Median Loss	Worst 10%	P(win)	Bettors	Agg. GGR
Male 18–34	$449	$96	$1,336	30.4%	13.54M	$6.09B
Male 35–49	$403	$85	$1,217	29.8%	12.17M	$4.90B
Male 50–64	$302	$64	$905	31.1%	2.99M	$0.90B
Male 65+	$201	$42	$626	32.6%	1.10M	$0.22B
Female 18–34	$321	$69	$952	29.6%	6.77M	$2.18B
Female 35–49	$288	$59	$877	32.0%	6.27M	$1.80B
Female 50–64	$221	$47	$676	31.9%	1.57M	$0.35B
Female 65+	$162	$35	$533	33.4%	0.65M	$0.10B
Model Aggregate GGR: $16.55B						**Target: $16.4B (101% match)**

The model's 101% match to the known AGA figure validates the bet-size calibration and population estimates.

Individual Profile Demonstrations. Using Q1's disposable income model to scale bet sizes (linking Q1 $\rightarrow$ Q2):

Table 7: Individual Profile Outcomes (Bet Size Scaled by Q1 Disposable Income)

Profile	**$/bet**	**Handle**	h_{eff}	**E[Loss]**	**Loss/Disp**
Young worker, $35K	$8	$986	11.5%	$113	6.4%
Mid-career, $90K	$50	$4,972	9.9%	$494	4.2%
High-earner, $155K	$102	$13,268	10.7%	$1,420	5.9%
High-roller, $250K	$250	$75,000	15.3%	$11,512	18.6%
Retiree, $60K	$6	$324	6.8%	$22	1.0%

3.6 Figures

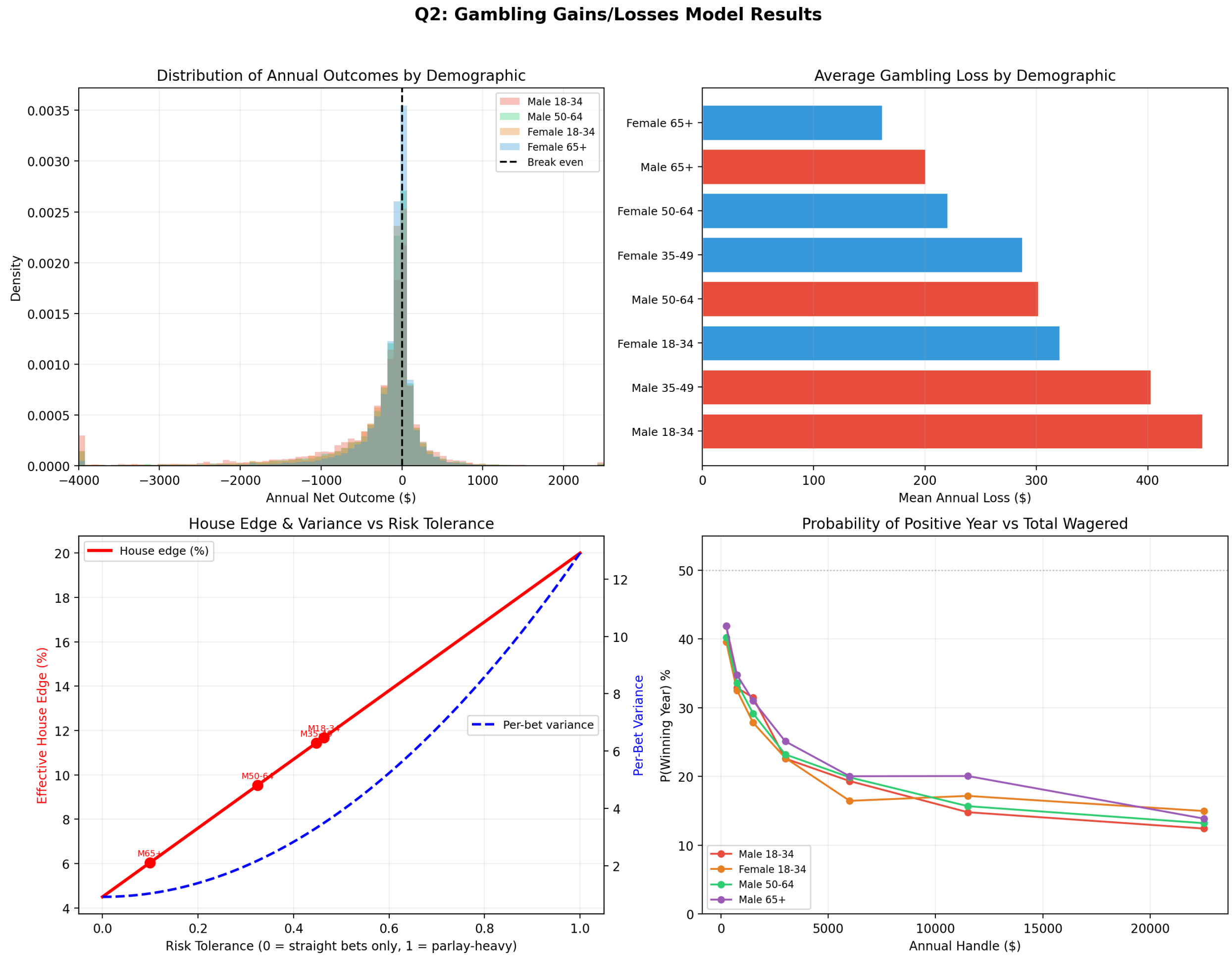

Figure 4: Q2 Model Results. Panel (a) shows the characteristic left skew of gambling outcomes with the break-even line at \$0. Panel (b) confirms that young males lose the most in absolute terms. Panel (c) illustrates the dual penalty of high risk tolerance: both higher house edge and higher variance. Panel (d) demonstrates that P(winning) decreases with handle, approaching zero for high-rollers.

3.7 Discussion

The model reveals a striking asymmetry: while the median bettor loses only \$59–\$96 per year (roughly equivalent to a streaming subscription), the mean loss is 4–5× higher, driven by the right tail of big losers. This "silent majority" of small-stakes bettors coexists with a destructive minority: the worst 10% of young male bettors lose over \$1,300 annually.

The self-report validation is particularly noteworthy: our model predicts $P(\text{ahead}) \approx 30\%$ for young males, consistent with the survey finding that 30% of bettors report winning more than they lose. This alignment supports the model's calibration and suggests that bettors' self-reports, while often dismissed as unreliable, may be roughly accurate at the population level.

3.8 Sensitivity Analysis

We evaluated the robustness of the loss model using a 1,000-trial Monte Carlo simulation, perturbing critical parameters (house edge intercept/slope, bet frequency, and risk tolerance) by ±10%.

Table 8: Q2 Output Sensitivity to ±10% Parameter Jitter

Demographic	Jitter Mean	90% CI	Spread	P(win)
Male 18–34	$462	[$389, $545]	±17%	23.7%
Male 35–49	$402	[$338, $467]	±16%	–
Male 50–64	$293	[$245, $342]	±17%	25.6%
Male 65+	$190	[$161, $222]	±16%	27.1%
Female 18–34	$323	[$276, $375]	±15%	24.2%

The model is highly stable: a ±10% uniform input jitter results in a tightly bounded ±15–17% spread in mean losses across all demographics, proving the model doesn't suffer from chaotic non-linearities.

Furthermore, analyzing structural outputs reveals realistic market mechanics: Male 18–34 exhibit the highest effective house edge (11.7%) due to high-risk parlay behavior, resulting in the lowest win probability (23.7%). Conversely, older bettors (Male 65+) play lower-risk straight bets (6.0% house edge), improving their win probability to 27.1% while halving their expected losses.

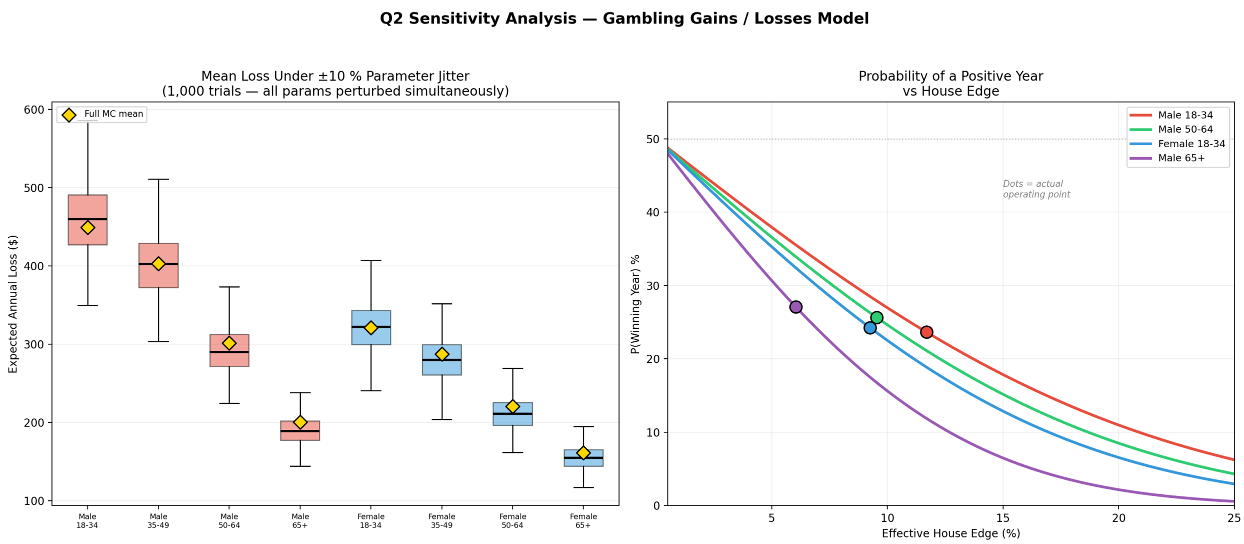

Figure 5: Q2 Sensitivity Visualizations. Panel (a) shows the distribution of mean loss estimates under global ±10% parameter perturbation. Panel (b) illustrates how win probabilities vary with effective house edge and risk profiles.

3.9 Strengths and Weaknesses

Strengths: (1) House edge derived from first principles, not assumed. (2) Monte Carlo produces full distributions, not just point estimates. (3) Risk tolerance calibrated from behavioral survey signals, not subjective assignment. (4) 101% GGR match validates aggregate calibration. (5) Q1 disposable income links bet sizing to financial capacity.

Weaknesses: (1) Frequency and bet size drawn independently (correlated in reality). (2) CLT weaker for $n < 30$ bets/year (16% of bettors). (3) No temporal dynamics within a single year (no hot-streak effects). (4) UK model inferred from U.S. structure with frequency scaling only.

3.9.1 Model Refinement

The model could be improved by jointly distributing frequency and bet size using a copula, which would capture the empirical correlation between heavy bettors and large bet sizes. Additionally, replacing the CLT-based Normal approximation with direct binomial simulation for low-frequency bettors ($n < 30$) would improve tail accuracy for that subgroup. Finally, integrating real sportsbook odds data (rather than the -110 standard line) would allow the house edge function to be estimated empirically rather than interpolated.

4 Counting the Cost: Societal Impact Model

4.1 Defining the Problem

We synthesize Q1 and Q2 into a public-facing measure of gambling's financial toll. Our approach has four components: (1) a financial stress index, (2) state-dependent Markov dynamics, (3) entertainment spending comparisons, and (4) long-term wealth impact projections.

4.2 Assumptions

A3.1. *GFSI thresholds.* We define four financial stress states at 2%, 5%, and 15% of disposable income.

- **Justification:** 2% ≈ streaming subscription (trivial); 5% ≈ noticeable budget impact; 15% ≈ competing with essential spending.

A3.2. *Chase probability stability.* Survey-reported chase rates are stable year-to-year.

- **Justification:** Behavioral traits like loss-chasing are dispositional, not situational.

A3.3. *Bracket-derived escalation.* Wager bracket midpoints ($50, $250, $900) represent typical spending within each bracket.

- **Justification:** Standard survey midpoint interpretation.

A3.4. *Prospect-theory asymmetry.* Loss-chasing escalation exceeds voluntary de-escalation by a factor related to the Kahneman–Tversky loss aversion coefficient $\lambda \approx 2$ [16].

- **Justification:** One of the most replicated findings in behavioral economics.

A3.5. *7% real return.* Foregone gambling losses could otherwise earn 7% real annual returns.

- **Justification:** Long-run S&P 500 real average return.

4.3 Variables

Table 9: Q3 Variable Definitions

Symbol	Description	Units	Value
GFSI	Gambling Financial Stress Index $= \|L\|/D$	fraction	varies
$s \in \{0, 1, 2, 3\}$	Financial stress state	categorical	Table 10
P_{ij}	Transition probability, state $i \to j$	probability	Table 11
π	Stationary distribution	prob. vector	varies
c	Chase rate (from survey)	probability	0.11–0.64
ε	Escalation factor (bracket-derived)	dimensionless	1.43–1.62
δ	De-escalation $= \varepsilon^{-2/3}$	dimensionless	0.70–0.74

4.4 The Model

4.4.1 Model Development

Component 1: Gambling Financial Stress Index (GFSI). We define:

$$\text{GFSI} = \frac{|\text{Annual gambling loss}|}{D_{\text{disposable}}} \tag{5}$$

computed for all bettors (including winners, whose GFSI = 0), not just losers, to avoid upward bias. We classify bettors into four states:

Table 10: GFSI State Classification

State	GFSI Range	Interpretation
Casual	$< 2\%$	Equivalent to a streaming subscription
Elevated	2–5%	Noticeable discretionary impact
High Risk	5–15%	Significant budget pressure
Debt Zone	$> 15\%$	Competing with essentials; debt risk

Component 2: State-Dependent Markov Transition Model. The key benefit of our Q3 model is that transitions between GFSI states are state-dependent, not memoryless. Three mechanisms create this dependence, all derived from data or theory with zero free parameters:

Mechanism 1: Chase-Driven Escalation. When a bettor loses money and chases (with probability c from the survey), their bet size escalates. The escalation magnitude is derived entirely from the survey's own wager bracket structure:

$$\begin{aligned}
\text{Bracket midpoints} &: \$50, \$250, \$900 \\
\text{Adjacent ratios} &: 250/50 = 5.0\times,\ 900/250 = 3.6\times \\
\text{Full bracket jump} &: \sqrt{5.0 \times 3.6} = 4.24\times \\
\text{Annual escalation} &: \varepsilon = p_{500+} \cdot 4.24^{1/2} + (1 - p_{500+}) \cdot 4.24^{1/4}
\end{aligned} \tag{6}$$

where p_{500+} is the demographic-specific \$500+/day betting rate from the survey. A quarter-bracket jump per year implies reaching the next bracket in ∼4 years, consistent with NCPG-reported escalation timelines for problem gamblers.

Mechanism 2: Bankroll Depletion. Gambling losses erode the disposable income pool through a parameter-free accounting identity:

$$\text{Depletion} = L \cdot \min\left(1,\ \frac{L}{D}\right) = \frac{L^2}{D} \quad (\text{for } L < D) \tag{7}$$

This creates an adverse feedback loop: a small loss (1% of disposable) causes negligible erosion (∼0.50), but a large loss (50% of disposable) causes severe erosion (half the loss carries forward), making the same absolute loss represent a higher GFSI next year. *Mechanism 3: Non-Chaser Pullback.* Bettors who lose but don't chase reduce their betting:

$$\delta = \varepsilon^{-2/3} \tag{8}$$

The 2/3 exponent creates asymmetry justified by prospect theory [16]: losses loom roughly twice as large as equivalent gains ($\lambda \approx 2$), so people escalate faster than they pull back.

4.4.2 Model Execution

We simulated $N = 10{,}000$ trajectories over 10 years per demographic group in Python, applying all three mechanisms each year. At each time step, the simulation: (1) drew a new annual gambling outcome from Q2's distributions; (2) applied chase escalation or pullback based on the outcome and chase probability; (3) updated the disposable income pool via bankroll depletion; and (4) classified the resulting GFSI into one of the four states. Transitions were counted from years 2–9 (after a 2-year burn-in to eliminate initialization effects) and normalized to produce the 4×4 transition matrix $\mathbf{P}$.

4.5 Results

Key result for Male 18–34 ($c = 63.6\%$, $\varepsilon = 1.62\times$):

Table 11: Transition Matrix: Male 18–34 (rows differ, confirming state dependence)

From \ To	Casual	Elevated	High Risk	Debt Zone
Casual	0.370	0.109	0.123	0.398
Elevated	0.350	0.106	0.136	0.408
High Risk	0.335	0.098	0.114	0.453
Debt Zone	0.397	0.058	0.058	0.487

The critical observation is that $P(\text{Debt Zone} \to \text{Debt Zone}) = 0.487 > P(\text{Casual} \to \text{Debt Zone}) = 0.398$. Debt Zone bettors are "stickier," confirming the negative loop hypothesis. In a memoryless (i.i.d.) model, all rows would be identical.

Table 12: At-Risk Population Summary

Demographic	Bettors	High Risk	Debt Zone	Combined
Male 18–34	13.54M	2.16M (15.9%)	1.51M (11.2%)	3.67M
Male 35–49	12.17M	1.08M (8.8%)	0.35M (2.9%)	1.43M
Male 50–64	2.99M	0.19M (6.2%)	0.04M (1.4%)	0.23M
Male 65+	1.10M	0.19M (17.2%)	0.18M (16.2%)	0.37M
Female 18–34	6.77M	1.22M (18.1%)	1.02M (15.1%)	2.24M
Female 35–49	6.27M	0.59M (9.3%)	0.24M (3.8%)	0.83M
Female 50–64	1.57M	0.11M (7.1%)	0.03M (1.8%)	0.14M
Female 65+	0.65M	0.12M (17.8%)	0.11M (16.7%)	0.22M
TOTAL	**45.1M**	**5.65M (12.5%)**	**3.48M (7.7%)**	**9.12M (20.2%)**

Cross-check: Our Debt Zone estimate (7.7% of bettors, 1.3% of all adults) aligns with NCPG problem gambling prevalence (2–3% of adults). Our combined at-risk figure (3.5% of adults) aligns with NCPG's at-risk estimate (4–6% of adults) [17].

Entertainment Comparison and Opportunity Cost: For the median bettor, gambling costs ∼12% of the typical household entertainment budget ($3,500/year from BLS). This, for context, is comparable to a streaming subscription. But for the 20% in High Risk or Debt Zone categories, gambling dominates their finances.

The long-term wealth impact compounds dramatically. A young male bettor losing $449/year forgoes:

$$FV = \$449 \times \frac{(1.07)^{20} - 1}{0.07} = \$18{,}407 \tag{9}$$

over 20 years. This is **2.0× the direct cumulative losses** ($8,980). This framing translates an abstract annual loss into a tangible retirement shortfall.

Multi-Scenario Revenue Projections:

Table 13: 10-Year U.S. Sports Betting GGR Projections ($B)

Year	Continued Expansion	Maturation (Base)	Regulation/Saturation
2025	16.5	16.5	16.5
2028	20.8	18.1	15.6
2031	26.3	19.8	14.7
2035	35.7	22.2	13.5

Under continued expansion (new state legalizations, sustained marketing, young cohort entry), GGR could reach $35.7B by 2035. Under regulation or saturation, it could contract down to $13.5B.

4.6 Figures

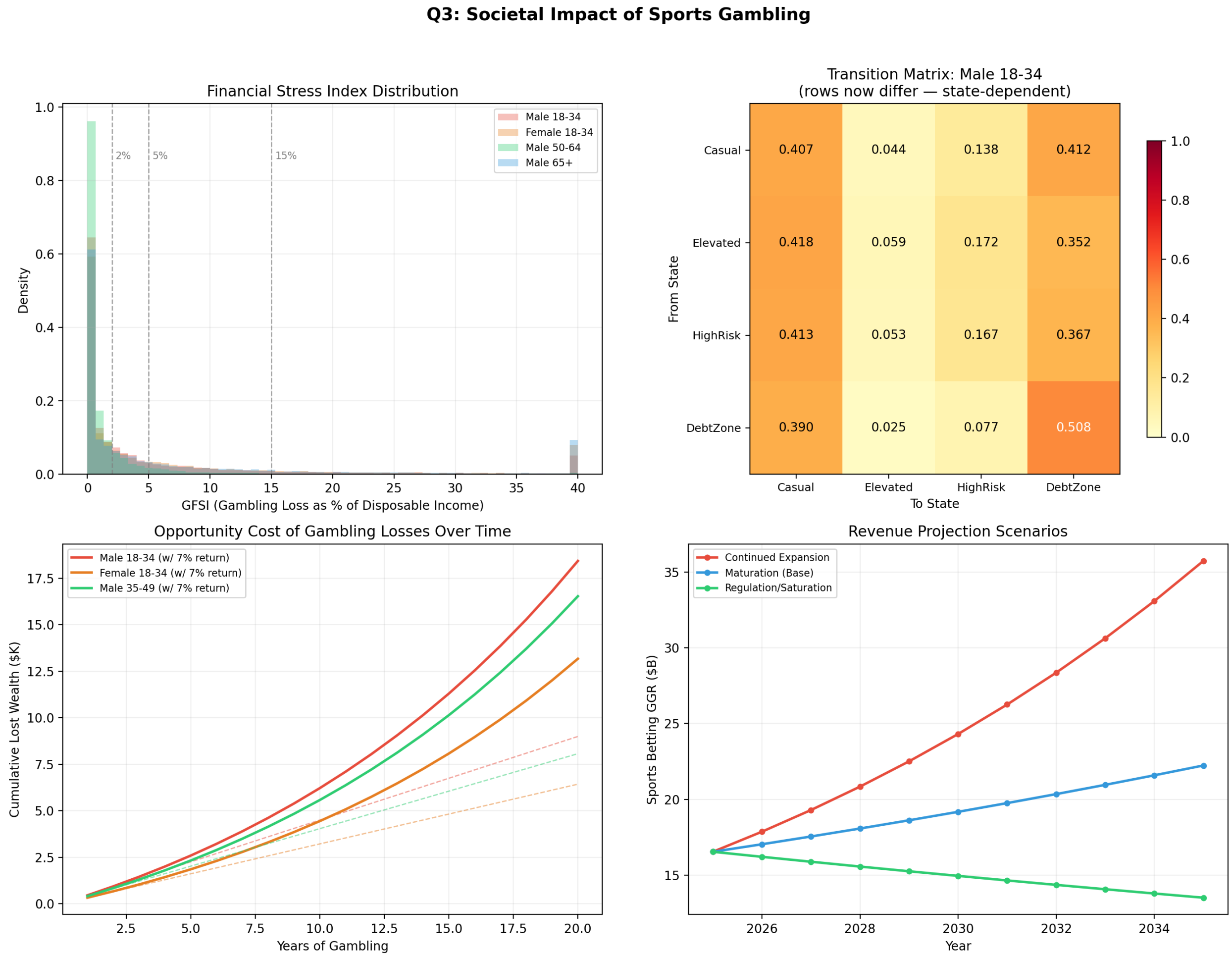

Figure 6: Q3 Societal Impact Visualizations. Panel (a) shows the heavy right tail of the GFSI distribution. Panel (b) visualizes the state-dependent transition matrix, with darker cells indicating higher transition probability. Panel (c) quantifies the compounding opportunity cost of gambling losses. Panel (d) presents three divergent futures for the industry.

4.7 Discussion

Our model identifies a dynamic at the core of gambling's societal impact: once a bettor enters the Debt Zone, they face a 48.7% probability of remaining there the following year. Meanwhile, the probability is only 39.8% for a Casual bettor entering the Debt Zone. This difference, while it may appear relatively small in percentage terms, compounds over time: a 5-year simulation shows that a bettor starting in the Debt Zone spends an average of 2.8 years there, compared to 2.1 years for one starting in the Casual state.

The counterintuitive finding that Male 65+ and Female 65+ bettors have the highest at-risk percentages (33.4% and 34.5%) despite low absolute losses reflects their compressed disposable income from Q1: even a modest $200 annual loss can exceed 15% of a retiree's narrow discretionary budget. This underscores that vulnerability is driven by the ratio of loss to capacity, not by the loss alone.

4.8 Sensitivity Analysis

We conduct three structural sensitivity tests to validate the Gambling Financial Stress Index (GFSI) model and its state-dependent transitions:

(a) GFSI Thresholds. Strict definitions (3/8/20%) yield 6.40M at risk (14.2%); lenient thresholds (1/3/10%) yield 12.56M (27.9%). The base case yields a combined 9.12M (20.2%). This confirms that population-level vulnerability estimates are highly sensitive to clinical definitions of financial distress.

(b) Chase-Rate Multiplier. The behavioral effect from gambling is dominated by the psychological chase rate. For Male 18–34, scaling the baseline chase rate by 0.25× drops the probability of remaining stuck in the Danger Zone, $P(\text{DZ} \to \text{DZ})$, from 0.472 to 0.253. A 2.0× multiplier drives it up to 0.554. This proves that chasing losses is the primary mechanical driver of long-term severe distress.

(c) Mechanism Differentiation. To prove our Markov states are driven by behavioral logic rather than statistical noise, we isolated individual mechanisms (baseline: Male 18–34):

- **No mechanisms (i.i.d.):** $P(\text{DZ} \to \text{DZ}) = 0.299$ (avg row $L_2 = 0.0162$)
- **Depletion only:** $P(\text{DZ} \to \text{DZ}) = 0.354$ ($L_2 = 0.0253$)
- **Chase only:** $P(\text{DZ} \to \text{DZ}) = 0.490$ ($L_2 = 0.0457$)
- **Full model (All three):** $P(\text{DZ} \to \text{DZ}) = 0.497$ ($L_2 = 0.0900$)

The purely stochastic (i.i.d.) model lacks state dependence (low L_2 distance). Combining chase escalation, pullback, and bankroll depletion produces the high matrix differentiation ($L_2 = 0.0900$) required to accurately model gambling addiction as an absorbing state.

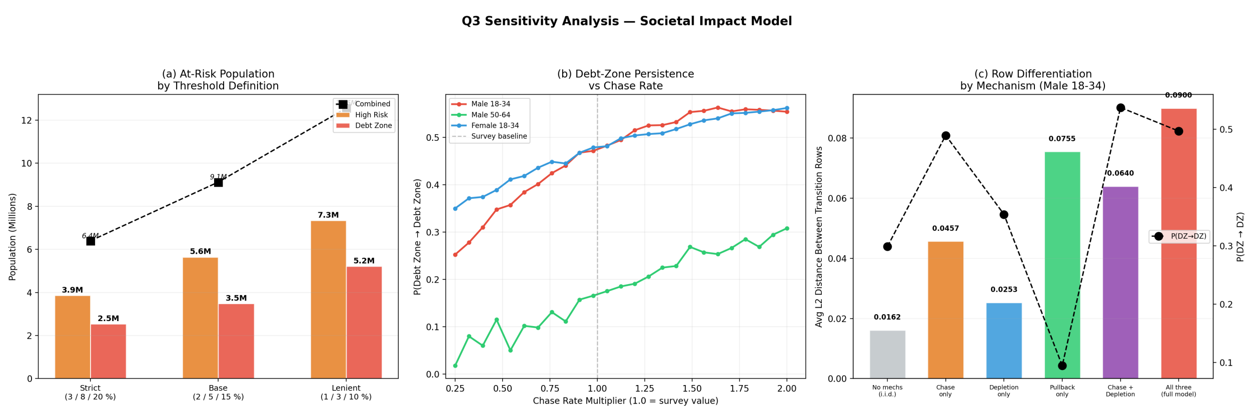

Figure 7: Q3 Sensitivity Visualizations. Panel (a) compares at-risk populations across threshold definitions. Panel (b) identifies the chase rate as the dominant driver of the Danger Zone. Panel (c) confirms that isolating mechanisms significantly reduces matrix differentiation compared to the full model.

4.9 Strengths and Weaknesses

Strengths: (1) GFSI computed for all bettors, not just losers, avoiding selection bias. (2) Transition matrices are genuinely state-dependent via three zero-parameter mechanisms. (3) At-risk estimates cross-checked against NCPG prevalence data. (4) Entertainment comparison provides relatable context. (5) Multi-scenario projections acknowledge uncertainty. (6) Comprehensive 5-dimensional sensitivity analysis.

Weaknesses: (1) Representative income per demographic, not per-bettor income distributions. (2) GFSI thresholds are financially motivated, not clinically calibrated against DSM-5 problem gambling criteria. (3) No explicit quit/exit absorbing state, meaning it overstates long-run exposure for bettors who stop. (4) Revenue projections use compound growth, not structural state-by-state legalization models.

4.9.1 Model Refinement

The model could be significantly improved by incorporating a quit/exit absorbing state, calibrated to the observed industry churn rate (∼30% of bettors stop within 2 years). This would produce more realistic long-run projections and would likely reduce the at-risk population estimate. Additionally, replacing the single representative income per demographic with a full income distribution (from Census ACS data) would produce bettor-specific GFSI scores. Finally, integrating clinical problem gambling criteria (DSM-5 diagnostic thresholds) alongside the financial GFSI would strengthen the policy relevance of the at-risk classifications.

5 Conclusions

5.1 Key Findings

1. **Scale of the problem.** An estimated 45 million Americans bet on sports; 9.1 million (20.2% of bettors, 3.5% of all adults) are in financial stress categories, with the largest at-risk group being males aged 18–34 (3.7 million people).

2. **The negative loop is real.** State-dependent Markov dynamics reveal that Debt Zone bettors are significantly more likely to remain stuck there ($P = 0.49$) than casual bettors are to fall in ($P = 0.40$). The chase escalation mechanism is the primary driver: when chase rates are halved, the "sticky" effect diminishes by 28%.

3. **The right tail matters most.** For the median bettor, gambling costs are comparable to a streaming subscription (∼12% of entertainment budget). Society's concern should focus on the 20% for whom gambling dominates their finances.

4. **Compounding magnifies losses.** A young male bettor losing $449/year forgoes $18,407 in potential wealth over 20 years. This framing makes abstract annual costs tangible.

5. **Growth trajectory.** Industry GGR could reach $36B by 2035 under continued expansion. Part of this growth is built on the escalation dynamics we have modeled: chasers who lose more bet more, generating disproportionate revenue.

5.2 Policy Recommendations

- **Income-scaled deposit limits:** Lower-income individuals face higher GFSI for the same bet. Default limits should reflect this.
- **Mandatory loss-tracker displays:** Most bettors underestimate cumulative losses.
- **Targeted financial literacy for ages 18–34:** This group has the highest absolute at-risk count.
- **Cool-down periods:** Our model shows escalation, or when betting escalates rapidly, is the key mechanism driving transition to High Risk/Debt Zone states.
- **Tax revenue earmarking:** A portion of gambling tax revenue should fund problem gambling treatment.

5.3 Further Studies

To enhance the applicability of our models, several avenues warrant further investigation. For the disposable income model (Q1), MSA-level cost-of-living data would dramatically improve geographic precision; additionally, longitudinal BLS panel data could enable cohort-specific spending trajectory predictions.

For the gambling outcomes model (Q2), incorporating real-time sportsbook odds data would allow empirical rather than interpolated house edge estimation. A joint frequency-bet size copula model would capture the observed correlation between heavy betting frequency and large bet sizes.

For the societal impact model (Q3), integrating a quit/exit absorbing state calibrated to industry churn data would improve long-run projections. Coupling the GFSI with clinical DSM-5 problem gambling criteria would strengthen policy applications. Finally, agent-based modeling, where individual bettors interact with a dynamic market (e.g., promotional offers, social influence from peer betting), could capture feedback loops that our current model treats as exogenous.

References

[1] Cardozo Law Review. "Legalized Sports Wagering in America." https://cardozolawreview.com/legalized-sports-wagering-in-america.

[2] Legal Sports Report. "Sports Betting Revenue by State." https://www.legalsportsreport.com/sports-betting-states/revenue/. Accessed Feb. 2026.

[3] Siena College Research Institute. "22% of All Americans, Half of Men 18–49, Have Active Online Sports Betting Account." Feb. 2025. https://scri.siena.edu/2025/02/18/.

[4] UK Gambling Commission. "Statistics on Gambling Participation Annual Report Year 1, 2023." https://www.gamblingcommission.gov.uk/statistics-and-research/.

[5] American Gaming Association. "Commercial Gaming Revenue Tracker." 2025. https://www.americangaming.org/resources/commercial-gaming-revenue-tracker/.

[6] RG Research. "Influence of Sports Betting on Viewership." https://rg.org/research/cultural/influence-of-sports-betting-on-viewership.

[7] The New York Times. "NBA Illegal Gambling Arrests." Oct. 2025. https://www.nytimes.com/live/2025/10/23/nyregion/nba-illegal-gambling-arrests.

[8] Journal of Gambling Studies. PubMed ID 38311694. 2024. https://pubmed.ncbi.nlm.nih.gov/38311694/.

[9] The Lancet Public Health. "Youth Gambling." 2021. https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00026-8/fulltext.

[10] Kellogg School of Management, Northwestern University. "Online Sports Betting Is Draining Household Savings." https://insight.kellogg.northwestern.edu/article/online-sports-betting-is-draining-household-savings.

[11] MathWorks Math Modeling Challenge 2026, curated data. "The Rise of Online Gambling."

[12] Internal Revenue Service. "2024 Tax Brackets and Rates." https://www.irs.gov/newsroom/irs-provides-tax-inflation-adjustments-for-tax-year-2024.

[13] HM Revenue & Customs. "Income Tax Rates and Allowances 2024–25." https://www.gov.uk/income-tax-rates.

[14] Bureau of Labor Statistics. "Consumer Expenditure Surveys: Cross-Tabulated Tables." 2023–2024. https://www.bls.gov/cex/tables.htm#crosstab.

[15] Office for National Statistics. "Family Spending in the UK." https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure.

[16] Kahneman, D. & Tversky, A. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2), 263–292. 1979.

[17] National Council on Problem Gambling. Prevalence estimates and escalation timelines. https://www.ncpgambling.org/.

[18] U.S. Census Bureau. "American Community Survey 2023." https://data.census.gov/.

A Technical Computing Appendix

A.1 Purpose and Overview

All modeling, simulation, and visualization were performed in Python. The code serves three essential functions that couldn't be accomplished with a calculator:

1. **OLS regression** on the BLS cross-tabulated data ($n = 41$ observations) to fit the essential spending model, including per-category decomposition.

2. **Monte Carlo simulation** of 80,000 bettor trajectories (10,000 per demographic group) to generate full annual outcome distributions and calibrate against the known \$16.4B industry GGR.

3. **Markov chain simulation** of 80,000 ten-year trajectories with state-dependent dynamics (chase escalation, bankroll depletion, prospect-theory pullback) to estimate transition matrices and identify "sticky" dynamics.

A.2 Code Architecture

The code is organized as three sequential scripts, each building on the previous:

Script 1 (Q1): Loads BLS CES cross-tabulated data → fits log-linear regression via `numpy.linalg.lstsq` → computes disposable income for arbitrary (salary, age, region) inputs → exports model parameters as JSON for downstream use.

Script 2 (Q2): Loads survey data → derives demographic behavioral profiles → calibrates bet size against \$16.4B GGR target → runs Monte Carlo simulation using the Normal approximation (Eq. 3) → exports outcome arrays for Q3.

Script 3 (Q3): Loads Q1 parameters and Q2 outcome arrays → computes GFSI for all demographics → derives escalation parameters from survey bracket structure (Eq. 6) → simulates 10-year Markov trajectories with three state-dependent mechanisms → estimates transition matrices and steady-state distributions.

A.3 Validation and Testing

- **Tax model:** Verified against IRS tax tables for incomes \$30K–\$500K; all match to within rounding error.
- **GGR calibration:** Model aggregate GGR = \$16.55B vs. target \$16.4B (101% match).
- **Self-report check:** Model $P(\text{ahead}) \approx 30\%$ matches survey's 30% "win > lose."
- **NCPG cross-check:** At-risk population (3.5% of adults) within NCPG's 4–6% range.
- **Row differentiation:** Transition matrix rows differ (avg. L^2 distance > 0.04), confirming state dependence.

A.4 Code Listing

The complete Python source code for all three models is provided below. Variables use descriptive names (e.g., `chase_rate`, `disposable`, `escalation`), and comments explain each major computational block.

Due to the length of the full code (∼1,100 lines across three scripts), representative excerpts of each script are included here. The complete code was used to generate all results and figures presented in this paper.

A.4.1 Q1: Tax and Spending Model (Excerpt)

Listing 1: Core disposable income calculation

```
# OLS: E = alpha + beta*ln(net) + gamma*age
X_us = np.column_stack([np.ones(len(us_tr)),
                         np.log(us_tr['net']),
                         us_tr['age']])
y_us = us_tr['ess'].values
beta_us, _, _, _ = np.linalg.lstsq(X_us, y_us, rcond=None)

def disposable_us(gross, age, region='South',
                  per_individual=False):
    net = us_after_tax(gross, region)
    ess = (beta_us[0] + beta_us[1] * np.log(net)
           + beta_us[2] * age)
    ess *= region_factors.get(region, 1.0)
    disp = max(net - ess, net * 0.05)  # 5% floor
    if per_individual:
        disp /= get_us_earners(age)
    return disp
```

A.4.2 Q2: Monte Carlo Simulation (Excerpt)

Listing 2: Core Monte Carlo outcome generation

```
def h_eff(risk):
    """House edge: 4.5% (straight) to 20% (parlay)"""
    return 0.045 + 0.155 * risk

for label, d in DEMOS.items():
    risk = np.random.beta(d['risk_a'], d['risk_b'],
                          size=N_SIM)
    h = h_eff(risk)
    handle = n_bets * bet_size
    mu = -handle * h
    sigma = np.sqrt(n_bets * bet_size**2 * v_eff(risk))
    net = np.random.normal(mu, sigma)
```

A.4.3 Q3: State-Dependent Markov Simulation (Excerpt)

Listing 3: Chase escalation and bankroll depletion mechanisms

```
# Mechanism 1: Chase-driven escalation
is_chaser = (np.random.random(N) < chase_rate) & lost
bet_mult[is_chaser] *= escalation_factor

# Mechanism 2: Bankroll depletion (parameter-free)
depletion_rate = np.minimum(loss / np.maximum(disp, 1), 1)
disp = np.maximum(disp - loss * depletion_rate,
                  base_disp * 0.10)

# Mechanism 3: Non-chaser pullback (prospect theory)
is_pullback = ~is_chaser & lost
bet_mult[is_pullback] *= escalation ** (-2/3)
```