House Always Wins: A Model of Gambling

Team #19068

MathWorks Math Modeling Challenge 2026

Online sports gambling has grown rapidly across the U.S. and U.K., changing how people spend their leisure money and raising significant concerns about financial harm. In the United States, the 2018 *Murphy v. NCAA* Supreme Court decision ended the federal ban on sports betting, allowing each state to set its own rules for online platforms. The industry expanded quickly: by 2025, American operators reported roughly $15 billion in revenue on $150 billion wagered [1]. Much of this growth came from mobile apps and younger adults who treat placing a bet as casually as checking a score. The U.K. has followed similar trends, with a long-standing regulatory framework supporting millions of active accounts and steadily increasing revenue. Together, these shifts represent a meaningful change in how discretionary income flows, and they raise an important question: who actually bears the financial cost?

This question drives the three parts of our analysis.

In **Q1**, we estimate how much money households realistically have left after paying taxes and covering necessities. Gambling losses only matter in the context of what someone can actually afford to lose. We model real disposable income, DI_{real}, by taking gross earnings, including side jobs, and subtracting taxes along with six unavoidable expense categories: housing, food, healthcare, transportation, education debt, and childcare. Costs are adjusted for regional price differences using BEA data for U.S. profiles and ONS survey data for U.K. profiles. The results were notable. Two of twelve U.S. profiles, a single mother in Georgia and an adjunct professor in Massachusetts, had negative disposable income, meaning their basic expenses already exceeded earnings. All U.K. profiles remained solvent, largely because NHS coverage removes a cost that can be crippling for lower-income Americans.

In **Q2**, we simulate individual gambling behavior over the course of a year. Instead of a simple yes/no measure, engagement is treated as a spectrum. Most people gamble lightly or not at all, while a small number wager heavily. This is captured using a lognormal distribution of engagement intensity, adjusted for age, gender, education, and legal-age restrictions. Wager amounts are tied to disposable income to keep the model financially grounded. The year is divided into two phases. After the first half, bettors who have exceeded a personal loss threshold stop gambling entirely. Those who continue tend to increase wagers and take larger risks, exhibiting loss-chasing behavior. Expected returns range from -4.5% to 4.5% for standard bets and up to -15% to 15% for riskier parlay bets, reflecting the house edge built into sportsbook operations.

Q3 examines the long-term financial consequences of these losses. We consider three measures: the ratio of losses to disposable income, the amount of retirement wealth forfeited, and the time required to recover financially after quitting. These are combined into a single Gambling Impact Score summarizing overall harm. Risk escalates sharply once annual losses approach available disposable income. At that point, debt compounds at rates around 22% APR, faster than typical savings can offset. Recovery becomes effectively impossible. Age magnifies this effect: an 18-year-old and a 40-year-old losing the same amount each year may appear similar in the short term, but by retirement, the younger person will have forfeited nearly five times as much wealth due to compounding.

Legal access and smartphone technology have made sports betting easier to stumble into than to walk away from, and the financial consequences can compound faster than most people expect. The risk is not evenly distributed. Younger bettors are particularly exposed: once losses cross a certain threshold relative to what they actually have available to spend, debt becomes hard to avoid and long-term wealth building stalls in ways that are difficult to reverse. Addressing this will require more than general awareness campaigns. Financial education needs to reach people before they are in trouble, and policy interventions need to be specific enough to target the groups most at risk rather than the gambling population as a whole.

1 Problem Restatement

Online sports gambling has expanded rapidly in both the U.S. and U.K. Since 2018, American platforms generated approximately \$15 billion in gross gaming revenue from \$150 billion in wagers by 2025 [1]. Gambling can be an expensive form of entertainment, often leading to addiction or problem gambling and negatively affecting financial health. Given this context, we are tasked with the following:

1. Modeling individual disposable income from demographics.
2. Modeling individual gambling gains/losses over one year.
3. Quantifying the societal financial impact of online gambling in terms accessible to the general public.

2 Global Assumptions

G-1. **Assumption:** Disposable income (DI) is defined as primary and secondary earned income minus taxes and essential expenses, adjusted for sales tax (U.S.) or VAT-inclusive pricing (U.K.). Investment returns, savings withdrawals, gifts, and informal income are excluded.

Justification: This definition follows the BLS Consumer Expenditure Survey framework and captures annual spendable cash flow, avoiding conflation with accumulated wealth.

G-2. **Assumption:** The analysis year is 2026, with tax brackets reflecting 2026 IRS values, including OBBBA amendments [2].

Justification: Using the current tax year ensures accuracy in disposable income calculations and aligns modeled outcomes with contemporary financial conditions.

G-3. **Assumption:** Single filer status is the default, and household composition is simplified to a reference person plus dependents; spousal income is excluded unless specified.

Justification: This avoids spousal income correlations and aligns with BLS CEX data. Modeling dual-income households would require additional complexity beyond the scope of this analysis.

G-4. **Assumption:** COVID-era economic anomalies have normalized.

Justification: BLS and ONS 2024 data indicate that spending patterns have returned to pre-pandemic trends, making historical averages representative for modeling.

G-5. **Assumption:** Exchange rate is \$1 = £0.75, used only for cross-country comparisons.

Justification: This allows direct comparison between U.S. and U.K. monetary values while leaving within-country calculations unaffected.

G-6. **Assumption:** Survey data approximates actual gambling behavior.

Justification: The M3 Challenge survey (n = 3,047) provides representative participation rates, allowing calibration of model behavior without requiring proprietary or inaccessible datasets [1].

3 Q1: Playing With House Money — Disposable Income

3.1 Q1 Assumptions

Q1-1. **Assumption:** Side income probability depends on age, education, and salary; based on BLS CPS data.

Justification: Approximately 16% of workers hold multiple jobs. Participation peaks for younger adults and declines with higher wages and education, reflecting opportunity cost of time.

Q1-2. Assumption: Essential expenditures are limited to six categories: housing, food, healthcare, transportation, education debt, and childcare, and are treated as non-discretionary and non-deferrable.

Justification: These categories account for roughly 75–85% of household expenditures. Treating them as non-deferrable ensures the model reflects true surplus income available for discretionary spending.

Q1-3. Assumption: Housing expenditures follow an Engel curve (sublinear in income with a hard floor), and all individuals are modeled as renters or mortgage holders; fully owned homes are not modeled separately.

Justification: Engel's Law is supported by BLS data, with housing share decreasing as income rises. Excluding outright ownership simplifies modeling but may overstate housing costs for older homeowners.

Q1-4. Assumption: Healthcare expenditures follow an exponential J-curve in age, modulated by an income-based insurance tier $\varphi(S)$.

Justification: Empirical data show healthcare spending increases sharply with age, with costs affected by insurance coverage type and income.

Q1-5. Assumption: Student loans follow a 10-year standard repayment plan at 5.5%.

Justification: The 10-year plan is the federal default, and the 5.5% rate approximates the weighted average of current Direct Loan rates.

Q1-6. Assumption: Regional cost-of-living differences are incorporated using BEA Regional Price Parities (U.S.) and ONS regional expenditure data (U.K.). U.K. expenses are interpolated directly from ONS data rather than using parametric models.

Justification: BEA RPPs provide comprehensive state-level price indices. Direct interpolation of ONS data preserves observed expenditure patterns without imposing U.S.-based functional forms.

Q1-7. Assumption: No major financial shocks (e.g., job loss, medical emergency, divorce) occur during the analysis year.

Justification: These events are unpredictable and individual-specific. The steady-state annual framework focuses on typical outcomes.

Q1-8. Assumption: Tax withholding equals actual tax liability.

Justification: Because the model is annual, only total yearly taxes matter. This simplifies the framework without affecting annual disposable income.

3.2 Model

We estimate real disposable income as:

$$\boxed{DI_{\text{real}} = \frac{S^* - T(S^*) - E(a, S^*, r, h, d, e)}{1 + \tau_{\text{sales}}(r)}} \tag{1}$$

where S^* is effective gross income (salary + expected side income), T is total taxes, E is essential expenses, and τ_{sales} is the combined sales tax rate. Dividing by $1 + \tau_{\text{sales}}$ adjusts for the fact that sales tax inflates purchase prices, so DI_{real} represents the pre-tax value of goods and services the remaining money can buy.

Effective Income. We model effective gross income as salary plus *expected* side income:

$$S^* = S + \pi(a, e, S) \cdot \bar{Y}_{\text{side}}(a, e),$$

where S is primary salary and $\bar{Y}_{\text{side}}(a, e)$ is the *average* side-income amount for individuals of age a and education level e and the overline denotes a mean. The factor $\pi(a, e, S) \in [0, 1]$ is the probability an individual holds a second job, modeled as a multiplicative combination of an education-dependent baseline rate, 22% for non-graduates down to 8% for doctoral holders, a Gaussian-shaped age effect peaked at 27 which shows that side income is most common in the late 20s, and a sigmoid salary-decay term, which means that higher salaries reduce the incentive for additional work. Parameters are chosen so that the population average probability is consistent with Bureau of Labor Statistics evidence that ∼16% of workers hold multiple jobs [5].

Taxes. We compute taxes in three components, each applied to the individual's effective income S^*. (1) **Federal income tax:** we first compute taxable income as $(S^* - \sigma)$, where $\sigma = \$16{,}100$ is the 2026 standard deduction. Federal tax is then the sum of the amounts owed in each of the seven marginal brackets. Finally, we subtract the $2,500 per-child tax credit (OBBBA) from the resulting liability [2]. (2) **Payroll taxes (FICA):** we add Social Security tax at 6.2% on wages up to the $184,100 cap, Medicare tax at 1.45% on all wages, and an additional 0.9% Medicare surcharge on earnings above $200,000. If side income is treated as self-employment income in our model, it is subject to self-employment (SE) payroll taxes as well. (3) **State taxes:** we apply state-specific income tax rules using published brackets, selecting ten representative states that span progressive systems (CA, NY, GA, OH, MA), flat-rate systems (IL, PA), and zero-income-tax states (TX, FL, NV) [3].

Essential Expenses. We model essential expenses as the sum of six required spending categories:

$$E = E_{\text{house}} + E_{\text{food}} + E_{\text{health}} + E_{\text{trans}} + E_{\text{edu}} + E_{\text{child}}.$$

Each component is chosen to match (i) how the cost *actually* scales with income/age (shape), and (ii) published benchmark values (level). Regional price differences are applied through the multiplier λ_r wherever the cost of living materially varies by location.

Component	**Model**	**Source**
Housing	Engel curve: $\lambda_r \cdot f(a) \cdot [\alpha(S^*)^{0.38} + F]$	[1],[2]
Food	USDA per-capita with economy-of-scale factor	[3]
Healthcare	J-curve: $\varphi(S) \cdot [15e^{0.055a} + 1200] \cdot h^{0.7}$	[4],[5]
Transport	Urban/rural split: $(1-u_r)E_{\text{car}} + u_r E_{\text{transit}}$	[6],[7]
Education	Loan PMT: $r = 5.5\%$, $n = 10$ yr, L by degree	[8]
Childcare	Regional: $\lambda_r \cdot \psi(S) \cdot \$11{,}582 \cdot d$	[9]

What these model forms mean

Engel curve for housing. An *Engel curve* describes how spending on a necessity changes with income: as income rises, people spend more dollars, but the fraction of income spent usually falls. That is why housing is modeled as a concave power law $(S^*)^{0.38}$, or sublinear growth. The constant F is a hard floor, meaning you cannot spend below a minimum to be housed, $f(a)$ allows housing needs to vary by age, and λ_r scales costs up/down by regional price levels.

Economy-of-scalefor food. Food costs do not scale linearly with household size. So, two people living together do not need exactly twice the food budget of one person due to shared staples or bulk buying. An economy-of-scale factor reduces per-capita food spending as household size increases.

J-curve for healthcare. A *J-curve* means costs rise slowly at first and then accelerate later. The exponential term $e^{0.055a}$ captures rapidly increasing healthcare costs with age. The multiplier $\varphi(S)$ represents an income-linked insurance tier, different effective out-of-pocket burden by income/coverage, and $h^{0.7}$ applies diminishing returns with household size.

Urban/rural split for transport. Transportation is modeled as a weighted average of car-heavy spending and transit-heavy spending. The parameter u_r is the local share of households that are transit-oriented or urban, so $(1 - u_r)$ weight goes to car costs and u_r weight goes to transit costs.

Loan PMT for education. Education cost is modeled as a standard fixed-payment amortized loan (PMT) using interest rate r, term n, and principal L (set by degree level). This converts "how much debt" into an annual required payment.

Childcare scaling. Childcare is approximately proportional to the number of dependents d, scaled by region λ_r. The factor $\psi(S)$ allows childcare burden to vary by income, for example, different childcare choices and subsidy eligibility. All expenses are scaled by BEA Regional Price Parities (λ_r) for state-level cost-of-living adjustment [6].

UK Model. For the U.K., we replace parametric expense models with **direct interpolation of ONS Family Spending Survey data** [1], indexed by age band and income quintile. Taxes are UK Income Tax (20/40/45% brackets with £12,570 personal allowance) plus National Insurance (8% on £12,570–50,270, 2% above) [9]. ONS data is VAT-inclusive, so no purchasing power correction is needed.

3.3 Results

We run the model on 12 U.S. profiles (9 states, both genders, \$15k–\$185k salary) and 8 U.K. profiles (£18k–£85k):

Table 1: Selected US results (full table: 12 profiles across TX, GA, NY, IL, CA, FL, MA, PA, NV, OH).

Profile	Salary	Tax+Exp	DI_{real}	DI%
Young man, no degree, TX	\$28,000	\$21,679	\$7,029	24.0%
Recent grad (M), NY	\$58,000	\$37,215	\$20,364	34.3%
Mid-career family, OH	\$88,000	\$64,364	\$22,607	25.5%
Single mother, GA	\$32,000	\$44,070	−\$9,909	−29.6%
High earner, CA	\$185,000	\$94,118	\$83,769	45.2%
Min wage, IL (F)	\$15,080	\$14,823	\$1,135	7.1%
Adjunct prof (F), MA	\$40,000	\$42,591	−\$1,142	−2.8%

Two profiles produce **negative DI**: the single mother in GA (−\$9,909) and the adjunct professor in MA (−\$1,142). These individuals cannot cover essentials from income alone—a critical finding for Q3's vulnerability assessment. In contrast, **no UK profiles produce negative DI**, as the NHS eliminates healthcare costs and UK housing outside London is substantially cheaper.

Figure 1 shows the full income breakdown for all 12 U.S. profiles.

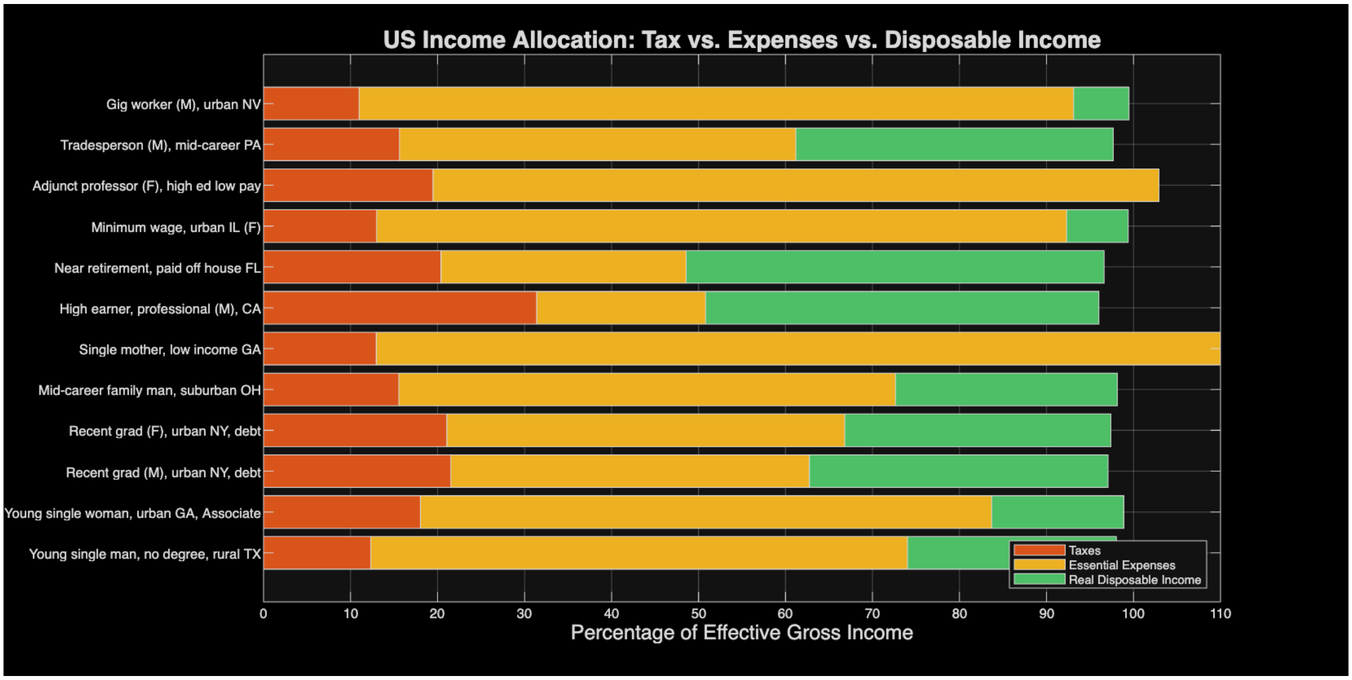

Figure 1: US income allocation by profile. Orange = taxes, yellow = essential expenses, green = disposable income.

Figure 2 shows the same breakdown for UK profiles. The absence of healthcare costs (NHS) and lower housing costs outside London mean all UK profiles retain positive DI.

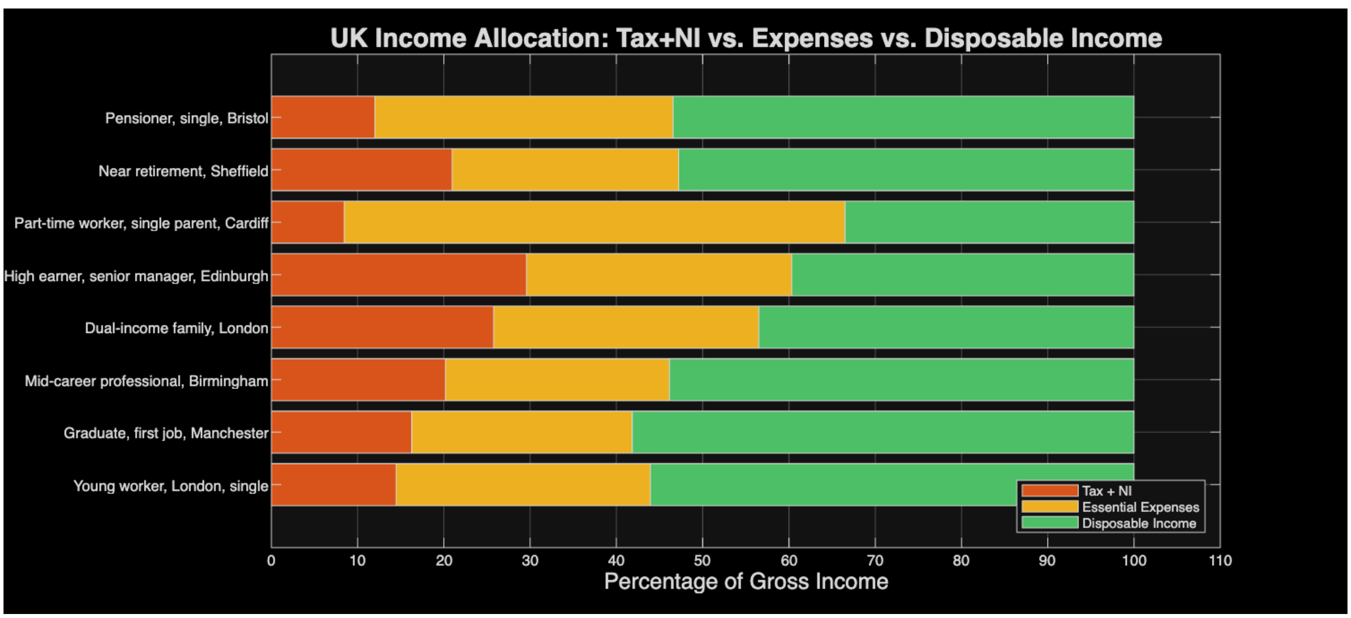

Figure 2: UK income allocation by profile. Orange = tax + NI, yellow = essential expenses, green = disposable income.

Figure 3 maps DI as a function of age and salary for a single, childless Bachelor's-degree holder in Ohio. The dashed black line marks $DI = 0$: anyone below this line cannot cover essentials from salary alone. The break-even salary rises with age due to increasing healthcare costs, then drops slightly near retirement as education debt is paid off.

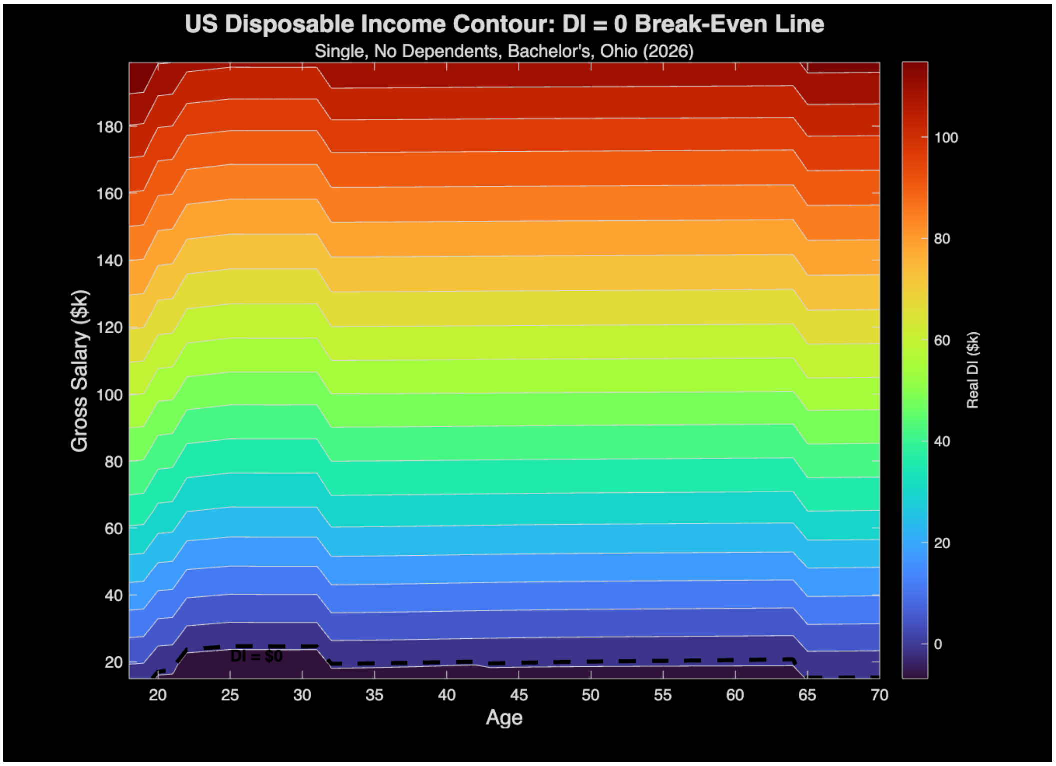

Figure 3: US disposable income contour: $DI(\text{age}, \text{salary})$ for a single Bachelor's-degree holder in Ohio. Black dashed line = break-even ($DI = 0$).

Figure 4 shows the equivalent UK contour. The break-even salary is lower across all ages, reflecting the NHS safety net and generally lower essential expenses.

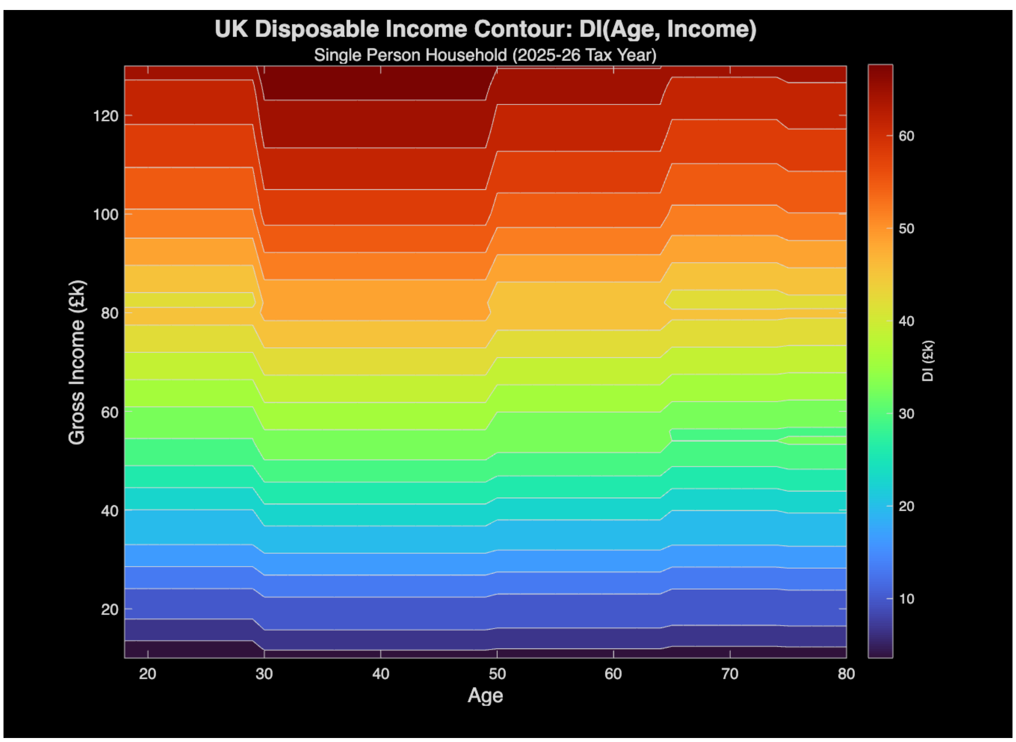

Figure 4: UK disposable income contour: $DI(\text{age}, \text{income})$ for a single-person household. Break-even salary is lower than in the US at all ages.

Figure 5 tracks cumulative lifetime disposable income by education level. Higher-degree holders start earning later (and carry more debt), so their cumulative DI initially lags behind high-school graduates. However, higher peak earnings eventually produce a crossover. A Bachelor's degree overtakes a high-school diploma around age 30; a Doctoral degree overtakes around age 38. This analysis informs Q3 by showing that young, educated gamblers have the most to lose in long-term opportunity cost.

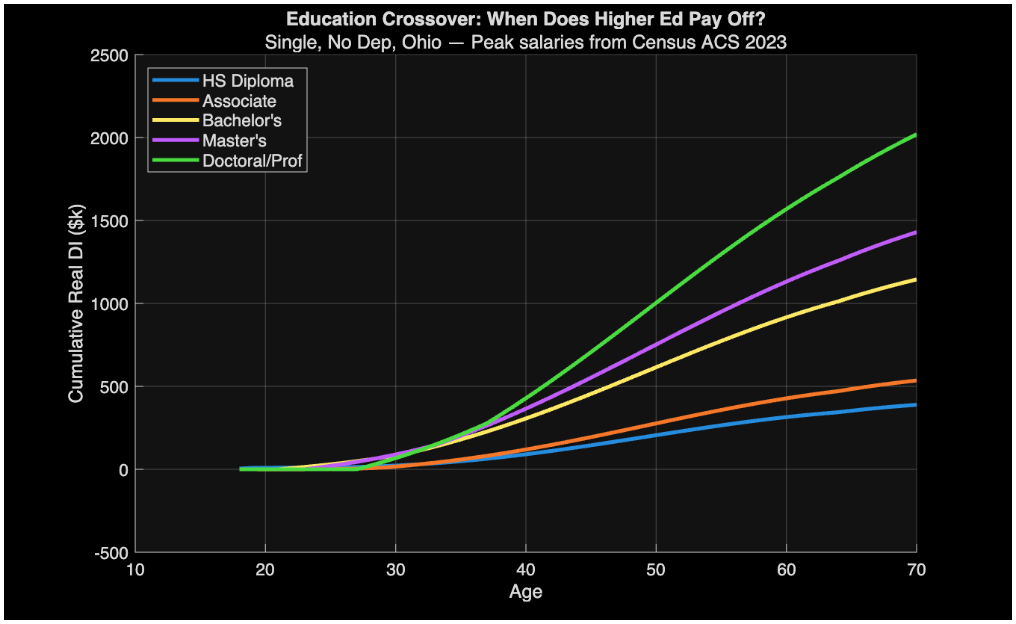

Figure 5: Cumulative real DI by education level. Peak salaries from Census ACS 2023 [7]; age-earnings profile from BLS CPS.

3.3.1 Validation

We validate by comparing predicted expenses to BLS CEX actual survey values at mean incomes per age group [4, 1]. The model tracks well, with the largest error in healthcare for ages 65+ (J-curve underestimates post-Medicare spending).

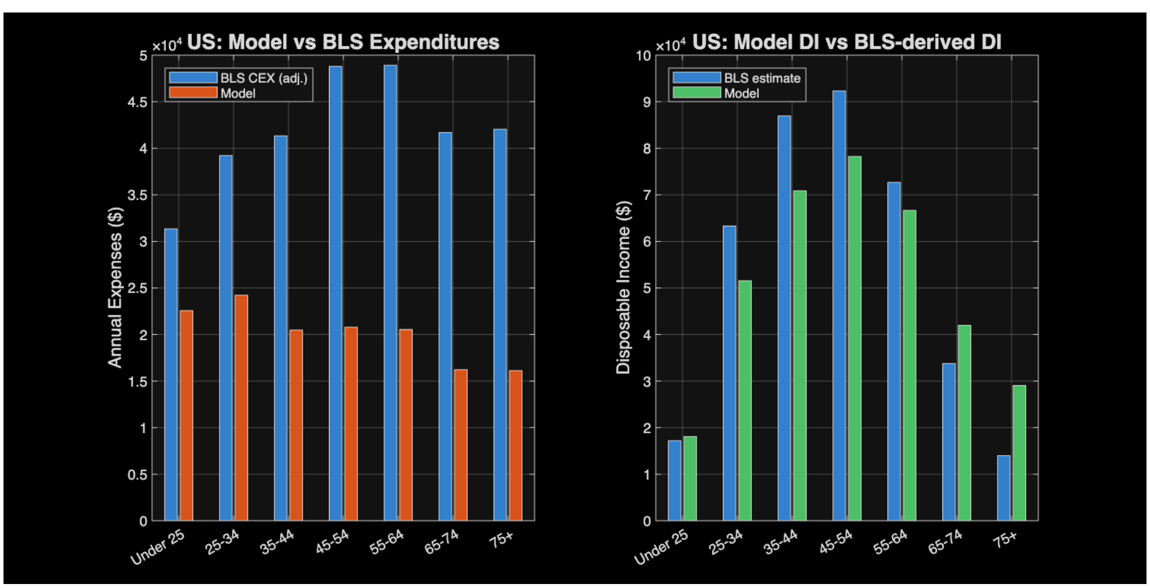

Figure 6: Model vs. BLS CEX 2024 actual expenditures by age group and category.

3.4 Q1 Strengths and Weaknesses

Strengths: Component-based architecture allows independent calibration; covers both U.S. (parametric) and U.K. (data-driven); validated against BLS/ONS data; captures 50-state variation via BEA price parities and actual tax brackets.

Weaknesses: Side-income parameters from aggregate statistics ; healthcare J-curve underestimates 65+ spending; standard deduction used universally; UK model cannot extrapolate beyond observed ONS income range.

Q2. Behavioral Gambling Trajectory (BGT) Model For Gain/Loss

N	Total number of simulated individuals (unitless)
W_i	Total annual wager for individual i (USD/year)
D_i	Disposable income for individual i (USD/year)
I_i	Gambling intensity for individual i (unitless)
f_0	Base gambling frequency constant (unitless)
$f_{\text{demo},i}$	Demographic multiplier for individual i, product of gender, age, and education factors (unitless)
Z_i	Latent heterogeneity variable, $Z_i \sim \mathcal{N}(0,1)$ (unitless)
$L(a_i)$	Legality sigmoid function, $L(a_i) = \left(1 + \exp\left(-(a_i - A_{\text{legal}})/\tau\right)\right)^{-1}$ (unitless)
ρ_i	Risk preference index for individual i, $\rho_i \in [0,1]$; 0 = straight bets, 1 = parlay-heavy (unitless)
α_i	Tilt sensitivity parameter for individual i, derived from survey loss-chasing rates (unitless)
B_i	Effective number of bets placed by individual i per year (bets/year)
R_i	Return per dollar wagered for individual i (unitless)
$C_{\text{mid},i}$	Net outcome at the end of Phase 1, $C_{\text{mid},i} = W_{i,1}R_{i,1}$ (USD)
θ_i	Mid-year performance ratio, $\theta_i = C_{\text{mid},i}/\max(W_{i,1},\,1)$ (unitless)
$W_{i,1},\ W_{i,2}$	Wager amounts for Phase 1 and Phase 2 respectively (USD)
$\lambda(D_i)$	Dynamic quitting threshold, $\lambda(D_i) = 0.05 + 0.03\ln(D_i/1000)$ (unitless)
AnnualNet_i	Final annual net gain or loss, $\text{AnnualNet}_i = C_{\text{mid},i} + W_{i,2}R_{i,2}$ (USD/year)

Table 2: Key variables used in the Behavioral Gambling Trajectory Model.

The baseline Q2 model treats betting as a collection of independent wagers with fixed odds. However, real bettors adapt their behavior after wins and losses (loss-chasing, risk shifting, and quitting). To incorporate these dynamics while keeping the model tractable, we introduce a behavioral extension: the **Behavioral Gambling Trajectory (BGT)** model. BGT preserves the same survey-grounded demographic inputs, but replaces the binary "active bettor" gate with a continuous engagement intensity, and introduces mid-year behavioral feedback.

3.5 Part I: Key Assumptions and Justifications

1. Population Representativeness

Assumption: The simulated population of size $N = 100{,}000$ reflects the U.S. adult demographic distribution in age, gender, and education.

Justification: The goal of the model is to approximate national annual gambling outcomes. Therefore, age proportions, gender splits (50/50), and college attainment rates are sampled to mirror broad U.S. population statistics. A large N ensures convergence of aggregate statistics by the Law of Large Numbers.

2. Disposable Income Determines Betting Capacity

Assumption: Annual gambling volume is proportional to disposable income:

$$W_i = D_i \cdot I_i,$$

where D_i is disposable income and I_i is gambling intensity.

Justification: Individuals cannot sustainably wager more than available discretionary resources. Using income minus expenditures ensures betting behavior is economically constrained and realistic.

3. Continuous Gambling Intensity (No Bernoulli Participation)

Assumption: Participation is not modeled as a binary coin flip. Instead, each individual has a continuous gambling intensity:

$$I_i = f_{\text{demo},i} \cdot \exp(\mu + \sigma Z_i), \quad Z_i \sim \mathcal{N}(0,1).$$

Justification: A lognormal structure:

- Produces many near-zero values (natural non-participants),
- Generates a heavy right tail ("whales"),
- Captures unobserved heterogeneity (sports interest, impulsivity, peer effects).

Thus, non-participation emerges endogenously rather than artificially.

4. Demographic Multipliers Affect Participation

Assumption: Gender, age, and education affect base gambling intensity via multiplicative factors:

$$f_{\text{demo},i} = f_0 \cdot m_{\text{gender}} \cdot m_{\text{age}} \cdot m_{\text{edu}} \cdot L(a_i),$$

where $L(a)$ is a legality sigmoid.

Justification: Survey data show systematic differences in account ownership and betting frequency by demographic group. A multiplicative structure preserves proportional relationships across categories.

5. Risk Tolerance Parameter (ρ)

Assumption: Each bettor has a risk preference parameter:

$$\rho_i \in [0,1],$$

where:

$$\rho_i = \frac{1}{1 + \exp\left(-(r_0 + r_{\text{age}} + r_{\text{gender}} + \varepsilon_i)\right)}.$$

Interpretation:

- $\rho = 0$: conservative bettor (straight bets),
- $\rho = 1$: aggressive bettor (parlays).

Justification: Survey evidence shows younger and male bettors are more likely to place large or high-risk wagers. A logistic transformation ensures ρ remains bounded between 0 and 1.

6. Tilt Sensitivity / Loss-Chasing (α)

Assumption: Behavioral response to losses is governed by:

$$\alpha_i = \sqrt{p_{\text{age}} \cdot p_{\text{gender}}},$$

where p_{age} and p_{gender} are empirical chase rates.

Justification: Loss-chasing is a documented behavioral bias. Using the geometric mean balances demographic influence without exaggerating extreme values.

7. Expected Return Depends on Risk Profile

Assumption: Expected return per dollar wagered is:

$$\mu(\rho) = -0.045 - 0.105\rho.$$

Interpretation:

- $\rho = 0$: $\mu = -4.5\%$ (straight bets),
- $\rho = 1$: $\mu = -15\%$ (parlays).

Justification: Sportsbooks earn higher margins on parlays than straight bets, so higher risk tolerance implies higher expected loss.

8. Central Limit Theorem Scaling

Assumption: If a bettor places B bets per year, volatility scales as:

$$\sigma_{\text{annual}} = \frac{\sigma_{\text{per bet}}}{\sqrt{B}}.$$

Justification: Averaging many independent wagers reduces variance of aggregate returns.

9. Two-Phase Behavioral Year

Assumption: The year is split into two halves. Phase 2 adapts based on Phase 1 results.

Justification: Real gambling behavior evolves dynamically — winners become cautious, losers chase losses.

10. Bankruptcy / Quit Condition

Assumption: If midyear losses exceed a dynamic fraction of disposable income:

$$C_{\text{mid}} < -\lambda(D_i)D_i,$$

the bettor quits.

Justification: Individuals have implicit loss limits and financial constraints.

Part II: How the Model Works

Layer 1: Demographic Layer

Step 1: Generate Population Each individual i is assigned:

- Age a_i,
- Gender g_i,
- Education level e_i,
- Disposable income D_i (lognormally perturbed).

Step 2: Compute Gambling Intensity

$$I_i = f_{\text{demo},i} \cdot \exp(\mu + \sigma Z_i).$$

Annual wager:

$$W_i = D_i \cdot I_i.$$

If $W_i < 10$, the individual is treated as a non-gambler.

This demographical layer generates a synthetic population representing U.S. adult gamblers. It assigns age, gender, education, and disposable income to each individual, which then determine their baseline gambling intensity. This layer uses survey data to account for observed correlations between demographics and participation in online sports gambling.

Layer 2: Stochastic Return Layer

Effective bet count:

$$B_i = \min\left(1000, \max\left(1, \left\lfloor 104 \cdot \frac{I_i}{\text{median}(I)} \right\rceil\right)\right).$$

Return per dollar:

$$R_i \sim \mathcal{N}\left(\mu(\rho_i), \frac{\sigma(\rho_i)}{\sqrt{B_i}}\right),$$

with truncation:

$$-1 \le R_i \le r_{\max}(\rho_i).$$

Annual net:

$$\text{Net}_i = W_i \cdot R_i.$$

This stochastic layer converts baseline intensity and disposable income into actual betting behavior. It simulates each individual's annual wager and introduces stochasticity in betting outcomes based on the number of bets placed and risk preferences. The stochastic layer captures both variability in bet sizes and the inherent house edge.

Layer 3: Behavioral Feedback Layer

Phase 1

$$W_{i,1} = \frac{W_i}{2}, \quad C_{\text{mid}} = W_{i,1} R_{i,1}.$$

Behavioral Adjustment Define performance ratio:

$$\theta_i = \frac{C_{\text{mid}}}{W_{i,1}}.$$

Second-half wager:

$$W_{i,2} = \frac{W_i}{2} \cdot \exp(-\alpha_i \theta_i).$$

- If $\theta_i < 0$ (loss), $W_{i,2}$ increases.
- If $\theta_i > 0$ (win), $W_{i,2}$ decreases.

Risk update:

$$\rho_{i,2} = \text{clip}(\rho_i - 0.3\theta_i, 0, 1).$$

Phase 2 Net

$$\text{Net}_{i,2} = W_{i,2}R_{i,2}.$$

Total annual outcome:

$$\text{AnnualNet}_i = C_{\text{mid}} + \text{Net}_{i,2}.$$

The behavioral layer models how individuals adapt their gambling over a year in response to wins and losses. The year is split into two phases: Phase 1 is baseline betting; Phase 2 adapts wagers and riskiness based on Phase 1 outcomes. This layer accounts for loss-chasing, quitting after large losses, and changes in bet types (straight vs. parlay).

Model Outputs

The model produces:

- Percentage of active gamblers,
- Mean and median annual net,
- Distribution percentiles,
- Outcome breakdown (profit, loss, near-even),
- Subgroup comparisons (age, gender),
- Loss-chasing behavioral effects.

3.6 Strengths and Weaknesses of the Q3 Model

Strengths

The Behavioral Gambling Trajectory (BGT) model derives its primary strength from its deep empirical grounding, utilizing datasets such as the *Siena College/St. Bonaventure American Sports Fanship Survey 2025* to parameterize demographic behavior, risk profiles, and loss-chasing tendencies. By eschewing binary "participation gates" in favor of a continuous lognormal intensity variable, the model allows non-participation to emerge endogenously, effectively simulating the "long tail" of gambling volume. Furthermore, the integration of the Central Limit Theorem to scale volatility relative to the effective number of bets B_i provides a mathematically rigorous way to differentiate the risk experiences of casual bettors versus frequent participants. The two-phase behavioral feedback mechanism—which dynamically adjusts wager sizes W and risk parameters ρ based on mid-year performance θ_i—successfully captures the psychological "tilt" phenomenon, resulting in a more realistic distribution of annual financial outcomes than static models.

3.6.1 Weaknesses

Conversely, the BGT model faces limitations regarding the complexity of human decision-making. The behavioral update rules are essentially linear, which may oversimplify the non-linear "sunk cost" biases and idiosyncratic psychological triggers that often dictate individual loss-chasing in real-world settings. Additionally, the model assumes that individual betting outcomes are independent, failing to account for social network effects, peer-group betting, or the impact of external promotional events like sign-up bonuses that could skew returns temporarily. The current risk/return function $\mu(\rho)$ treats the house edge as a static property of the risk profile, which overlooks the potential for "sharp" bettors to shift lines or exploit specific market inefficiencies that deviate from the standard house hold. Finally, while the model captures a "heavy right tail" of high-volume bettors, it lacks the nuance to distinguish between pathological "whales" and sophisticated professional bettors, potentially misrepresenting the financial impact of the extreme end of the distribution.

Summary

The Behavioral Gambling Trajectory (BGT) Model is a three-layer system:

1. **Demographic Layer:** Determines who gambles and how much.
2. **Stochastic Layer:** Simulates market returns with realistic house edge and volatility.
3. **Behavioral Layer:** Introduces dynamic adaptation, loss-chasing, and quitting.

The structure allows the model to:

- Match empirical participation rates,
- Match industry-level average losses,
- Produce heavy-tailed outcome distributions,
- Capture psychologically realistic behavioral feedback.

This results in a calibrated, behaviorally grounded annual prediction of net gambling gain or loss for each individual based on demographics and behavioral traits.

4 Q3: Don't Break the Bank — Societal Impact

4.1 Nomenclature

Table 3: Nomenclature for Q3 (units in USD/year unless noted).

Symbol	Meaning
a	Starting age of the individual (years)
t	Year index, $t = 0, 1, \ldots$
T	Years of active gambling before quitting (years)
A_R	Retirement benchmark age (years), e.g., $A_R = 65$
DI_{real}	Real disposable income after essentials (USD/year)
L	Expected net annual gambling loss (USD/year)
β	Gambling burden ratio, $\beta = L/DI_{\text{real}}$ (unitless)
r	Annual investment growth rate (unitless)
i	Annual debt interest rate / APR (unitless)
D_t	Debt balance at start of year t (USD)
S	Annual recovery contribution after quitting (USD/year), $S = s \cdot DI_{\text{real}}$
OC_{end}	Opportunity cost at end of gambling window (USD)
OC_{A_R}	Opportunity cost compounded to retirement age A_R (USD)
n_{debt}	Years to pay off debt after quitting (years)
n_{wlth}	Years to rebuild missed wealth after debt payoff (years)
n_{tot}	Total recovery time, $n_{\text{tot}} = n_{\text{debt}} + n_{\text{wlth}}$ (years)
$a_{\text{break-even}}$	Break-even age, $a_{\text{break-even}} = a + T + n_{\text{tot}}$ (years)
GIS	Gambling Impact Score (0–100), public-facing severity index

4.2 Q3 Assumptions

Q3-1. Assumption: The model uses an annual time step. Individuals are tracked from current age a through T years of gambling and any subsequent recovery period.

Justification: Annual discretization aligns with income, tax, and expenditure reporting, simplifying computations while capturing long-term financial effects.

Q3-2. Assumption: Annual gambling loss is constant, L, equal to the expected loss estimated in Q2 for the individual's demographic profile.

Justification: A fixed annual loss isolates the long-run impact of gambling without introducing stochastic variability or escalating bet behavior.

Q3-3. Assumption: After quitting, individuals allocate a fixed fraction of real disposable income toward recovery: $S = 0.50 \cdot DI_{\text{real}}$.

Justification: A fixed 50% savings rate provides a transparent benchmark for debt repayment and rebuilding wealth.

Q3-4. Assumption: Assets grow at $r = 7\%$, and debt compounds at $i = 22\%$ APR.

Justification: Distinct growth rates reflect empirical differences between long-term investment returns and high-cost consumer debt.

Q3-5. Assumption: Debt occurs only when annual gambling losses exceed real disposable income; otherwise, surplus is used to pay down existing debt.

Justification: This enforces liquidity constraints consistent with household behavior and prevents unrealistic debt accumulation.

Q3-6. Assumption: Retirement age is a fixed benchmark, $A_R = 65$.

Justification: A standardized retirement age allows for consistent evaluation of long-term wealth and opportunity costs across individuals.

4.3 Model

Q3 combines Q1 disposable income with Q2 expected losses to produce three public-facing metrics.

(1) Gambling burden ratio

$$\boxed{\beta = \frac{L}{DI_{\text{real}}}} \tag{2}$$

$\beta \geq 1$ means losses exceed disposable income and the individual must borrow.

(2) Missed wealth at retirement

$$\boxed{OC_{A_R}(a, T) = L \cdot \frac{(1+r)^T - 1}{r} \cdot (1+r)^{\max\{0,\, A_R - (a+T)\}}} \tag{3}$$

Starting younger is more damaging because losses compound longer.

(3) Debt mechanics and recovery time

Debt evolves as:

$$\boxed{D_{t+1} = \max(0,\ (1+i)D_t + L - DI_{\text{real}})} \tag{4}$$

After quitting with debt D_T, recovery requires $S > i \cdot D_T$ (otherwise: **debt spiral**). If solvent:

$$n_{\text{debt}} = \frac{\ln\left(\frac{S}{S-iD_T}\right)}{\ln(1+i)}, \qquad n_{\text{wlth}} = \frac{\ln\left(1 + \frac{r\,OC^\star}{S}\right)}{\ln(1+r)} \tag{5}$$

where $OC^\star = OC_{\text{end}}(1+r)^{n_{\text{debt}}}$. Total recovery: $n_{\text{tot}} = n_{\text{debt}} + n_{\text{wlth}}$.

(4) Gambling Impact Score (GIS)

$$\boxed{GIS = 100\Big(0.30 \cdot \widetilde{\beta} + 0.40 \cdot \widetilde{n} + 0.30 \cdot \widetilde{D}\Big)} \tag{6}$$

4.4 Results

Table 4: Q3 outputs for representative profiles.

Profile	Age	DI	L	D_T	OC_{65}	n_{debt}	n_{wlth}	GIS
Young Male (21, high risk)	21	\$6,000	\$4,800	\$0	\$120,619	0.0	19.8	34
Young Female (24, moderate)	24	\$7,500	\$2,400	\$0	\$24,624	0.0	5.6	12
Working Male (35, moderate)	35	\$15,000	\$6,000	\$0	\$82,899	0.0	8.5	19
Working Female (40, low risk)	40	\$12,000	\$1,800	\$0	\$10,351	0.0	1.7	6
Mid-career Male (50, high)	50	\$18,113	\$9,600	\$0	\$132,638	0.0	10.5	39
Senior (65, low risk)	65	\$10,000	\$1,200	\$0	\$3,858	0.0	0.8	33

Tipping point

When $L \gtrsim DI_{\text{real}}$, debt begins and recovery time can explode.

Table 5: Start age sensitivity (fixed DI, L, T).

Start Age	n_{tot}	Break even Age	Missed Wealth at 65
18	11.4	39	\$1,013,322
21	11.4	42	\$827,173
25	11.4	46	\$631,046
30	11.4	51	\$449,927
35	11.4	56	\$320,792
40	11.4	61	\$228,720
50	11.4	71	\$116,270
60	11.4	81	\$82,899

Table 6: Loss sensitivity: tipping point where L crosses DI.

Annual Loss L	n_{tot}	Missed Wealth @ 65	GIS
\$1,000	2.6	\$105,174	5
\$3,000	6.8	\$315,523	13
\$5,000	10.0	\$525,872	20
\$8,000	13.8	\$841,395	30
\$10,000	15.9	\$1,051,743	36
\$12,000	INF	\$1,262,092	94
\$15,000	INF	\$1,577,615	100

4.5 Strengths and Weaknesses of the Q3 Model

Strengths

The model produces public-facing metrics that non-technical audiences can easily understand: the fraction of disposable income consumed by gambling (β), missed wealth at retirement (OC_{A_R}), and recovery time (n_{tot}). It also connects Q1 and Q2 into a single long-run impact measure by combining real disposable income DI_{real} with expected annual loss L, translating short-run gambling behavior into long-run financial consequences. Finally, because both opportunity cost and recovery are modeled with compound growth, the framework captures the key time effect: losses earlier in life are more damaging than the same losses later, simply because they have more years to compound.

Weaknesses / Limitations

The model assumes a constant annual loss over the entire gambling window ($L_t = L$), even though real gamblers may escalate their betting, take breaks, or quit early; as a result, the model can understate harm for escalating bettors and overstate harm for short-term bettors. It also relies on single "average" interest rates, one investment return r and one debt APR i, while in reality households face a mix of returns and multiple debts with different rates and repayment structures. Finally, post-quit savings behavior is stylized: we assume a fixed recovery contribution $S = s \cdot DI_{\text{real}}$, but actual savings rates vary substantially across individuals and can shift after major life events such as rent increases, childcare costs, or unexpected medical bills.

5 Sensitivity Analysis

5.1 Q1

We perform Monte Carlo sensitivity analysis ($N = 5{,}000$) on Q1, perturbing individual parameters by $\pm 10\%$. Base profile: mid-career family man in Ohio (age 38, \$88k, Bachelor's, 2 dependents). Salary dominates because it affects both income and income-linked expenses. Housing is second, consistent with being the largest BLS CEX category.

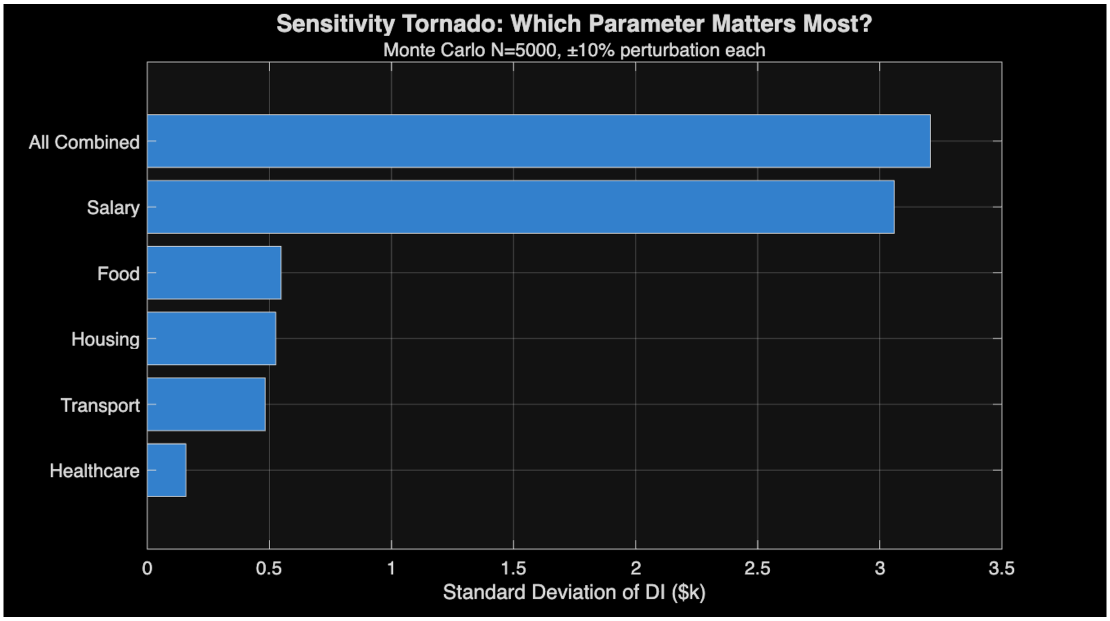

Figure 7: Sensitivity tornado: SD of DI under ±10% perturbation of each parameter.

5.2 Q2

Table 7: Q2 sensitivity using elasticity of key outputs to ±10% parameter perturbation.

Parameter	**Gambler %**	**Mean Loss**	**P95 Loss**
Latent mean μ_ℓ	1.42	1.31	0.89
Latent std σ_ℓ	0.74	0.92	1.53
Base frequency f_0	0.98	0.95	0.61
House edge $\mu(\rho)$	0.00	0.87	0.73
Chase sensitivity α	0.00	0.34	0.58
Quit threshold $\lambda(D)$	0.11	0.18	0.09

The latent mean μ_ℓ is the dominant driver: it simultaneously controls how many people gamble and how much they wager, producing the largest elasticities across all three outputs. The latent standard deviation σ_ℓ has a moderate effect on participation and mean loss but the strongest effect on 95th-percentile loss. House edge $\mu(\rho)$ shifts losses proportionally but does not affect who gambles. Loss-chasing sensitivity α has a small effect on the mean but a larger effect on tail losses. The quit threshold $\lambda(D)$ has the weakest influence overall, since most gamblers do not hit their ruin gate within a single year.

5.3 Q3

A key goal of Q3 is to communicate how gambling harm changes when inputs shift. Because our outputs (missed wealth and recovery time) compound over time, the model is *nonlinear*: small increases in annual loss can translate into multi-decade increases in recovery, and crossing a threshold can make recovery effectively infeasible under the same behavioral assumptions.

5.3.1 Starting age vs. recovery (fixed DI, loss, and duration)

Table 5 holds disposable income at $10,000, annual loss at $6,000, and gambling duration at $T = 10$ years. Across starting ages 18–60, the model reports `INF` recovery time, meaning the individual cannot catch up under the current recovery capacity assumption. In this scenario, the missed wealth at the end of the gambling period is \$82,899 (the future value of investing \$6,000 per year for 10 years at $r = 7\%$), which then continues to grow if the person delays saving. The practical interpretation is that when losses are large relative to disposable income, the compounding gap can outpace the individual's ability to rebuild wealth, so *starting age no longer "rescues" the outcome*: the situation is dominated by the size of the annual loss compared to available recovery savings.

5.3.2 Annual loss vs. recovery (fixed age, DI, and duration)

Table 6 varies annual loss while holding age at 25, disposable income at \$10,000, and duration at $T = 10$ years. The relationship is strongly nonlinear:

- At \$1,000/year loss (10% of DI), recovery is 3.2 years, with missed wealth \$13,816.
- At \$3,000/year loss (30% of DI), recovery rises to 12.8 years, with missed wealth \$41,449.
- At \$5,000/year loss (50% of DI), recovery explodes to 50.5 years, with missed wealth \$69,082.
- At \$8,000/year loss (80% of DI) and above, the model reports `INF`: recovery becomes infeasible, and the Gambling Impact Score (GIS) saturates at high-risk values.

This sensitivity identifies a clear tipping point for public communication: once annual losses approach a large fraction of disposable income (roughly the 50–80% range in our fixed-DI experiment), recovery time can jump from "a few years" to "multiple decades," and then to "effectively never" under unchanged assumptions. This is the boundary where gambling stops behaving like discretionary entertainment and begins behaving like a long-term financial crisis.

The sensitivity results support our main societal takeaway: harm from online sports gambling is *concentrated and nonlinear*. Many individuals at low loss levels can recover, but higher-loss scenarios create compounding damage that is disproportionately large, especially when losses consume a major share of disposable income and crowd out saving.

6 Conclusion

In **Q1**, we modeled real disposable income by subtracting demographically scaled taxes and essential expenditures from total earnings. Our results identified break-even salary thresholds that varied widely for different U.S. profiles compared to UK profiles, with certain U.S. profiles operate at or below zero disposable income, demonstrating that financial vulnerability depends heavily on age, income level, and regional cost structure.

For **Q2**, we developed the Behavioral Gambling Trajectory (BGT) model to simulate one year of gambling using continuous engagement intensity, demographic multipliers, and behavioral feedback such as loss-chasing and quitting thresholds. The results produced heavy-tailed loss distributions and showed that while most participants lose moderately, a small high-intensity subgroup accounts for disproportionately large losses.

In **Q3**, we translated expected gambling losses into long-term financial consequences by modeling burden ratios, debt accumulation, recovery time, and missed retirement wealth. The results revealed a nonlinear tipping point when losses approach disposable income, after which debt can escalate rapidly and long-term opportunity costs grow dramatically, especially for younger individuals.

Combined, our models demonstrate that the financial risk associated with online gambling is not uniform, but rather depends heavily on disposable income, which itself is heavily dependent on demographic factors, behavioral adaptation, and compounding over time. Additionally, our model identifies which demographic groups may be more vulnerable and heavily affected by online gambling. This framework provides policymakers and educators with a quantitative tool to identify high-risk groups, evaluate intervention thresholds, and publicly communicate the long-term financial significance of sustained gambling losses.

References

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

[2] IRS, "Tax Inflation Adjustments for Tax Year 2026, Including Amendments from the One Big Beautiful Bill," https://www.irs.gov/newsroom/irs-releases-tax-inflation-adjustments-for-tax-year-2026

[3] Tax Foundation, "State Individual Income Tax Rates and Brackets" and "State and Local Sales Tax Rates," 2024. https://taxfoundation.org

[4] Bureau of Labor Statistics, Consumer Expenditure Survey, 2024. https://www.bls.gov/cex/

[5] Bureau of Labor Statistics, Current Population Survey, Table A-16: Multiple Jobholding, 2023.

[6] Bureau of Economic Analysis, Regional Price Parities by State, 2022. https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area

[7] U.S. Census Bureau, ACS "Earnings in the Past 12 Months," 2023.

[8] Pew Research Center, "The State of Gig Work in 2023."

[9] HMRC, "Income Tax Rates and Personal Allowances," 2025–26 tax year. https://www.gov.uk/income-tax-rates

[10] Office for National Statistics, "Family Spending in the UK," 2024.

[11] Cookson, Engelberg, Mullins, "Online Sports Betting Is Draining Household Savings," Kellogg Insight.

[12] Siena College Research Institute, "22% of All Americans Have Active Online Sports Betting Account," Feb 2025.

[13] U.S. Census Bureau, "American Community Survey (ACS)," https://www.census.gov/programs-surveys/acs

[14] U.S. Census Bureau, "ACSBR-025: (ACS Brief)," 2025. https://www2.census.gov/library/publications/2025/demo/acsbr-025.pdf

[15] Bureau of Labor Statistics, Current Population Survey, "Table A-36," https://www.bls.gov/cps/cpsaat36.htm

[16] "State Tax Brackets (50 states) — compiled spreadsheet," Google Sheets (accessed Mar. 2, 2026). https://docs.google.com/spreadsheets/d/1KYujtuqw4-SvMLvGXSXUYCRwbPZyn0CB6M1UPf1yDn0/edit?usp=sharing

[17] Office for National Statistics, "Household disposable income and inequality, financial year ending 2024." https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2024

[18] UK Government (DWP), "Households below average income: an analysis of the UK income distribution (FYE 1995 to FYE 2024)." https://www.gov.uk/government/statistics/households-below-average-income-for-financial-years-ending-1995-to-2024/households-below-average-income-an-analysis-of-the-uk-income-distribution-fye-1995-to-fye-20

[19] Casino.org, "Sports Betting: The Juice, House Edge, Magic Number." https://www.casino.org/blog/sports-betting-house-edge/

Appendix of MatLab Code

```
function Q2_BGT_Model()
    clc; close all;
    rng(42);

    N = 100000;

    % Population
    genders    = randsample({'Male','Female'}, N, true, [0.50 0.50]);
    edu_levels = randsample({'College','NoCollege'}, N, true, [0.33 0.67]);

    age_probs = [0.08 0.18 0.20 0.20 0.20 0.14];
    age_groups_drawn = randsample(1:6, N, true, age_probs);
    age_ranges = [16 20; 21 29; 30 44; 45 54; 55 64; 65 80];
    ages = zeros(N,1);
    for i = 1:N
        lo = age_ranges(age_groups_drawn(i), 1);
        hi = age_ranges(age_groups_drawn(i), 2);
        ages(i) = randi([lo, hi]);
    end

    disposable_by_age = [1231, 28019, 37056, 40794, 36625, 10106, 194];
    disposable_income = zeros(N,1);
    for i = 1:N
        a = ages(i);
        if a < 25,     base_d = disposable_by_age(1);
        elseif a < 35, base_d = disposable_by_age(2);
        elseif a < 45, base_d = disposable_by_age(3);
        elseif a < 55, base_d = disposable_by_age(4);
        elseif a < 65, base_d = disposable_by_age(5);
        elseif a < 75, base_d = disposable_by_age(6);
        else,          base_d = disposable_by_age(7);
        end
        disposable_income(i) = max(500, base_d * exp(0.4 * randn()));
    end

    TARGET_MEAN_LOSS = -349;

    % Intensity
    intensity = zeros(N, 1);
    mu_latent  = -2.5;
    sig_latent = 1.2;

    f0 = 0.025;
    A_legal = 21; tau = 0.5;

    for i = 1:N
        a = ages(i);
        g = genders{i};
        e = edu_levels{i};

        if strcmpi(g, 'Male'), m_gender = 30/22; else, m_gender = 15/22; end

        if a < 18,       m_age = 0.01;
        elseif a <= 34,  m_age = 35/22;
        elseif a <= 49,  m_age = 34/22;
        elseif a <= 64,  m_age = 10/22;
        else,            m_age = 4/22;
        end

        if strcmpi(e, 'College'), m_edu = 25/22; else, m_edu = 21/22; end

        L = 1 / (1 + exp(-(a - A_legal) / tau));

        f_demo = f0 * m_gender * m_age * m_edu * L;
        intensity(i) = f_demo * exp(mu_latent + sig_latent * randn());
    end

    % Annual wager + active flag
    annual_wager = zeros(N, 1);
    for i = 1:N
        W = disposable_income(i) * intensity(i);
        if W < 10, W = 0; end
        annual_wager(i) = W;
    end
    is_gambler = annual_wager > 0;

    % Effective bet counts
    med_int = median(intensity(is_gambler));
    if isempty(med_int) || isnan(med_int)
        med_int = median(intensity(intensity > 0));
    end

    eff_bets = zeros(N, 1);
    for i = 1:N
        if is_gambler(i)
            B = max(1, round(104 * intensity(i) / med_int));
            eff_bets(i) = min(B, 1000);
        end
    end
```

```
function Q2_BGT_Model()
    rho = zeros(N, 1);
    for i = 1:N
        a = ages(i); g = genders{i};
        r0 = -0.3;

        if a <= 34,     r_age = 0.5;
        elseif a <= 49, r_age = 0.2;
        elseif a <= 64, r_age = -0.2;
        else,           r_age = -0.6;
        end

        if strcmpi(g, 'Male'), r_gender = 0.25; else, r_gender = -0.25; end

        sigma_r = 0.7;
        rho(i) = 1 / (1 + exp(-(r0 + r_age + r_gender + sigma_r * randn())));
    end

    % Tilt sensitivity (alpha)
    alpha = zeros(N, 1);
    for i = 1:N
        a = ages(i); g = genders{i};

        if a <= 34,     a_age = 0.58;
        elseif a <= 49, a_age = 0.54;
        elseif a <= 64, a_age = 0.38;
        else,           a_age = 0.15;
        end

        if strcmpi(g, 'Male'), a_gen = 0.57; else, a_gen = 0.41; end

        alpha(i) = sqrt(a_age * a_gen);
    end

    % Two-phase simulation
    annual_net    = zeros(N, 1);
    total_wagered = zeros(N, 1);
    quit_midyear  = false(N, 1);
    C_mid_all     = zeros(N, 1);

    for i = 1:N
        if ~is_gambler(i), continue; end

        W1 = annual_wager(i) / 2;
        B1 = max(1, round(eff_bets(i) / 2));

        mu1  = -0.045 - 0.105 * rho(i);
        sig1 = (0.15 + 0.60 * rho(i)) / sqrt(B1);
        rmax1 = 0.91 * (1 - rho(i)) + 6.0 * rho(i);
        R1 = min(max(mu1 + sig1 * randn(), -1), rmax1);

        C_mid = W1 * R1;
        C_mid_all(i) = C_mid;

        dynamic_limit = 0.05 + 0.03 * log(disposable_income(i) / 1000);
        if C_mid < -dynamic_limit * disposable_income(i)
            annual_net(i) = C_mid;
            total_wagered(i) = W1;
            quit_midyear(i) = true;
            continue;
        end

        theta = C_mid / max(W1, 1);

        W2 = (annual_wager(i) / 2) * exp(-alpha(i) * theta);
        W2 = min(W2, 3 * (annual_wager(i) / 2));
        if W2 < 5
            annual_net(i) = C_mid;
            total_wagered(i) = W1;
            continue;
        end

        wager_ratio = W2 / max(annual_wager(i) / 2, 1);
        B2 = max(1, round((eff_bets(i) / 2) * wager_ratio));
        B2 = min(B2, 1000);

        rho2 = max(0, min(1, rho(i) - 0.3 * theta));

        mu2  = -0.045 - 0.105 * rho2;
        sig2 = (0.15 + 0.60 * rho2) / sqrt(B2);
        rmax2 = 0.91 * (1 - rho2) + 6.0 * rho2;
        R2 = min(max(mu2 + sig2 * randn(), -1), rmax2);

        annual_net(i) = C_mid + W2 * R2;
        total_wagered(i) = W1 + W2;
    end

    % Results
    fprintf('============================================================\n');
```

```
function Q2_BGT_Model()
    pct_gambler = 100 * sum(is_gambler) / N;
    fprintf('Population simulated:            %d\n', N);
    fprintf('Active gamblers (wager >= $10):  %d (%.1f%%)\n', sum(is_gambler), pct_gambler);
    fprintf('Quit midyear (hit loss limit):   %d (%.1f%% of gamblers)\n', ...
        sum(quit_midyear), 100*sum(quit_midyear)/max(1,sum(is_gambler)));

    g_net = annual_net(is_gambler);
    g_wag = total_wagered(is_gambler);

    fprintf('\n--- Among Active Gamblers ---\n');
    fprintf('Mean annual net:          $%.2f\n', mean(g_net));
    fprintf('  Target:                 $%.2f\n', TARGET_MEAN_LOSS);
    fprintf('Median annual net:        $%.2f\n', median(g_net));
    fprintf('Std dev:                  $%.2f\n', std(g_net));
    fprintf('Mean total wagered:       $%.2f\n', mean(g_wag));
    fprintf('Mean eff. bets/year:      %.0f\n', mean(eff_bets(is_gambler)));

    mo = g_wag / 12;
    fprintf('\nMonthly wager distribution\n');
    fprintf('  $1-100/mo:   %.1f%%\n', 100*mean(mo >= 1 & mo <= 100));
    fprintf('  $101-500/mo: %.1f%%\n', 100*mean(mo > 100 & mo <= 500));
    fprintf('  $500+/mo:    %.1f%%\n', 100*mean(mo > 500));

    pct_profit = 100 * sum(g_net > 0) / numel(g_net);
    pct_loss   = 100 * sum(g_net < 0) / numel(g_net);
    pct_even   = 100 * sum(abs(g_net) < 0.02 * g_wag) / numel(g_net);
    fprintf('\nOutcome breakdown\n');
    fprintf('  Profitable:    %.1f%%\n', pct_profit);
    fprintf('  Net loss:      %.1f%%\n', pct_loss);
    fprintf('  Near even:     %.1f%%\n', pct_even);
    fprintf('  Lost > $500:   %.1f%%\n', 100*sum(g_net < -500)/numel(g_net));
    fprintf('  Lost > $1000:  %.1f%%\n', 100*sum(g_net < -1000)/numel(g_net));
    fprintf('  Lost > $5000:  %.1f%%\n', 100*sum(g_net < -5000)/numel(g_net));

    fprintf('\nPercentiles of annual net\n');
    for p = [1, 5, 10, 25, 50, 75, 90, 95, 99]
        fprintf('  %3dth: $%.2f\n', p, safe_prctile(g_net, p));
    end

    % Plots
    figure('Position', [50 50 950 500]);
    histogram(g_net, 100, 'FaceColor', [0.20 0.60 0.86], 'EdgeColor', 'w', 'FaceAlpha', 0.85);
    hold on;
    xline(mean(g_net), 'r--', 'LineWidth', 2.5);
    xline(median(g_net), 'b--', 'LineWidth', 2);
    xline(0, 'Color', [0.3 0.7 0.3], 'LineWidth', 1.5);
    title('Annual Net Gain/Loss Distribution (Active Gamblers)');
    xlabel('Annual Net Gain/Loss ($)'); ylabel('Count');
    legend(sprintf('Mean: $%.0f', mean(g_net)), sprintf('Median: $%.0f', median(g_net)), ...
        'Break Even', 'Location', 'northeast');
    grid on; hold off;

    figure('Position', [50 600 950 400]);
    age_labels = {'16-20','21-29','30-44','45-54','55-64','65+'};
    age_bins = [16 21 30 45 55 65 81];
    means_age = zeros(6,1); medians_age = zeros(6,1);
    for k = 1:6
        mask = is_gambler & ages >= age_bins(k) & ages < age_bins(k+1);
        if any(mask)
            means_age(k) = mean(annual_net(mask));
            medians_age(k) = median(annual_net(mask));
        end
    end
    b = bar([means_age, medians_age]);
    b(1).FaceColor = [0.85 0.33 0.10]; b(2).FaceColor = [0.20 0.60 0.86];
    set(gca, 'XTickLabel', age_labels);
    xlabel('Age Group'); ylabel('Annual Net ($)');
    title('Net Outcome by Age Group'); legend('Mean', 'Median'); grid on;

    figure('Position', [50 1050 500 400]);
    m_mask = is_gambler & strcmpi(genders, 'Male');
    f_mask = is_gambler & strcmpi(genders, 'Female');
    bar_g = [mean(annual_net(m_mask)), median(annual_net(m_mask)); ...
             mean(annual_net(f_mask)), median(annual_net(f_mask))];
    b2 = bar(bar_g);
    b2(1).FaceColor = [0.85 0.33 0.10]; b2(2).FaceColor = [0.20 0.60 0.86];
    set(gca, 'XTickLabel', {'Male','Female'});
    ylabel('Annual Net ($)'); title('Net Outcome by Gender');
    legend('Mean','Median'); grid on;

    figure('Position', [600 1050 600 400]);
    mid_losers  = is_gambler & C_mid_all < -100;
    mid_winners = is_gambler & C_mid_all > 100;
    if any(mid_losers) && any(mid_winners)
        edges = linspace(-5000, 3000, 60);
        histogram(annual_net(mid_losers), edges, 'FaceColor', [0.85 0.33 0.10], 'FaceAlpha', 0.6, 'EdgeColor', 'none');
        hold on;
        histogram(annual_net(mid_winners), edges, 'FaceColor', [0.20 0.60 0.86], 'FaceAlpha', 0.6, 'EdgeColor', 'none');
        title('Loss-Chasing Effect: Outcomes by Mid-Year Status');
```

```
function Q3_DontBreakTheBank()
    clc;

    fprintf('============================================================\n');
    fprintf('  M3 CHALLENGE 2026 — Q3: DON''T BREAK THE BANK\n');
    fprintf('  Quantifying the Long-Run Cost of Online Sports Gambling\n');
    fprintf('============================================================\n\n');


    r = 0.07;   % annual investment return (keep consistent: nominal vs real)
    i = 0.22;   % debt APR (credit card-like)
    retirement_age = 65;

    ent_netflix   = 180;    % Netflix ~ $15/mo
    ent_concerts  = 600;    % ~4 concerts/year at $150 avg
    ent_dining    = 3500;   % avg dining out spending (illustrative)
    ent_videogames = 500;   % hardware + games
    ent_sports_tix = 1200;  % ~10 games/year at $120 avg


    profiles = {
        'Young Male (21, high risk)',     21,   7000,   4800,    15;
        'Young Female (24, moderate)',    24,   7500,   2400,     8;
        'Working Male (35, moderate)',    35,  15000,   6000,    10;
        'Working Female (40, low risk)',  40,  12000,   1800,     5;
        'Mid-career Male (50, high)',     50,  18000,   9600,    10;
        'Senior (65, low risk)',          65,  10000,   1200,     3;
    };

    N = size(profiles, 1);

    results = struct('label',{}, 'age',{}, 'DI',{}, 'loss',{}, 'T',{}, ...
                     'debt',{}, 'DC_end',{}, 'n_debt',{}, 'n_wealth',{}, ...
                     'n_total',{}, 'retirement_delay',{}, 'GIS',{}, ...
                     'loss_pct_DI',{}, 'break_even_age',{}, 'FV_missed_ret',{});


    savings_frac = 0.50;


    fprintf('%-35s | %5s | %10s | %10s | %7s | %7s | %7s | %5s\n', ...
        'Profile', 'Age', 'Debt($)', 'DC_end($)', 'n_debt', 'n_wlth', 'n_tot', 'GIS');
    fprintf('%s\n', repmat('-', 1, 115));

    for k = 1:N
        label = profiles{k, 1};
        age   = profiles{k, 2};
        DI    = profiles{k, 3};
        L_yr  = profiles{k, 4};
        T     = profiles{k, 5};

        % Annual savings available for recovery after quitting
        S = savings_frac * DI;

        [n_total, DC_end, D, n_d, n_p] = compute_recovery(DI, L_yr, T, S, r, i);

        % Missed wealth :(
        FV_end = future_value_annuity(L_yr, r, T); % value at end of gambling period
        years_to_ret = max(0, retirement_age - (age + T));
        FV_missed_ret = FV_end * (1 + r)^years_to_ret;

        %Catch up to non gambler!!!!!!!!!! :DDDDDD
        break_even_age = age + T + n_total;

        retirement_delay = n_total;


        loss_ratio    = min(L_yr / max(DI, 1), 1.5) / 1.5;
        remaining_yrs = max(retirement_age - age, 1);

        if isinf(n_total)
            recovery_burden = 1.0;
        else
            recovery_burden = min(n_total / remaining_yrs, 1.0);
        end

        debt_exposure = min(D / max(DI * 5, 1), 1.0);

        GIS = round(100 * (0.30 * loss_ratio + ...
                            0.40 * recovery_burden + ...
                            0.30 * debt_exposure));

        % Store results
        results(k).label          = label;
        results(k).age            = age;
        results(k).DI             = DI;
        results(k).loss           = L_yr;
        results(k).T              = T;
        results(k).debt           = D;
        results(k).DC_end         = DC_end;
```

```
function Q3_DontBreakTheBank()
    for k = 1:N
        results(k).n_debt          = n_d;
        results(k).n_wealth        = n_o;
        results(k).n_total         = n_total;
        results(k).retirement_delay = retirement_delay;
        results(k).GIS             = GIS;
        results(k).loss_pct_DI     = 100 * L_yr / max(DI, 1);
        results(k).break_even_age  = break_even_age;
        results(k).FV_missed_ret   = FV_missed_ret;

        % Print summary row
        fprintf('%-35s | %5d | %10.0f | %10.0f | %7.1f | %7.1f | %7.1f | %5d\n', ...
            label, age, D, OC_end, n_d, n_o, n_total, GIS);
    end


    fprintf('\n==============================================================\n');
    fprintf('  ENTERTAINMENT SPENDING COMPARISON (Annual, per person)\n');
    fprintf('==============================================================\n');
    fprintf('  %-25s  $%6.0f\n', 'Netflix subscription', ent_netflix);
    fprintf('  %-25s  $%6.0f\n', 'Concerts (4/yr)', ent_concerts);
    fprintf('  %-25s  $%6.0f\n', 'Video games', ent_videogames);
    fprintf('  %-25s  $%6.0f\n', 'Sports tickets (10/yr)', ent_sports_tix);
    fprintf('  %-25s  $%6.0f\n', 'Dining out', ent_dining);
    fprintf('  %s\n', repmat('-', 1, 40));

    for k = 1:N
        L = results(k).loss;
        fprintf('  %-25s  $%6.0f  (= %.0fx Netflix, %.1fx dining out)\n', ...
            results(k).label, L, L / ent_netflix, L / ent_dining);
    end



    for k = 1:N
        R = results(k);

        fprintf('>> %s\n', R.label);
        fprintf('   Gambling eats %.0f%% of disposable income ($%,.0f of $%,.0f)\n', ...
            R.loss_pct_DI, R.loss, R.DI);

        if isinf(R.n_debt) || isinf(R.n_total)
            fprintf('   CRITICAL: Debt spiral - yearly savings cannot beat interest under assumptions.\n');
        else
            fprintf('   After %d years of gambling, recovery takes %.1f years\n', R.T, R.n_total);
            fprintf('   Break-even age (catches up to non-gambler): %.0f years old\n', R.break_even_age);
            fprintf('   Missed wealth at age %d (if invested instead): %s\n', retirement_age, format_dollars(R.FV_missed_ret));
            fprintf('   Retirement delayed by ~%.1f years (approx.)\n', R.retirement_delay);
        end

        fprintf('   GAMBLING IMPACT SCORE: %d / 100', R.GIS);
        if R.GIS >= 70
            fprintf('  [HIGH RISK]\n');
        elseif R.GIS >= 40
            fprintf('  [MODERATE RISK]\n');
        else
            fprintf('  [LOW RISK]\n');
        end
        fprintf('\n');
    end



    fprintf('  %8s | %10s | %12s | %18s\n', 'StartAge', 'n_total', 'BreakEvenAge', 'MissedWealth@65');
    fprintf('  %s\n', repmat('-', 1, 62));

    ages_test = [18, 21, 25, 30, 35, 40, 50, 60];
    DI_fixed  = 10000;
    L_fixed   = 6000;
    T_fixed   = 10;
    S_fixed   = savings_frac * DI_fixed;

    for j = 1:length(ages_test)
        a = ages_test(j);
        [nt, ~, ~, nd, ~] = compute_recovery(DI_fixed, L_fixed, T_fixed, S_fixed, r, i);

        FV_end = future_value_annuity(L_fixed, r, T_fixed);
        years_to_ret = max(0, retirement_age - (a + T_fixed));
        FV_ret = FV_end * (1 + r)^years_to_ret;

        if isinf(nd) || isinf(nt)
            fprintf('  %8d | %10s | %12s | %18s\n', a, 'INF', 'INF', format_dollars(FV_ret));
        else
            fprintf('  %8d | %10.1f | %12.0f | %18s\n', a, nt, a + T_fixed + nt, format_dollars(FV_ret));
        end
    end

    fprintf('\n  KEY INSIGHT: Starting younger pushes the break-even age later,\n');
    fprintf('  and increases the missed-wealth-at-retirement due to compounding.\n\n');
```

```
function Q3_DontBreakTheBank()

    fprintf('  %12s | %10s | %18s | %5s\n', 'AnnualLoss', 'n_total', 'MissedWealth@65', 'GIS');
    fprintf('  %s\n', repmat('-', 1, 55));

    losses_test = [1000, 3000, 5000, 8000, 10000, 12000, 15000];
    age_fixed = 25;

    for j = 1:length(losses_test)
        Lj = losses_test(j);

        [nt_j, ~, Dj, nd_j, ~] = compute_recovery(DI_fixed, Lj, T_fixed, S_fixed, r, i);

        % Missed wealth at retirement for age_fixed
        FV_end = future_value_annuity(Lj, r, T_fixed);
        years_to_ret = max(0, retirement_age - (age_fixed + T_fixed));
        FV_ret = FV_end * (1 + r)^years_to_ret;

        % Compute GIS (same structure as main)
        lr = min(Lj / max(DI_fixed, 1), 1.5) / 1.5;
        if isinf(nt_j)
            rb = 1.0;
        else
            rb = min(nt_j / max(retirement_age - age_fixed, 1), 1.0);
        end
        de = min(Dj / max(DI_fixed * 5, 1), 1.0);
        gis_j = round(100 * (0.30*lr + 0.40*rb + 0.30*de));

        if isinf(nd_j) || isinf(nt_j)
            fprintf('  $%10.0f | %10s | %18s | %5d\n', Lj, 'INF', format_dollars(FV_ret), gis_j);
        else
            fprintf('  $%10.0f | %10.1f | %18s | %5d\n', Lj, nt_j, format_dollars(FV_ret), gis_j);
        end
    end

    fprintf('\n  NOTE: Once annual losses exceed disposable income, debt can accumulate.\n');
    fprintf('  If post-quit savings cannot beat debt interest (S <= i*Debt), recovery becomes impossible.\n\n');

    fprintf('==============================================================\n');
    fprintf('  END OF Q3 ANALYSIS\n');
    fprintf('==============================================================\n');
end


function [n_total, OC_final, D_final, n_d, n_o] = compute_recovery(DI, L, T, S, r, i)
    % Simulate gambling period and compute recovery time with:
    % - debt that compounds at APR i
    % - wealth gap (opportunity cost) that compounds at return r
    %
    % Returns:
    %   n_total: total recovery years after quitting
    %   OC_final: opportunity cost at end of gambling period
    %   D_final: debt at end of gambling period
    %   n_d: years to pay off debt
    %   n_o: years to rebuild wealth gap

    % --- Simulate gambling years ---
    OC = 0;
    D  = 0;

    for t = 1:T
        % opportunity cost grows at r, plus this year's loss
        OC = OC * (1 + r) + L;

        shortfall = max(0, L - DI);   % needs borrowing
        slack     = max(0, DI - L);   % can repay debt

        if shortfall > 0
            D = D * (1 + i) + shortfall;
        else
            D = max(0, D * (1 + i) - slack);
        end
    end

    if D > 0
        if S <= 0 || S <= i * D
            n_d = Inf; % true debt spiral condition (payment can't beat interest)
        else
            n_d = log(S / (S - i * D)) / log(1 + i);
        end
    else
        n_d = 0;
    end

    if isinf(n_d) || S <= 0
        n_o = Inf;
    else
        OC_star = OC * (1 + r)^n_d; % gap grows while paying debt
        if abs(r) < 1e-12
            n_o = OC_star / S;
        else
```

```
function [n_total, OC_final, D_final, n_d, n_o] = compute_recovery(DI, L, T, S, r, i)
    OC_final = OC;
    D_final  = D;
end

function fv = future_value_annuity(payment, r, T)
    % FV of investing "payment" each year for T years at return r
    if abs(r) < 1e-12
        fv = payment * T;
    else
        fv = payment * ((1 + r)^T - 1) / r;
    end
end

function s = format_dollars(x)
    % Format number with commas + leading $
    if x < 0
        s = ['-$', format_pos(-x)];
    else
        s = ['$', format_pos(x)];
    end
end

function s = format_pos(x)
    s = sprintf('%.0f', x);
    n = length(s);
    if n > 3
        groups = {};
        while n > 3
            groups{end+1} = s(n-2:n); %#ok<AGROW>
            s = s(1:n-3);
            n = length(s);
        end
        groups{end+1} = s; %#ok<AGROW>
        groups = fliplr(groups);
        s = strjoin(groups, ',');
    end
end
```