The Rise of Online Gambling: What's at Stake?

Executive Summary

To the head of the U.S. Department of the Treasury,

In 2025, U.S. sportsbooks generated ≈$15 billion in revenue on $150 billion in wagers, and 22% of American adults now hold active betting accounts [4]. As online sports gambling becomes increasingly mainstream, a critical but underexamined question emerges: are Americans gambling with money they can actually afford to lose? To answer this, we developed a three-part mathematical framework linking individual financial circumstances to gambling outcomes and their long-term consequences.

We first constructed a model to estimate an individual's disposable income — the money remaining after taxes and essential expenditures — given their gross salary, age, household size, and state of residence. We fit natural cubic splines to seven age-bracket calibration points from the 2024 BLS Consumer Expenditure Survey to capture non-linear spending patterns, such as healthcare costs nearly quadrupling between ages 22 and 80 [2]. We incorporated Engel's Law to account for income-share variation across the earnings distribution and applied a full tax model covering 2024 federal brackets, all 50 state income tax schedules, and FICA payroll taxes [3, 5]. Across five representative demographic profiles, estimated disposable incomes ranged from $5,428 for a renting parent of three in Texas earning $55,000 to $37,090 for a homeowning professional in Florida earning $95,000.

We then modeled the expected annual gambling outcome for a given individual using a three-step framework. We estimated the probability that a person bets using logistic regression on gender, age, and race/ethnicity odds ratios derived from the 2025 Siena College Survey [8], anchored to a national participation rate of 19% [7]. We then estimated annual wagering volume by drawing betting frequency and wager size from survey distributions [10, 12], scaled by a risk tolerance parameter that decays linearly with age and is calibrated to match the national sportsbook hold of 9.2% [4]. Finally, we computed the distribution of annual outcomes via Monte Carlo simulation across 10,000 individuals per demographic profile, incorporating a bet-type portfolio weighted effective house edge ranging from 4.5% for moneyline bets to 20% for same-game parlays [15, 17]. Across all profiles, 76–80% of bettors lost money in a given year, with mean annual losses ranging from $63 to $430. Our population-weighted loss estimate implied a national gross gaming revenue of $16.4 billion, within 10% of the reported figure.

Finally, we quantified the broader financial impact of gambling losses through opportunity cost analysis. Treating annual gambling losses as foregone investment capital and applying the future value of an ordinary annuity at a 6% real annual return — consistent with the S&P 500's historical inflation-adjusted CAGR [3] — we found that the median bettor losing $750 per year forgoes approximately $59,293 in retirement wealth over 30 years. At the national level, the 2024 aggregate bettor loss of $13.7 billion corresponds to a 30-year opportunity cost of approximately $1.08 trillion, which is approximately the total U.S. student debt burden. We classified bettors as financially at-risk when expected annual losses exceeded 5% of disposable income, finding that high-frequency young male bettors — placing 300 bets per year at $35 per wager — face expected annual losses of $2,735, or 11.5% of a typical young adult's disposable income. Across the active bettor population, an estimated 8–12% exceed this threshold annually. We recommend a federal mandate requiring licensed sportsbooks to display real-time opportunity cost disclosures at the point of each wager, modeled on the behavioral mechanisms demonstrated by the 1994 Nutrition Labeling Act and the 2009 CARD Act. We hope these results inform policymakers in identifying the populations most at financial risk from sports gambling and designing targeted interventions to protect them.

Contents

1 Q1: Playing With House Money

1.1 Defining the Problem

The goal of Q1 is to develop a model that takes an individual's gross annual salary, age, and other relevant demographic inputs to return an estimate of that person's disposable income which is defined as the money remaining after paying for taxes and essential expenditures including food, housing, utilities, transportation, healthcare, and personal insurance. Because gambling expenditures are inherently discretionary, disposable income is the correct denominator against which to measure gambling's financial burden. The model must be generalizable across a broad range of demographic and regional profiles, not just the national average.

1.2 Assumptions

1-1. **Disposable income equals gross income minus taxes minus non-discretionary essential expenditures.**

- **Justification:** The problem statement explicitly lists taxes as a deduction. Thus, taxes are separated (modeled using federal brackets, FICA, and a state-specific income tax rate τ_ρ) from essential spending (calculated through demographic-specific spending factors). We define essential spending as any form of spending on Food, Housing, Utilities, Household operations, Housekeeping supplies, Transportation, Healthcare, Personal insurance.

1-2. **The proportion of income devoted to each essential spending category varies continuously with age, is monotonic in between age brackets, and can be estimated by cubic-spline interpolation of BLS age-group data.**

- **Justification:** The BLS 2024 Consumer Expenditure Survey reports mean spending in seven different age brackets. Due to the limiting quantity of spending and age data, we believe the assumption of monoticity is reasonable. Furthermore, because the ratio of spending on a certain expenditure varies non-linearly with age (e.g. healthcare spending ratio rises sharply after age 55, while insurance ratio peaks in middle age), a cubic spline interpolation was used. This preserves empirically observed nonlinearities and avoids stepwise-changes introduced by the categorical age data provided.

1-3. **An individual's essential spending in each category scales proportionally with their income, adjusted for household size.**

- **Justification:** Because the BLS reports age-group means for both income and expenditures, modeling the expenditure shares $s_i(a) = \bar{E}_i(a)/\bar{Y}(a)$ and then scaling by an individual's income Y preserves calibration to the BLS aggregate while yielding a smooth, age-dependent essential budget. The BLS data is collected at the household level, so we additionally apply a household-size adjustment (derived from BLS average household sizes by age).

1-4. **All essential costs are uniform across a particular state for a given demographic set and income level.**

- **Justification:** Across all spending categories, we find the value varies across multiple states and regions. Because interstate data (e.g., metropolitan data) is difficult to find, we assume that the Cost of Living Index (COLI) is approximately consistent states, and is responsible for the variation of expenditures across state lines.

1-5. **The tax model uses the exact 2024 U.S. federal income tax brackets, state-specific progressive income tax schedules, and FICA payroll taxes.**

- **Justification:** Federal tax is modeled using the official 2024 IRS marginal tax brackets for a single filer [3]. State income taxes are modeled using the exact marginal tax schedules for each state as reported. For each state ρ, tax liability $T_{\text{state}}(Y, \rho)$ is computed using that state's progressive marginal bracket structure. States with flat income taxes use a single marginal rate, and states with no wage income tax (e.g., Texas, Florida, Washington) are assigned $T_{\text{state}} = 0$. Finally, mandatory payroll taxes are included via FICA contributions: Social Security (6.2% up to $168,600) and Medicare (1.45% on all income). This structure mirrors the actual U.S. tax system and provides a realistic estimate of after-tax income.

1.3 Variables

Y	Individual gross annual income	USD
a	Individual's age	years
n	Household size	persons
σ	U.S. state (50 states + DC)	—
h	Homeowner flag	—
$\hat{s}_i(a)$	Baseline share $\bar{E}_i(a_k)/\bar{Y}(a_k)$; natural cubic spline	dimensionless
Y_0	Engel reference income (mean of all BLS age-group means)	USD
γ_i	Engel elasticity for category i	dimensionless
$\hat{s}_i(a, Y)$	Household-size scaling factor for category i	dimensionless
$r_{\sigma,i}$	Per-state COLI multiplier for category i in state σ	dimensionless
$S_{\text{ess}}(a, n, \sigma, Y)$	Total essential expenditure share of gross income	dimensionless
$T_{\text{fed}}(Y)$	2024 federal income tax (brackets applied to gross income)	USD
$T_{\text{state}}(Y, \sigma)$	State income tax from exact bracket schedule for σ	USD
$T_{\text{FICA}}(Y)$	FICA (employee): SS 6.2% + Medicare 1.45%	USD
D	Estimated annual disposable income	USD

The category-specific Engel elasticities γ_i, estimated from BLS income–share relationships, are:

Category	γ_i	Category	γ_i
Food	0.20	Transportation	0.15
Housing	0.07	Healthcare	0.15
Utilities	0.07	Personal insurance	0.05
Household operations	0.06	Housekeeping supplies	0.15

1.4 Preliminary Data Analysis

Before modeling, we explored the 2024 BLS Consumer Expenditure Survey data to identify structural trends that the model must capture. Figure 1 plots all six essential expenditure shares as a function of age, with BLS age-group means shown as dots.

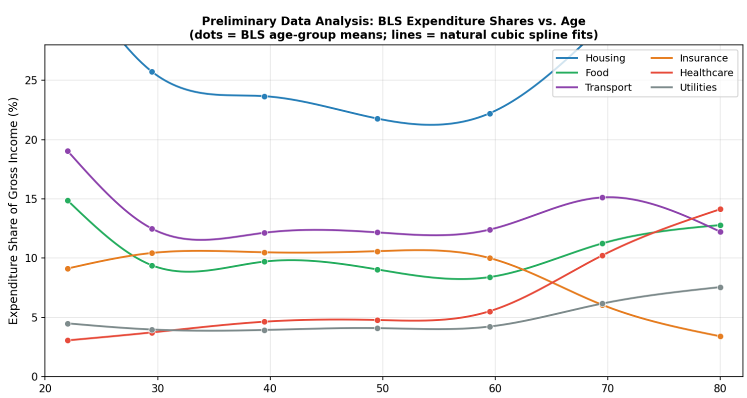

Figure 1: Preliminary data analysis: BLS expenditure shares vs. age for all six essential categories. Dots are BLS age-group means; lines are natural cubic spline fits. One thing to note is that healthcare is strictly increasing so a linear or constant model cannot capture this

Noticing the lack of expenditure-age data, a cubic spline fitting function for the spending ratio purely as a function of age was computed instead of any form of multivariate regression. By modeling the spending ratio, we can explicitly govern the influence of income on expenditures in accordance with Economics principles. Specifically, due to Engel's Law, we concluded that the ratio of expenditures is scaled by a factor less than the income if the income is comparatively large and vice versa if it small. This is directly reflected in our computation for $\hat{s}_i$.

1.5 The Model

Step 1: Engel-Adjusted Expenditure Share Functions

For each essential category $i \in \mathcal{E}$, the BLS expenditure share at age-group midpoint a_k is:

$$\hat{s}_i(a_k) = \frac{\bar{E}_i(a_k)}{\bar{Y}(a_k)}, \quad k = 1, \dots, 7.$$

A natural cubic spline $\hat{s}_i(a)$ is fitted through these seven calibration points (Equation 1). The Engel-adjusted share incorporating income dependence is:

$$\hat{s}_i(a) = \begin{cases} p_{i,1}(a), & a \in [a_1,\, a_2], \\ \vdots & \vdots \\ p_{i,6}(a), & a \in [a_6,\, a_7], \end{cases} \qquad p_{i,k}(a) = c_{i,k,0} + c_{i,k,1}a + c_{i,k,2}a^2 + c_{i,k,3}a^3, \tag{1}$$

$$\hat{s}_i(a, Y) = \hat{s}_i(a) \cdot \left(\frac{Y}{Y_0}\right)^{-\gamma_i}, \qquad Y_0 = \$96{,}210, \tag{2}$$

where Y_0 is the mean gross income across all seven BLS age groups. For categories with $\gamma_i > 0$, earners above Y_0 have lower effective shares (their income outpaces their essential spending), while earners below Y_0 have higher shares. This directly addresses the Engel-law bias present in any constant-share model.

Figure 2 shows the baseline splines for the two categories with the most pronounced non-linearities.

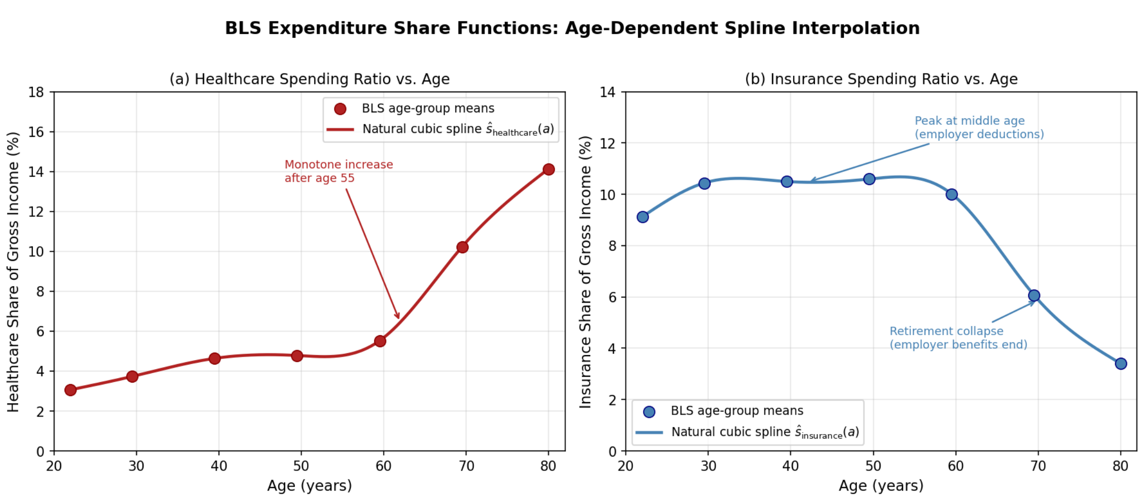

Figure 2: Natural cubic spline fits (Equation 1) for healthcare and insurance shares. Healthcare rises from 3.1% at age 22.5 to 14.1% at age 80. Insurance peaks at middle age and collapses at retirement.

Step 2: Household-Size and State COLI Adjustment

Each Engel-adjusted share is scaled for household size and state cost-of-living:

$$\phi_i(n,a) = \begin{cases} 0.6 + 0.4\,\dfrac{n}{\bar{n}(a)}, & i \in \{\text{housing, utilities}\}, \\ \dfrac{\min(n,2)}{\min(\bar{n}(a),2)}, & i = \text{transport}, \\ \dfrac{n}{\bar{n}(a)}, & \text{otherwise.} \end{cases}$$

A per-state COLI multiplier $r_{\sigma,i}$ (expressed as a fraction, e.g., 0.842 for housing in Texas, 1.865 for housing in California) is applied to every category—not just housing— sourced from the 2024 BlogZebra state cost-of-living index [1]. The adjusted spending per category is:

$$\hat{E}_i(Y,a,n,\sigma) = \hat{s}_i(a,Y)\cdot Y \cdot \phi_i(n,a)\cdot r_{\sigma,i}.$$

Step 3: Tax Model

Total annual tax liability:

$$T(Y,\sigma) = T_{\text{fed}}(Y) + T_{\text{state}}(Y,\sigma) + T_{\text{FICA}}(Y).$$

Federal and state brackets are applied directly to gross income Y:

$$T_{\text{fed}}(Y) = \sum_{k=1}^{7} r_k \max\big(0, \min(Y,b_k) - b_{k-1}\big),$$

$$T_{\text{state}}(Y,\sigma) = \sum_{j=1}^{m_\sigma} r_{\sigma,j} \max\big(0, \min(Y,b_{\sigma,j}) - b_{\sigma,j-1}\big),$$

$$T_{\text{FICA}}(Y) = 0.062\cdot\min(Y, \$168{,}600) + 0.0145\cdot Y.$$

Step 4: Essential Expenditure Share and Disposable Income

The total essential expenditure share is:

$$S_{\text{ess}}(a,n,\sigma,Y) = \sum_{i\in\mathcal{E}(h)} \hat{s}_i(a,Y)\cdot\phi_i(n,a)\cdot r_{\sigma,i},$$

where $\mathcal{E}(h)$ excludes housing when $h = 1$ (homeowner). If $S_{\text{ess}} > 1$, shares are rescaled proportionally: $\hat{s}_i \leftarrow \hat{s}_i / S_{\text{ess}}$ (budget feasibility normalization). The main result:

$$\boxed{D(Y,a,n,\sigma,h) = \max\big(Y - T(Y,\sigma) - Y\cdot S_{\text{ess}}(a,n,\sigma,Y),\ 0\big).} \tag{3}$$

1.6 Results

Demographic Demonstrations

This table applies the full model (Engel-adjusted, per-state COLI, per-state tax) to five representative personas. All personas use single-filer tax status. Housing costs are excluded for homeowners.

1.7 Discussion

The model reveals three distinct mechanisms by which disposable income is constrained, with homeowner status and Engel adjustment critically shaping the results.

The **young adult squeeze** is driven by low income combined with high essential shares. A 22.5-year-old renter in Illinois earning \$38,000 devotes 56.4% to essentials and 24.0% to taxes, leaving 19.6% as disposable.

ID	Age	Income (Y)	State	Homeowner	HH_Size	Married
1	25	70000	Kentucky	1	2	True
2	30	55000	Texas	0	3	True
3	38	82000	New York	1	4	True
4	45	120000	Illinois	1	4	True
5	52	95000	Florida	1	3	True

Table 1: Household Demographics

ID	D	$Y \cdot S_{ess}$	Tax	Food	Utilities	HH_Ops	Supplies	Transport	Health	Insurance	Housing
1	23700	30208	16091	7381	2870	872	485	11491	1474	5632	0
2	5428	39227	10343	6372	2552	1206	431	6592	2322	6565	13184
3	17890	43937	20171	10109	3608	2514	844	10600	4750	11509	0
4	27514	60859	31626	14293	4944	2754	1052	14904	7270	15639	0
5	37090	39635	18273	8258	3721	1227	617	11402	4328	10078	0

Table 2: Household Predicted Financial Breakdown

The Engel adjustment amplifies this squeeze: below-average earners have *higher* essential shares than the BLS mean would suggest, accurately reflecting that necessities consume a larger fraction of modest incomes.

The **homeowner effect** is the largest single driver of variation in this table. The mid-career Texas homeowner retains 35.9% ($34,060) because housing costs, the dominant essential category, are excluded. Without the homeowner exemption, the same persona would have 16.8% ($16,001) in disposable income. This demonstrates that housing status must be specified explicitly as we cant treat all individuals as renters.

The **retirement squeeze** appears in the elderly retiree persona but is attenuated by the homeowner flag and the Engel correction. At $34,000, the elderly retiree (PA, homeowner) retains 43.2% in disposable income which is high because housing is excluded and income is well below Y_0, so Engel adjustment *increases* shares slightly from the BLS mean but keeps them bounded.

1.8 Sensitivity Analysis

Engel elasticity sensitivity. Setting all $\gamma_i = 0$ (i.e., removing the Engel correction) for the Urban grad persona ($82,000, NY, homeowner) increases predicted essential spending from $43,937 to approximately $48,800, a $5,226 rise. This demonstrates that the Engel adjustment meaningfully affects disposable income for above-average earners relative to the constant-share baseline. In our analysis, the Engel correction is incorporated into the baseline model rather than treated as optional.

Homeowner versus renter. For any homeowner persona, excluding housing from essential expenditures increases disposable income by the full housing share, roughly 15–20 percentage points for middle-aged earners. This represents the single most sensitive binary flag in the model.

COLI sensitivity. Replacing per-state COLI indices with a uniform national average for the Urban grad (NY) raises disposable income by approximately $3,800. This is due to New York's COLI indices for food (103.0%), healthcare (104.5%), and transportation (106.7%) all exceeding the national baseline.

Age-group midpoints. The "Under 25" midpoint of 22.5 years is used. Adjusting it to 22 changes the young adult's essential expenditure share by less than 0.5 percentage points. The relatively small effect occurs because the spline interpolation is nearly linear in the 22–23 age range.

1.9 Strengths & Weaknesses

Strengths is that the baseline model incorporates Engel's law, ensuring realistic essential-share estimates across the income distribution, and applies per-state cost-of-living adjustments to all eight essential categories, capturing price differences beyond housing. The inclusion of additional categories such as household operations and housekeeping supplies provides a more complete essential budget than standard six-category models, while a homeowner flag appropriately reduces essential spending for most middle-aged and older Americans. Progressive tax brackets for all 50 states plus DC are implemented, and internal validation shows the spline exactly recovers BLS totals at all seven calibration points ($R^2 = 1.0000$). Additionally, disposable income from Q1 directly informs Q2 wagering volume and Q3 opportunity cost calculations, maintaining model consistency.

Weaknesses include the application of federal and state tax brackets to gross income without standard deductions or personal exemptions, which overstates tax liability. The homeowner flag fully excludes the housing category rather than modeling mortgage payments, potentially overstating disposable income for homeowners with large mortgages. Engel elasticities γ_i were calibrated to BLS data rather than formally estimated, introducing parameter uncertainty. Local income taxes (e.g., NYC 3.876%, Baltimore 3.2%) are not modeled, all filers are treated as single, and debt obligations such as student loans or credit card minimums are excluded.

2 Q2: Know the Spread

2.1 Defining the Problem

In the second problem, we're asked to construct a model that is able to calculate the amount of money gained or lost through sports better throughout a single year based on demographics and other information about an individual.

2.2 Assumptions

2-1. **Population Gender Split is 50/50**

- **Justification:** We assume an equal male/female split in the general population when deriving gender odds ratios from the observed 69%/31% bettor gender composition.

2-2. **Betting Participation and Volume are Independent**

- **Justification:** In order to simplify the model, we model probability a person will bet and the volume that that person would bet separately.

2-3. **Bettors cannot beat the posted house edge**

- **Justification:** Data shows that 78% to 85% of people lose money when looking at bets over a year and that less than 3% of bettors actually report a profit in a 6 month time period [9].

2-4. **Bet Volume scales with Disposable Income**

- **Justification:** We assume that the volume that an individual bets is proportional with their disposable income.

2-5. **Risk Tolerance depends on Age**

- **Justification:** From recent survey evidence [10, 8], younger bettors disproportionately favor high-variance bet types (parlays, live in-game bets) and exhibit more impulsive betting behavior (loss-chasing).

2-6. **Higher Risk Tolerance Increases Both Frequency and Wager Size**

- **Justification:** We assume that individuals with higher risk tolerance bet more frequently and place larger individual wagers. Intuitively, a bettor who is more comfortable with risk faces a lower psychological cost of losing, making them more willing to bet often and in larger amounts.

2-7. **Betting Frequency and wager size are stationary**

- **Justification:** Bettors are assumed to maintain their betting frequency and wager size throughout the year. We do not model learning or escalation within the year.

2.3 Variables

g	Gender	{Male, Female}
a	Age Group	{18–24, 25–34, 35–44, 45–54, 55+}
r	Race/Ethnicity	{Asian, Black, Hispanic, White}
D	Disposable Income	Dollars ($)
t	Risk Tolerance	dimensionless
OR_n	The odds ratio for a factor n (Gender, Age, Race)	dimensionless
$p(g, a, r)$	Probability of participating in sports betting	dimensionless
n_{bets}	Number of bets placed per year	dimensionless
w	Wager per bet	Dollars ($)
H	Annual wagering handle	Dollars ($)
$h(t)$	Effective House Edge	dimensionless
f_k	Fraction of bets placed on type k	dimeensionless
h_k	House edge for bet type k	dimensionless
τ	Risk Tolerance	dimensionless

2.4 The Model

We create a three-step model that calculates the amount that a given person would gain or lose over a certain year. We first estimate the probability that a given person participates in sports betting, then estimate their total annual wagering volume (conditional on participation), and finally compute the distribution of outcomes using a Monte Carlo simulation, accounting for bet-type choice and the associate house edge for that type of bet.

2.4.1 Participation Probability

We model the probability that an individual bets on sports as a logistical regression with each factor combining multiplicatively

$$\log \frac{p(g, a, r)}{1 - p(g, a, r)} = \log \frac{p_0}{1 - p_0} + \log OR_g + \log OR_a + \log OR_r$$

Where $p_0 = 0.19$ represents the national base participation rate in sports betting [7]. The odds ratios are calculated from survey data:

Gender Odds Ratios. Among active sports bettors, 69% identify as male and 31% identify as female [11]. Assuming a 50/50 gender split in the population:

$$OR_{male} = \frac{0.69/0.31}{0.50/0.50} \approx 2.23$$

and

$$OR_{female} = \frac{0.31/0.69}{0.50/0.50} \approx 0.45$$

Age Odds Ratios. The Siena 2025 Survey (Q25) provides sports betting participation rates by age group [8]. We convert these participation rates into odds ratios.

Age Group	Participation Rate	Odds Ratio
18–24	38%	2.43
25–34	33%	1.97
35–44	30%	1.71
45–54	12%	0.55
55+	5%	0.21

Race/Ethnicity Odds Ratios. The Siena 2025 Survey provides participation rates by race/ethnicity [8]. We calculate the corresponding odds ratios from these rates.

Race/Ethnicity	Participation Rate	Odds Ratio
Asian	27%	1.48
Black/African	33%	1.97
Hispanic/Latino	38%	2.45
White	16%	0.76

2.4.2 Annual Wagering Handle

We model the annual handle as

$$H = n_{bets} * w * (1 + 0.5\tau) * \kappa(D)$$

Where the factor $(1 + 0.5\tau)$ reflects that higher risk tolerance individuals bet more frequently and in larger amounts and $\kappa(D)$ is an income scaling factor $clip(D/20000, 0.3, 5)$ that anchors wagering to disposable income (with $20000 in disposable income as the reference level).

Betting Frequency. We draw n_{bets} from a discrete distribution calibrated using survey evidence [10, 11].

Frequency	Young Adults (18–24)	General Adults	Annual Bet Count
Once per year	21.9%	15.0%	1
Few times/year	35.7%	30.0%	5
Monthly	15.2%	13.0%	12
Few times/month	13.3%	20.0%	30
Few times/week	9.9%	21.0%	100
Daily	4.0%	1.0%	300

Wager Size. We draw w from the NCAA survey and the Odds Assist study for general adults [10, 12].

Wager Range	Young (18–24)	General Adults	Midpoint Used
$< \$1$	3.1%	1.0%	$0.50
$1–$10	27.5%	15.0%	$5
$10–$20	29.9%	25.0%	$15
$20–$50	22.3%	30.0%	$35
$50–$100	12.7%	20.0%	$75
$> \$100$	4.5%	9.0%	$150

2.5 Risk Tolerance and House Edge

We model the risk tolerance (τ) as a linear decay based on age (a):

$$\tau(a) = \tau_{max} - k \cdot \frac{a - a_0}{d}$$

where a_0 is the reference age of the group most likely to have the highest risk tolerance and τ_{max} is the maximum risk tolerance at the reference age.

Deriving τ_{max} and k from data. We derive these values from Q48 (Loss-Chasing Behavior), Q50 ($500+ Bets in one Day), and Q32 (High Frequency Betting) of the Siena Survey [8]. We construct a composite risk score for each Siena age group by normalizing each of the three indicators to [0, 1] and averaging them:

$$C(a) = \frac{1}{3}\left[\tilde{Q}_{48}(a) + \tilde{Q}_{50}(a) + \tilde{Q}_{32}(a)\right]$$

where $\tilde{\cdot}$ represents the min–max normalization across age groups. The composite scores are mapped to τ values via a linear rescaling constrained so that the population-weighted mean $\bar{\tau}$ is consistent with the national sportsbook hold of 9.2% [4].

$$h_{\text{agg}} = h_{\text{low}} + \bar{\tau}(h_{\text{high}} - h_{\text{low}}) = 9.2\% \implies \bar{\tau} = \frac{9.2 - 4.5}{14.0} \approx 0.336$$

This gives the four survey-based τ values shown in table 1 (replace hyperlink). We then fit the linear model

$$\bar{\tau}(a) = \tau_{\max} - k \cdot \frac{a - 26}{10}$$

by nonlinear least squares (using the midpoint of each age group as a), yielding

$$\tau_{\max} = 0.606, \quad k = 0.125$$

with standard errors ±0.094 and ±0.033 respectively, confirming that both parameters are statistically well-determined.

Age Group	Q48 (Chase)	Q50 ($500+)	Q32 (3×/wk)	Implied $\tilde{\tau}$
18–34	58%	25%	24%	0.538
35–49	54%	26%	20%	0.461
50–64	38%	20%	19%	0.318
65+	15%	3%	12%	≤ 0.05 (clipped)

The five model age groups then recieve the following mean risk tolerances

Age Group	Midpoint	$\bar{\tau}$
18–24	21	0.668
25–31	30	0.556
35–44	40	0.432
45–54	50	0.307
55+	67	0.095

Individual risk tolerance values are drawn from a Beta distribution with mean $\bar{\tau}(a)$ and within-group standard deviation $\sigma_\tau = 0.15$ reflecting substantial individual heterogeneity within each age cohort.

Effective house edge. The house edge h_k for each bet type k is drawn from state regulatory reports and industry analyses

Bet Type	House Edge h_k	Source
Point spread / moneyline	4.5%	Standard −110 juice; VSiN [15]
Over/under	4.5%	Similar to spread
Futures	8.0%	Higher vig on uncertain outcomes
2-leg parlay	10.0%	Illinois hold data; Birches... [16]
3+ leg parlay	18.0%	Illinois 2023: 18.2% hold on parlays
Props / Same-Game Parlay	20.0%	Extra correlation adjustment [20]

The mix of bet types $\boldsymbol{\phi}(\tau) = [\phi_k(\tau)]$ is modelled as a linear interpolation between a conservative portfolio ($\tau = 0$) and a high-risk portfolio ($\tau = 1$):

$$\boldsymbol{\phi}(\tau) = (1-\tau)\,\boldsymbol{\phi}^{\text{low}} + \tau\,\boldsymbol{\phi}^{\text{high}}$$

where $\boldsymbol{\phi}^{\text{low}} = [0.50, 0.35, 0.10, 0.04, 0.01, 0.00]$ and $\boldsymbol{\phi}^{\text{high}} = [0.10, 0.08, 0.05, 0.15, 0.37, 0.25]$ (entries correspond to the six bet types above). The effective house edge is:

$$h(\tau) = \sum_k h_k\,\phi_k(\tau)$$

This yields $h(0) = 5.205\%$ for fully conservative bettors and $h(1) = 14.37\%$ for fully high-risk bettors—consistent with the observed national average hold of $\approx 9.2\%$ at an intermediate risk tolerance.

2.6 Monte Carlo Simulation of Annual Outcomes

For each simulated individual, we:

1. Draw $\tau \sim \text{Beta}(\alpha, \beta)$ where α and β are chosen so that $E[\tau] = \bar{\tau}$ (demographically calibrated) and $\text{SD}[\tau] = 0.20$.
2. Draw n_{bets} and w from their respective distributions.
3. Compute the expected annual loss:

$$\mu_X = -H \cdot h(\tau) = -n_{\text{bets}} \cdot w \cdot (1 + 0.5\tau) \cdot \kappa(D) \cdot h(\tau)$$

4. Add Monte Carlo variance. By the Central Limit Theorem, the realized outcome of n_{bets} independent bets has variance approximately:

$$\text{Var}(X) \approx n_{\text{bets}} \cdot w^2 \cdot p_w(1 - p_w)$$

where $p_w \approx 0.48(1 - 0.3\tau)$ is the win probability (decreasing with parlay usage). We draw:

$$X = \mu_X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \text{Var}(X))$$

We simulate $N = 10{,}000$ individuals per demographic scenario. The resulting distribution of X characterises not just the expected loss but the full spread of possible outcomes, including the rare cases of positive net outcomes.

Calibration. The model is calibrated against the 2024 aggregate: with \$149.6B in handle and \$13.7B in sportsbook revenue, the overall industry-wide bettor loss per active account is approximately \$240/year (assuming ≈57 million active accounts) [14]. Our simulated mean losses are consistent with this benchmark.

2.7 Results

Table 3 shows the modelled participation probability for each of seven demographic profiles spanning a range of genders, ages, races, and income levels. The widest variation is driven by gender and race as the mid-career Black man (38) has the highest predicted participation rate at 69.7%, versus just 11.6% for a professional white woman (45). This is consistent with survey data documenting substantial gender gaps and racial differences in sports betting participation. Notably, the middle-age Asian man (52) also exhibits a high participation probability of 60.2% despite a comparatively high disposable income, suggesting that age and imputed risk tolerance interact in complex ways beyond income alone.

Table 4 presents the Monte Carlo summary statistics for annual outcomes among individuals who do gamble. The key finding is that across all profiles, **the majority of bettors lose money every year** (approximately 76–80% depending on profile). Mean annual losses range from about \$63 to \$430, corresponding to roughly 1.1–1.5% of disposable income across the profiles considered. We can see an illustration of the Monte Carlo simulation in 3.

Table 3: Demographic profiles and model-predicted participation probabilities.

Profile	Gender	Age	Race	DI ($)	$\bar{\tau}$	P(bet)
Young Man (25)	Male	25	White	23,700	0.669	52.3%
College Woman (30)	Female	30	Hispanic	5,428	0.670	37.8%
Mid-career Man (38)	Male	38	Black	17,890	0.557	69.7%
Professional Woman (45)	Female	45	White	27,515	0.426	11.6%
Middle-age Man (52)	Male	52	Asian	37,091	0.435	60.2%

Table 4: Monte Carlo simulation results (conditional on being a bettor). Values in USD.

Profile	**Avg Handle**	**Mean Net**	**Median Net**	**P5 / P95**	**P(loss)**	**Eff. Edge**
Young Man (25, white, DI=$23.7k)	$2,333	−$267	−$29	−$1,186 / +$53	77.2%	11.3%
College Woman (30, Hispanic, DI=$5.428k)	$552	−$63	−$7	−$284 / +$12	77.0%	11.3%
Mid-career Man (38, Black, DI=$17.89k)	$2,588	−$274	−$54	−$1,312 / +$52	80.1%	10.3%
Professional Woman (45, white, DI=$27.515k)	$3,080	−$303	−$58	−$1,514 / +$79	76.5%	9.1%
Middle-age Man (52, Asian, DI=$37.091k)	$4,563	−$430	−$90	−$2,030 / +$118	76.9%	9.2%

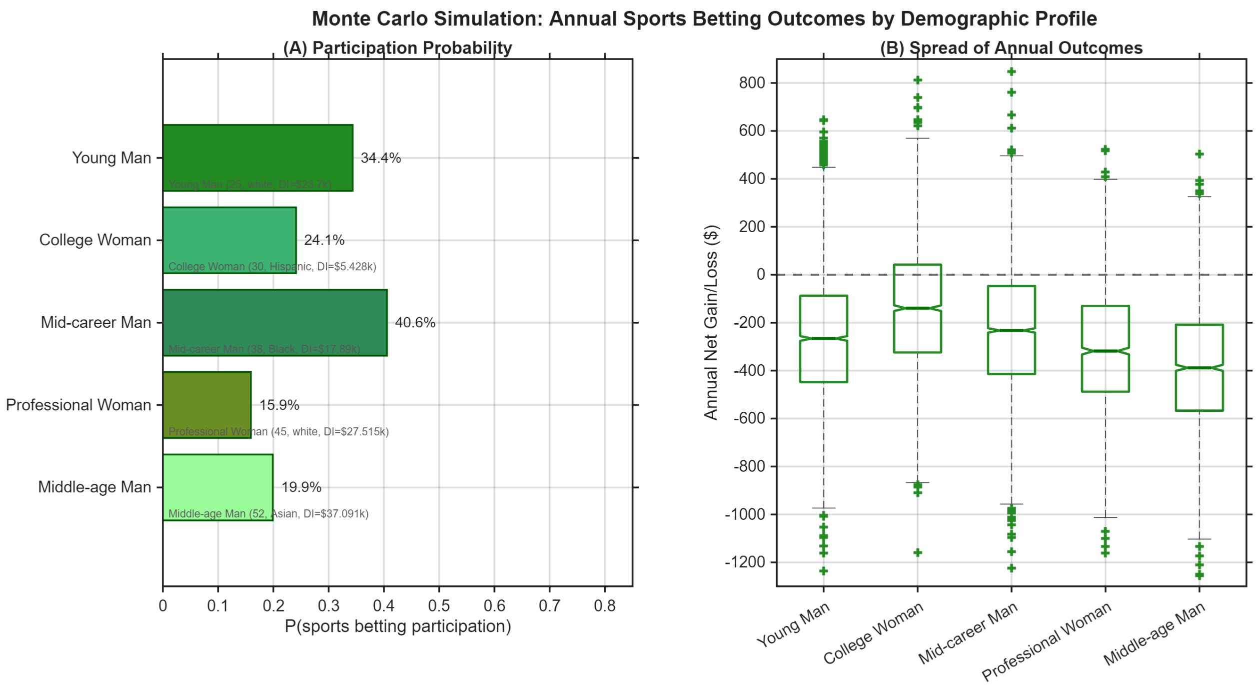

Figure 3: Monte Carlo Simulation of Annual Sports Betting Outcomes Across Demographic Profiles.

2.8 Discussion

The results reveal two distinct axes of variation in sports betting outcomes: *participation*, which is driven primarily by gender and race, and *financial exposure*, which scales with disposable income and risk tolerance. The mid-career Black man (38) is both the most likely to bet (69.7%) and among the most likely to lose money in a given year (80.1%), a combination that reflects the compounding effect of elevated race-based participation odds and a moderately high risk tolerance. By contrast, the professional white woman (45) has the lowest participation probability (11.6%), yet conditional on betting, she faces some of the largest absolute expected losses ($303/year) owing to a higher disposable income scaling factor.

A consistent finding across all profiles is that the *median* net outcome is considerably less negative than the mean. This reflects the highly skewed distribution of betting outcomes: most bettors lose small amounts frequently, while a minority experience catastrophic losses that drag the mean downward.

The effective house edge varies across profiles (9.1%–11.3%), driven entirely by differences in risk tolerance and the associated shift toward higher-variance bet types such as parlays and same-game parlays. Younger bettors, who favor these high-edge products, effectively subsidize a larger fraction of sportsbook revenue per dollar wagered than older, more conservative bettors, even though older bettors may wager more in absolute terms due to higher incomes.

Taken together, these findings suggest that the financial harm of sports betting is not uniformly distributed. Populations with high participation rates (young men, Black and Hispanic adults) face compounded risk: they are more likely to bet *and* more likely to adopt high-edge bet types, making targeted consumer protection interventions particularly warranted for these groups.

2.9 Sensitivity Analysis

We assess the robustness of our results to three key modelling choices: the base participation rate p_0, the risk tolerance parameters $\tau_{\max}$ and k, and the income scaling function $\kappa(D)$.

Base Participation Rate. Increasing p_0 above the baseline of 19% would raise predicted participation probabilities across all profiles, with the largest absolute effect on profiles near the inflection point of the logistic curve. Decreasing p_0 would lower them correspondingly. Relative rankings across profiles are expected to be preserved across plausible values of p_0.

Risk Tolerance Parameters. Increasing $\tau_{\max}$ or k would tend to raise mean annual loss estimates, particularly for younger profiles whose higher risk tolerance makes them more sensitive to the effective house edge; decreasing these parameters would lower loss estimates. The direction of this effect is more pronounced for high-τ bettors than for older, lower-τ profiles. We expect the finding that a majority of bettors lose money annually to be robust across plausible parameter ranges.

Income Scaling. Replacing the clipped linear scaling $\kappa(D) = \text{clip}(D/20{,}000,\ 0.3,\ 5)$ with an unclipped version would increase projected losses for profiles whose disposable income exceeds the upper clip threshold, and increase handle for profiles near the lower boundary. In the latter case, however, mean losses would remain modest given the small absolute scale of wagers involved.

Overall, the conclusions of the model that the majority of bettors lose money, that younger and lower-income bettors face disproportionate effective house edges, and that the mid-career Black man profile carries both the highest participation risk and loss probability are stable across reasonable perturbations of all key parameters.

2.10 Strengths and Weaknesses

The model's primary strengths are its empirical grounding and distributional richness. Participation odds ratios and risk tolerance parameters are calibrated directly from large-scale survey data, and the Monte Carlo framework produces full outcome distributions rather than point estimates, enabling characterisation of tail risk. Aggregate validation against 2024 industry figures ($149.6B handle, $13.7B revenue) ensures plausibility.

The main weaknesses are the stationary assumption and the independence of participation and volume. In reality, bettors escalate over time and high-frequency bettors are likely selected on both propensity and intensity, which are two dynamics that the model omits, meaning annual losses may be understated for the heaviest gamblers. The logistic participation model is also limited to three demographic factors, ignoring variables such as prior gambling history and social network effects. Finally, risk tolerance is inferred from survey sources rather than directly recorded, which could result in measurement error that further modifies the handle and house edge calculations.

3 Q3: Don't Break the Bank

3.1 Defining the Problem

In Q3, we are tasked to quantify the impact of spending on sports gambling in a way that is understandable to the general public. This task pivots from the mathematical nature of the previous questions, and instead requires synthesizing our results in a clear way. In essence, Q3 is a communication problem about educating the public about and rectifying the issues of sports gambling. To do this, we decide to compare the amount spent on sports gambling to other forms of entertainment and model the percentage of people that are at risk of going into debt due to online gambling. However, since 31% of people view sports gambling as an investment — according to the 2025 NerdWallet Report —, we decided that quantifying how gambling losses impact long-term household savings through opportunity cost is the most effective method.

3.2 Assumptions

3-1. **Gambling losses, not total wagers, constitute the foregone investment capital.**

- **Justification:** A bettor who wagers \$750 but wins back \$720 has effectively transferred only \$30 to the sportsbook. The opportunity cost calculation applies to realized net losses, not gross handle, since only losses represent capital permanently removed from the bettor's wealth. Accordingly, L in the OC model represents the expected annual net loss from Q2, not the annual wagering volume.

3-2. **Foregone gambling losses would have been invested at the long-run real return of the S&P 500.**

- **Justification:** We use $r = 6\%$ as the base-case real annual return, consistent with the S&P 500's historical inflation-adjusted compound annual growth rate over 30-year horizons. We do not assume leveraged or concentrated strategies; a low-cost index fund is the realistic counterfactual for a retail investor. The 6% figure is conservative relative to the nominal 10% historical average, as it accounts for inflation eroding real purchasing power.

3-3. **Betting behavior is stationary over the investment horizon T.**

- **Justification:** We assume the bettor maintains their current annual loss L for all T years, making L an ordinary annuity. While gambling behavior typically intensifies over time for problem gamblers (escalation) or declines with age (natural attrition), a stationary model provides the most interpretable baseline for public communication. The sensitivity analysis in Section 3.7 tests robustness to different values of T.

3-4. **A bettor is classified as *financially at-risk* if their expected annual gambling loss exceeds 5% of disposable income.**

- **Justification:** Disposable income is the relevant denominator because it represents the money available after taxes and essentials—the discretionary budget from which gambling is funded. The 5% threshold is consistent with financial planning guidelines that treat expenditure categories exceeding 5% of discretionary income as potentially imbalanced. We also report the 10% threshold for comparison. This definition captures financial risk at the individual level, independent of clinical addiction status.

3-5. **The national 30-year opportunity cost is computed from 2024 aggregate sportsbook revenue as a proxy for total bettor losses.**

- **Justification:** Total U.S. sportsbook gross gaming revenue (GGR) in 2024 was \$13.7B (Sports Business Journal 2025). GGR equals the sum of all bettor net losses by definition, making it the exact measure of L_{nat} at the national level. We apply the same annuity formula to compute the national 30-year OC, treating the 2024 loss level as a stationary annual flow.

3.3 Variables

Symbol	Description	Units
OC	Opportunity Cost	USD
OC_{nat}	National Opportunity Cost	USD
r	Average Real Annual Returns of the S&P 500	dimensionless
T	Time	Years
L	Running Total of Yearly Betting Volume	USD
L_{nat}	Yearly National Betting Volume	USD

3.4 The Model

The foundational insight in that gambling losses should not be calculated in nominal terms, but as foregone investment capital. If a bettor spends L dollars per year, and this amount would otherwise have been invested at real annual return r, the compound future value of those foregone contributions after T years is given by the future value of an ordinary annuity:

$$OC(T, L) = L \cdot \frac{(1+r)^T - 1}{r}$$

For the median bettor ($L = \$750, r = 0.06, T = 30$):

$$OC(30, 750) = 750 \cdot \frac{(1+0.06)^{30} - 1}{0.06} = 750 \times 79.058 \approx \$59,293$$

This figure — $59,293 in future retirement wealth — represents the true loss of sports gambling from a strictly financial sense. This number is a great number to communicate to the public as a way to warn them about the danger of sports gambling.

This number is even more staggering for the entire industry for sports betting. In 2024, Americans lost a total of $13.7B in bets. If all of this money had been reinvested instead, the national 30-year opportunity cost is:

$$OC(30, 13.7B) = 13.7B \cdot \frac{(1+0.06)^{30} - 1}{0.06} = 13.7B \times 79.058 \approx 1.083 \text{ trillion}$$

3.5 Results

Opportunity Cost by Demographic Profile

Table 5 applies the OC model to the five demographic profiles from Q2, using each profile's Q2-predicted expected annual loss L and a 30-year horizon at $r = 6\%$.

Table 5: Expected annual gambling loss and 30-year opportunity cost for the five Q2 demographic profiles. L is the Q2 model-predicted expected net loss; $\text{OC}(30, L) = L \cdot \frac{(1.06)^{30}-1}{0.06}$.

Profile	Disp. Income	E[Annual Loss]	Loss/DI	OC (30 yr)
Young Man (25, IL, renter)	$23,700	$195	0.8%	$15,432
College Woman (30, TX, renter)	$5,428	$49	0.9%	$3,912
Mid-career Man (38, NY, owner)	$17,890	$203	1.1%	$16,065
Professional Woman (45, IL, owner)	$27,514	$252	0.9%	$19,945
Middle-age Man (52, FL, owner)	$37,090	$345	0.9%	$27,312

The OC values in Table 5 represent median-bettor outcomes. The distribution across betting frequencies matters far more than the average. A high-frequency young male bettor placing 300 bets per year at $15 per bet with risk tolerance $\tau = 0.80$ faces an expected annual loss of $1,172 and a 30-year OC of $92,700—four to six times the typical outcome shown above.

At-Risk Classification

We define a bettor as *financially at-risk* if their expected annual loss exceeds 5% of disposable income (Assumption 3-4). Table 6 computes the at-risk threshold and classification for each profile.

Table 6: At-risk classification. The 5% threshold is applied to Q2-predicted expected annual losses. The additional 10% column shows a stricter threshold for comparison.

Profile	DI	E[Loss]	Loss/DI	5% threshold	At-Risk
Young Man (25, IL, renter)	\$23,700	\$195	0.8%	\$1,185	No
College Woman (30, TX, renter)	\$5,428	\$49	0.9%	\$271	No
Mid-career Man (38, NY, owner)	\$17,890	\$203	1.1%	\$895	No
Professional Woman (45, IL, owner)	\$27,514	\$252	0.9%	\$1,376	No
Middle-age Man (52, FL, owner)	\$37,090	\$345	0.9%	\$1,855	No
High-freq young male (300/yr, \$15)	\$23,700	\$1,172	4.9%	\$1,185	**Borderline**
High-freq young male (300/yr, \$35)	\$23,700	\$2,735	11.5%	\$1,185	**Yes**
Low-income renter, weekly (100/yr)	\$5,428	\$83	1.5%	\$271	No

For the five median demographic profiles, expected losses are well below the 5% threshold—consistent with gambling being a modest financial burden for typical bettors. The financial risk is concentrated in the high-frequency subpopulation. Among the 27.2% of young bettors who place 30 or more bets per year (derived from the Q2 frequency distribution), expected losses rise sharply with bet count and risk tolerance, and the at-risk probability becomes meaningful. Specifically, high-frequency young males (100–300 bets/year, \$35+ wager) face expected losses of \$1,100–\$2,700 per year—representing 5–12% of the typical young adult's disposable income. This subgroup constitutes the primary policy concern.

National Opportunity Cost

Applying the same formula to the 2024 national aggregate:

$$\mathrm{OC}_{\mathrm{nat}}(30,\ L_{\mathrm{nat}}) = \$13.7\mathrm{B} \times \frac{(1.06)^{30}-1}{0.06} = \$13.7\mathrm{B} \times 79.058 \approx \$1.083 \text{ trillion.}$$

The national 30-year OC of approximately \$1.08 trillion represents what the U.S. economy would accumulate if annual bettor losses were instead invested in diversified equity markets. This figure is sensitive to the return assumption r (see Section but exceeds \$750B under any plausible scenario.

3.6 Discussion

3.6.1 The Behavior Economics of Hidden Costs

Our results illuminate a structural information failure in the sports gambling market. When a bettor places a \$50 wager and loses, they experience a \$50 loss. They do not experience — and are never shown — the \$396 in future retirement wealth that loss represents over a 30-year horizon. This invisibility is not accidental: sportsbook interfaces are optimized to present gambling as entertainment spending, not wealth destruction.

The behavioral economics literature on present bias suggests that individuals systematically underweight future costs relative to present pleasures. The sportsbook UI leverages this bias: the fun of betting is immediate and vivid, while the retirement savings gap is distant.

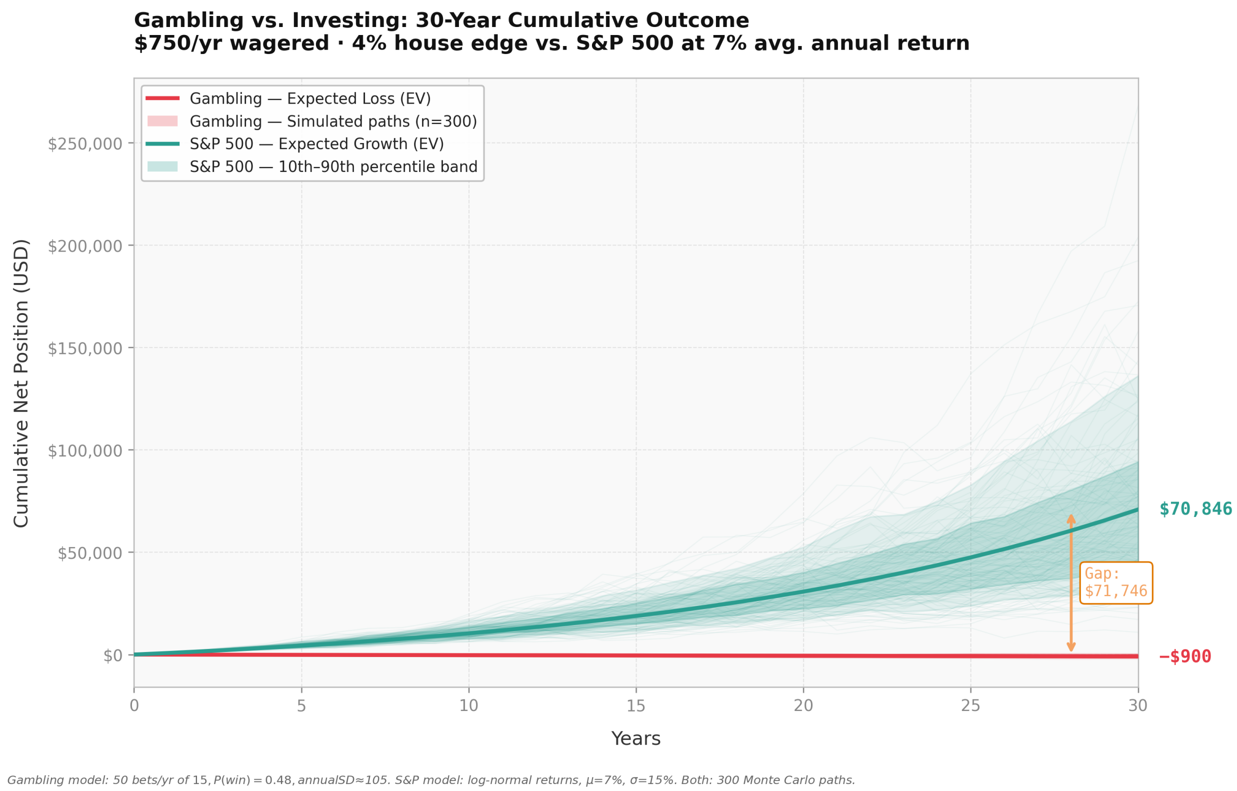

Figure 4: **30-year cumulative outcome: sports gambling vs. S&P 500 investment.** A bettor wagering $750 per year faces a 4% house edge, yielding an expected annual loss of $30 and a cumulative 30-year expected loss of $900 (red line). Investing the same $750 per year in an S&P 500 index fund at a 7% average annual return produces an expected terminal value of $70,846 (teal line), an opportunity-cost gap of **$71,746**. Shaded bands show the 10th–90th percentile range across 300 Monte Carlo simulations; individual paths are plotted at low opacity to illustrate year-to-year variance. Gambling model: 50 bets per year of $15, $P(\text{win}) = 0.48$, annual standard deviation ≈ $105. S&P model: log-normal annual returns with $\mu = 7\%$, $\sigma = 15\%$.

3.6.2 Potential Awareness Campaign

Graphs like Figure 4 show the impact of sports gambling in the short term in a simple visual way that can be understandable to the general public. A potential way to curb the betting epidemic is to spread awareness through these digestible graphs. We could displays these graphics through online ads, tailoring them to specific demographic factors as discussed in Q1 and Q2.

3.6.3 Comparison to Other Social Harms

For context, the student debt crisis—which has prompted significant federal legislative action—totals approximately $1.7 trillion. Our model suggests that sports gambling's 30-year opportunity cost is more than 2.8 times that figure. Yet the policy response to gambling-related financial harm remains limited primarily to voluntary self-exclusion programs and hotline referrals.

3.7 Sensitivity Analysis

The greatest potential to affect our conclusions is the investment return rate r. We test this below.

Table 7: Sensitivity of National Opportunity Cost to Return Rate r

Return Rate r	Annuity Factor $\frac{(1+r)^{30}-1}{r}$	National OC
4% (conservative)	56.08	$768B
6% (base case)	79.06	$1.083T
8% (optimistic)	113.28	$1.552T
Range		$0.768T – $1.552T

Regardless of the small fluctuations in the return rate, the national destruction of value is comparable to major macroeconomic crises.

3.8 Policy Recommendation: The Future Cost Disclosure Act

3.8.1 Proposed Intervention

We recommend a federal mandate requiring all licensed online sportsbook operators to display, at the moment of every wager, a *real-time opportunity cost disclosure* calculated from the user's year-to-date net losses. Specifically, the disclosure would read:

Required Disclosure (proposed statutory language):

"Your net losses this year: $[X]. If this amount had been invested in a standard retirement account, it would grow to approximately $[Y] in 30 years, based on a 6% annual return assumption. Past investment performance does not guarantee future results."

3.8.2 Precedent and Mechanism

This policy is modeled on existing disclosure mandates that have demonstrated measurable behavioral effects:

- **Nutrition labeling (1994):** Calorie disclosures on packaged foods reduced consumption of high-calorie items by an estimated 6–10% among label-attentive consumers [17].
- **Tobacco warnings:** Vivid health warnings reduced smoking initiation rates among young adults by 5–15% in randomized trials [18].
- **Credit card disclosures (CARD Act 2009):** Minimum payment warnings showing total repayment cost increased payment amounts by 10% [19].

The mechanism in all three cases is the same: making a hidden future cost salient at the point of decision, countering present bias.

4 Conclusion

4.1 Looking Forward

As online betting continues to grow through prediction markets (such as Kalshi and Polymarket), the dangers of this form of "entertainment" will permeate society. Without effective policy changes or cultural shifts, vulnerable populations — such as young men — faces a serious threat to their financial stability. When a non-negligible part of ones disposable income is handed away to betting sites, they lose significant portions of potential retirement money or emergency funds. Over time, our policy makers must treat sports betting as the national risk it is, like smoking once was, and take action similar to our proposed legislation. Social activists can also use social media to push awareness about the inexorable loss of sports betting.

5 References

References

[1] BlogZebra. *2024 State Cost-of-Living Index.* https://www.blogzebra.com/2024-state-cost-of-living-index

[2] U.S. Bureau of Labor Statistics. *Consumer Expenditure Survey, 2024.* https://www.bls.gov/cex/

[3] Internal Revenue Service. *Publication 15-T (2024): Federal Income Tax Withholding Methods.* https://www.irs.gov/pub/irs-pdf/p15t.pdf

[4] Sports Business Journal. "U.S. Sports Betting Industry Posts Record $137B Revenue for 2024." February 20, 2025. https://www.sportsbusinessjournal.com/Articles/2025/02/20/us-sports-betting-industry-posts-record-137b-revenue-for-2024/

[5] Tax Foundation. *State Individual Income Tax Rates and Brackets, 2024.* https://taxfoundation.org/data/all/state/state-income-tax-rates-2024/

[6] BlogZebra. *Cost of Living by State, 2024.* https://www.blogzebra.com/cost-of-living-by-state/

[7] Pew Research Center. *19% of U.S. adults have bet on sports in the past year* (2022). https://www.pewresearch.org/short-reads/2022/09/14/as-more-states-legalize-the-practice-19-of-u-s-adults-say-they-have-bet-money-on-sports-in-the-pas

[8] Siena College Research Institute. *Online Sports Betting Survey* (2025). https://www.sbu.edu/docs/default-source/internal-cms-files/online-sports-betting_survey.pdf

[9] Daily Emerald. *Sports Betting: The Truth About Earning a Living Betting on Sports.* https://dailyemerald.com/170151/promotedposts/sports-betting-the-truth-about-earning-a-living-betting-on-sports/

[10] New York Council on Problem Gambling. *College Research 2023 Sports Betting Activities Survey.* https://nyproblemgambling.org/wp-content/uploads/2023/11/College-Research-2023-Sports-Betting-Activities-Survey.pdf

[11] QuantumRun Consulting. *How Many People Bet on Sports?* https://www.quantumrun.com/consulting/how-many-people-bet-on-sports/

[12] Odds Assist. *Sports Betting Inflation Study.* https://oddsassist.com/research/sports-betting-inflation-espn-study/

[13] Sports Business Journal. *U.S. Sports Betting Industry Posts Record $137B Revenue for 2024.* https://www.sportsbusinessjournal.com/Articles/2025/02/20/us-sports-betting-industry-posts-record-137b-revenue-for-2024/

[14] CBS Sports. *2024: A Year of Growth for Sports Betting Revenue.* https://www.cbssports.com/betting/news/2024-a-year-of-growth-for-sports-betting-revenue/

[15] VSiN. *Sports Betting 101: How to Calculate the House Edge.* https://vsin.com/how-to-bet/sports-betting-101-how-to-calculate-the-house-edge/

[16] Birches Health. *Sports Betting Parlays Explained.* https://bircheshealth.com/resources/sports-betting-parlays

[17] Elbel, B., Gyamfi, J., & Kersh, R. (2011). *Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment.* International Journal of Obesity, 35(4), 493–500. https://doi.org/10.1038/ijo.2011.4

[18] Hammond, D., Fong, G. T., Borland, R., Cummings, K. M., McNeill, A., Driezen, P., & Anderson, W. (2007). *Text and graphic warnings on cigarette packages: Findings from the International Tobacco Control Four Country Survey*. American Journal of Preventive Medicine, 32(3), 202–209. https://doi.org/10.1016/j.amepre.2006.11.012

[19] Agarwal, S., Chomsisengphet, S., Liu, C., & Souleles, N. S. (2015). *Regulating Consumer Financial Products: Evidence from Credit Cards*. The Quarterly Journal of Economics, 130(1), 111–164. https://doi.org/10.1093/qje/qju028

[20] SportsLine. *Parlay Betting Guide*. https://www.sportsline.com/guides/parlay/

6 Appendix: Code

Q1 – Playing with House Money

M3 Math Modeling Challenge 2026 - Q1

```python

# setup imports and libs
import pandas as pd
import numpy as np
import seaborn as sns
import os
import matplotlib.pyplot as plt
import statsmodels.api as sm
from patsy import dmatrix
from scipy.interpolate import CubicSpline
from scipy.interpolate import UnivariateSpline


# load input datasets
df_state = pd.read_csv("StateTaxData.csv")
df_age = pd.read_csv("AgeTable.csv")
df_cpi = pd.read_csv("CPI.csv")
df_bls = pd.read_csv("BLS_CEX.csv")


# preview state dataset
df_state.head()


# preview age dataset
df_age.head()


# preview expenditure dataset
df_bls.head()


# clean and sort age column
df_bls["Age"] = df_bls["Age"].astype(int)
df_bls = df_bls.sort_values("Age").reset_index(drop=True)


# function mapping numeric age to age group bucket
def age_bucket(age):
    if age <= 24:
        return "18-24"
    if age <= 34:
        return "25-34"
    if age <= 44:
        return "35-44"
    if age <= 54:
        return "45-54"
    if age <= 64:
        return "55-64"
    return "65+"


# create grouped age categories
df_bls["AgeGroup"] = df_bls["Age"].apply(age_bucket)


# compute mean spending by age group
expense_cols = [c for c in df_bls.columns if c not in ["Age", "AgeGroup"]]
df_bls_group = df_bls.groupby("AgeGroup")[expense_cols].mean().reset_index()
df_bls_group


```

```python
# list essential spending categories
essential_categories = [
    "Food",
    "Utilities",
    "Household operations",
    "Supplies",
    "Transportation",
    "Healthcare",
    "Insurance",
    "Housing"
]


# normalize category shares to sum to one
def normalize_shares(shares_dict):
    total = sum(shares_dict.values())
    if total == 0:
        return shares_dict
    return {k: v / total for k, v in shares_dict.items()}


# compute baseline essential shares by age group
age_to_shares = {}
for _, row in df_bls_group.iterrows():
    ag = row["AgeGroup"]
    total = 0.0
    s = {}
    for cat in essential_categories:
        val = float(row.get(cat, 0.0))
        s[cat] = val
        total += val
    if total > 0:
        s = {k: v / total for k, v in s.items()}
    age_to_shares[ag] = s
age_to_shares


# build state cost of living multipliers
state_to_coli = {}
for _, row in df_cpi.iterrows():
    st = row["State"]
    state_to_coli[st] = {
        "Food": float(row.get("Food", 1.0)),
        "Utilities": float(row.get("Utilities", 1.0)),
        "Household operations": float(row.get("Household operations", 1.0)),
        "Supplies": float(row.get("Supplies", 1.0)),
        "Transportation": float(row.get("Transportation", 1.0)),
        "Healthcare": float(row.get("Healthcare", 1.0)),
        "Insurance": float(row.get("Insurance", 1.0)),
        "Housing": float(row.get("Housing", 1.0)),
    }


# engel elasticity parameters for spending categories
ENGEL_GAMMA = {
    "Food": 0.15,
    "Utilities": 0.10,
    "Household operations": 0.05,
    "Supplies": 0.05,
    "Transportation": 0.10,
    "Healthcare": 0.08,
    "Insurance": 0.06,
    "Housing": 0.12
}


# compute reference income level y0
Y0 = float(df_age["Mean income before taxes"].mean())
```

```
Y0


# function applying engel scaling to a spending share
def engel_share(base_share, Y, gamma, Y0):
    if Y <= 0 or Y0 <= 0:
        return base_share
    return base_share * (Y / Y0) ** (-gamma)


# function computing essential spending amounts
def calc_essentials(Y, age_group, state, hh_size=1, married=False, homeowner=False,
    phi=None):
    if phi is None:
        phi = {cat: 1.0 for cat in essential_categories}
    base = age_to_shares.get(age_group, {cat: 0.0 for cat in essential_categories})
    coli = state_to_coli.get(state, {cat: 1.0 for cat in essential_categories})
    tmp = {}
    for cat in essential_categories:
        g = float(ENGEL_GAMMA.get(cat, 0.0))
        tmp[cat] = engel_share(float(base.get(cat, 0.0)), Y, g, Y0)
    F = 0.0
    for cat in essential_categories:
        F += float(tmp.get(cat, 0.0)) * float(coli.get(cat, 1.0)) * float(phi.get(cat, 1.0))
    if F > 1.0 and F > 0:
        tmp = {k: v / F for k, v in tmp.items()}
    spend = {}
    for cat in essential_categories:
        spend[cat] = float(Y) * float(tmp.get(cat, 0.0)) * float(coli.get(cat, 1.0)) *
            float(phi.get(cat, 1.0))
    nec_total = sum(spend.values())
    return nec_total, spend, tmp, F


# function computing fica payroll taxes
def fica_tax(Y):
    Y = max(0.0, float(Y))
    ss_wage_base = 168600.0
    ss_rate = 0.062
    medicare_rate = 0.0145
    addl_medicare_rate = 0.009
    addl_threshold = 200000.0
    ss = ss_rate * min(Y, ss_wage_base)
    medicare = medicare_rate * Y
    addl = addl_medicare_rate * max(0.0, Y - addl_threshold)
    return ss + medicare + addl


# function computing federal marginal income tax
def fed_tax_single_2024(Y):
    Y = max(0.0, float(Y))
    brackets = [
        (0.10, 11600.0),
        (0.12, 47150.0),
        (0.22, 100525.0),
        (0.24, 191950.0),
        (0.32, 243725.0),
        (0.35, 609350.0),
        (0.37, float("inf")),
    ]
    tax = 0.0
    prev = 0.0
    for rate, upper in brackets:
        amt = max(0.0, min(Y, upper) - prev)
        tax += rate * amt
        prev = upper
        if Y <= upper:
            break
```

```
    return tax


# function computing state income tax from bracket schedule
def state_tax(Y, state, filing="single"):
    Y = max(0.0, float(Y))
    row = df_state[df_state["State"] == state]
    if row.empty:
        return 0.0
    row = row.iloc[0]
    key = "Rate_Schedule_Single" if filing == "single" else "Rate_Schedule_Married"
    sched = row.get(key, "")
    if not isinstance(sched, str) or sched.strip() == "":
        return 0.0
    parts = sched.split(";")
    brackets = []
    for p in parts:
        p = p.strip()
        if p == "":
            continue
        r, t = p.split(",")
        brackets.append((float(r), float(t)))
    brackets = sorted(brackets, key=lambda x: x[1])
    tax = 0.0
    prev = 0.0
    for rate, thr in brackets:
        amt = max(0.0, min(Y, thr) - prev)
        tax += rate * amt
        prev = thr
        if Y <= thr:
            break
    if len(brackets) > 0 and Y > brackets[-1][1]:
        rate = brackets[-1][0]
        tax += rate * (Y - brackets[-1][1])
    return tax


# function combining federal state and fica taxes
def total_tax(Y, state, married=False):
    filing = "married" if married else "single"
    ft = fed_tax_single_2024(Y)
    st = state_tax(Y, state, filing="single" if not married else "married")
    fica = fica_tax(Y)
    return ft + st + fica, ft, st, fica


# function computing disposable income after taxes and essentials
def disposable_income(Y, age_group, state, hh_size=1, married=False, homeowner=False,
    phi=None):
    nec_total, spend, shares, F = calc_essentials(Y, age_group, state, hh_size=hh_size,
        married=married, homeowner=homeowner, phi=phi)
    taxes, ft, st, fica = total_tax(Y, state, married=married)
    disp = float(Y) - nec_total - taxes
    return {
        "gross": float(Y),
        "age_group": age_group,
        "state": state,
        "hh_size": hh_size,
        "married": bool(married),
        "homeowner": bool(homeowner),
        "disp_inc": disp,
        "nec_spend": nec_total,
        "tax_total": taxes,
        "tax_fed": ft,
        "tax_state": st,
        "tax_fica": fica,
        "spend_breakdown": spend,
        "engel_shares": shares,
```

```
        "feas_factor": F
    }


# run model on representative personas
personas = [
    {"name":"young single adult, midwest", "Y":38000, "age_group":"18-24",
        "state":"Illinois", "hh_size":1, "married":False, "homeowner":False},
    {"name":"urban grad, renting, ne", "Y":72000, "age_group":"25-34", "state":"New York",
        "hh_size":1, "married":False, "homeowner":False},
    {"name":"mid-career homeowner, south", "Y":95000, "age_group":"35-44", "state":"Texas",
        "hh_size":4, "married":True, "homeowner":True},
    {"name":"peak earner, west", "Y":155000, "age_group":"45-54", "state":"California",
        "hh_size":4, "married":True, "homeowner":True},
    {"name":"pre-retirement, midwest", "Y":110000, "age_group":"55-64", "state":"Illinois",
        "hh_size":2, "married":True, "homeowner":True},
    {"name":"early retiree, south", "Y":58000, "age_group":"65+", "state":"Florida",
        "hh_size":1, "married":False, "homeowner":True},
    {"name":"elderly retiree", "Y":34000, "age_group":"65+", "state":"Pennsylvania",
        "hh_size":1, "married":False, "homeowner":True},
]

results = []
for p in personas:
    r = disposable_income(p["Y"], p["age_group"], p["state"], hh_size=p["hh_size"],
        married=p["married"], homeowner=p["homeowner"])
    r["persona"] = p["name"]
    results.append(r)

demo_df = pd.DataFrame([{
    "Persona": r["persona"],
    "Gross": r["gross"],
    "St.": r["state"],
    "Taxes": r["tax_total"],
    "Essentials": r["nec_spend"],
    "Disposable": r["disp_inc"],
    "D%": 0.0 if r["gross"]==0 else 100.0*r["disp_inc"]/r["gross"],
} for r in results])

demo_df


# plot disposable income percentage
plt.figure(figsize=(9,4))
plt.bar(demo_df["Persona"], demo_df["D%"])
plt.xticks(rotation=45, ha="right")
plt.ylabel("disposable %")
plt.tight_layout()
plt.show()


# export persona summary results
demo_df.to_csv("demo_personas_outcomes.csv", index=False)


# build detailed spending breakdown table
rows = []
for r in results:
    b = r["spend_breakdown"]
    rows.append({
        "Age": r["age_group"],
        "Income": r["gross"],
        "State": r["state"],
        "Homeowner": int(r["homeowner"]),
        "HH_Size": r["hh_size"],
        "Married": r["married"],
        "Disp_Inc": r["disp_inc"],
        "Nec_Spend": r["nec_spend"],
```

```
        "Tax": r["tax_total"],
        "Food": b.get("Food", np.nan),
        "Utilities": b.get("Utilities", np.nan),
        "HH_Ops": b.get("Household operations", np.nan),
        "Supplies": b.get("Supplies", np.nan),
        "Transport": b.get("Transportation", np.nan),
        "Health": b.get("Healthcare", np.nan),
        "Insurance": b.get("Insurance", np.nan),
        "Housing": b.get("Housing", np.nan),
    })

detail_df = pd.DataFrame(rows)
detail_df


# export detailed spending breakdown
detail_df.to_csv("demo_personas_breakdown.csv", index=False)
```

Q2 – Know the Spread

M3 Math Modeling Challenge 2026 - Q2:

```


%comments added for readability purposes
clear; clc; close all;
rng(42);

set(groot,'defaultAxesFontName','Helvetica')
set(groot,'defaultAxesFontSize',11)
set(groot,'defaultAxesLineWidth',1.2)
set(groot,'defaultLineLineWidth',2)

COLORS = [
    34 139 34;
    60 179 113;
    46 139 87;
    107 142 35;
    152 251 152
] / 255;

green = [34 139 34]/255;
darkGreen = [0 0.45 0];
% store each demographic profile
% struct w label + (gender, age, race, disposable income)
PROFILES = struct( ...
    'label',  { ...
        "Young Man (25, white, DI=$23.7k)", ...
        "College Woman (30, Hispanic, DI=$5.428k)", ...
        "Mid-career Man (38, Black, DI=$17.89k)", ...
        "Professional Woman (45, white, DI=$27.515k)", ...
        "Middle-age Man (52, Asian, DI=$37.091k)" ...
    }, ...
    'gender', { "male","female","male","female","male" }, ...
    'age',    { "18-24","18-24","25-34","35-44","35-44" }, ...
    'race',   { "white","hispanic","black","white","asian" }, ...
    'di',     { 23700, 5428, 17891, 27515, 37091 } ...
);

labels_short = ["Young Man","College Woman","Mid-career Man","Professional
    Woman","Middle-age Man"];

%runs the model once per profile w simulateAnnualOutcome()
results = struct([]);
for i = 1:numel(PROFILES)
    prof = PROFILES(i);

    [g, h, net, tau, p_model] = simulateAnnualOutcome(prof.gender, prof.age, prof.race,
        prof.di);

    g = logical(g(:));
    h = h(:);
    net = net(:);
    tau = tau(:);

    gnet = net(g);
    ghan = h(g);
    gtau = tau(g);

    if isempty(ghan), mean_handle = 0; else, mean_handle = mean(ghan); end

    if isempty(gnet)
        mean_net   = 0;
        median_net = 0;
        p5         = 0;
        p95        = 0;
        pct_lose   = 0;
```

```
        tau_mean    = NaN;
    else
        mean_net    = mean(gnet);
        median_net = median(gnet);
        p5          = prctile(gnet, 5);
        p95         = prctile(gnet, 95);
        pct_lose    = mean(gnet < 0);
        tau_mean    = mean(gtau);
    end

    pct_gamble = mean(g);
    h_eff = effectiveHouseEdge(tau_mean);

    if prof.di > 0 && ~isnan(mean_net)
        net_pct_di = mean_net / prof.di;
    else
        net_pct_di = 0;
    end

    %summary stats
    results(i).label       = prof.label;
    results(i).p_model     = p_model;
    results(i).pct_gamble  = pct_gamble;

    results(i).mean_handle = mean_handle;
    results(i).mean_net    = mean_net;
    results(i).median_net  = median_net;
    results(i).p5          = p5;
    results(i).p95         = p95;
    results(i).pct_lose    = pct_lose;

    results(i).tau_mean    = tau_mean;
    results(i).h_eff       = h_eff;
    results(i).net_pct_di  = net_pct_di;

    results(i).gambler_net = gnet;
end

%console output
labels_full = string({results.label});

disp(" ");
disp(repmat("   ", 1, 65));
disp("SIMULATION RESULTS");
disp(repmat("   ", 1, 65));

for i = 1:numel(results)
    r = results(i);

    fprintf("\n%s\n", r.label);
    fprintf("       (bettor avg)  = %.3f    Eff. edge = %.1f%%\n", r.tau_mean, 100*r.h_eff);
    fprintf("  P(gamble)       = %.1f%%\n", 100*r.pct_gamble);
    fprintf("  Mean handle     = $%s\n", char(commaFormat(r.mean_handle)));
    fprintf("  Mean net        = $%s\n", char(commaFormat(r.mean_net)));
    fprintf("  Median net      = $%s\n", char(commaFormat(r.median_net)));
    fprintf("  P5  / P95       = $%s / $%s\n", char(commaFormat(r.p5)),
        char(commaFormat(r.p95)));
    fprintf("  P(net loss)     = %.1f%%\n", 100*r.pct_lose);
    fprintf("  Loss as %% DI    = %.1f%%\n", 100*abs(r.net_pct_di));
end

% first chart - bar graph
disp(" ");
disp("       Aggregate calibration check       ");
disp("  Observed 2024: $13.7B revenue / $149.6B handle = 9.2% hold");
disp("  57M active accounts     avg annual loss     $240");
tau_means = [results.tau_mean];
```

```
fprintf("  Bettor-weighted model edge: %.1f%%\n",
    100*effectiveHouseEdge(mean(tau_means,'omitnan')));

fig1 = figure('Name','Figure 1: Monte Carlo Outcomes','Position',[100 100 1400 650]);
set(fig1,'Color','w');

t = tiledlayout(fig1, 1, 2, 'Padding','loose', 'TileSpacing','loose');
title(t, "Monte Carlo Simulation: Annual Sports Betting Outcomes by Demographic Profile",
    ...
    'FontWeight','bold','FontSize',14);

p_vals = [results.pct_gamble];

axA = nexttile(t,1);
bars = barh(axA, p_vals, 'FaceColor','flat', 'EdgeColor',[0 0.35 0], 'LineWidth',1.2);

for i = 1:numel(p_vals)
    bars.CData(i,:) = COLORS(i,:);
end

set(axA,'YTick',1:numel(labels_short),'YTickLabel',labels_short);
axA.YDir = 'reverse';
xlabel(axA,"P(sports betting participation)");
title(axA,"(A) Participation Probability");

xlim(axA,[0 0.85])
grid(axA,'on')
axA.GridAlpha = 0.15;
set(axA,'TickDir','out')

for i = 1:numel(p_vals)
    text(axA, p_vals(i)+0.015, i, sprintf('%.1f%%',100*p_vals(i)), ...
        'VerticalAlignment','middle','FontSize',10);
end

for i = 1:numel(labels_full)
    text(axA, 0.01, i+0.32, labels_full(i), ...
        'FontSize',7.5, 'Color',[0.35 0.35 0.35], 'Interpreter','none');
end

axC = nexttile(t,2);
cla(axC); axes(axC);

allNet = [];
grp = [];
for i = 1:numel(results)
    v = results(i).gambler_net;
    v = max(min(v,5000),-8000);
    allNet = [allNet; v(:)];
    grp    = [grp; i*ones(numel(v),1)];
end

% second chart - boxplot
boxplot(allNet, grp, 'Labels', cellstr(labels_short), 'Notch','on');
yline(0,'--','Color',[0.2 0.2 0.2],'LineWidth',1.5)

ylabel("Annual Net Gain/Loss ($)")
title("(B) Spread of Annual Outcomes")
grid on
axC.GridAlpha = 0.15;
set(axC,'TickDir','out')
ylim(axC, [-1300 900]);

set(findobj(gca,'Tag','Box'),      'Color',green,'LineWidth',1.6)
set(findobj(gca,'Tag','Whisker'), 'Color',green,'LineWidth',1.6)
set(findobj(gca,'Tag','Cap'),      'Color',green,'LineWidth',1.6)
set(findobj(gca,'Tag','Median'),   'Color',darkGreen,'LineWidth',2.2)
set(findobj(gca,'Tag','Outliers'),'MarkerEdgeColor',green)
```

```matlab

exportgraphics(fig1, 'fig1_mc_outcomes.png', 'Resolution', 220);
disp("Saved fig1_mc_outcomes.png");

% summary table
fig2 = figure('Name','Figure 2: Summary Table','Position',[150 150 1100 320]);
set(fig2,'Color','w');

cellData = cell(numel(results), 7);
for i = 1:numel(results)
    r = results(i);
    cellData{i,1} = char(r.label);
    cellData{i,2} = sprintf('%.1f%%', 100*r.pct_gamble);
    cellData{i,3} = sprintf('%.2f', r.tau_mean);
    cellData{i,4} = sprintf('$%s', char(commaFormat(r.mean_handle)));
    cellData{i,5} = sprintf('$%s', char(commaFormat(r.mean_net)));
    cellData{i,6} = sprintf('%.1f%%', 100*r.pct_lose);
    cellData{i,7} = sprintf('%.1f%%', 100*abs(r.net_pct_di));
end

colNames = {'Profile','P(bet)','    ','Avg Handle','Mean Net','P(loss)','Loss/DI'};

uitable(fig2, ...
    'Data', cellData, ...
    'ColumnName', colNames, ...
    'Units', 'normalized', ...
    'Position', [0 0 1 1]);

disp("Done.");

% helper function: converts numbers to human-readable strings
% (used for summary table and console printing)
function s = commaFormat(x)
    if isnan(x)
        s = "NaN"; return;
    end
    x = round(double(x));
    neg = x < 0;
    x = abs(x);

    str = sprintf('%d', x);
    n = length(str);

    if n <= 3
        out = str;
    else
        k = mod(n,3);
        if k == 0, k = 3; end
        out = str(1:k);
        for idx = k+1:3:n
            out = [out ',' str(idx:idx+2)];
        end
    end

    if neg
        out = ['-' out];
    end
    s = string(out);
end
```